



ABORDAGEM ESPAÇO-TEMPORAL APLICADA À ESTIMAÇÃO DE  
DEMANDA DE GÁS NATURAL

Jorge Henrique de Castro

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Alexandre Pinto Alves da Silva

Rio de Janeiro

Março de 2012

ABORDAGEM ESPAÇO-TEMPORAL APLICADA À ESTIMAÇÃO DE  
DEMANDA DE GÁS NATURAL

Jorge Henrique de Castro

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ  
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA  
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS  
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM  
CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

---

Prof. Alexandre Pinto Alves da Silva, Ph.D.

---

Prof. Carlos Eduardo Pedreira, Ph.D.

---

Prof. Luiz Fernando Loureiro Legey, Ph.D.

---

Prof. Luiz Pereira Calôba, Dr.Ing.

---

Prof. Nelson Kagan, Ph.D.

RIO DE JANEIRO, RJ - BRASIL

MARÇO DE 2012

Castro, Jorge Henrique de

Abordagem Espaço-temporal Aplicada à Estimação de Demanda de Gás Natural / Jorge Henrique de Castro – Rio de Janeiro: UFRJ / COPPE, 2012.

XIX, 300 p.: il.; 29,7 cm.

Orientador: Alexandre Pinto Alves da Silva

Tese (doutorado) – UFRJ / COPPE / Programa de Engenharia Elétrica, 2012.

Referências Bibliográficas: p. 206 – 220.

1. Previsão de Demanda. 2. Gás Natural. 3. Modelos Espaço-Temporais. I. Silva, Alexandre Pinto Alves da. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título

*Dedico esta tese  
à minha querida, amada e dedicada  
esposa, Cristiana e aos meus amados  
filhos Caio e Carol pelo carinho,  
força, paciência, apoio e  
compreensão ao longo desta jornada.*

## **AGRADECIMENTOS**

Agradeço primeiramente a Nosso Senhor Jesus Cristo que em sua infinita bondade e misericórdia me permitiu chegar com sucesso ao fim desta jornada, bem como propiciou encontrar o apoio de outras pessoas que me ajudaram e tiveram contribuição valorosa na consecução desta tese.

Agradeço a minha mãe pelo apoio, dedicação e incentivo ao longo de toda a minha vida, pois sem este, se quer, teria a condição de realizar este trabalho.

Agradeço a minha esposa pelo apoio incondicional, paciência com minha ausência e dedicação para cuidar da nossa família durante o período de estudos.

Agradeço aos meus filhos pelos momentos em que me fizeram ver que algumas coisas têm mais importância do que a busca pela prova técnica aqui inserida.

Agradeço ao meu orientador pela sua dedicação e apoio na execução desta jornada, bem como pelo auxílio na minha formação profissional.

Agradeço a Universidade Federal do Estado do Rio de Janeiro e ao Programa de Pós-graduação em Engenharia Elétrica pela oportunidade de realizar o Doutorado.

Agradeço a Petrobras pela liberação para cumprir as exigências do curso, bem como aos colegas de trabalho que ao longo desta tese incentivaram sua execução.

Agradeço aos colegas da Companhia Paranaense de Gás, COMPAGAS, e da Gerência de Marketing e Comercialização de Gás Canalizado no Espírito Santo, da BR Distribuidora, pela atenção e apoio durante a execução da tese.

Agradeço aos amigos que ao longo destes anos ajudaram, opinaram e contribuíram para execução desta tese, pois sem suas valorosas participações tornariam o caminho mais árduo.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

## ABORDAGEM ESPAÇO-TEMPORAL APLICADA À ESTIMAÇÃO DE DEMANDA DE GÁS NATURAL

Jorge Henrique de Castro

Março/2012

Orientador: Alexandre Pinto Alves da Silva

Programa: Engenharia Elétrica

De acordo com a orientação de diversificação das fontes de geração de energia, o Gás Natural vêm, ao longo dos últimos anos, aumentando a sua participação na matriz energética nacional. Neste segmento, típica indústria de rede, o planejamento dos dutos de distribuição é um dos fatores críticos de sucesso, pois o acesso ao produto só é possível quando existem gasodutos que permitam atender aos clientes. Considerando as particularidades envolvidas neste crescimento, e a mitigação dos riscos de incertezas futuras, torna-se necessário estabelecer cenários para o apoio à tomada de decisão, no qual, requer-se a análise de um amplo conjunto de variáveis de naturezas distintas. Neste sentido, a visão espacial da infraestrutura de gasodutos, associada às variáveis de consumo e socioeconômicas ampliam a capacidade de avaliação logística do negócio e, conseqüentemente, o planejamento e as tomadas de decisão sobre vendas, engenharia e operação. Em particular, ao se observar a atividade de previsão de demanda por Gás Natural, fundamental ao estabelecimento da rede de distribuição, percebe-se que os modelos empregados são primordialmente baseados na análise da dimensão temporal dos dados; com pouco desenvolvimento no sentido de incorporar a dimensão espacial ao processo de modelagem. Nesta lacuna, esta tese propõe-se a desenvolver um modelo para a previsão de demanda, no contexto da atividade da distribuição de Gás Natural, a partir das séries temporais de consumo e de dados socioeconômicos espacialmente distribuídos; observando a influência da dimensão espacial, no que tange suas representação espaço-tempo e a dependência e heterogeneidade espacial.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

THE EMPLOYMENT OF SPATIO-TEMPORAL APPROACH IN NATURAL GAS  
DEMAND FORECASTING

Jorge Henrique de Castro

March/2012

Advisor: Alexandre Pinto Alves da Silva

Department: Electrical Engineering

Following the goal of diversifying the energy generation matrix, natural gas has played an important role in recent years. Therefore, its distribution network planning is a critical factor for success, since natural gas demand exists only if the ducts reach consumers. Considering the particularities involved in the expansion of a natural gas distribution network and aiming at mitigating the impact of uncertainties, scenarios are created to support decision making. These scenarios take into account a wide set of exogenous variables. In addition to that, the spatial analysis of the gas pipeline network, associated with socioeconomic variables and consumption, improve business evaluation, and, consequently, the planning and decision making on sales, engineering, and operation. However, it can be noticed that the natural gas demand forecasting has been primarily based on temporal information, especially by using time series econometric models. Only a few researchers have tried to embody the spatial dimension in the modeling process. Therefore, this thesis, based on the local distribution of natural gas, introduces the application of the space-time perspective in the development of natural gas demand forecasting models. The history of natural gas consumption and socioeconomic spatially distributed data are used as input variables, investigating the effect of the spatial dimension regarding space-time representation and spatial dependency / heterogeneity.

## SUMÁRIO

|   |           |
|---|-----------|
| ÍNDICE DE FIGURAS. . . . .  | xii       |
| ÍNDICE DE TABELAS. . . . .  | xv        |
| LISTA DE SIGLAS . . . . .   | xvii      |
| <b>1 – INTRODUÇÃO. . . . .</b>  | <b>1</b>  |
| 1.1 – A IMPORTÂNCIA DO TEMA ESTUDADO. . . . .                         | 1         |
| 1.2 - DESCRIÇÃO DO PROBLEMA. . . . .                                  | 5         |
| 1.3 – MOTIVAÇÃO E CONTRIBUIÇÃO. . . . .                               | 6         |
| 1.4 - OBJETIVO DA TESE. . . . .                                       | 8         |
| 1.4.1 – OBJETIVOS SECUNDÁRIOS. . . . .                                | 8         |
| 1.5 – ESCOPO . . . . .  | 8         |
| 1.6 - ESTRUTURA DA TESE . . . . .                                     | 9         |
| <b>2 – FUNDAMENTAÇÃO TEÓRICA. . . . .</b>                             | <b>10</b> |
| 2.1 – A MODELAGEM DA DEMANDA COMO APOIO À DECISÃO. . . . .            | 10        |
| 2.1.1 – MODELOS PARA PROJEÇÕES DE DEMANDA E CENÁRIOS DE ENERGIA . . . | 13        |
| 2.1.2 – MODELOS PARA PROJEÇÕES DE DEMANDA E CENÁRIOS DE GN . . . . .  | 15        |
| 2.2 – A ÓTICA DA DEMANDA PELO PLANEJAMENTO. . . . .                   | 17        |
| 2.2.1 – A ABORDAGEM DO PLANEJAMENTO E A DEMANDA DE ENERGIA. . . . .   | 20        |
| 2.3 - A VISÃO DA DEMANDA PELA ECONOMIA. . . . .                       | 20        |
| 2.4 - A VISÃO GEOGRÁFICA DA DEMANDA. . . . .                          | 22        |
| 2.4.1 – A ECONOMETRIA ESPACIAL. . . . .                               | 24        |
| 2.4.1.1 – A DEPENDÊNCIA ESPACIAL. . . . .                             | 25        |
| 2.4.1.2 – A HETEROGENEIDADE ESPACIAL. . . . .                         | 26        |
| 2.4.2 – INFORMAÇÃO ESPACIAL (IE). . . . .                             | 27        |
| 2.4.2.1 – A NATUREZA DO DADO GEOGRÁFICO. . . . .                      | 27        |
| 2.4.2.2 – ESTRUTURA DE DADOS ESPACIAIS. . . . .                       | 29        |
| 2.4.2.3 – A REPRESENTAÇÃO ESPACIAL DE DADOS SÓCIO-ECONÔMICOS. . .     | 31        |
| 2.4.3 – A ABORDAGEM ESPAÇO-TEMPORAL. . . . .                          | 31        |
| 2.4.3.1 – A ORDEM TEMPORAL. . . . .                                   | 32        |
| 2.4.3.2 – AS RELAÇÕES ESPAÇO-TEMPORAIS. . . . .                       | 33        |
| 2.4.3.3 – A REPRESENTAÇÃO DOS FENÔMENOS ESPAÇO-TEMPORAIS. . . . .     | 35        |
| 2.4.4 – A IE E SEU EMPREGO NAS ATIVIDADES EMPRESARIAIS . . . . .      | 36        |
| 2.4.4.1 – A IE EM ATIVIDADES DE PREVISÃO . . . . .                    | 38        |



|   |           |
|---|-----------|
| 2.5 – AS METODOLOGIAS PARA PREVISÃO DA DEMANDA. . . . .                           | 41        |
| 2.5.1 – ANÁLISE DE SÉRIES TEMPORAIS. . . . .                                      | 42        |
| 2.5.1.1 – PREVISÃO DE DEMANDA DE GN EMPREGANDO SÉRIES TEMPORAIS                   | 43        |
| 2.5.2 - PREVISÃO DE DEMANDA DE GN EMPREGANDO O AM. . . . .                        | 45        |
| 2.5.3 – A ANÁLISE DE DADOS ESPACIAIS. . . . .                                     | 50        |
| 2.5.3.1 – AS ANÁLISES DE EVENTOS PONTUAIS. . . . .                                | 52        |
| 2.5.3.2 – AS ANÁLISES DE ÁREAS. . . . .   | 54        |
| 2.5.3.3 – OS MODELOS DE REGRESSÃO ESPACIAL – <i>CROSS SECTION</i> . . . . .       | 57        |
| 2.5.3.4 – O MODELO DE REGRESSÃO ESPACIAL PONDERADA. . . . .                       | 60        |
| 2.5.3.5 – OS MODELOS DE REGRESSÃO ESPAÇO-TEMPORAL - <i>PANEL DATA</i> .           | 61        |
| 2.5.3.6 – EMPREGO DE ANÁLISE ESPACIAL NA PREVISÃO DE DEMANDA GN                   | 66        |
| 2.6 – CONSIDERAÇÕES DA FUNDAMENTAÇÃO TEÓRICA E DIRETRIZES À TESE. . . . .         | 67        |
| 2.6.1 – A NATUREZA DOS DADOS APLICADOS AO PROBLEMA. . . . .                       | 68        |
| 2.6.2 – O TERMO RESIDUAL. . . . .   | 71        |
| 2.6.3 – ESCOLHA DO MODELO. . . . .  | 72        |
| <b>3 – METODOLOGIA. . . . .</b>   | <b>74</b> |
| 3.1 - ESTRUTURAÇÃO E ANÁLISE DE DADOS. . . . .                                    | 76        |
| 3.2 - DIRETRIZES PARA A MODELAGEM. . . . .  | 78        |
| 3.2.1 - DEFINIÇÃO DA REPRESENTAÇÃO ESPACIAL DO CONSUMO. . . . .                   | 78        |
| 3.2.2 - ANÁLISE DE DEPENDÊNCIA ESPACIAL GLOBAL E PRÉ-SELEÇÃO DE                   |           |
| VARIÁVEIS EXPLICATIVAS. . . . .   | 80        |
| 3.2.3 - ANÁLISE DE DEPENDÊNCIA ESPACIAL LOCAL E DEFINIÇÃO DAS UE PARA             |           |
| EMPREGO DO MODELO SAR <i>PANEL</i> . . . . .                                      | 83        |
| 3.2.4 – SELEÇÃO DAS VARIÁVEIS PARA ESTIMAÇÃO DO MODELO SAR <i>PANEL</i> . . . . . | 87        |
| 3.3 – DESENVOLVIMENTO DO MODELO E PREVISÃO DO CONSUMO DE GN. . . . .              | 92        |
| 3.3.1 – MECANISMO DE ESTIMAÇÃO DO MODELO . . . . .                                | 92        |
| 3.3.2 – FORMA FUNCIONAL E MÉTODOS DE OTIMIZAÇÃO DO MODELO. . . . .                | 95        |
| 3.3.3 – ENTRADA DE DADOS PARA ESTIMAÇÃO DO MODELO. . . . .                        | 97        |
| 3.3.4 – PREVISÃO. . . . .   | 100       |
| 3.4 – TESTES DO MODELO E ANÁLISES DOS RESULTADOS. . . . .                         | 102       |
| 3.4.1 – TESTES DE ESTRUTURA SAR <i>PANEL</i> . . . . .                            | 103       |
| 3.4.2 – ANÁLISE DOS RESULTADOS DA PREVISÃO SAR <i>PANEL</i> . . . . .             | 103       |
| 3.4.3 – TESTES DE ESTABILIDADE DO MODELO. . . . .                                 | 105       |

|   |            |
|---|------------|
| 3.5 – COMPARAÇÃO COM OUTRAS ESTRUTURAS DE PREVISÃO. . . . .                 | 106        |
| 3.5.1 – SIMULAÇÃO DOS PREVISORES LINEARES. . . . .                          | 106        |
| 3.5.2 – SIMULAÇÃO DO PREVISOR NÃO-LINEAR. . . . .                           | 107        |
| 3.5.2.1 – SIMULAÇÃO DOS PREVISORES POR RN SEM A VIZINHANÇA. . . . .         | 110        |
| <b>4 – DESENVOLVIMENTO DA TESE . . . . .</b>                                | <b>111</b> |
| 4.1 – ESTRUTURAÇÃO E ANÁLISE DE DADOS. . . . .                              | 111        |
| 4.1.1 – ESCOLHA DA CDL E DEFINIÇÃO DOS LIMITES TEMPORAL E ESPACIAL. . . . . | 111        |
| 4.1.1.1 - LIMITE TEMPORAL E DA ÁREA DE CONCESSÃO DAS CDL. . . . .           | 112        |
| 4.1.1.2 - LIMITE ESPACIAL DA AE DA COMPAGAS E DA BR. . . . .                | 113        |
| 4.1.2 – SELEÇÃO, COLETA E FORMAÇÃO DA BASE DE DADOS. . . . .                | 118        |
| 4.1.2.1 - DADOS ESPACIAIS. . . . .  | 118        |
| 4.1.2.2 – VARIÁVEL DE INTERESSE. . . . .                                    | 118        |
| 4.1.2.3 - VARIÁVEL EXPLICATIVA - ÁLCOOL E DERIVADOS DE PETRÓLEO. . . . .    | 119        |
| 4.1.2.4 - VARIÁVEL EXPLICATIVA - CONSUMO DE EE. . . . .                     | 119        |
| 4.1.2.5 – VARIÁVEIS EXPLICATIVAS - PREÇO DOS ENERGÉTICOS. . . . .           | 120        |
| 4.1.2.6 – VARIÁVEIS EXPLICATIVAS - ECONOMIA E DEMOGRAFIA. . . . .           | 121        |
| 4.1.3 – ANÁLISE GRÁFICA E DE CORRELAÇÃO CRUZADA DOS DADOS. . . . .          | 124        |
| 4.1.4 – CONSIDERAÇÕES A CERCA DA BASE DE DADOS. . . . .                     | 125        |
| 4.2 – DIRETRIZES PARA MODELAGEM. . . . .                                    | 128        |
| 4.2.1 – ESCOLHA DA UGR PARA A VARIÁVEL DEPENDENTE. . . . .                  | 128        |
| 4.2.1.1 – ANÁLISE DA DISPOSIÇÃO DOS CLIENTES NAS UGR DEFINIDAS. . . . .     | 130        |
| 4.2.1.2 – ANÁLISE DAS RESTRIÇÕES DO USO DO SOLO. . . . .                    | 132        |
| 4.2.2 – ÍNDICE DE AUTOCORRELAÇÃO ESPACIAL GLOBAL (MORAN GLOBAL). . . . .    | 132        |
| 4.2.3 – PRÉ-SELEÇÃO DAS VARIÁVEIS EXPLICATIVAS. . . . .                     | 134        |
| 4.2.4 – ÍNDICE DE AUTOCORRELAÇÃO ESPACIAL LOCAL (GETIS E ORD*). . . . .     | 136        |
| 4.2.5 – DEFINIÇÃO DAS ÁREAS PARA EMPREGO DO MODELO ESPACIAL. . . . .        | 138        |
| 4.2.6 – SELEÇÃO DAS VARIÁVEIS EXPLICATIVAS. . . . .                         | 141        |
| 4.2.6.1 - PCA DEFASADA COM A VARIÁVEL DEPENDENTE. . . . .                   | 143        |
| 4.2.6.2 – DEFASAGEM ENTRE VARIÁVEIS EXPLICATIVAS E DE INTERESSE. . . . .    | 144        |
| 4.2.6.3 — SELEÇÃO DAS VARIÁVEIS EXPLICATIVAS. . . . .                       | 145        |
| 4.2.7 – ANÁLISE DA AUTOCORRELAÇÃO DAS VARIÁVEIS DEPENDENTES. . . . .        | 147        |
| 4.3 – ESTIMAÇÃO DO MODELO. . . . .  | 148        |
| 4.3.1 – RESULTADOS DA PREVISÃO. . . . .                                     | 150        |

|  |            |
|--|------------|
| 4.4 – ANÁLISE DOS RESULTADOS DAS PREVISÕES SAR. . . . .                            | 156        |
| 4.4.1 – CONCLUSÕES ACERCA DOS TESTES DE ESTRUTURA. . . . .                         | 157        |
| 4.4.2 – ANÁLISE DOS ERROS DA PREVISÃO. . . . .                                     | 158        |
| 4.4.2.1 – ANÁLISE DAS PREVISÕES INDIVIDUAIS POR UE. . . . .                        | 158        |
| 4.4.2.2 – CONCLUSÕES COM BASE NA ANÁLISE DAS PREVISÕES. . . . .                    | 167        |
| 4.4.3 – TESTES DE ESTABILIDADE DO MODELO. . . . .                                  | 172        |
| 4.4.3.1 – CONCLUSÕES COM BASE NOS TESTE DE ESTABILIDADE. . . . .                   | 175        |
| 4.5 – COMPARAÇÃO ENTRE ESTRUTURAS DE PREVISÃO. . . . .                             | 175        |
| 4.5.1 – COMPARAÇÃO E CONCLUSÕES SOBRE OS RESULTADOS SAR E ARIMA. .177              |            |
| 4.5.2 – SIMULAÇÃO COM PREVISOR POR ALISAMENTO EXPONENCIAL. . . . .                 | 178        |
| 4.5.2.1 - COMPARAÇÃO ENTRE SAR E ALISAMENTO EXPONENCIAL. . . . .                   | 180        |
| 4.5.3 – SIMULAÇÃO DO PREVISOR NÃO-LINEAR POR RN. . . . .                           | 184        |
| 4.5.3.1 – RESULTADOS DOS PREVISORES POR RN SEM A VIZINHANÇA. . . .189              |            |
| 4.5.3.2 – ANÁLISE E CONCLUSÕES A CERCA DOS RESULTADOS RN. . . . .                  | 192        |
| <b>5 – CONCLUSÕES E RECOMENDAÇÕES. . . . .</b>                                     | <b>199</b> |
| 5.1 – PRINCIPAIS CONTRIBUIÇÕES. . . . .  | 202        |
| 5.2 – TRABALHOS FUTUROS. . . . .   | 204        |
| <b>6 – REFERÊNCIAS BIBLIOGRÁFICAS. . . . .</b>                                     | <b>206</b> |
| <b>7 – APÊNDICES</b>   |            |
| APÊNDICE (A) CONTEXTUALIZAÇÃO DA INDÚSTRIA DO GN. . . . .                          | 221        |
| APÊNDICE (B) PROCEDIMENTO DE ESPACIALIZAÇÃO DE CLIENTES. . . . .                   | 227        |
| APÊNDICE (C) COLETA DE DADOS - INTERPOLAÇÃO DE VARIÁVEIS. . . . .                  | 230        |
| APÊNDICE (D) GRÁFICOS DE QUANTIDADE DE PONTOS DE CONSUMO . . . . .                 | 231        |
| APÊNDICE (E) GRÁFICOS DE ÍNDICE MORAN GLOBAL. . . . .                              | 235        |
| APÊNDICE (F) RESUMO DA DETERMINAÇÃO DAS PCA. . . . .                               | 237        |
| APÊNDICE (G) AUTOCORRELAÇÃO PARCIAL VARIÁVEL DE INTERESSE. . . . .                 | 242        |
| APÊNDICE (H) PREVISÃO SAR <i>PANEL</i> – GRÁFICOS DE ERRO. . . . .                 | 250        |
| APÊNDICE (I) PREVISÃO ARIMA POR UE. . . . .  | 265        |
| APÊNDICE (J) PREVISÕES SAR <i>PANEL</i> X ARIMA / ALISAMENTO EXPONENCIAL. . . .270 |            |
| APÊNDICE (K) PREVISÃO ALISAMENTO EXPONENCIAL POR UE. . . . .                       | 280        |
| APÊNDICE (L) PREVISÃO RN POR UE. . . . .   | 285        |
| APÊNDICE (M) PREVISÕES SAR <i>PANEL</i> X RN POR UE. . . . .                       | 297        |
| APÊNDICE (N) PRODUÇÃO TEXTUAL. . . . .   | 300        |

## ÍNDICE DE FIGURAS

|  |
|--|
| Figura 1.1 – OIE por Fonte – Comparação Brasil e Mundo – Extraída de EPE (2008 / 2009). . . 2          |
| Figura 1.2 – Projeção de OIE por Fonte – Extraída de TOMALSQUIM <i>et al.</i> (2007) . . . . . 3       |
| Figura 2.1 – Esquemático dos Estados de Escopo – Adaptada de KIRKWOOD (1997). . . . . 12               |
| Figura 2.2 – Dinâmica do Mercado Mundial de GN – Adaptada de EGGING <i>et al.</i> (2008). . . 15       |
| Figura 2.3 – Sobreposição empregando Dados <i>raster</i> - Extraída de ARONOFF (2004). . . . . 30      |
| Figura 2.4 – Comparação dos Modelos Vetorial e <i>raster</i> - Extraída de ARONOFF (2004). . . 30      |
| Figura 2.5 – Fluidez do Tempo – Adaptada de WORBOYS e DUCKHAM (2004). . . . . 32                       |
| Figura 2.6 – Variável Contínua no Tempo - Extraída de PEDROSA e CAMARA (2007) . . . 33                 |
| Figura 2.7 – Variável Discreta no Tempo - Extraída de PEDROSA e CAMARA (2007) . . . 33                 |
| Figura 2.8 – Perspectivas de Mudança - Extraída de HORNSBY e EGENHOFER (1997) . . . 35                 |
| Figura 2.9 – Despadronização da Representação Espacial – Extraída de SHORT (2007). . . . 40            |
| Figura 2.10 – Estrutura de Camadas Proposta em MISHRA e PATRA (2008a) . . . . . 48                     |
| Figura 3.1 – Fluxograma da Metodologia Proposta. . . . . 76  |
| Figura 3.2 – Significância $G_i^*$ (07/ 2008) - Curitiba Residencial - Vizinhos de 1ª ordem . . . 86   |
| Figura 3.3 – Significância $G_i^*$ (07/ 2008) - Curitiba Residencial - Vizinhos por distância . . . 86 |
| Figura 3.4 – Dependência Espacial Local - Exemplo da Matriz de Estacionariedade . . . . . 87           |
| Figura 3.5 – Vizinhança da UE 207 definida para Modelagem . . . . . 90                                 |
| Figura 3.6 – Estrutura da Rede Neural . . . . . 107  |
| Figura 4.1 – RM de Curitiba – Fonte: IBGE. . . . . 113   |
| Figura 4.2 – Consumo Total de GN Municipal x Municípios da RM – Fonte: Compagas. . . 114               |
| Figura 4.3 – Classificação Municipal - Representatividade % do PIB – Fonte: IPPUC . . . . 114          |
| Figura 4.4 – AE Compagas . . . . . 115   |
| Figura 4.5 – RM de Vitória – Fonte: IBGE. . . . . 116  |
| Figura 4.6 – Consumo Municipal de GN x Municípios da RM – Fonte: BR Distribuidora . . 116              |
| Figura 4.7 – Classificação da Representatividade.% do PIB - Fonte: IJSN. . . . . 116                   |
| Figura 4.8 – AE BR . . . . . 117   |
| Figura 4.9 – Consumo Mensal de Gás $MM^3$ por Segmentos – Fonte: Compagas. . . . . 125                 |
| Figura 4.10 – Evolução Mensal do N <sup>o</sup> de Clientes de GN por Segmentos – Fonte: Compagas 126  |
| Figura 4.11 – Consumo Mensal de Gás $MM^3$ por Segmentos – Fonte: BR. . . . . 127                      |
| Figura 4.12 – Qtd. Mensal de Clientes GN por Segmentos – Fonte: BR Distribuidora . . . . 127           |
| Figura 4.13 – Qtd. Mensal de Clientes GN por Segmentos (Log) – Fonte: BR Distribuidora . 127           |
| Figura 4.14 – Distribuição Espacial dos Clientes por Segmento – Fonte: Compagas . . . . . 129          |
| Figura 4.15 – Distribuição Espacial dos Clientes por Segmento – Fonte: BR Distribuidora. . 129         |
| Figura 4.16 – Segmento Veicular x Setor Censitário – Curitiba . . . . . 130                            |
| Figura 4.17 – Segmento Comercial x Setor Censitário – Curitiba . . . . . 131                           |

|  |     |
|--|-----|
| Figura 4.18 – Autocorrelação Local Dez/08 - Vizinhança de 1ª ordem – Vitória Residencial.  | 137 |
| Figura 4.19 – Regiões de Significância e UE com Estacionariedade - Curitiba Residencial. . | 139 |
| Figura 4.20 – Regiões de Significância e UE com Estacionariedade - Vitória Residencial . . | 139 |
| Figura 4.21 – Regiões de Significância e UE com Estacionariedade - Vitória Comercial . . . | 140 |
| Figura 4.22 – Comportamento da Variância Individual das Principais Componentes . . . . .   | 142 |
| Figura 4.23 (a) – Previsão UE 624 - Segmento Residencial de Curitiba – Opção (A) . . . . . | 151 |
| Figura 4.23 (b) – Previsão UE 624 - Segmento Residencial de Curitiba – Opção (B) . . . . . | 151 |
| Figura 4.23 (c) – Previsão UE 624 - Segmento Residencial de Curitiba – Opção (C) . . . . . | 151 |
| Figura 4.24 (a) – Previsão UE 638 - Segmento Residencial de Curitiba – Opção (A) . . . . . | 152 |
| Figura 4.24 (b) – Previsão UE 638 - Segmento Residencial de Curitiba – Opção (B) . . . . . | 152 |
| Figura 4.24 (c) – Previsão UE 638 - Segmento Residencial de Curitiba – Opção (C) . . . . . | 152 |
| Figura 4.25 (a) – Previsão UE 207 - Segmento Comercial de Vitória – Opção (A). . . . .     | 153 |
| Figura 4.25 (b) – Previsão UE 207 - Segmento Comercial de Vitória – Opção (B). . . . .     | 153 |
| Figura 4.25 (c) – Previsão UE 207 - Segmento Comercial de Vitória – Opção (C). . . . .     | 153 |
| Figura 4.26 (a) – Previsão UE 247 - Segmento Residencial de Vitória – Opção (A). . . . .   | 154 |
| Figura 4.26 (b) – Previsão UE 247 - Segmento Residencial de Vitória – Opção (B). . . . .   | 154 |
| Figura 4.26 (c) – Previsão UE 247 - Segmento Residencial de Vitória – Opção (C). . . . .   | 154 |
| Figura 4.27 (a) – Previsão UE 258 - Segmento Residencial de Vitória – Opção (A). . . . .   | 155 |
| Figura 4.27 (b) – Previsão UE 258 - Segmento Residencial de Vitória – Opção (B). . . . .   | 155 |
| Figura 4.27 (c) – Previsão UE 258 - Segmento Residencial de Vitória – Opção (C). . . . .   | 155 |
| Figura 4.28 – Previsão UE 624 – Erro Relativo Mensal – Opção (A) . . . . .                 | 159 |
| Figura 4.29 – Previsão UE 638 – Erro Relativo Mensal – Opção (A) . . . . .                 | 159 |
| Figura 4.30 – Previsão UE 624 – Erro Relativo Mensal – Opção (C) . . . . .                 | 159 |
| Figura 4.31 – Previsão UE 638 – Erro Relativo Mensal – Opção (C) . . . . .                 | 160 |
| Figura 4.32 – Previsão UE 624 – Erro Acumulado Total – Opção (C) . . . . .                 | 162 |
| Figura 4.33 – Previsão UE 638 – Erro Acumulado Total – Opção (C) . . . . .                 | 162 |
| Figura 4.34 – Previsão UE 207 – Erro Acumulado Total – Opção (B) . . . . .                 | 163 |
| Figura 4.35 – Previsão UE 247 – Erro Relativo Mensal – Opção (C) . . . . .                 | 164 |
| Figura 4.36 – Previsão UE 247 – Erro Acumulado Total – Opção (C). . . . .                  | 165 |
| Figura 4.37 – Previsão UE 247 – Erro Absoluto Mensal – Opção (C). . . . .                  | 166 |
| Figura 4.38 – Previsão UE 258 – Erro Acumulado Total – Opção (B). . . . .                  | 166 |
| Figura 4.39 – Previsão UE 258 – Erro Relativo Mensal – Opção (C) . . . . .                 | 167 |
| Figura 4.40 – Previsão UE 258 – Erro Acumulado Total – Opção (C) . . . . .                 | 167 |
| Figura 4.41 – Erro Relativo Mensal SAR UE 624 – Valores Originais x Degradados. . . . .    | 172 |
| Figura 4.42 – Erro Relativo Mensal SAR UE 638 – Valores Originais x Degradados. . . . .    | 173 |
| Figura 4.43 – Erro Relativo Mensal SAR UE 207 – Valores Originais x Degradados. . . . .    | 173 |

|   |     |
|---|-----|
| Figura 4.44 – Erro Relativo Mensal SAR UE 247 – Valores Originais x Degradados . . . . .    | 173 |
| Figura 4.45 – Erro Relativo Mensal SAR UE 258 – Valores Originais x Degradados . . . . .    | 174 |
| Figura 4.46 – Previsão ARIMA UE 624 – Erro Relativo Mensal . . . . .                        | 175 |
| Figura 4.47 – Previsão ARIMA UE 638 – Erro Relativo Mensal . . . . .                        | 176 |
| Figura 4.48 – Previsão ARIMA UE 207 – Erro Relativo Mensal . . . . .                        | 176 |
| Figura 4.49 – Previsão ARIMA UE 247 – Erro Relativo Mensal . . . . .                        | 176 |
| Figura 4.50 – Previsão ARIMA UE 258 – Erro Relativo Mensal . . . . .                        | 177 |
| Figura 4.51 – Previsão Alisamento Exponencial - UE 624 – Erro Relativo Mensal . . . . .     | 178 |
| Figura 4.52 – Previsão Alisamento Exponencial - UE 638 – Erro Relativo Mensal . . . . .     | 179 |
| Figura 4.53 – Previsão Alisamento Exponencial - UE 207 – Erro Relativo Mensal . . . . .     | 179 |
| Figura 4.54 – Previsão Alisamento Exponencial - UE 247 – Erro Relativo Mensal . . . . .     | 179 |
| Figura 4.55 – Previsão Alisamento Exponencial - UE 258 – Erro Relativo Mensal . . . . .     | 180 |
| Figura 4.56 – Alisamento Exponencial x SAR - UE 624 – Comparativo da Previsão . . . . .     | 182 |
| Figura 4.57 – Alisamento Exponencial x SAR - UE 638 – Comparativo da Previsão . . . . .     | 182 |
| Figura 4.58 – Alisamento Exponencial x SAR - UE 247 – Comparativo da Previsão . . . . .     | 182 |
| Figura 4.59 – Alisamento Exponencial x SAR - UE 247 – Erro Acumulado Total. . . . .         | 183 |
| Figura 4.60 – Alisamento Exponencial x SAR - UE 624 – Erro Absoluto Mensal . . . . .        | 183 |
| Figura 4.61 – Previsão RN - UE 624 – Método de Treinamento LM GD / GDM. . . . .             | 185 |
| Figura 4.62 – Previsão RN - UE 624 – Método de Treinamento GDM GD / GDM . . . . .           | 185 |
| Figura 4.63 – Previsão RN - UE 624 – Método de Treinamento GDX GD / GDM. . . . .            | 186 |
| Figura 4.64 – Previsão RN - UE 624 – Método de Treinamento BR GD / GDM. . . . .             | 186 |
| Figura 4.65 – Previsão RN - UE 207 – Método TRAINGDX GD / GDM. . . . .                      | 187 |
| Figura 4.66 – Previsão RN - UE 207 – Método TRAINBR GD / GDM. . . . .                       | 187 |
| Figura 4.67 – Previsão RN - UE 258 – Método TRAINGDX GD / GDM. . . . .                      | 188 |
| Figura 4.68 – Previsão RN - UE 258 – Método TRAINBR GD / GDM. . . . .                       | 188 |
| Figura 4.69 – Previsão RN - UE 624 – Método TRAINGDX GD / GDM (TEMP) . . . . .              | 190 |
| Figura 4.70 – Previsão RN - UE 624 – Método TRAINBR GD / GDM (TEMP) . . . . .               | 190 |
| Figura 4.71 – Previsão RN - UE 207 – Método TRAINGDX GD / GDM (TEMP) . . . . .              | 190 |
| Figura 4.72 – Previsão RN - UE 207 – Método TRAINBR GD / GDM (TEMP) . . . . .               | 191 |
| Figura 4.73 – Previsão RN - UE 258 – Método TRAINGDX GD / GDM (TEMP) . . . . .              | 191 |
| Figura 4.74 – Previsão RN - UE 258 – Método TRAINBR GD / GDM (TEMP) . . . . .               | 191 |
| Figura 4.75 – UE 624 – Erro Absoluto Mensal – Experimento RN c/ vizinhança . . . . .        | 193 |
| Figura 4.76 – UE 207 – Erro Absoluto Mensal – Experimento RN c/ vizinhança . . . . .        | 193 |
| Figura 4.77 – UE 258 – Erro Absoluto Mensal – Experimento RN c/ vizinhança . . . . .        | 194 |
| Figura 4.78 – RN x RN (TEMP) x SAR <i>Panel</i> - UE 624 – Comparativo da Previsão. . . . . | 196 |
| Figura 4.79 – RN x RN (TEMP) x SAR <i>Panel</i> - UE 207 – Comparativo da Previsão. . . . . | 196 |
| Figura 4.80 – RN x RN (TEMP) x SAR <i>Panel</i> - UE 258 – Comparativo da Previsão. . . . . | 196 |

## ÍNDICE DE TABELAS

|  |  |
|--|--|
| Tabela 1.1 – OIE em Ton. Equivalentes de Petróleo - Brasil 1940-2007 – Fonte: EPE (2008). . . 1              |  |
| Tabela 1.2 – OIE – Participação Individual – Fonte: EPE (2008). . . . . 2                                    |  |
| Tabela 2.1 – Análises Espaciais - Tipos de Dados e Modelos - Extraída de FISHER (2006). . 51                 |  |
| Tabela 3.1 – Exemplo de priorização de Variáveis Explicativas. . . . . 82                                    |  |
| Tabela 3.2 – Comprimento do Vetor de dados de Consumo de GN – UE 207 e Vizinhança . . 91                     |  |
| Tabela 3.3 – Casamento Dimensional ( $T$ ) do Conjunto de Variáveis . . . . . 98                             |  |
| Tabela 3.4 – Relação de UE de Interesses e Respectivas Séries Históricas por Variável . . . . 98             |  |
| Tabela 3.5 – Matrizes de Entrada de Dados para RN puramente Temporais. . . . . 110                           |  |
| Tabela 4.1 – Quantidade de Clientes, por Segmento e por CDL – Fonte: Petrobras . . . . . 111                 |  |
| Tabela 4.2 – Consumo Total Anual de GN em $MM^3$ – Fonte: Compagas. . . . . 112                              |  |
| Tabela 4.3 – Consumo Total Anual de GN em $MM^3$ – Fonte: BR Distribuidora . . . . . 113                     |  |
| Tabela 4.4 – Quantidade Total de Clientes – Base Dezembro 2008 – Fonte: Compagas. . . . 115                  |  |
| Tabela 4.5 – Quantidade Total de Clientes – Base Dez – Fonte: BR Distribuidora. . . . . 117                  |  |
| Tabela 4.6 - Configuração das feições da Base de dados espacial à RM de Curitiba. . . . . 118                |  |
| Tabela 4.7 - Configuração das feições da Base de dados espacial à RM de Vitória . . . . . 118                |  |
| Tabela 4.8 – Característica das Variáveis Brutas . . . . . 124   |  |
| Tabela 4.9 – Característica das Variáveis Brutas Extratificadas . . . . . 124                                |  |
| Tabela 4.10 – Quantidade de Correlações . . . . . 125  |  |
| Tabela 4.11 – Distâncias Euclidianas empregadas nas matrizes de proximidade. . . . . 133                     |  |
| Tabela 4.12 – Distribuição de Variáveis Explicativas por UGR – Frequência Mensal. . . . . 134                |  |
| Tabela 4.13 – Variáveis Explicativas Pré-selecionadas para o Modelo – Curitiba Residencial 135               |  |
| Tabela 4.14 – Variáveis Explicativas Pré-selecionadas para o Modelo – Curitiba Comercial . 135               |  |
| Tabela 4.15 – Variáveis Explicativas Pré-selecionadas para o Modelo – Vitória Residencial. 135               |  |
| Tabela 4.16 – Variáveis Explicativas Pré-selecionadas para o Modelo – Vitória Comercial. . 136               |  |
| Tabela 4.17 – Relação Final de UE e Respectiva Vizinhança . . . . . 141                                      |  |
| Tabela 4.18 – Tabela com os Resultados das PCAs para as Variáveis Explicativas . . . . . 142                 |  |
| Tabela 4.19 – Correlação PC $c/$ a Variável de Interesse ( $K = 0, \dots, 12$ ) – Curitiba Residencial . 143 |  |
| Tabela 4.20 – Correlação PC $c/$ a Variável de Interesse ( $K = 0, \dots, 12$ ) – Vitória Residencial . 143  |  |
| Tabela 4.21 – Correlação PC $c/$ a Variável de Interesse ( $K = 0, \dots, 12$ ) – Vitória Comercial. . 144   |  |
| Tabela 4.22 – Correlações Variáveis Explicativas x Interesse ( $K = 0, \dots, 12$ ) – Curitiba Res. . 144    |  |
| Tabela 4.23 – Correlações Variáveis Explicativas x Interesse ( $K = 0, \dots, 12$ ) – Vitória Res. . 145     |  |
| Tabela 4.24 – Correlações Variáveis Explicativas x Interesse ( $K = 0, \dots, 12$ ) – Vitória Com. . 145     |  |
| Tabela 4.25 – Correlações PC e Var. Explicativas x Var. de Interesse - Curitiba Residencial . 146            |  |

|  |     |
|--|-----|
| Tabela 4.26 – Correlações PC e Var. Explicativas x Var. de Interesse - Vitória Residencial .   | 146 |
| Tabela 4.27 – Correlações PC e Var. Explicativas x Var. de Interesse - Vitória Comercial . .   | 146 |
| Tabela 4.28 – Relação Final de Polígonos de Interesses e Respectiva Vizinhança. . . . .        | 147 |
| Tabela 4.29 – Estimação UGR 624 - Segmento Residencial de Curitiba . . . . .                   | 148 |
| Tabela 4.30 – Estimação UGR 638 - Segmento Residencial de Curitiba . . . . .                   | 149 |
| Tabela 4.31 – Estimação UGR 207 - Segmento Comercial de Vitória. . . . .                       | 149 |
| Tabela 4.32 – Estimação UGR 247 - Segmento Residencial de Vitória . . . . .                    | 149 |
| Tabela 4.33 – Estimação UGR 258 - Segmento Residencial de Vitória . . . . .                    | 149 |
| Tabela 4.34 – Resultados do Teste de Estrutura – Entrada de dados opção (A). . . . .           | 156 |
| Tabela 4.35 – Erro Acumulado Total e MAPE por UE . . . . .                                     | 158 |
| Tabela 4.36 – Teste de Hipótese – UE 624 – Previsões por Opção de Entrada de Dados . . .       | 160 |
| Tabela 4.37 – Teste de Hipótese – UE 638 – Previsões por Opção de Entrada de Dados . . .       | 161 |
| Tabela 4.38 – Teste de Hipótese – UE 624 – PSAR c/ Opção (C) de Entrada de Dados. . . .        | 161 |
| Tabela 4.39 – Teste de Hipótese – UE 247. . . . .  | 164 |
| Tabela 4.40 – Teste de Hipótese – UE 247 – Experimentos para a Convergência de $p$ . . . . .   | 165 |
| Tabela 4.41 – Teste de Hipótese – Comparações Entre Formas Funcionais. . . . .                 | 169 |
| Tabela 4.42 – Comparativo dos Resultados de Degradação . . . . .                               | 174 |
| Tabela 4.43 – ARIMA - Erro Acumulado Total e MAPE por UE . . . . .                             | 177 |
| Tabela 4.44 – ARIMA x SAR - Erro Acumulado Total e MAPE por UE. . . . .                        | 178 |
| Tabela 4.45 – Teste de Hipótese – Comparações entre Previsores ARIMA x SAR <i>Panel</i> . . .  | 178 |
| Tabela 4.46 – Alisamento Exponencial - Erro Acumulado Total e MAPE por UGR. . . . .            | 180 |
| Tabela 4.47 – Alisamento x SAR - Erro Acumulado Total e MAPE por UGR . . . . .                 | 180 |
| Tabela 4.48 – Teste de Hipótese – Comparação de Previsores - Alisamento x SAR <i>Panel</i> . . | 181 |
| Tabela 4.49 – RN x SAR – UE 624 - Erro Acumulado Total e MAPE . . . . .                        | 186 |
| Tabela 4.50 – Teste de Hipótese UE 624 – Comparações entre Previsores RN x SAR <i>Panel</i> .  | 187 |
| Tabela 4.51 – RN - UE 624, 207 e 258 - Erro Acumulado Total e MAPE . . . . .                   | 189 |
| Tabela 4.52 – RN - UE 624, 207 e 258 - Erro Acumulado Total e MAPE – RN (TEMP) . . .           | 189 |
| Tabela 4.53 – Erro Acumulado Total e MAPE – Comparação de Métodos de Treinamento .             | 192 |
| Tabela 4.54 – Erro Acumulado Total e MAPE – Comparação de Funções de Adaptação. . .            | 193 |
| Tabela 4.55 – Erro Acumulado Total e MAPE por UE e Condição de Vizinhança. . . . .             | 194 |
| Tabela 4.56 – Teste de Hipótese – Comparação RN C/ x S / Vizinhança . . . . .                  | 195 |
| Tabela 4.57 – RN x SAR - Erro Acumulado Total e MAPE por UE. . . . .                           | 197 |
| Tabela 4.58 – Teste de Hipótese – Comparações entre Previsores RN C/ x S / Vizinhança. .       | 198 |



## LISTA DE SIGLAS

- AE - Área de Estudo
- AIE - Agência Internacional de Energia
- AM – Aprendizado de Máquina
- ANFAVEA - Associação Nacional dos Fabricantes de Veículos Automotores
- ANP - Agência Nacional do Petróleo
- ArcGIS – Ferramenta SIG empregada na tese
- ARIMA – *Autoregressive Integrated Moving Average Model* (Modelo de estrutura autoregressiva integrada à parcela média móvel)
- ARX – *Autoregressive Model with Exogenous Terms* (Modelo de estrutura autoregressiva com parcela regressiva de variáveis exógenas)
- BACEN – Banco Central do Brasil
- BD – Banco de Dados
- BDE - Banco de Dados Espaciais
- BEN – Balanço Energético Nacional
- BP – *Back Propagation*
- CDL – Companhia Distribuidora Local de Gás Natural
- CNI - Confederação Nacional da Indústria
- COPEL - Companhia Paranaense de Energia Elétrica
- CP – Componentes Principais
- CSR - *Complete Spatial Randomness* (aleatoriedade espacial completa)
- EE – Energia Elétrica
- ELCELSA - Centrais Elétricas do Espírito Santo
- EPE – Empresa de Pesquisa Energética
- FAR - *First Order Spatial AR* (Modelo autoregressivo espacial de primeira ordem)
- FIEP – Federação das Indústrias de São Paulo
- GA – *Genetic Algorithm* (algoritmos genéticos)
- GD – Função de adaptação de treino das RN por gradiente descendente
- GDM - Função de adaptação de treino das RN por gradiente descendente com momento
- GLP – Gás Liquefeito de Petróleo
- GN – Gás Natural
- GWR - *Geographically Weighted Regression* (regressão geograficamente ponderadas)
- IBGE – Instituto Brasileiro de Geografia e Estatística

IEA – Agência Internacional de Energia

IGN – Indústria de Gás Natural

IJSN - Instituto de Pesquisa Jones Santos Neves

IPEA – Instituto de Pesquisa Econômica Aplicada

IPPUC - Instituto de Pesquisa e Planejamento Urbano de Curitiba

*Lndet full* – Método de determinação completo para o grid de possíveis valores de  $p$

*Lndet MC* – Método de determinação via algoritmo de Monte Carlo para o grid de possíveis valores de  $p$

LTLF - *Long Term Load Forecasting* (previsão de carga de longo prazo)

MAPE – *Mean Absolute Percent Error* (Erro percentual médio absoluto)

MAUP - *Modifiable Area Unit Problem* (problema da unidade de área modificada)

MTE - Ministério do Trabalho e Emprego

MUB - Mobiliário Urbano Básico

OIE – Oferta Interna de Energia

OLS - *Ordinary Least Squares* (mínimos quadrados ordinários)

ONS - Operador Nacional do Sistema

$p$  – Escalar de maximização dos modelos de abordagem espacial / espaço-temporal

PC – Principal Componente

PCA – *Principal Component Analysis* (análise de componentes principais)

PIB – Produto Interno Bruto

PSAR – *Pooled Panel Data Spatial Autoregressive - Regressive Model* (Modelo autoregressivo - regressivo espacial para dados em painel - forma funcional aninhada)

$pWy$  - Termo autoregressivo dos modelos de abordagem espacial / espaço-temporal

RM - Região Metropolitana

RN – Redes Neurais ou NN (*Neural Networks*)

RSVGM - *Recurrent Support Vector Machines with Genetic Algorithms* (máquinas de vetor suporte com algoritmos genéticos)

SAD - Sistema de Apoio à Decisão

SADE - Sistema de Apoio à Decisão Espacial

SAR - *Spatial Autoregressive-Regressive Model* (Modelo autoregressivo-regressivo espacial).

SAR Panel - *Panel Data Spatial Autoregressive - Regressive Model* (Modelo autoregressivo - regressivo espacial para dados em painel)

SARMA - *Spatial Autoregressive Moving Average Model* (Modelo espacial autoregressivo e média móvel)

SARSFE – *Fixed Effects Panel Data Spatial Autoregressive - Regressive Model* (Modelo autoregressivo - regressivo espacial para dados em painel - forma funcional de efeitos fixos)

SCN - Sistemas de Contas Nacionais

SDM - *Spatial Durbin Model* (Modelo autoregressivo espacial de Durbin)

SEM - *Spatial Error Model* (Modelo autoregressivo espacial de erros)

SEM Panel - *Panel Data Spatial Error Model* (Modelo autoregressivo espacial de erros autocorrelacionados para dados em painel)

SI – Sistemas de Informações

SIG - Sistema de Informações Geográficas

SMA - *Spatial Moving Average Error Model* (Modelo espacial de erros por média móvel)

SpVAR – *Spatial Vector Autoregressive Model* (Modelo espacial de vetores autoregressivos)

STARMA – *Space-time Autoregressive Moving Average* (Modelo espacial autoregressivos e médias móveis)

STLF - *Short Term Load Forecasting* (previsão de carga de curto prazo)

SUR - *Seemingly Unrelated Regressions Model* (Modelo de regressões aparentemente relacionadas)

SVM – *Support Vector Machine* (Máquinas de Vetor Suporte)

TRAINBR - Método de otimização de treinamento de *Levenberg-Marquardt* com regularização bayesiana

TRAINGDM - Método de otimização de treinamento de gradiente descendente com taxa adaptativa de aprendizado

TRAINGDX - Método de otimização de treinamento de gradiente descendente com momento e taxa adaptativa de aprendizado na *backpropagation*

TRAINLM – Método de otimização de treinamento de *Levenberg-Marquardt*

UE - Unidades Espaciais

UGR - Unidade Geográfica de Referência

$W_{ij}$  – Matriz de Contiguidade Espacial

$X\beta$  - Termo regressivo dos modelos de abordagem espacial / espaço-temporal

# 1 – INTRODUÇÃO

Esta tese aborda o desenvolvimento de um modelo de previsão de demanda de Gás Natural considerando o emprego da perspectiva espaço-temporal. Inicialmente, neste capítulo, sintetiza-se a importância e as principais particularidades do tema, bem como as limitações causadas pela não adoção de tal perspectiva na formulação dos modelos empregados nesta atividade. Neste contexto apresenta a (o): motivação, definição do problema, objetivo e uma sucinta descrição da estrutura da tese.

## 1.1 – A IMPORTÂNCIA DO TEMA ESTUDADO

O aumento da necessidade de energia levou a uma diversificação da matriz energética de forma global. Para atender esta necessidade, surgiram cadeias de produção e suprimento de energia mais complexas, bem como, a crescente pressão da sociedade, sobre a questão ambiental, que exige uma melhor definição do aproveitamento dos potenciais energéticos.

Em decorrência disto, a matriz energética nacional, que se desenvolveu historicamente calcada no petróleo e no potencial hídrico, experimentou participações, em forma crescente, de outros energéticos.

Na tabela 1, extraída do Balanço Energético Nacional (BEN)<sup>1</sup>, de 2008 (EPE, 2008), é possível verificar que cerca de 50% da Oferta Interna de Energia (OIE)<sup>2</sup>, ainda encontra-se sobre a forma de petróleo, derivados e GN.

Tabela 1.1 – OIE em Ton. Equivalentes de Petróleo - Brasil 1940-2007 – Fonte: EPE (2008).

| Identificação                     | 10 <sup>3</sup> tep |        |        |        |         |         |         |         |         |         |
|-----------------------------------|---------------------|--------|--------|--------|---------|---------|---------|---------|---------|---------|
|                                   | 1940                | 1950   | 1960   | 1970   | 1980    | 1990    | 2000    | 2005    | 2006    | 2007    |
| Petróleo, Gás Natural e Derivados | 1.522               | 4.280  | 12.668 | 25.420 | 56.485  | 62.085  | 96.999  | 105.079 | 107.261 | 111.438 |
| Carvão Mineral e Derivados        | 1.520               | 1.583  | 1.412  | 2.437  | 5.902   | 9.615   | 13.571  | 13.721  | 13.537  | 14.356  |
| Hidráulica e Eletricidade         | 352                 | 536    | 1.580  | 3.420  | 11.063  | 20.051  | 29.980  | 32.379  | 33.537  | 35.505  |
| Lenha e Carvão Vegetal            | 19.795              | 25.987 | 31.431 | 31.852 | 31.083  | 28.537  | 23.060  | 28.468  | 28.589  | 28.628  |
| Produtos da Cana                  | 563                 | 892    | 2.131  | 3.593  | 9.217   | 18.988  | 20.761  | 30.147  | 32.999  | 37.847  |
| Outras *                          | 0                   | 0      | 0      | 223    | 1.010   | 2.724   | 6.245   | 8.869   | 10.421  | 10.985  |
| Total                             | 23.752              | 33.278 | 49.222 | 66.945 | 114.761 | 142.000 | 190.615 | 218.663 | 226.344 | 238.758 |

(\*) Inclui Outras Fontes Primárias Renováveis e Urânio

Entretanto, é perceptível o crescimento da participação de outros energéticos, em especial os derivados da cana-de-açúcar. Nesta tabela observa-se que estes derivados, na

<sup>1</sup> Elaborado pela Empresa de Pesquisa Energética (EPE), contabiliza a oferta e o consumo de energia no Brasil, bem como dos processos de conversão de produtos energéticos e do comércio exterior. Reúne as séries históricas destas operações, bem como informações sobre reservas e capacidades instaladas.

<sup>2</sup> Grandeza que representa a energia total demandada por um país ou região, para satisfazer as suas necessidades de transformação, distribuição e uso final.

década atual, ultrapassaram a energia hidráulica / eletricidade e a lenha / carvão vegetal, na OIE do país, permanecendo atrás apenas do petróleo e seus derivados.

Comparado a este último, no mesmo período, os derivados da cana obtiveram um crescimento na casa dos 80%, contra cerca de 15% do petróleo e seus derivados.

Especificamente o Gás Natural (GN) foi o energético que apresentou o maior crescimento na OIE, como é possível verificar respectivamente na tabela 2 extraída de EPE (2008) e, na figura 1 extraída de EPE (2008, 2009), onde sua participação aumentou de 3,7%, em 1998, para 9,3%, em 2007 e, 10,3% em 2008.

Tabela 1.2 – OIE – Participação Individual – Fonte: EPE (2008).

| Identificação                 | 1998        | 1999        | 2000        | 2001        | 2002        | 2003        | 2004        | 2005        | 2006        | 2007        | % |
|-------------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|---|
| <b>Energia Não Renovável</b>  | <b>57,7</b> | <b>57,7</b> | <b>59,0</b> | <b>60,7</b> | <b>58,8</b> | <b>56,3</b> | <b>56,2</b> | <b>55,5</b> | <b>55,0</b> | <b>54,1</b> |   |
| Petróleo e Derivados          | 46,5        | 46,2        | 45,5        | 45,4        | 43,0        | 40,1        | 39,1        | 38,7        | 37,8        | 37,4        |   |
| Gás Natural                   | 3,7         | 4,1         | 5,4         | 6,5         | 7,4         | 7,7         | 8,9         | 9,4         | 9,6         | 9,3         |   |
| Carvão Mineral e Derivados    | 6,7         | 6,7         | 7,1         | 6,9         | 6,5         | 6,7         | 6,7         | 6,3         | 6,0         | 6,0         |   |
| Urânio (U308) e Derivados     | 0,8         | 0,7         | 0,9         | 2,0         | 1,9         | 1,8         | 1,5         | 1,2         | 1,6         | 1,4         |   |
| <b>Energia Renovável</b>      | <b>42,3</b> | <b>42,3</b> | <b>41,0</b> | <b>39,3</b> | <b>41,2</b> | <b>43,7</b> | <b>43,8</b> | <b>44,5</b> | <b>45,0</b> | <b>45,9</b> |   |
| Hidráulica e Eletricidade (*) | 15,3        | 15,1        | 15,7        | 13,6        | 14,0        | 14,6        | 14,4        | 14,8        | 14,8        | 14,9        |   |
| Lenha e Carvão Vegetal        | 11,5        | 11,7        | 12,1        | 11,6        | 11,9        | 12,9        | 13,2        | 13,0        | 12,6        | 12,0        |   |
| Derivados da Cana-de-Açúcar   | 13,6        | 13,3        | 10,9        | 11,8        | 12,8        | 13,4        | 13,5        | 13,8        | 14,6        | 15,9        |   |
| Outras Renováveis             | 1,9         | 2,1         | 2,3         | 2,4         | 2,5         | 2,8         | 2,7         | 2,9         | 3,0         | 3,2         |   |
| <b>Total</b>                  | <b>100</b>  | <b>100</b>  | <b>100</b>  | <b>100</b>  | <b>100</b>  | <b>100</b>  | <b>100</b>  | <b>100</b>  | <b>100</b>  | <b>100</b>  |   |

Apesar disto, a participação do GN na matriz energética brasileira ainda é baixa quando comparado ao cenário mundial, como pode ser visto na figura 1.

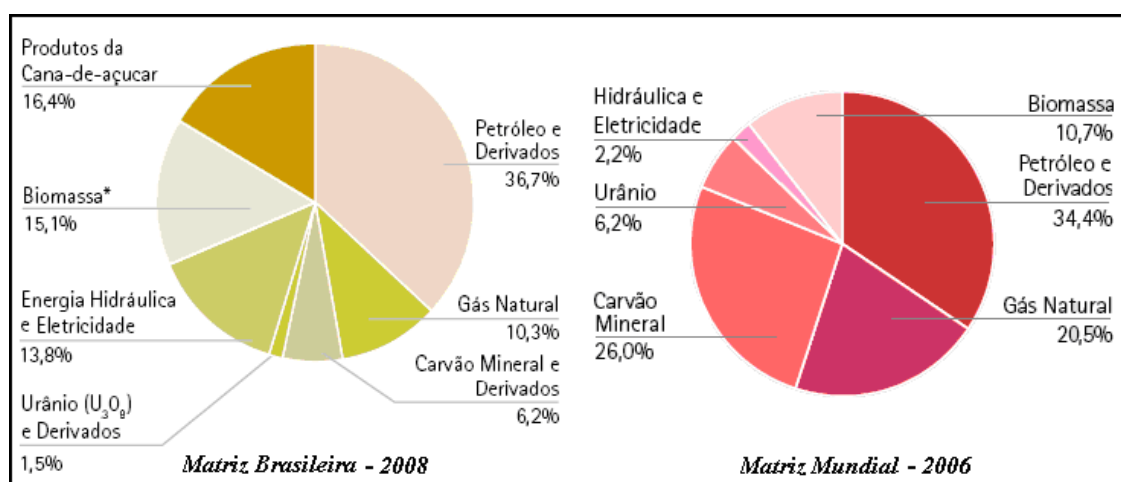


Figura 1.1 – OIE por Fonte – Comparação Brasil e Mundo – Extraída de EPE (2008 / 2009).

Entretanto, este cenário tende a se modificar nas décadas vindouras. A *World Energy Outlook 2000*, da IEA – Agência Internacional de Energia, prevê que a participação do GN na matriz energética nacional aumentará para cerca de 25% em 2020 (SANTOS *et al.*, 2002).

Tomalsquim *et al.* (2007) prevêem a diversificação da matriz energética brasileira, onde, pelas suas projeções, em 2030 serão necessárias quatro fontes para satisfazer 77% do consumo: petróleo e energia hidráulica, cana-de-açúcar e GN.

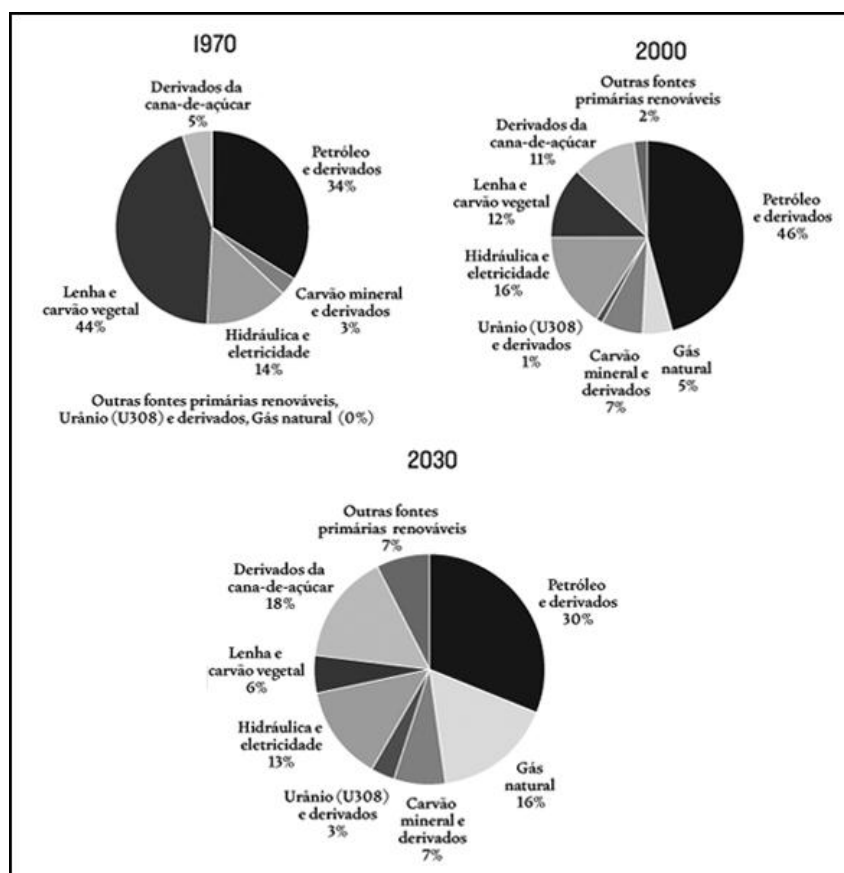


Figura 1.2 – Projeção de OIE por Fonte – Extraída de TOMALSQUIM *et al.* (2007).

Sobre tal oferta ressalta-se que o contínuo investimento na exploração e produção de GN poderá elevar a produção para mais de 250 milhões de m<sup>3</sup> por dia em 2030, com uma taxa de crescimento média de 6,3% ao ano no período da projeção.

Ou seja, nos últimos anos, o GN vem aumentando gradativamente a sua participação na matriz energética e, se mostrado um recurso muito versátil em função de seu custo competitivo, baixo impacto ambiental, crescimento das reservas e suas múltiplas aplicações, como por exemplo, matéria-prima.

Tal crescimento insta a Indústria do GN (IGN), considerando a complexidade da cadeia produtiva e mercados cada vez mais competitivos, ao estabelecimento de planejamentos, estratégicos e operacionais, com menor exposição ao risco. Isto se deve ao volumoso aporte de capital requerido, com retornos financeiros a maior prazo, associado a uma complexa engenharia de financiamento para desenvolver a produção.

Particularmente, o planejamento do segmento de energia requer um conjunto de variáveis amplo e de naturezas distintas, por exemplo: desenvolvimento econômico,

crescimento da demanda, disponibilidade de recursos energéticos, entre outros. Em decorrência, torna-se necessário averiguar e acompanhar o comportamento das relações entre estas variáveis ao longo do tempo. Tal particularidade é imperativa para estabelecer o modelo e os métodos de exploração e manutenção da base de dados.

Historicamente, os avanços da tecnologia de informação nas últimas décadas, disponibilizaram novas estruturas e metodologias que potencializam a descoberta de conhecimento, gestão e decisão a partir de banco de dados. Entre estas, a adoção da informação de natureza espacial apresenta-se como elemento que permite aos gestores ampliar tanto a visão do ambiente onde os empreendimentos encontram-se inseridos, como das relações de troca com a sociedade, constituindo elemento primordial ao entendimento da dinâmica da organização como sistema aberto.

Em particular, no segmento de GN, típica indústria de rede, a disponibilização do energético é fator crítico de sucesso, pois não se materializa o acesso ao produto sem sua rede física. Além disto, o acesso se dá de forma diferenciada em capacidade da rede e necessidade de volume em função do segmento de negócio que se pretende atender.

Neste sentido, a visão espacial da infra-estrutura da rede associada às variáveis de consumo, população, renda, entre outros, amplia a capacidade de análise logística. Em particular, as disponibilizações do consumo e da demanda no espaço celebram o amplo entendimento do problema de engenharia relativo à distribuição da rede, pois possibilitam avaliar de forma integrada o acesso ao produto, o atendimento ao consumidor e a segmentação do mercado (clientes, consumo e posição geográfica).

Em linhas gerais, o problema de engenharia relacionado com a provisão da rede de distribuição passa primeiro pela decisão entre investir ou resguardar, questão básica e central, quando se entende que o capital é um item escasso. Em um segundo momento, a decisão é sobre a melhor forma de remunerar o capital que, no caso da companhia distribuidora local (CDL), constitui-se na definição dos melhores locais e segmentos para desenvolver a rede de distribuição.

Tal definição baseia-se principalmente na projeção de consumo de GN para um tempo  $t$  futuro, de forma que este direcionamento, que inclui a análise dos segmentos de consumo, aporte as melhores condições para o retorno do capital; sendo assim, para otimizar o projeto da rede, é fundamental analisar o equilíbrio entre oferta e demanda.

Em especial, conforme apresentado em 2.4, quando observada à relação entre a atividade de previsão de demanda de energia, inclui-se aqui o GN, e a dimensão espacial, verifica-se que os modelos econômicos pouco se desenvolveram no sentido de

incluir a questão da localização. Excetuando o ramo da economia intitulado economia geográfica, em especial a economia urbana, ramo específico onde se tornou necessário considerar tal informação, a dimensão espacial é pouco empregada pelos economistas.

O estudo dos fenômenos de concentração, aglomeração e vizinhança, os quais favorecem economicamente a formação e o acúmulo de renda e, conseqüentemente do consumo, é pouco explorado. Quando o são, na grande maioria têm abordagem estatística, e de forma estática no tempo, ou seja, *cross section*. Embora tenham efetiva aplicação, não são tratados de forma integrada quanto à relação espaço e tempo, suprimindo assim a percepção espaço temporal do problema.

Coletar, manipular e armazenar dados de recursos energéticos, de variáveis econômicas e de consumo de energia, distribuídos espacialmente, ao longo do tempo, é tarefa importante para inferir o comportamento do consumo futuro e, a respectiva necessidade de oferta de um produto, garantindo assim o acesso ao mesmo.

Ou seja, para a previsão de demanda de GN, objeto desta tese, entende-se que desenvolver modelos de simulação, com emprego da informação espacial e de técnicas correlatas, combinadas com técnicas de análise de séries temporais, apresenta-se como possível solução para constituição de cenários para apoio a decisão.

## **1.2 - DESCRIÇÃO DO PROBLEMA**

O assunto de previsão de demanda no segmento de energia desenvolveu-se, como poderá ser visto em 2.3 e 2.5, prioritariamente pela orientação temporal, principalmente para o curto e o longo prazo. O primeiro aplicado às atividades de operação, e o segundo as atividades relacionadas à economia nacional, principalmente a avaliação de infra-estrutura, onde os energéticos são encarados como *commodities*. Com relação aos modelos, estes empregam prioritariamente as abordagens:

- Econômica do tipo *top down*, na qual se consideram variáveis puramente econômicas e a demanda é projetada para um alto nível de agregação e desmembrada para as camadas inferiores (setores e segmentos); e,
- Extrapolativa, a partir da replicação do comportamento de consumo, pelo entendimento que a série de demanda futura é função direta do consumo recente, sendo no máximo influenciada por questões de ciclo, principalmente os produtivos e de origem ambiental.

É possível afirmar que, na maioria das vezes, tais abordagens reduzem o desenvolvimento do modelo à busca pelo melhor ajustamento matemático das



previsões, relegando a análise do problema de engenharia e das premissas para constituição do modelo a planos secundários.

Como salientado, a infra-estrutura de rede para distribuir o GN é fator crítico de sucesso para a CDL. Estabelecer este ativo constitui um problema de engenharia complexo, pois trata as questões técnicas da infra-estrutura do duto (dimensionamento, equipamentos) à luz de premissas do mercado consumidor, tais como: o segmento e a localização do cliente, os volumes ótimos de pressão e vazão, entre outros.

Isto representa um aspecto técnico forte, pois um ramal da rede atende a particularidades de pressão e vazão específicas; desta forma este tipo de projeto tem que considerar as limitações máximas de carga e pressão possíveis em cada ramal.

Equacionar estes pilares está no cerne deste problema, pois quando um projeto atende um ponto específico, um cliente, por exemplo, as constituintes técnicas levam em consideração a necessidade somente daquele cliente. Entretanto, quando a atividade envolve a espacialização da rede em um ambiente urbano, atendendo clientes com requisitos diferenciados, a opção de se criar vários ramais específicos não é uma condição ótima de engenharia e investimento.

Neste ponto é diferencial para um projeto eficiente poder avaliar no espaço a demanda total por segmentos, fruto da soma das necessidades individuais dos clientes, neste caso as demandas de GN em uma determinada área de interesse.

Para tal, a abordagem exclusivamente temporal não é suficiente, pois é preciso incluir, para a correta determinação dos modelos e posteriores métodos de projeção, a avaliação dos fenômenos e inter-relações associados com a perspectiva espacial.

Sua não observância na composição dos modelos de projeção de demanda, seja pelo emprego de técnicas de modelagem computacional, de análise espacial, ou de ambas, reduz a capacidade de discretização do ambiente de decisão envolvido.

Com isto a projeção da demanda ao longo da superfície do território é comprometida em sua resolução, reduzindo a capacidade de análise, apoio à decisão e determinação das premissas para a solução do problema de engenharia de rede.

### **1.3 – MOTIVAÇÃO E CONTRIBUIÇÃO**

Face à investigação do estado da arte em modelos de previsão aplicados à demanda de GN, explicitados ao longo da revisão bibliográfica, elenca-se como motivação para esta tese, o fato de que tal atividade, assim como para outros energéticos, não se desenvolveu no sentido de incluir a perspectiva espaço-temporal no

desenvolvimento de modelos de previsão; e conseqüentemente, como elemento empregado na tomada de decisão.

Entende-se que a principal contribuição desta tese é desenvolver um modelo de previsão que integra de forma ampla a abordagem temporal e a espacial, a partir do emprego de dados em painel e da existência de dependência espacial estacionada no tempo, como condição necessária e suficiente para definir uma região candidata a um modelo de previsão com cunho espaço-temporal.

Ressalta-se, até esta tese, a aplicação de tais abordagens de forma não integrada; e, no tocante à questão espacial, particularmente aos modelos autoregressivos espaciais, a inexistência de referências do trato de questões temporais associadas à variável de interesse ou mesmo, exemplos que ilustrem como: determinar a composição de variáveis na entrada de dados do modelo e estabelecer um mecanismo de previsão que comprove a eficiência da estimação dos parâmetros em previsões propriamente ditas.

A não observação das condições espaciais, seja na contribuição à alocação do meio físico, neste caso a rede de distribuição, seja como ferramenta de previsão, neste caso um modelo para determinar a demanda, reduz tal perspectiva exclusivamente à representação da previsão do consumo do GN a partir de modelos temporais.

Com isto a questão espacial tangencia somente o uso e a ocupação do solo, sem caracterizar a influência do ambiente no qual o próprio consumo e suas variáveis explicativas, encontram-se inseridos. Fatores estes que, observados na forma de dependência espacial introduzem ao problema de previsão a influência que os eventos de interesse nas regiões vizinhas têm sobre a região de estudo.

Com isto, esta tese ao estudar e desenvolver um modelo de previsão que contempla a evolução espaço-temporal da demanda contribui com a CDL na medida em que melhora:

- A determinação da demanda de GN, pois permite contratar volumes de GN de forma mais apurada;
- A qualidade da atividade de alocação das redes, em especial pela mitigação dos riscos dos investimentos realizados; e,
- O planejamento, fortalecendo a atividade de comercialização de GN e, conseqüentemente a solidificação dos fluxos de capital e caixa da empresa.

Some-se a isto, que o resultado da proposta contribui também com a redução do hiato existente no estudo da IGN, uma vez que se percebem no âmbito dos energéticos,

que o GN ao longo do tempo é o menos estudado mesmo com o aumento de sua importância na matriz energética nacional.

#### **1.4 - OBJETIVO DA TESE**

Desenvolver um modelo de previsão de demanda de GN, por segmento de consumo, para um horizonte de previsão de 12 meses, a partir do emprego de séries temporais de consumo e de dados sócio-econômicos espacialmente distribuídos.

##### **1.4.1 – OBJETIVOS SECUNDÁRIOS**

Como objetivos secundários têm-se:

- i. Avaliar o emprego da informação espacial e das técnicas de análise e projeção correlatas, nas atividades de projeção de demanda de energia, contribuindo assim como arcabouço para o desenvolvimento / aperfeiçoamento de modelos de simulação para o segmento de GN;
- ii. Avaliar a influência da dimensão espacial para a sensibilidade do modelo no que tange: as relações de vizinhança, a dependência espacial e a estabilidade da estrutura espacial, ao longo do tempo
- iii. Avaliar o ganho de qualidade na previsão da demanda de GN, a partir da comparação dos resultados obtidos com modelos espaciais frente a outras abordagens de cunho puramente temporal; e,
- iv. Avaliar e identificar, no âmbito da atividade da distribuição de GN, um conjunto de variáveis públicas, de caráter sócio-econômico, que subsidiem o desenvolvimento de um modelo de previsão de demanda.

##### **1.5 – ESCOPO**

O presente estudo tem os seguintes balizadores a serem observadas:

- a) Foco - previsão de demanda de GN, para apoio a tomada de decisão no âmbito das CDL, ator final da IGN, contextualizada em sua cadeia produtiva e mercados consumidores no Apêndice A;
- b) Horizonte de previsão – previsão de 12 meses visto o reduzido histórico de dados de consumo e, emprego de frequência mensal visto a disponibilidade proporcionada pelas variáveis, de interesse e explicativas;
- c) Segmentos - não serão estudados os segmentos de co-geração, matéria-prima e geração elétrica a partir de GN. O primeiro, pois a co-geração encontra-se diluída na carteira de clientes do segmento comercial, sem contrato específico

de fornecimento não permitindo sua correta apropriação. No caso de matéria-prima, ou na geração de energia elétrica, sua distribuição é pontual e não influenciada por questões relacionadas com a dependência espacial.

## **1.6 - ESTRUTURA DA TESE**

O resultado final desta tese é apresentado em 5 capítulos, sendo o primeiro relativo à Introdução, que versa sobre a importância do tema, o problema, as contribuições, o escopo e os objetivos.

No Capítulo 2, de Fundamentação Teórica, revisa-se a literatura com foco nos objetivos da tese; em especial, nos tópicos 2.4 e 2.5, introduz-se o conceito de dimensão espacial, ao objeto de estudo, com ênfase nas relações espaço-tempo. Para tal, aborda o papel dos modelos e métodos empregados na previsão de GN, a análise de séries temporais e o emprego de aprendizado de máquina (AM); bem como, as particularidades dos modelos baseados em regressões espaço-temporais.

Na Metodologia, Capítulo 3, apresenta-se a descrição da proposta de desenvolvimento do modelo de previsão. Aborda-se a escolha do modelo vis a vis as características do problema e das variáveis disponíveis, descrevendo ainda, as etapas e ferramentas a serem empregadas na solução do problema.

No Capítulo 4, o de Desenvolvimento, descreve-se a execução e a avaliação da previsão. Inicialmente aborda-se a definição do conjunto amostral com base na dimensão espacial (dependência espacial) e temporal (série histórica de consumo), a formação da base de dados, o pré-processamento das séries (dependentes e independentes), a análise de suas interdependências e a redução da dimensão do conjunto de dados. No tocante à estruturação e estimação do modelo, descrevem-se: (a) a validação dos segmentos sensíveis a questão espacial e a priorização das variáveis explicativas; (b) a análise das dependências espacial global e local da variável dependente; (c) a definição da forma funcional e do método de estimação do modelo; (d) a aplicação do modelo espaço-temporal escolhido; (e) a análise dos resultados e dos erros das previsões; e, (f) os testes de estrutura e de comparação com modelos temporais e de redes neurais.

Por último, no Capítulo 5, Conclusões e Recomendações, apresenta-se a análise comparativa dos resultados obtidos no estudo, frente aos objetivos e motivadores à proposição do estudo; bem como, as sugestões de avanços da tese.

## **2 – FUNDAMENTAÇÃO TEÓRICA**

A fundamentação teórica relativa à atividade da modelagem de previsão de demanda de Gás Natural aborda inicialmente, nos tópicos 2.1 à 2.3, o seu emprego como ferramenta para apoio a decisão e planejamento, bem como o tratamento dado ao assunto pelas ciências econômicas, tradicional e espacial.

Em continuação, no tópico 2.4, investiga-se a perspectiva espacial e suas particularidades, com ênfase nas relações espaço-tempo e, como ultima etapa de fundamentação, são apresentadas no tópico 2.5 as contribuições baseadas em: análise de séries temporais, ferramentas de aprendizado de máquina e auto-regressão espacial.

Ressalta-se ao leitor que os tópicos 2.4 e 2.5, junto ao 2.6, relativo à escolha do modelo aplicado ao problema, constituem o arcabouço específico à questão de modelagem espaço-temporal.

### **2.1 – A MODELAGEM DA DEMANDA COMO APOIO À DECISÃO**

No tocante a modelagem como forma de representar a realidade, DOS SANTOS (1994) apresenta uma proposição satisfatória, ao englobar três itens centrais a constituição de um modelo simplificado de representação: o relacionamento entre entidades, o sistema e o modelo mental. Ou seja:

*“A modelagem é o processo de estabelecer relações entre entidades importantes de um sistema, salientando que para cada modelador existe um modelo mental básico, isto é, uma imagem do sistema real a partir da qual um modelo simplificado é construído.”*

SKOVSMOSE (1988) aprofunda estes três itens introduzindo, na estruturação do pensamento, a identificação de padrões e as relações de causa e efeito. Para o autor, conceber a realidade passa por estruturar o pensamento de maneira que os padrões possam ser identificados, e para tal é necessário selecionar elementos importantes da realidade e decidir quais as relações, entre os elementos, são importantes. Quanto a isto cabem as seguintes considerações quanto à relevância do modelo e seu emprego:

- Relações de causa e efeito – as relações entre variáveis explicam, ou podem explicar determinados fenômenos. Estas relações podem ser avaliadas pelo emprego de técnicas específicas, por exemplo, as correlações e regressões;
- Sistema – apesar de poder ter várias dimensões (a empresa ou um departamento como o sistema) e tipos de relação de troca com o meio (fechado ou aberto), demanda relacionamento entre as partes; e

- Capital intelectual - ao questionar a essência, os modelos contribuem para o ganho de capital intelectual, pois encapsulam e transmitem conhecimento.

Para mitigar riscos de interpretação, o termo sistema será empregado para designar o agrupamento de partes que operam juntas e com propósito comum. No caso das estruturas de informações para armazenar, tratar, processar e recuperar dados empregar-se-á o termo Sistemas de Informações (SI).

Tal diferenciação é necessária, pois na identificação do sistema, conforme observam ANDRADE *et al.* (2006), empregam-se ferramentas de modelagem computacional como apoio ao pensamento sistêmico<sup>3</sup>. Por meio destas ferramentas são construídos os micro mundos do sistema real, de maneira que se possa avaliar as relações de causa e efeito no tempo e no espaço.

COUGO (1997) salienta que cada ambiente tem suas características particulares. Por mais que as experiências anteriores sejam suficientes para conhecer um novo ambiente, deve-se estar preparado para suas particularidades, tornando cada modelo semelhante em essência, mas único em detalhes.

As observações de ANDRADE *et al.* (2006) e COUGO (1997) conferem movimento ao ato de modelar, pois precisam ser atualizados e melhorados com o tempo. Na verdade os sistemas são dinâmicos no tempo, e acompanhá-los consiste em manter a representação dos processos que definem seu comportamento.

Percebe-se com isto que os sistemas não podem ser vistos como um conjunto de partes funcionando isoladas umas das outras, mas sim como uma estrutura complexa onde as interações entre elas originam propriedades, que não são explicáveis pela simples soma das partes (VIANNA, 2005).

Apesar do exposto, a modelagem não é um ato sem fim, pois desta forma não iria conferir conhecimento e propósito. A questão fundamental é que para construir um modelo é necessário definir e limitar um escopo. A não ser que o mapa sistêmico<sup>4</sup> seja muito simples, raramente se modela computacionalmente 100% do mapeamento.

Conforme a representação esquemática da visão dos estados do escopo, na figura 2.1, em geral o modelo é menor do que o mapa total, porém alguns podem ser sobrepostos (mesmas variáveis) e outros contemplarem algumas variáveis externas.

---

<sup>3</sup> Conjunto de idéias sistêmicas gerais, bem como teorias e abordagens aplicadas a uma nova estrutura conceitual para desenvolvimento, alternativa à estrutura clássica (pensamento analítico).

<sup>4</sup> Diagrama de apresentação de um modelo contendo os recursos do sistema, o estado dos recursos, as operações que transformam recursos entre estados e a quantificação das relações através de funções matemáticas ou empíricas.

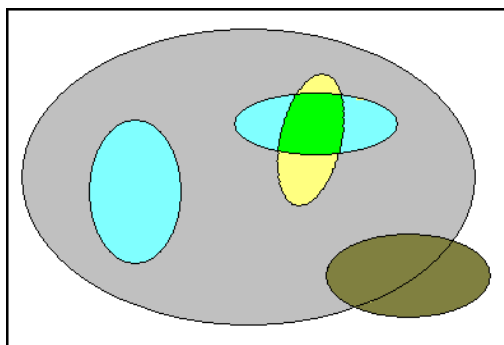


Figura 2.1 – Esquemático dos Estados de Escopo – Adaptada de KIRKWOOD (1997).

Ainda sobre o escopo, ANDRADE *et al.* (2006) cita outros aspectos relevantes quanto ao objetivo, à definição das variáveis e a condição de aprendizado. À luz destas particularidades reafirma-se que o objetivo de desenvolver um modelo de previsão de demanda de GN, observando os aspectos espaço-temporais, é o seu uso como ferramenta de simulação, de modo a permitir que existam opções excludentes ou de comparação para a tomada de decisão.

KIRKWOOD (1998) observa que um aspecto crítico para criar modelos de simulação consiste em modelar adequadamente os processos de decisão, pois envolvem decisões sobre quais dados coletar e como usar e processar as informações, determinando seu fluxo e impacto na organização. Desta forma entende-se que o processo de decisão é o elo entre as informações e seus fluxos junto à organização.

Entretanto, mesmo com aplicação crescente de SI nas decisões, muitas destas continuam sendo feitas pelo homem; por isto a necessidade de modelar a decisão humana como se fosse um modelo real constituído sobre a forma de um processo. Isto é complexo, pois existem decisões que apresentam alternativas e a escolha por uma delas envolve situações que podem gerar diferentes conseqüências ou resultados

Neste ponto, a informação assume dois papéis preponderantes na decisão: reduzir as incertezas sobre o assunto ou cenário que se vai analisar e, a partir destas análises produzir conhecimento que pode ser transferido e empregado em outras situações.

Como a distribuição do GN é uma atividade que acontece no espaço, a não incorporação desta perspectiva ao processo decisório subtrai capacidade do modelo em fornecer alternativas, uma vez que as variáveis envolvidas não mais reduzem incertezas e o objetivo do modelo é comprometido. Além de que, o espaço e o tempo possuem dinâmica própria tornando-se assim um aspecto crítico para que o modelo de simulação seja útil, visto a necessidade de definir como coletar, relacionar e processar as informações que impactam a demanda para a operação da CDL.

### 2.1.1 – MODELOS PARA PROJEÇÕES DE DEMANDA E CENÁRIOS DE ENERGIA

Inicialmente, cita-se JANNUZI e SWISHER (1997) como exemplo de modelo voltado à atividade de previsão de demanda no segmento de energia, particularmente na energia elétrica (EE). Com foco no detalhamento e qualidade das informações, os autores abordam a importância de empregar projeções de demanda de EE e carga no Plano Integrado de Recursos Energéticos, visto sua ajuda para avaliar a necessidade de novos recursos e, a priorização dos investimentos por usos finais e setores.

Na verdade esta abordagem é empregada pelo autor para introduzir o que ele intitula “*A Estrutura Tecnológica das Projeções e dos Cenários da Demanda de Energia*” e que trata da apresentação dos modelos, no sentido de processo estruturado para apoio a tomada de decisão, no âmbito do setor elétrico.

Cabe ressaltar que o modelo mental aplicado aos setores de EE e de GN tem grande similaridade quando se observa estes segmentos pelos prismas de: tipos de mercado consumidor, infra-estrutura de indústria de rede, migração inter-segmentos (substituição de energéticos), entre outros aspectos.

JANNUZI e SWISHER (1997) primeiramente identificam dois principais procedimentos de uso corrente pelas agências de planejamento e companhias de eletricidade: modelos econométricos e modelos técnico-econômicos. O primeiro procura representar o consumo de energia através de equações, empregando dados do passado para estimação estatística. Geralmente estudam uma classe completa e homogênea de consumidores não considerando a sua estrutura tecnológica e o uso final da energia. Serão abordados com mais profundidade em tópico futuro.

Já os modelos técnico-econômicos, ou de projeção de usos finais, são mais detalhados. A demanda de energia para cada atividade é função dos fatores de intensidade (parte relativa à eficiência energética) e o nível de atividade (parte relativa à necessidade de energia dependente da população, da renda e da produção econômica).

Os modelos técnico-econômicos empregam uma estratégia de análises intitulada *bottom-up*, ou seja, estimam o consumo dos usos finais na estrutura dos segmentos. E os modelos econométricos, na linha contrária, geralmente adotam uma estratégia de análise *top-down*, ou seja, estimam os valores em função de variáveis macro-econômicas e depois “rateiam” os resultados para baixo até o nível dos segmentos.

JANNUZI e SWISHER (1997) citam também o emprego de cenários de referência para compor modelos de decisão para demanda de energia:



*“Os modelos de usos-finais (bottom-up) não são completos, por isso é necessário se apoiar em informações da estrutura e crescimento econômico, ou seja, num modelo macro-econômico top-down. Frequentemente, começa-se com um cenário socioeconômico que fornece projeções de população... podemos trabalhar como mais de um cenário, por exemplo, um cenário de alto crescimento e outro de baixo crescimento.”*

Mesmo sem estabelecer uma estrutura formal de tomada de decisão, esta proposição guarda relação com as peculiaridades apresentadas, principalmente por empregar modelos de estratégia pré-concebida (*bottom up e top-down*) para constituir cenários de apoio à decisão a partir da seleção de variáveis e formulações matemáticas para representar o problema

Ainda sobre este assunto, CAIO e BERMAN (1998) citam os cenários como base fundamental para exercício de planejamento, na medida em que congregam um conjunto coerente e plausível, não necessariamente exato, de acontecimentos endógenos e exógenos, aos quais estão associados determinados atores e uma ordem no tempo.

Os autores salientam que, face às fragilidades dos modelos econométricos na decomposição quantitativa do peso dos principais atores, supondo simplesmente que as variáveis se mantenham inalteradas, surgem os métodos técnico-econômicos, com o objetivo de melhor compreender os fatores determinantes na demanda de energia.

Entre a maturação dos modelos técnico-econômicos e o continuado emprego dos econométricos, surgem os modelos de equilíbrio como solução intermediária a negação absoluta da validade deste último como instrumento de previsão e, da afirmação alusiva a importância dos modelos de segregação para atingir os objetivos da previsão.

Na verdade existem vários modelos para prever demanda de energia, que empregam na essência abordagens variantes de modelos econômicos ou técnico-econômicos, destacando-se entre outros: MEDEE<sup>5</sup>, MESSAGE<sup>6</sup>, MARKAL<sup>7</sup> e os baseados em insumo-produto<sup>8</sup>.

---

<sup>5</sup> Château, B., L.apillonne, B. La prévision à long terme de la demande d'énergie: Essai de renouvellement des méthodes, Institut Economique et Juridique de l'Energie, Université de Grenoble, France (1977).

<sup>6</sup> Schrattenholzer, L. THE ENERGY SUPPLY MODEL MESSAGE - International Institute for Applied Systems Analysis, Laxenburg, Austria. (1981).

<sup>7</sup> Loulou, R. Goldstein, G. Noble, K. Standart MARKAL User Manual”, Volume-I / II / III. <http://www.etsap.org/documentation.asp> (2009)

<sup>8</sup> Leontief, W. Input-Output Economics. 2nd ed., New York: Oxford University Press, 1986.

### 2.1.2 – MODELOS PARA PROJEÇÕES DE DEMANDA E CENÁRIOS DE GN

A modelagem para a previsão da demanda de GN, salvo pequenas variações, segue a linha do equilíbrio de mercado. Esta abordagem é intitulada complementar, no sentido matemático do termo, e é encarada como um problema de otimização, no qual o equilíbrio se dá a partir da modelagem da participação de cada ator no mercado.

Isto acontece por que o GN é comercializado como *commodity*, com contratos de fornecimento de longo prazo e com isto, torna-se necessário estabelecer a sua participação em uma cadeia maior de suprimento. Na figura 2.2, apresenta-se um exemplo da dinâmica do equilíbrio, conforme proposição de EGGING *et al.* (2008).

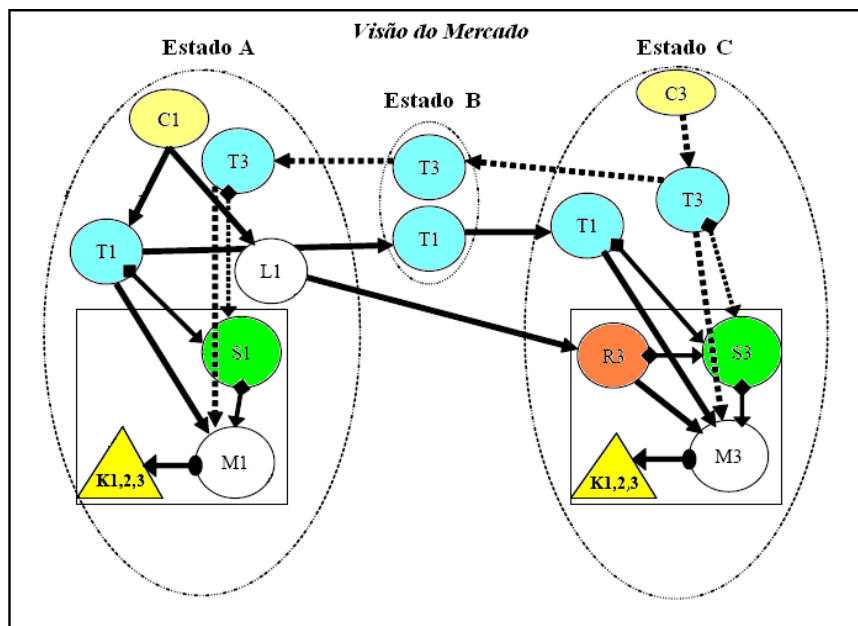


Figura 2.2 – Dinâmica do Mercado Mundial de GN – Adaptada de EGGING *et al.* (2008).

Nesta figura, identificam-se os seguintes atores:

- Operando na produção: os produtores C1 e C3, os transportadores T1 e T3, operando entre produção, transporte e consumo e, o liquefador L1.
- Operando nos pontos de consumo: o regaseificador R3, os armazenadores S1 e S3, os comercializadores M1 e M3 e, por último, os consumidores K1, K2 e K3; respectivamente, setores residencial / comercial, industrial e elétrico.

Esta característica peculiar torna o mercado de suprimento de GN naturalmente integrado, e sua modelagem mais complexa por incorporar peculiaridades de transações internacionais, como por exemplo, arbitragem, cabotagem, frete, entre outros.

Em outro exemplo da abordagem de equilíbrio, neste caso de suprimento entre regiões ou organizações, AL-FATTAH e STARTZMAN (2000) propõem modelar as

tendências de produção por modelos de previsão por ciclos, visto que os países industrializados apresentam ciclos de produção adicionais e complementares.

Os autores ressaltam ainda que o modelo convencional de Hubbert,<sup>9</sup> com um ciclo de produção completo provou ser útil para prever óleo-produção, mas por não considerar as flutuações na produção de GN, não é apropriado para prever suas tendências de produção.

Em outro exemplo na abordagem de equilíbrio, HÖFFLER e KÜBLER (2007) propõem aplicar uma abordagem *top-down* para prever a demanda de GN flexível para 2030 no nordeste da Europa, adotando a demanda e o suprimento de GN como variáveis exógenas e, derivando os resultados dos balanços de demanda e de suprimento de GN, de produção e importação, para extrapolação da infra-estrutura atual.

Sobre uma perspectiva *top-down*, HUNTINGTON (2007) desenvolveu um modelo de consumo de GN industrial e combustível fóssil, no qual as variáveis, excetuando a porcentagem de capacidade utilizada, foram expressas em logaritmos.

O ponto forte do modelo é a capacidade de comparar a demanda em função do nível de substituição inter-combustíveis. As transformações logarítmicas permitiram interpretar os coeficientes de produção como uma elasticidade, com base nas variações na porcentagem da demanda de energia causadas pelas mudanças de porcentagem das variáveis independentes.

O modelo emprega a série histórica do consumo e de preços do GN industrial e do óleo combustível destilado, a estrutura de produção, o aquecimento diário e a capacidade de utilização. Uma vez que estas variáveis são estacionárias, a análise adotou relacionamento auto-regressivo com atraso distribuído, conforme:

$$dY_t = \beta_0 + \beta_1(dX_t) + (\beta_1 + \beta_2)X_{t-1} - (1 - \beta_3)Y_{t-1} + \mu_t \quad (2.1)$$

Onde,

$Y$  e  $dY$  - nível de consumo e substituição de gás natural industrial;

$X$  e  $dX$  - nível de troca no conjunto de variáveis explicativas; e,

$t$  - índice que indica o ano.

Em outro exemplo, PERSAUD e KUMAR (2001) propõem um o modelo econométrico com abordagem *top-down* para prever o suprimento de GN e óleo, com

---

<sup>9</sup> Nos anos 50, o geólogo M. K. Hubbert desenvolveu o conceito que a produção do petróleo é dividida em três fases: rápido crescimento, pico de produção e declínio gradual; seguindo uma curva em forma de sino. Acompanhando a idéia da evolução da produção, existe a hipótese de que a curva de evolução das reservas siga a mesma forma de sino. Com isto durante a primeira fase, as novas descobertas tendem a compensar a diminuição das reservas em produção.

base na gestão de recursos de naturais e com grande enfoque nos fluxos de caixa e de investimentos, por serem itens determinantes para estas indústrias.

Por último cita-se SÁNCHEZ-ÚBEDA e BERZOSA (2007), que propõem um modelo de previsão para o médio prazo aplicando algoritmos para descobrir padrões em séries históricas e permitir sua extrapolação para o futuro. Este modelo emprega decomposição estatística para reconhecer padrões dentro de vários perfis históricos de consumo industrial de GN.

A idéia fundamental é fazer a separação dos diferentes componentes determinísticos (tendência e sazonalidade) e isolar os sem padrões. Porém, em vez de remover previamente o efeito destes componentes, para modelar por transformações, propõe fazê-lo diretamente. Um ponto interessante é que a abordagem vai do curto para o médio prazo, pois é possível verificar no modelo o uso de variáveis normalmente empregadas na previsão de energia de curto prazo.

## **2.2 – A ÓTICA DA DEMANDA PELO PLANEJAMENTO**

Um aspecto de grande relevância na formulação dos modelos é o tempo, pois tem variada finalidade, visto o objetivo e as nuances do estudo. Por exemplo, na perspectiva econômica tem papel fundamental na determinação de limites, onde se realizam as investigações das relações entre as variáveis, por exemplo, a elasticidade ao longo do tempo. No enfoque sistêmico, passa a ter a função de marcar os acontecimentos ou as evoluções, como a de uma organização.

No tocante à dimensão espacial é preponderante na identificação de relacionamento entre as informações espaciais, suas modificações de representação, em função da dinâmica do espaço e, da cronologia destes acontecimentos (ordem de fluxo).

Na previsão de demanda, como em qualquer atividade empresarial, o tempo é fundamental para constituir o planejamento, que é um processo e uma ferramenta de gestão, com objetivo de orientar a organização ao alcance das metas, de forma eficiente e eficaz, otimizando capital, recursos humanos e meios produtivos.

O tempo atua no planejamento como elemento delineador dos planos da organização guardando relação direta com a natureza hierárquica e podem ser assim relacionados (CHIAVENATO e SAPIRO, 2009):

- Planejamento Estratégico – é o de mais alto escalão, lida com a missão e a formulação das estratégias da organização. Relaciona-se com os objetivos de longo prazo e busca inserir a organização no ambiente de mercado;

- Planejamento Tático – é executivo e mais detalhado. Alinha o nível operacional a estratégia, pelo desdobramento do plano estratégico em planos que permeiam toda a organização. Busca a eficiência na gestão dos recursos.
- Planejamento Operacional – detalha os recursos de cada atividade, com o objetivo da gestão do curto prazo. Busca otimizar as condições de operação.

Em particular para o planejamento a constituição de cenários de decisão é de fundamental importância, sendo que em algumas linhas de pensamento está fortemente associada ao nível estratégico.

Conforme KIRKWOOD (1997) a constituição de um cenário segue um processo pré-determinado<sup>10</sup>, e é típico de longo prazo por lidar com variáveis e incertezas de difícil modelagem, que não têm como ser tratadas visando-se o plano diário. Por exemplo, substituir um energético como insumo de produção não busca no primeiro momento otimizar a produção do dia ou do mês seguinte, e sim a de anos à frente; bem como, têm uma vasta gama de desdobramentos associados a esta substituição.

Segundo GODET (2000) um cenário é a descrição de uma situação futura e, do curso dos eventos que permitem um progresso da situação original à situação desejada.

São identificadas duas categorias principais de cenários:

- Exploratório – os futuros prováveis são baseados em fatos no passado e em tendências presentes; e,
- Antecipativo ou normativo - construído com visões alternativas do futuro, podem ser desejáveis ou, pelo contrário, temidos.

Encontra-se ainda na literatura citação a outros tipos de cenários. Por exemplo, MAKRIDAKIS *et al.* (1998) subdividem os cenários em métodos quantitativos<sup>11</sup> e qualitativos<sup>12</sup>; classificação esta similar à empregada por ALCAMO (2001).

Estas divergências advêm de diferentes interpretações sobre a aplicabilidade e prestação dos cenários.

---

<sup>10</sup> O processo de cenarização conforme KIRKWOOD (1997): (a) Definir variáveis chaves à decisão; (b) Classificar os maiores atores envolvidos na decisão; (c) Listar tendências que afetam a decisão; (d) Especificar incertezas; (e) Construir dois cenários extremos; (f) Consistir e avaliar as possibilidades e plausibilidade de acontecer; (g) Avaliar a influência e o impacto da ação do maior ator no cenário; e, (h) Criar até quatro cenários alternativos que cubram um amplo espectro de possíveis condições futuras, e que sejam inseridos no processo de planejamento estratégico.

<sup>11</sup> É aplicado quando existe quantidade suficiente de dados históricos, a informação pode ser quantificada numericamente e é possível identificar padrões de comportamento, que podem ocorrer no futuro. Condição esta chamada de “pressuposto da continuidade” que aparece em todos os modelos quantitativos e em alguns qualitativos.

<sup>12</sup> Podem ser empregados com uma pequena quantidade de dados, em contrapartida necessitam de bom conhecimento sobre o problema em análise, a partir da incorporação da opinião de especialistas ou experts sobre o assunto.

GRUMBACH (2000), que advoga o emprego dos cenários no âmbito do plano estratégico, vê seu emprego no suporte à previsão prospectiva. Tal linha tem como premissa que o futuro pode ser alterado pela ação do homem e por isto, está em constante construção, não podendo ser previsto por simples extrapolação, analogia, casualidade ou presságios, como os métodos empregados na previsão clássica.

Apesar das colocações aparentarem certa inutilidade dos métodos matemáticos e estatísticos tradicionais à constituição dos cenários, ao se observar o próprio trabalho de GRUMBACH (2010), verifica-se que na fase de diagnóstico estratégico cita-se o emprego de ferramentas estatísticas para mineração de dados (*data mining*).

Esta tese não tem o objetivo de avaliar a teoria de cenários, todavia, é clara sua importância no apoio à tomada de decisão; e, por conseguinte, na forma como os modelos são úteis para a organização. Por este prisma, a visão prospectiva reduz a importância desta ferramenta, como útil somente às análises de longo prazo.

Não obstante, tal abordagem cria um disparate com a evolução das tecnologias de informações e das técnicas de previsão baseadas em inteligência computacional. Não olhar a questão de cenários por este prisma é ignorar o aumento exponencial na capacidade de processamento e o desenvolvimento de ferramentas de previsão pelas mais variadas tendências: *Neural Networks* (NN), a partir desta intituladas RN (Redes Neurais), *Support Vector Machine* (SVM), reconhecimento de padrões, entre outros.

Em função do exposto e para evitar dissonância cognitiva, nesta tese a visão de planejamento segue a tipologia e o padrão hierárquico concebidos em CHIAVENATO e SAPIRO (2009) e os cenários seguem a proposição de MAKRIDAKIS *et al.* (1998) por entender que:

- Os métodos quantitativos permitem identificar padrões comportamentais nas variáveis, bem como se encontram em constante evolução técnica;
- Uma vez identificadas, as dinâmicas destas alterações podem indicar as necessidades de melhorias nos modelos de decisão e na modelagem dos SI empregados para isto;
- A modelagem da demanda fornece cenários que além de apoiar a decisão, são insumos à construção de planos: operacionais, de mercado, de engenharia, entre outros; e que tais planos estão no horizonte limite desta tese; e,
- Uma vez que existem níveis específicos de planejamento, existem decisões específicas englobando modelos e ferramentas voltados para cada nível de decisão e que são suportados por cenários diferenciados.

### **2.2.1 – A ABORDAGEM DO PLANEJAMENTO E A DEMANDA DE ENERGIA**

O exemplo mais claro da interação entre níveis de planejamento e demanda é o que trata das previsões de carga, de EE e de GN.

No caso da EE, via de regra, no curto prazo ou STLF (*Short Term Load Forecasting*), cuida-se da previsão da carga na operação, ou seja, o suprimento horário, diário, em horizontes extremos de seis meses. O longo prazo ou LTLF (*Long Term Load Forecasting*) está mais orientado a estruturação do setor, ou seja, prever a demanda futura visando prover capacidade à infra-estrutura, neste caso à rede.

FU e NGUYEN (2003) ressaltam que, apesar da previsão da demanda de pico de carga e de energia para os anos à frente ser vital ao segmento, percebe-se que os modelos para este horizonte não se desenvolveram, na mesma quantidade e profundidade do que os para o curto prazo.

No segmento de EE podem ser citados como exemplo de modelos os trabalhos de PPALEXOPOULOS e HESTERBERG (1990), BAKIRTZIS *et al.* (1996), KHOTANZAD *et al.* (1998), YOO e PIMMEL (1999), YANG *et al.* (2001), SEETHA e SARAVANAN (2007), MISHRA e PATRA (2008a e 2008b). Na previsão de demanda de GN de curto prazo, podem ser citados LIU *et al.* (2004a), LIU *et al.* (2004b) e LAURET *et al.* (2008).

No caso do GN é mais comum encontrar abordagens de longo prazo com horizonte anual, provavelmente em função da sua visão de *commodity*. Como exemplo, agora pelo prisma do planejamento, cabe lembrar a abordagem complementar e de equilíbrio de suprimento de EGGING *et al.* (2008) e, a proposição para previsão da demanda por GN flexível de HÖFFLER e KÜBLER (2007).

No setor de GN nota-se ainda a previsão de médio prazo, como por exemplo, SÁNCHEZ-ÚBEDA e BERZOSA (2007) que propõe um modelo para prever a demanda no prazo de 1 a 5 anos, e LYNESS (1984) que descreve a estrutura de previsão de demanda da *British Gás Company* baseada em ciclos regulares de: 24 h, semanal, anual e de 1 a 5 anos.

### **2.3 - A VISÃO DA DEMANDA PELA ECONOMIA**

O pressuposto econômico decorre da capacidade de modelar, no sentido de abstrair a realidade em forma de modelo, constituindo um conjunto de regras que expliquem um fenômeno econômico, independente de casualidades; ou seja, o modelo será melhor tão quanto mais completa for a abstração da realidade.

De certo o tratamento dado aos fenômenos econômicos observados e as leis deles resultantes não alcançam o rigor numérico típico das ciências experimentais. Porém, o autor não quer dizer com isto que as leis da economia sejam destituídas de fundamento, mas que não se pode equiparar a precisão de tais leis com as alcançadas pelas experimentais.

Isto indica que o tratamento econômico antes de tudo deve ser entendido como de caráter hipotético e estatístico, válido dentro dos limites das hipóteses que as simplifica, e com um número estatisticamente significativo de observações que a validem.

Complementarmente, PINDYCK e RUBINFELD (2002) colocam que a economia como qualquer ciência preocupa-se com a explicação e a previsão de fenômenos observados, e que para isto baseiam-se em teorias. Os autores ressaltam que as teorias podem ser explicadas por técnicas estatísticas e/ou econométricas, que apresentam caráter dinâmico, ou seja, o contínuo teste das observações pode modificá-las, ou mesmo descartá-las.

Tal proposição vai ao encontro dos pressupostos do enfoque sistêmico, quando se coloca que os modelos são sistemas dinâmicos no tempo e que, o acompanhamento desta dinâmica é inerente às organizações e ao aprendizado.

No tocante à econometria, pode-se entender que é um procedimento especial, embutido na análise econômica, que combinando à teórica clássica, que no geral não fornece medidas de relação entre as variáveis, com pressupostos matemáticos e estatísticos, permite medir os fenômenos econômicos por meio de análise de variáveis.

Apesar dos recentes desenvolvimentos de ferramentas para emprego nos estudos econométricos, a de uso mais comum é a análise de regressão. Especialmente, ressalta-se o desenvolvimento do conceito de co-integração, que gerou uma maior atenção à modelagem das relações macro-econômicas e, de suas dinâmicas, embora continuem com foco em aspectos particulares da economia (VERBEEK, 2004).

O objetivo da econometria é especificar e quantificar as relações entre variáveis, formulando modelos, norteados pela teoria econômica, no sentido de estabelecer um arranjo que maximize os resultados, minimizando os desvios. Os parâmetros desconhecidos no arranjo são estimados a partir de amostra dos dados disponíveis.

Neste sentido, Verbeek (VERBEEK, 2004) aponta quatro grandes grupos de relações a serem observadas nos modelos econométricos:



- Presente e passado - tipicamente intituladas séries temporais, geralmente são desprovidos de teoria econômica. Tem o objetivo principal de apoiar a previsão de valores futuros mitigando a incerteza ou volatilidade;
- Quantidades econômicas por período de tempo – fornecem informações agregadas de como as quantidades econômicas flutuam, em relação umas com as outras, com o passar do tempo;
- Relações *cross sectional* - diferentes variáveis observadas em diferentes unidades, do espaço e no tempo, podem explicar por que estas são diferentes ou comportam-se de forma diferente; e,
- Relações *cross sectional* sobre um tempo maior (dados em painel) - descrevem simultaneamente as diferenças entre indivíduos diferentes e as diferenças de comportamento de um indivíduo com o passar do tempo.

De certo, no contexto econômico, independente da demanda ser caracterizada via mecanismos de equilíbrio ou técnicas econométricas, é amplo o emprego de modelos econômicos na previsão de energia, principalmente pelas vertentes de substituição de energéticos e de elasticidade do preço e da renda. Entre outros, podem ser citados como exemplos, os trabalhos de MADDALA *et al.* (1997), SCHMIDT e LIMA (2004) e MALLIARIS e MALLIARIS (2005).

Entretanto, independente da vertente, ressalta-se a importância de investigar os fenômenos e as conseqüentes variáveis e relações que explicam o comportamento do consumo de um bem, como atividade inerente ao desenvolvimento de qualquer modelo de previsão.

Isto posto pela percepção de que alguns destes trabalhos podem ter sua validade questionada, em função de privilegiarem a busca da acuracidade estatística, no método de previsão, em detrimento à análise do conjunto de variáveis explicativas.

#### **2.4 - A VISÃO GEOGRÁFICA DA DEMANDA**

Apesar de não existir uma ciência específica, em geral, a literatura emprega os termos economia espacial e geografia econômica como o estudo dos efeitos espaciais aplicados ao contexto desta atividade.

KRUGMAN *et al.* (2002) são mais específicos ao ressaltar que a economia espacial está mais ligada as teorias de localização, como por exemplo, as teorias de

aluguel e uso da terra de Von Thunen<sup>13</sup> e da região central de Christaller<sup>14</sup>; enquanto a geografia econômica foca o estudo sobre onde e o porquê ocorre a atividade econômica.

Já RICHARDSON (1975) adota uma linha mais centrada no estudo do efeito da distância sobre o preço e o transporte, abordando com mais ênfase o equilíbrio espacial entre mercados geograficamente separados. Percebe-se com isto, de uma forma mais clara, que o preço de um bem não é homogêneo em diferentes pontos do espaço:

“Uma vez que os mercados estejam separados espacialmente e, as curvas de oferta e demanda locais se cruzam a um nível de preços mais elevados em um mercado do que nos demais, o preço mais elevado tende a persistir enquanto os produtores dos outros mercados não considerarem compensador transportar seus bens para o local de preços mais altos... Na outra mão o equilíbrio permanece se os consumidores acreditarem não ser vantajoso adquirir o bem em outro local.”

Independente destas nuances, os estudos sobre localização, como por exemplo, os de localização industrial, ofereceram bases para entender o espaço como forma contínua, em vez de regiões pré-estabelecidas, e contribuem ou explicam em parte, a função de potencial de mercado<sup>15</sup>, empregada pelos geógrafos para descrever as vantagens entre diferentes localizações ou para prever tendências na localização real.

No caso do GN o espaço é naturalmente um contínuo para exploração, visto que o tipo de concorrência para a distribuição do GN é monopolista. Sendo assim, quando compradores e vendedores estão regularmente dispersos, não existindo concorrência próxima, o produtor tende a estabelecer o preço e maximizar o lucro e, a região de vendas seria circular e a mais lucrativa do mercado.

Porém, outros produtores tendem a entrar nos segmentos consumidores enquanto existir algum padrão regional que permita maior adensamento de vendedores, com a concorrência tendendo a um lucro pelo menos igual a zero, no momento da entrada. Com isto, o livre ingresso, neste caso dos energéticos substitutos, comprimiria o mercado circular ideal do GN para a formação hexagonal com as regiões de mercado cobrindo todo espaço, como uma colméia, até que não existam locais sem atendimento.

---

<sup>13</sup> Von Thunen, J. H. Der isoliert staat in beziehung auf landschaft und natinalokonomie. Hamburgo. 1826.

<sup>14</sup> Christaller, W. Central places in southern Germany. Jena, Alemanha, 1933.

<sup>15</sup> Mede o potencial de algum local  $i$  como soma ponderada do poder de compra de todos os outros locais  $j$ , como função decrescentes da distância  $M_i = \sum_j \frac{1}{D_{ij}} P_{jt}$ , onde  $D_{ij}$  é a distância de  $i$  e  $j$  e  $P_{jt}$  é o poder de compra em  $j$ .

Como a concorrência monopolista não é a que predomina em um mercado espacial, pois existe a inércia locacional fruto dos altos custos de mobilidade e de infraestrutura, é possível um equilíbrio em longo prazo, com um número menor de vendedores e um lucro maior do que zero na região atendida (RICHARDSON, 1975).

Isto corrobora o que se apresentou como substituição de bens, pois é possível que na substituição em uma região já atendida, o vetor de crescimento não seja zero, anulando-se em preço os energéticos e constituindo regiões de atendimento bem delimitadas quanto ao padrão de uso. Por exemplo, o GN na cocção em densos aglomerados urbanos e a manutenção do GLP em regiões periféricas.

#### **2.4.1 – A ECONOMETRIA ESPACIAL**

O termo econometria espacial foi cunhado por Jean Paelink nos anos 70 para designar um campo da ciência regional orientado a estimação e aos testes de problemas encontrados na implementação de modelos econométricos multiregionais. Desde a sua proposição, o termo encontrou aceitação, entretanto, não é claro o que diferencia a econometria espacial, da econometria tradicional.

ANSELIN (1999a) propõe que a questão espacial seja vista como subárea da econometria, lidando com aspectos específicos de dados e modelos espaciais aplicados às ciências regionais, com foco no tratamento da interação espacial (autocorrelação espacial) e da estrutura espacial (heterogeneidade espacial), em modelos de regressão.

Apesar de não existir um tratado que as limitem, várias questões metodológicas são igualmente abordadas por ambas; bem como, o crescente emprego da informação espacial ampliou o uso da dimensão espacial em um número muito maior de campos do que os previstos originalmente por Anselin.

Para exemplo, citam-se análises nas áreas de finanças (VAONA, 2008), de crescimento econômico (MELICIANI e PERACCHI, 2006), de mercado imobiliário (ANSELIN e LOZANO-GRACIA, 2008), entre outros.

De forma mais direta, LESAGE (1999) distingue que a diferença entre a econometria espacial e a tradicional é a dependência espacial entre as observações e a heterogeneidade espacial que ocorre nas relações modeladas.

Tanto LESAGE (1999) quanto ANSELIN (1999a) concordam que no contexto de regressão, base da econometria, os efeitos espaciais relacionam-se com a dependência espacial e a heterogeneidade espacial; deve-se ressaltar que assim como observado por KRUGMAN *et al.*, (2002) e RICHARDSON (1975), LESAGE (1999) afirma que os

econometristas tradicionais têm ignorado em grande medida estas duas questões, por violarem os pressupostos de *Gauss-Markov* aplicados na modelagem de regressão.

#### 2.4.1.1 – A DEPENDÊNCIA ESPACIAL

No âmbito das ciências regionais a dependência espacial é motivada pela primeira lei da geografia, definida por Waldo Tobler, como:

“Todas as coisas são parecidas, mas coisas mais próximas se parecem mais do que coisas distantes”.

Em caráter geral é possível afirmar que a maior parte dos fenômenos, sejam estes naturais ou sociais, apresentam entre si uma relação que depende da distância; e que existe dependência espacial entre amostras de dados quando as observações no local  $i$  dependem de outras observações em locais  $j \neq i$ , ou formalmente:

$$y_i = f(y_j), i = 1, \dots, n \quad j \neq i \quad (2.2)$$

Este pressuposto determina que, para escolher a estrutura de previsão mais adequada ao problema de previsão, não basta correlacionar os atributos no tempo, mas também no espaço; desta forma, torna-se necessário executar um processo exploratório nos dados, com objetivo de investigar os padrões comportamentais dos mesmos.

LESAGE (1999) argumenta que existem duas razões básicas para que ocorra dependência entre valores de observações amostrais de dados entre locais distintos:

- Erro de medida – imprecisão da distribuição da amostra, ocasionado pela disparidade entre valor observado e a sua correta posição espacial e / ou problemas com a escala da representação dos dados; e,
- Modelagem - as ciências regionais têm como premissa que a localização e a distância são forças importantes para emprego na geografia humana e na atividade de mercado; desta forma o estudo da dimensão espacial da atividade sócio-demográfica, econômica ou regional é primordial à modelagem do problema desta natureza.

CÂMARA *et al.* (2004a) salientam que a dependência espacial e a sua formulação matemática, a autocorrelação espacial, são peças chaves na análise dos fenômenos espaciais, a partir da qual pode ser entendido como uma situação em que as observações mais próximas no espaço possuem associações similares (correlação de atributos).

ANSELIN (1999a) observa que, apesar de serem geralmente empregados como sinônimos estes termos não são idênticos. O autor considera a autocorrelação como

forma fraca de representação, pois mede o momento de segunda ordem de um conjunto de distribuições, enquanto a dependência seria a forma mais forte, pois observa a densidade do conjunto.

Na prática a observação do todo é realizada pela análise da dependência global, e a segunda ordem, ou segundo momento, investiga a dependência local. Ou seja, no senso global é possível não existir dependência espacial, mas existir na sua forma “mais fraca”, ou local. Isto caracteriza no senso espacial a existência de aglomerados de padrões similares, visto um determinado atraso espacial.

O atraso espacial está relacionado com a ordem da investigação da dependência. É análogo ao conceito do atraso temporal, só que na dimensão espacial. Este atraso é avaliado seguindo uma regra de contiguidade espacial que geralmente considera as relações topológicas, ou a distância entre os eventos espaciais.

#### **2.4.1.2 – A HETEROGENEIDADE ESPACIAL**

A heterogeneidade espacial relaciona-se com a instabilidade estrutural sob a forma de variação inconstante dos erros (heterocedasticidade) ou nos coeficientes do modelo (coeficientes de variáveis, regimes espaciais) (ANSELIN, 1999a).

No caso mais geral espera-se que aconteçam diferentes relacionamentos para todos os pontos no espaço e, sendo assim, a heterogeneidade espacial influencia nestes relacionamentos. Formalmente é descrita como uma relação linear da seguinte forma:

$$y_i = X_i\beta_i + \varepsilon_i \quad (2.3)$$

Onde:

$i$  - observações coletadas  $i = 1, \dots, n$  pontos no espaço;

$X_i$  - vetor ( $1 \times k$ ) de variáveis explicativas;

$\beta_i$  - conjunto de parâmetros associados ao vetor  $X_i$ ;

$y_i$  - variável dependente para observação, ou localização  $i$ ; e,

$\varepsilon_i$  - erro estocástico dos relacionamentos lineares.

Existem três razões importantes para considerá-la de forma explícita:

- A "estrutura" por detrás da instabilidade é espacial (ou geográfica) no sentido que a localização das observações é crucial para determinar sua forma;
- Pelo fato da estrutura ser espacial, a heterogeneidade acontece frequentemente em conjunto com a autocorrelação espacial, e as técnicas padrões de econometria não são mais apropriadas; e,
- Em um simples cruzamento, a autocorrelação e a heterogeneidade espacial podem apresentar observações equivalentes.

Sobre isto, ANSELIN (1988) observa que ambos os aspectos devem ser cuidadosamente estruturados, em conjunto, para obter capacidade de identificar os parâmetros do modelo e que, não se pode considerar aspectos de forma isolada.

#### **2.4.2 – INFORMAÇÃO ESPACIAL (IE)**

Segundo ARONOFF (2004) os fenômenos geográficos apresentam três características fundamentais:

- Possuem dimensões ou classes que podem descrevê-los, por exemplo, a população de uma cidade e tipo de uma vegetação respectivamente;
- Podem ser localizado no espaço, onde usualmente emprega-se um sistema de coordenadas conhecido; e,
- Podem ter sua ocorrência determinada no tempo.

SILVA (2001) salienta que todo fenômeno apresenta relacionamentos, não existindo fenômeno totalmente isolado. Por isto, identificar, classificar e registrar os fenômenos, associados à investigação das relações causais entre as variáveis componentes do problema, constitui passo fundamental no procedimento científico.

Ao afirmar isto, SILVA (2001) insere o conceito de que é sempre possível explicar um fenômeno a partir de suas relações, ou de uma forma mais sistêmica, esta afirmação introduz o conceito de causa e efeito; ou seja, a ocorrência de um fenômeno registra no ambiente um efeito, em função de sua ocorrência.

Isto acontece porque os dados são a evidência mais básica para uma investigação, e guardam os aspectos dos fenômenos que estão sendo estudados; por conseguinte, uma vez coletados, registram os acontecimentos com certo grau de precisão e qualidade.

Desta forma, uma vez registrado o fenômeno, este deve ser referenciado a um referencial conveniente, e pode ser então descrito na forma de dados geográficos.

##### **2.4.2.1 – A NATUREZA DO DADO GEOGRÁFICO**

Um Sistema de Informações Geográficas<sup>16</sup> (SIG) permite armazenar dados geográficos para representar as feições do mundo real através de (ARONOFF, 2004):

- Pontos – empregados na localização de fenômenos geográficos ou feições pequenas. Por exemplo, a localização de um cliente;

---

<sup>16</sup> Sistema baseado em computador que provê um conjunto de quatro capacidades para manusear dados georreferidos: entrada, manutenção, manipulação / análise e saída de dados (ARONOFF, 2004).

- Linhas - cadeia de pontos ordenados e conectados que representam feições cuja característica é linear. Por exemplo, um limite administrativo; e,
- Polígonos - regiões limitadas por feições do tipo linha. Por exemplo, a extensão geográfica de uma cidade, o contorno de um lago, entre outros.

As feições geográficas apresentam quatro componentes principais:

- A posição geográfica – localização especificada conforme um sistema de coordenadas;
- Os atributos – informações que qualificam o dado geográfico. Na maioria das vezes são termos não espaciais, ou seja, particularmente não representam uma posição geográfica;
- As relações espaciais entre dados geográficos – podem ser complexas e em grande número. Apesar de intuitivas para uma pessoa observando um mapa, devem ser expressas em uma forma utilizável por um SIG; e,
- O tempo – a informação geográfica está associada há um tempo ou um período. Conhecer o instante que foi coletada pode ser crítico para seu correto emprego, visto que as características do fenômeno variam ao longo do tempo.

No campo da representação, um dado discreto assume comportamento único e bem definido relacionando-se a um fenômeno. Apesar de poder variar no tempo e, em sua forma, sempre haverá uma relação causal bem definida, ou seja, o valor do atributo está associado a uma feição a qual o representa.

Por exemplo, a população de um bairro é única para aquele bairro e, com certeza irá variar com o tempo, mas sempre poderá se dizer que no determinado bairro “A” a população é de “B” habitantes. Porém, isto não é verdadeiro quanto à distribuição de sua densidade, pois seu comportamento varia em função da superfície, ou seja, em uma determinada região “A1” deste bairro pode-se encontrar 10 habitantes por km<sup>2</sup>, em outra região 500 habitantes por km<sup>2</sup>, respeitando-se sempre o número total de habitantes do referido bairro.

Conceitualmente, as representações do espaço e das propriedades espaciais de um fenômeno geográfico, podem ser formalizadas a partir dos modelos baseados em campos ou em objetos (BURROUGH e MCDONNELL, 1998):

- Objeto – a forma mais comum de se ver o espaço é como objetos (entidades). Uma vez definida e reconhecida à entidade espacial, definem-se os atributos, os limites e a localização. Podem ser representados por pontos, linhas ou

polígonos, bem como permitem fácil manuseio dos fenômenos antropogênicos (casas, divisas, rodovias).

- Campo - representa a distribuição espacial da variável que possui valores pontuais ao longo de uma dada região geográfica, em um dado tempo  $t$  (duas ou três dimensões lineares e o tempo). Os atributos e suas variações espaciais assumem variações contínuas e suaves ao longo do espaço.

#### 2.4.2.2 – ESTRUTURA DE DADOS ESPACIAIS

Existem duas abordagens principais para representar os dados geográficos:

- Estrutura de dados vetorial – os objetos são representados por pontos e linhas, que definem seus limites. Sua localização espacial é definida a partir de um posicionamento dado por um sistema de coordenadas, o que faz com que cada posição seja única. Nesta estrutura assume-se que a área do mapa é um espaço coordenado contínuo, onde as posições podem ser definidas com tanta precisão quanto desejada; por isto, entende-se que são matematicamente exatas.
- Estrutura de dados *raster* - também conhecida como estrutura matricial, subdivide o espaço em células ou pixel (*Picture element*) homogêneos, formando um *grid* regular, onde a posição de cada uma das células é rígida e a sua localização é definida pela linha e a coluna que ocupam. Um arquivo *raster* pode conter inúmeras células e o valor de cada uma indica o valor do atributo representado, bem como sua área define a resolução espacial.

Dentre as possíveis formas de implementação matricial, as com células quadradas tem melhor manuseio, pois permitem adotar uma seqüência numérica; bem como, é um formato que interage facilmente com os equipamentos utilizados para a entrada e saída de dados espaciais (ARONOFF, 2004).

Cada célula *raster* assinala um único valor, por isto diferentes atributos são armazenados em arquivos separados; com isto processar dados envolvendo múltiplos *raster* demanda a correta sobreposição das células dos diferentes arquivos. Conceitualmente, é como empregar uma estaca vertical nos valores das células para analisar cada localização. Este procedimento, apresentado na figura 2.3, é chamado de análise de sobreposição (*overlay analysis*).



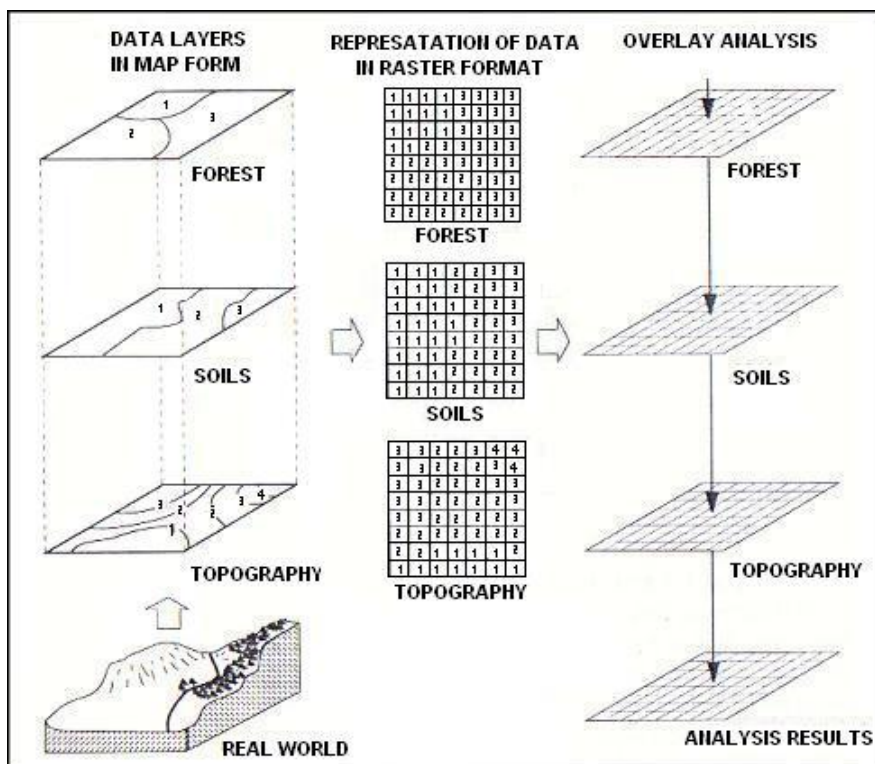


Figura 2.3 – Sobreposição empregando Dados *raster*- Extraída de ARONOFF (2004).

Cabe ressaltar que, a despeito da precisão ser limitada pelo número de *bits* utilizado na representação, a estrutura vetorial pode apresentar um melhor desempenho quando comparada a *raster*, visto o tamanho das células geralmente empregadas nesta última. Para exemplificar, na figura 2.4 tem-se a porção de terra (A) representada em *raster* (B) e em vetores (C).

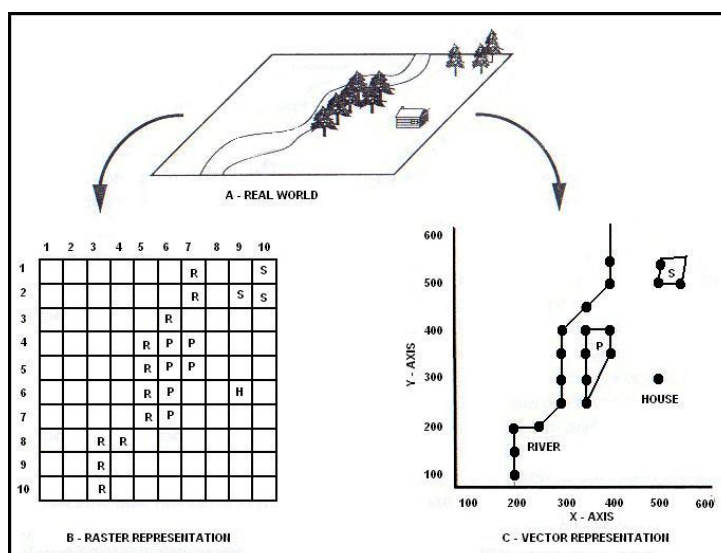


Figura 2.4 – Comparação dos Modelos Vetorial e *raster*- Extraída de ARONOFF (2004).

Como as técnicas de análise espacial são associadas ao tipo de representação, contínua (*raster*) ou discreta (vetorial), é fundamental identificar a estrutura na qual as variáveis explicativas do modelo são melhor representadas, permitindo assim a correta

definição da estrutura do modelo e das ferramentas / métodos para analisar a dependência.

#### **2.4.2.3 – A REPRESENTAÇÃO ESPACIAL DE DADOS SÓCIO-ECONÔMICOS**

Visto as características de confidencialidade e de tratamento estatístico, os dados sócio-econômicos normalmente são totalizados e disponibilizados para análise em polígonos (setores censitários, municípios); salvo quando referem-se a registros individualizados (dados de unidades produtoras e / ou consumidoras) que são descritas em formato de ponto.

CÂMARA *et al.* (2004b) ressaltam existir a suposição de que as regiões apresentadas em polígonos são homogêneas e por isto as mudanças significativas no valor das variáveis ocorrem nos limites; entretanto, isto não é correto, pois as unidades de levantamento são definidas por critérios operacionais e / ou políticos, não garantindo assim que o comportamento das variáveis seja homogêneo dentro das unidades. Isto inclusive é uma diferença significativa quando se consideram as representações pontuais, pois estas identificam a real posição, ou a posição individual da variável.

Ressalta-se que os objetos para representar os dados devem ter resolução adequada ao fenômeno em estudo, definida a partir da menor área para a qual existam dados disponíveis; ou seja, quanto menor a escala, maior a área em estudo e sua população, portanto, menor a homogeneidade e a capacidade de distinguir diferenças.

Por outro lado, aumentar a escala e a resolução ocasionam a redução da área e da ocorrência do fenômeno em estudo. O problema da interação entre a área de estudo (AE) e os resultados intitula-se MAUP - *modifiable area unit problem* (problema da unidade de área modificada); ou seja, alterar a escala e as dimensões da unidade de análise pode produzir conclusões dispares (CÂMARA *et al.*, 2004b).

#### **2.4.3 – A ABORDAGEM ESPAÇO-TEMPORAL**

Apesar de muitas vezes implícita, a observação do tempo é fundamental para que a informação geográfica descreva o fenômeno associando a localização e o tempo específico.

A representação do tempo nos SIG adiciona um considerável nível de complexidade dada à dificuldade de manuseá-lo e, em função disto, tal fator geralmente não tem sido acrescentado de forma sofisticada (HORNSBY e EGENHOFER, 2000; ARONOFF, 2004; PEUQUET, 2001).

Sobre o raciocínio temporal é relevante a contribuição e as perspectivas iniciais de ALLEN (1983), na qual observa-se que o problema de representar o conhecimento e o raciocínio temporal, surge em uma vasta gama de disciplinas, incluindo a informática, a filosofia, a psicologia, e a lingüística.

ALLEN (1983) aborda ainda, que um modelo para representar a informação temporal é naturalmente mais baseado em intervalos do que em pontos. Para tal propõe e descreve um método de representar as relações entre os intervalos temporais (uma noção de intervalos de referência) de uma maneira hierárquica com emprego de técnicas de propagação de limites.

Historicamente a perspectiva espaço-temporal tem sido conduzida por três ângulos distintos: cartográfica, modelos de dados e banco de dados espaciais (BDE). Em especial, cita-se Peuquet (PEUQUET, 2001), que descreve esta evolução histórica, relatando as iniciativas para representar o dado temporal e as abordagens iniciais no âmbito dos bancos de dados (BD).

#### 2.4.3.1 – A ORDEM TEMPORAL

São três as perspectivas a partir das quais pode-se representar o tempo: a ordem temporal, a variação temporal e a granularidade.

A ordem temporal é o modo como o tempo flui. Na figura 2.5 apresenta-se o esquemático deste conceito a partir da obra de WORBOYS e DUCKHAM (2004).

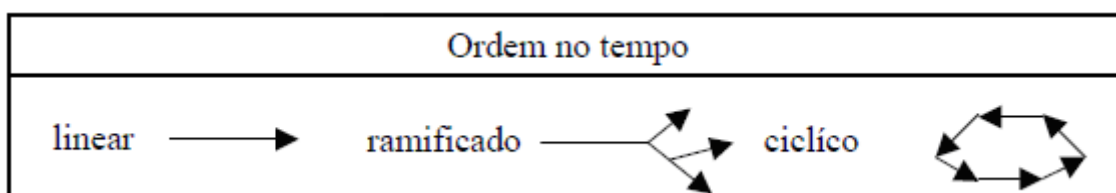


Figura 2.5 – Fluidez do Tempo – Adaptada de WORBOYS e DUCKHAM (2004).

Linearmente o tempo flui de forma seqüencial, com ordem de precedência entre os pontos no tempo, ou seja, cada ponto tem apenas um sucessor e um antecessor. No fluxo ramificado os pontos podem ser os sucessores ou, os antecessores imediatos de um mesmo ponto; e no cíclico, o tempo é utilizado para modelar eventos e processos recorrentes.

Na variação temporal, o tempo pode ser contínuo ou discreto. Emprega-se a variável temporal contínua em processos que demandam medidas de tempo com níveis arbitrários de precisão. Por exemplo, a figura 2.6 apresenta a expansão da área de desmatamento de uma floresta em dois instantes  $t$  e  $t'$ .

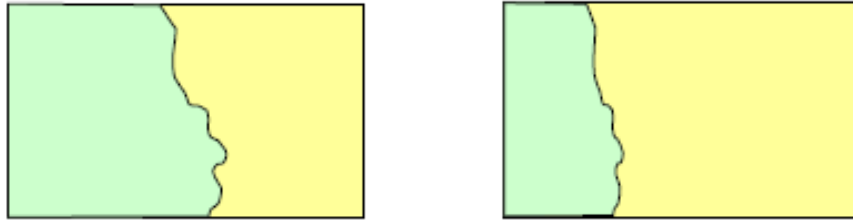


Figura 2.6 – Variável Contínua no Tempo - Extraída de PEDROSA e CÂMARA (2007).

Como este processo é contínuo no tempo, caso necessário poder-se-ia gerar um novo mapa, a partir de interpolação, para representar a área desmatada entre os instantes  $t$  e  $t'$ .

Uma variável temporal discreta é usada quando o tempo é medido em certos pontos ou intervalos e, não é necessário avaliar se a variação é descontínua entre estes, ou seja, registrar a posição intermediária entre  $t$  e  $t'$ . Por exemplo, na figura 2.7 existem 4 lotes no instante  $t$ ; já no instante  $t'$  os lotes 2 e 3 são unidos formando um novo lote 5.

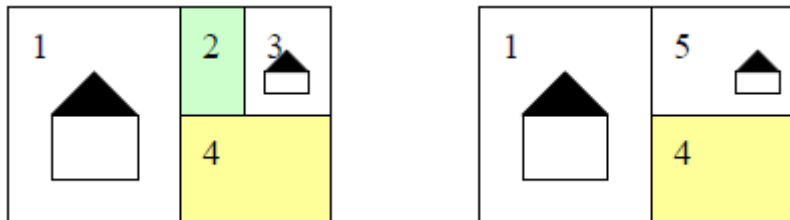


Figura 2.7 – Variável Discreta no Tempo - Extraída de PEDROSA e CÂMARA (2007).

A granularidade trata da forma como o tempo é medido, ou seja, em instantes pontuais  $t_0$ , ou em intervalos entre instantes  $t_0$  e  $t_1$  ou em períodos, que são na verdade seqüências de intervalos  $t_0 - t_1, t_1 - t_2, \dots, t_{N-1} - t_N$ .

#### 2.4.3.2 – AS RELAÇÕES ESPAÇO-TEMPORAIS

Os geógrafos distinguem os conceitos de espaço absoluto e relativo. O primeiro, também chamado Cartesiano, é um conjunto de eventos representados em uma estrutura que permite localizar trajetórias e objetos. O segundo, também chamado Leibnitziano, é o espaço que contempla as relações espaciais entre os objetos e eventos (COUCLELIS 1999).

WORBOYS E HORNSBY (2004) apontam que a característica que distingue se a entidade espacial é objeto ou evento é o ambiente onde esta se encontra. Cada objeto ou evento espacial está em um ambiente, que pode ser no espaço, tempo, ou uma combinação de ambos; porém, um objeto ou evento não pode estar situado em mais de um lugar ao mesmo tempo.

Este ambiente não se refere a uma localização pontual, pois assim é possível para um objeto, ou evento, situar-se sobre uma localização estendida (como uma região, ou um período de tempo). Podem-se elencar os seguintes os tipos de ambientes:

- Puramente espacial – são os mais estudados e podem ter até três dimensões; freqüentemente estão embutidos (explícita ou implicitamente) em um espaço (Euclidiano, métrico ou topológico) (WORBOYS e DUCKHAM, 2004);
- Puramente temporal – podem ser zero ou unidimensional, estão embutidos (explícita ou implicitamente) em um espaço, como um conjunto ordenado (linear ou parcialmente, discreto ou contínuo), ou em uma estrutura cíclica (momento, intervalo, período). Tem sido pesquisado na inteligência artificial e em BD puramente temporais (SNODGRASS, 2000) e (ALLEN, 1983); e,
- Espaço-temporal - formalmente, o ambiente espaço-temporal é função de um ambiente temporal para um espacial. Podem ser chamados de trajetórias (PARTSINEVELOS *et al.*, 2001), histórias (GRIFFITHS *et al.*, 2002), (GALTON, 2004), ou linhas do tempo geo-espaciais (HORNSBY e EGENHOFER, 2002). Como os objetos ou eventos não podem estar situados em mais de um local ao mesmo tempo, não permitem um completo produto cartesiano de espaço e tempo.

No desenvolvimento de um modelo, identificar as relações espaço-temporais deve ser entendido como parte integrante da etapa de definição de variáveis e de suas relações de causa e efeito. Esta identificação valida a série temporal para emprego nas operações de agregação de dados e, de avaliação da autocorelação espacial.

De forma prática, é necessário garantir que a série temporal das variáveis explicativas para a tese esteja corretamente relacionada com a unidade geográfica de referência (UGR) às quais estão associadas.

Em eventos tipo ponto, é difícil ocorrer essa situação, pois a relação de pertinência entre o atributo e o ponto é mais forte. Já para as operações com polígonos, isto é mais plausível, visto a maior probabilidade do valor do atributo ser resultado de operações matemáticas ou estatísticas (interpolação). Caso isto ocorra é necessário investigar o processo de geração do valor do atributo e sua constituição no tempo e no espaço.

Por exemplo, antes do ano 2000 era comum a emancipação de distritos para municípios. Como resultado, as séries de indicadores econômicos e sociais ficaram comprometidas, pois a partir da emancipação não existia histórico para a nova feição

gerada. Na mesma linha, os dados para o município original ficaram comprometidos, uma vez que não mais representavam aquela porção territorial como um todo.

Recomenda-se para aprofundamento, a leitura de EGENHOFER e GOLLEDGE (1988) na qual são apresentadas as fases de infusão entre os relacionamentos espaciais e os temporais, e é possível identificar a ação do tempo sobre os eventos espaciais.

#### 2.4.3.3 – A REPRESENTAÇÃO DOS FENÔMENOS ESPAÇO-TEMPORAIS

A dinâmica do relacionamento entre o tempo e o espaço motivou o desenvolvimento de elementos que permitissem representar e manter estas relações dentro dos SIG. De certa forma o entendimento desta questão traz à tona um novo integrante: o movimento.

Partindo do pressuposto que os objetos podem se mover de um lugar para outro no espaço e, mudarem sua forma ao longo do tempo, os eventos (fenômenos) ocorridos ao longo deste, passam a ser semanticamente relacionados a objetos em potencial processo de transformação. Na verdade passa a existir o movimento, ou a noção intuitiva de que as coisas se movem e/ou mutam ao longo do tempo.

Na figura 2.8 são apresentados, a partir de HORNSBY e EGENHOFER (1997), três, entre uma série de perspectivas de mudança do objeto ao longo do tempo, onde:

- *Merge* (a) - os objetos A e B fundem-se formando o objeto C e deixam de existir no tempo;
- *Generate* (b) - os objetos A e B continuam existindo no tempo, mas geram um objeto C; e,
- *Mix* (c) - o objeto A continua no tempo, mas parte deste funde-se com o objeto B formando o objeto C.

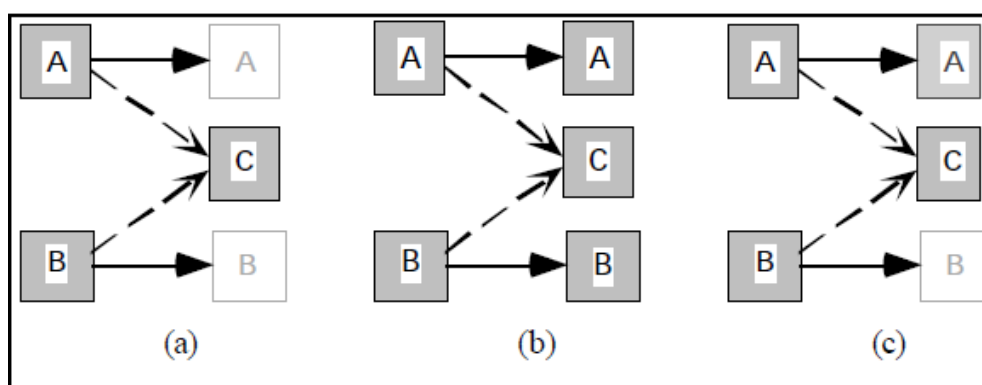


Figura 2.8 – Perspectivas de Mudança - Extraída de HORNSBY e EGENHOFER (1997).

Sob a perspectiva de mudança, HORNSBY E EGENHOFER (2000) propuseram representar o conhecimento espaço-temporal baseado na explícita descrição das

mudanças nos fenômenos geográficos. Com isto obtêm-se um melhor entendimento das modificações no estado das feições ao longo do tempo.

Cabe ressaltar que para dados representados em áreas isto é mais problemático, visto sua representação não ser homogênea, conforme citado em 2.4.2.3. As alterações no estado das feições, ou na sua forma, impactam também no resultado das medidas de autocorrelação espacial, pois inferem diretamente na determinação das matrizes de contiguidade que compõem a formulação de tais procedimentos.

Ainda sobre o assunto, ressalta-se existir esforço significativo para desenvolver estruturas de dados de espaço-tempo robustas e efetivas, das quais citam-se como exemplo:

- WORBOYS (1994) - modelo conceitual de objetos espaço-temporais como complexos espaço-bitemporal (é bi temporal no sentido de que ambos os tempos, de validade e do BD, estão representados);
- PEUQUET e DUAN (1995) - propõe o ESTDM (*Event-Based Spatiotemporal Data Model*) que registra as mudanças provocadas pelos eventos em ordem temporal (*life line*) listando os componentes de cada evento e a informação dos locais que sofreram modificações em função da ocorrência do evento;
- ERWIG *et al.* (1999) – propõe uma visão de nível-conceitual, ou nível-abstrato, baseada em objetos com coordenadas observadas ou projetadas (x; y; t), como um conjunto de atributos, no espaço 3-D,
- FORLIZZI *et al.* (2000) – subsequente ao modelo de ERWIG *et al.* (1999) descreve um nível de implementação para representação para movimento contínuo. Este é o mais alto-nível de estrutura de dados que pode ser implementado em um DBMS temporal; e,
- GATALSKY *et al.* (2004) - representação em  $R^3$ , onde o movimento do objeto passa a ser representado em um cubo espaço temporal.

#### **2.4.4 - A IE E SEU EMPREGO NAS ATIVIDADES EMPRESARIAIS**

Segundo NETO (2000), o problema espacial é aquele que para a solução deve se considerar a informação espacial, ou seja, a informação relativa às características geométricas, o posicionamento dos objetos ou fenômenos. Seu emprego deve auxiliar na exploração do espaço-problema (etapa de inteligência), na formulação de alternativas de solução (etapa de projeto) e na escolha de uma alternativa preferida (etapa de escolha).

A capacidade do SIG para suportar a decisão baseia-se no emprego do BDE e de suas ferramentas de manipulação e análise, de forma individual, mantendo o acervo de dados espaciais e analisando o comportamento dos fenômenos, gerando assim subsídio para a tomada de decisão.

A instituição de modelos de processos de decisão corrobora a visão do SIG como um Sistema de Apoio à Decisão (SAD), visto que criar e utilizar bancos de modelos são à base de estruturação deste tipo de sistema.

A este tipo de emprego cunhou-se o termo Sistema de Apoio à Decisão Espacial (SADE), para designar os SI com capacidade de auxiliar o ser humano a tomar decisões baseadas em dados georreferidos.

A partir deste enfoque são várias as possibilidades de emprego dos SIG em um ambiente empresarial, como por exemplo, as citadas por CASTRO (2007):

- Gestão de processos – emprego da informação espacial na gestão de processos de distribuição, atendimento, logística, entre outros;
- Gestão de relacionamento com o cliente - emprego da informação espacial na localização de clientes, bem como o que pode ser fornecido e em quanto tempo; evitando o impacto negativo na expectativa do cliente. Permite ainda analisar o comportamento de compra (frequência e quantidade); e,
- Planejamento estratégico – constituição de cenários para análise estratégica e estruturação do plano estratégico propriamente dito.

Ressalta-se que apesar da capacidade dos SIG, estes apresentam limitações para atuar em todas as fases dos processos decisórios, destinados a solução de problemas espaciais, pois no domínio dos SAD a referência espacial da informação usualmente é negligenciada, ou assume papel secundário no processo de tomada de decisão (NETO e RODRIGUES, 2001).

Mesmo existindo modelos científicos<sup>17</sup> que buscam representar pequenas porções do espaço, em sua maioria não consideram a referência espacial da informação. Sendo assim, os processos geográficos ficam em um elevado nível de generalização, os diversos sub-processos ficam reduzidos a um conjunto mínimo de parâmetros matemáticos e, conseqüentemente, o modelo tem baixa representatividade e risco para a decisão.

---

<sup>17</sup> Representações matemáticas de processos de domínios do mundo real. No domínio dos SIG tais processos representam o fluxo de massa, ou energia, entre os componentes físicos do sistema e, tem total inerência com a representação necessária destes fenômenos.



Em especial, no âmbito da análise estratégica, o emprego da informação espacial agrega valor pela capacidade de, entre outros, modelar, manipular, armazenar e analisar dados, independente da temporalidade e da posição espacial, suportando a constituição de cenários para atender uma tomada de decisão específica (CASTRO, 2005).

#### **2.4.4.1 – A IE EM ATIVIDADES DE PREVISÃO**

Quanto ao emprego da informação espacial e a aplicação da abordagem espaço-temporal, percebe-se uma maior ênfase no uso voltado ao planejamento e ocupação de solo (DENG *et al.*, 2009; JANSEN *et al.*, 2006; LIU *et al.*, 2005).

Em MENNIS e LIU (2005), empregou-se o SIG em conjunto com técnicas de mineração de dados para a identificação de padrões socioeconômicos e de mudança de uso do solo, para gerar hipótese relacionada às mudanças socioeconômicas e o crescimento urbano, na região de Denver, Colorado, EUA, de 1970 a 1990.

Com objetivo parecido DEAL e SCHUNK (2004) empregam autômatos celulares para apoio à decisão em problemas de cálculo do custo de dispersão urbana. Ao contrário de outros modelos de crescimento urbano baseados em análises de tendências, esta solução identifica e modela localmente os mecanismos causais de mudança.

Na questão de previsão de demanda os exemplos mais típicos abordam questões sócio-econômicas para apoio a projeção de demanda de mercados consumidores. Comumente intitulado demografia de negócios, refere-se ao conjunto de técnicas e métodos que tem como ponto central aplicações comerciais.

Por exemplo, em HAKKERT (2006), observa-se a necessidade de adaptar técnicas demográficas para emprego em situações comerciais. O autor cita o emprego de *Housing Units Method* (método das unidades de moradia)<sup>18</sup>, para prever a população em função da dinâmica da moradia, mais particularmente no estoque de domicílios.

AZEVEDO (2006), empregando informações sócio-econômicas aborda o emprego da informação espacial como perspectiva de análise da prestação do serviço de telefonia, para identificar a demanda social pela universalização de tal serviço.

Em outra contribuição, com um sentido mais temporal, GIVISIEZ *et al.* (2006) empregam o método das taxas de chefia (*headship rate*), em conjunto com modelos Idade-Período-Coorte, para projeção habitacional segmentada por regiões da grande Belo Horizonte.

---

<sup>18</sup> Estudo de Starsinic e Zitter publicado em *Accuracy of the housing unit method in preparing population estimates for cities*. DEMOGRAPHY, v. 5, n. 1, p.475-484, 1968.

Especificamente, na projeção de demanda de energia identificou-se as proposições de BANDEIRA (2005), MASSARA (2007), JOHANNESSON *et al.* (2006) e QUAN *et al.* (2002), que se valem da informação espacial como suporte para a referida atividade.

BANDEIRA (2005) propõe um método com abordagem *bottom-up* à orientação da expansão da rede de distribuição do GN no segmento residencial com base na substituição de energéticos e nos padrões de consumo de equipamentos domésticos. As agregações espaciais e os totalizadores partem de extrapolações, visto que não existem dados sobre o número de equipamentos domésticos.

A proposição apresenta como fragilidade, o fato dos pressupostos matemáticos e estatísticos empregados na modelagem, serem empíricos. Como virtude, cita-se o emprego de mapas como elemento de representação da demanda projetada.

MASSARA (2007) propõe distribuir a rede de gasodutos, empregando o conceito de demanda por atratividade baseada em pesquisa e seleção de áreas de interesse. Seu modelo de previsão espacial calcula os elementos de forma isolada e agrupa os resultados via SIG. A identificação e priorização das áreas de maior potencial de consumo têm base em indicadores e critérios de hierarquização.

O conjunto de variáveis tem safra única, o que compromete a projeção pela falta de dados históricos que validem a projeção. Além disto, segue os padrões empregados em outros modelos; entretanto, não estuda a correlação entre os fatores, bem como infere a contribuição real de cada parâmetro no mecanismo de priorização.

Com relação aos trabalhos de MASSARA (2007) e BANDEIRA (2005) cabe uma ressalva relacionada ao erro espacial. Uma vez que preconizam um modelo de decisão de alocação das redes de distribuição por priorização e seleção de áreas de interesse, de forma dinâmica, necessitam de assertividade no tempo e detalhamento pormenorizado com relação ao objeto de estudo, o que aumenta os requisitos de manutenção do SIG.

JOHANNESSON *et al.* (2006) salientam que os procedimentos espaciais exigem informações e dados com refinamento específico devido à sua variabilidade natural; no entanto, se não existe a informação na escala necessária, torna-se necessário extraí-la dos dados disponíveis, com posterior avaliação do erro, caracterizando um MAUP.

No exemplo “*National Electric Load and Demand Data by Control Areas*”, que emprega variáveis de carga na forma discreta, propõe-se sua projeção a partir de variáveis explicativas, por meio do emprego de técnicas de regressão clássica baseadas na correlação espacial entre a variável população e consumo de água na indústria.

Contudo, ressalta-se que não existe na proposta evidência da investigação da dependência espacial entre estas no tempo, ou seja, se estas se mantêm correlacionadas.

Desta forma, tal proposição não segue um processo espaço-temporal para cada evento identificado, porém tem como contribuição relevante a apresentação da utilidade das técnicas de regressão espacial para inferir dados.

Em QUAN *et al.* (2002), propõe-se aplicar abordagem espaço-temporal para prever a STLF com posterior apoio para a determinação da localização e da capacidade ótima de subestações. O SIG é empregado na integração espacial da carga histórica com os elementos que influenciam sua distribuição, particularmente o tipo de uso do solo e a caracterização dos empreendimentos.

Ainda sobre modelos de cálculo de demanda de energia<sup>19</sup>, cabe ressaltar a crítica de SHORT (2007) sobre a despadronização de representação espacial destes modelos conforme pode ser visto nas interfaces apresentadas na figura 2.9.

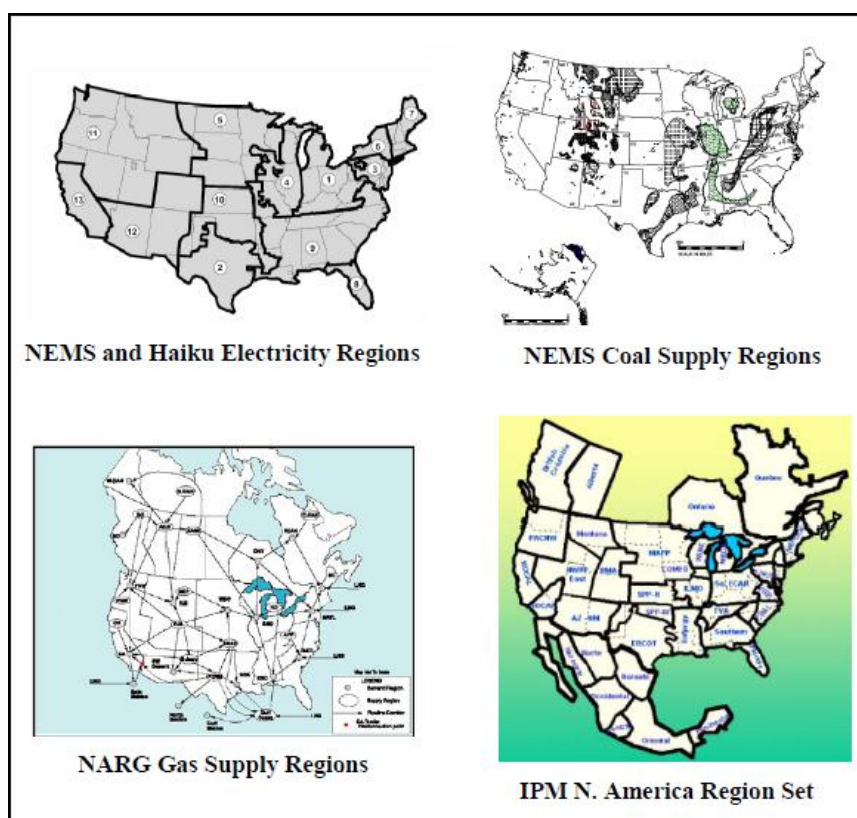


Figura 2.9 – Despadronização da Representação Espacial – Extraída de SHORT (2007).

<sup>19</sup> Integrated Planning Model (IPM); ICF International's North American Natural Gas Analysis System (NANGAS) model; All-Modular Industry Growth Assessment (AMIGA) model; Brookhaven National Laboratory's MARKET ALlocation (MARKAL) model; Resource For the Future's Haiku model; Altos' North American Regional Gas (NARG) model; Energy 2020 e ABB's Grid View transmission and optimal power flow model.

Os modelos citados não apresentam a mesma estrutura regional<sup>20</sup>, bem como a resolução espacial é em grande parte determinada pelo propósito planejado para estes; por outro lado, sua empregabilidade é reduzida por esta mesma resolução.

De certo tal situação compromete ainda a composição de análises empregando os resultados dos modelos, o que poderia em muito contribuir para seu aprimoramento.

No geral as proposições citadas neste tópico acompanham as variáveis no tempo, com sua respectiva projeção no espaço, o que poderia se dizer de uma variação temporal no espaço, ou acompanham as mutações dos objetos espaciais ao longo do tempo, o que estaria mais próximo de uma abordagem espacial no tempo.

Conclui-se que é menos usual a aplicação puramente espaço-temporal, na qual as variáveis apresentam inter-relacionamento em função de uma dependência espacial. Percebe-se com isto que, tanto as particularidades da informação espacial (resolução, escala), quanto à questão de relacionamento entre variáveis (dependência espacial, representação, modelagem), precisam ser melhor introduzidas no setor de energia, buscando com isto seu emprego como processo estruturado de conhecimento e ferramenta de modelagem e entendimento da dinâmica do setor.

## **2.5 – AS METODOLOGIAS PARA PREVISÃO DA DEMANDA**

Autores como MAKRIDAKIS *et al.* (1998) e ALCAMO (2001) citam a existência de dois métodos de previsão:

- Qualitativo – com base na opinião de especialistas, é útil na ausência de informações históricas ou diante de alterações significativas na realidade; e,
- Quantitativo – tem o objetivo de identificar padrões comportamentais que podem se repetir no futuro e geralmente empregam análises de séries temporais.

A abordagem mais aplicada à análise de séries temporais é a baseada em regressões; entretanto, isto confere classificação tênue, visto que está também é utilizada em outras técnicas, como por exemplo, as espaciais.

Na mesma linha, citam-se as ferramentas baseadas em abordagem *bayesiana*, largamente empregada em métodos de reconhecimento de padrões, assim como as regressões, igualmente empregada nas ferramentas espaciais.

---

<sup>20</sup> NEMS - múltiplos conjuntos de regiões sobrepostos, NERC - regiões do censo para modelo de uso-final, e estruturas regionais para suprimento de combustíveis fósseis, IPM - 80 regiões que podem ser agregadas conforme necessidade do usuário, Brookhaven MARKAL tem só uma região elétrica para todo o EUA.

Em função disto, para organizar o tópico empregar-se-á a seguinte divisão: a primeira tem foco na análise de séries temporais e no AM e a segunda abordará a visão de análise espacial e dos modelos de regressão espacial.

### 2.5.1 – ANÁLISE DE SÉRIES TEMPORAIS

A série temporal é um conjunto de observações seqüenciais acerca de um fenômeno. Formalmente é representada por um conjunto de observações  $[X(t), t \in T]$  de uma variável  $X$ , onde  $T$  é um conjunto de índices (tempo, espaço, profundidade). Dependendo da natureza de  $T$  e de  $X$ , a série temporal pode ser:

- Discreta – quando  $T$  é um conjunto finito de pontos,  $T = 1, 2, \dots, T$ ;
- Contínua – quando  $T$  é um intervalo finito,  $T = [t: 0 < t < T]$ ;
- Multivariada (discreta ou contínua) –  $[X_1(t), \dots, X_2(t), t \in T]$ ; e,
- Multidimensional – quando  $\underline{t}$  é um vetor  $[X(t), \underline{t} \in T]$ .

O objetivo da análise da série temporal é sumarizar suas propriedades e caracterizar seu comportamento, identificando ou sugerindo um modelo adequado. A proposição clássica decompõe as variações internas destas séries em componentes:

- Sazonal – trata da identificação da frequência de repetições de padrões ou comportamentos dentro da série;
- Tendência – trata da identificação do constante e sucessivo padrão de acréscimo / decréscimo apresentado pela série;
- Ciclos – trata da identificação das variações cíclicas regulares com frequência diferente de 1 ano; e,
- Flutuações irregulares – resíduo encontrado após a retirada de tendência, sazonalidade e ciclos. Como tal, podem ser aleatórios e neste caso não podem ser incluídos em um processo de previsão.

Há dois enfoques básicos para realizar a análise da série temporal:

- Domínio do tempo – investiga a magnitude com que os eventos ocorrem em determinado instante de tempo e a relação entre observações em diferentes instantes do tempo, com base em um número finito de parâmetros (paramétrica); e,
- Domínio da frequência – investiga a frequência com que certos eventos ocorrem em determinado período de tempo, com base em um número infinito de parâmetros (não-paramétrica).

Os métodos clássicos apresentam uma melhor estimação quando a variação apresenta tendência linear regular e / ou sazonalidade regular. Porém, perdem capacidade de estimação quando a tendência e / ou efeitos sazonais mudam no tempo, ou quando os valores sucessivos das flutuações irregulares são correlacionados.

Os métodos paramétricos mais usados são os de regressão, os auto-regressivos e de médias móveis, de memória longa, estruturais e não-lineares. Já os não-paramétricos mais utilizados são a função de autocovariância (autocorrelação) e a transformada de Fourier (MORETTIN e TOLOI, 2004).

### **2.5.1.1 - PREVISÃO DE DEMANDA DE GN EMPREGANDO SÉRIES TEMPORAIS**

O problema básico da previsão baseada em séries temporais consiste em estimar o termo  $X_{N+k}$  de uma série  $X_1, X_2, \dots, X_N$ , onde o índice  $k$  é denominado próximo passo (*lead time*).

A análise da série fornece os parâmetros que melhor caracterizam seu comportamento para desenvolvimento do método de previsão. Apesar da grande variedade destes métodos, não existe um que seja universalmente aplicável, cabendo ao analista definir o procedimento mais apropriado.

Os métodos baseados em séries temporais de maior emprego à previsão são os de: regressão, autoregressão, análise espectral e relações não-lineares. Sugere-se ainda uma classe de modelos intitulados econométricos (CHATFIELD, 2000); porém, entendendo que a econometria se utiliza destes métodos mais usuais na consecução de seus objetivos, não cabe uma classificação específica com relação a este ramo.

Observa-se que não existe uma linha tênue dividindo o emprego das ferramentas na modelagem da previsão das séries temporais. Por exemplo, os modelos auto-regressivos são empregados em metodologias paramétricas, ou em abordagens não-paramétricas, por exemplo, combinados com análises espectrais (série de Fourier).

Para aprofundamento recomenda-se a leitura de CHATFIELD (2000) e MORETTIN e TOLOI (2004), no qual são detalhados os métodos citados, bem como apresenta outros métodos relacionados ao assunto.

No tocante a previsão de GN com base em séries temporais, o primeiro exemplo citado é a proposição de LYNESS (1984), na qual são descritos os métodos de previsão de demanda de GN adotado pela British Gás Company com emprego de métodos regressivos e auto-regressivos. No curto prazo (previsão diária) empregavam-se

regressões simples para derivar a relação entre oferta e temperatura, com apoio de um modelo de *Box-Jenkins* para modelar a relação dinâmica dos atrasos.

Para o médio e longo prazo a companhia empregava o mesmo modelo de regressões adicionando componentes e variáveis específicas; por exemplo: até 1 ano a *Season Normal Temperature* para investigar a correspondência de oferta e procura nos meses do inverno, e de 1 a 5 anos as séries de vendas e econômicas.

Em outro modelo com séries temporais, SIEMEK *et al.* (2003) empregaram curvas logísticas, originalmente desenvolvidas para prever extração de óleo e adaptadas para o GN, com base na suposta equivalência entre estas atividades. Apesar de ter origem na lei de Hubbert esta proposição trata da adaptação do citado modelo ajustado à realidade polonesa, com base na proposição de modelo multi-ciclo de Startzman.

A proposição de SIEMEK *et al.* (2003) descreve a demanda hipotética de GN, baseada na tendência média de desenvolvimento econômico durante as décadas recentes. Na opinião dos autores o experimento apresentou boa compatibilidade com a demanda de GN no período de 1995 a 2000, e alcançou erro médio de 20% sobre a média calculada para o período de 1994 a 2000.

Merece destaque o trabalho de HUNTINGTON (2007), citado no tópico 2.1.2, que aborda a aplicação de um relacionamento geral auto-regressivo com atraso distribuído (ADL) para previsão do consumo industrial de GN nos EUA.

KAMERSCHEN E PORTER (2004) aplicaram séries temporais com ajustamento parcial e equações simultâneas para prever a demanda de EE segmentada e total, considerando substituição inter-combustíveis, entre eles o GN. A partir de um modelo de fluxo-ajuste<sup>21</sup>, assumiram que as expectativas de consumo poderiam ser descritas por um processo de média móvel de primeira ordem que, substituída em um processo de atrasos geométricos, descreveu o seguinte modelo de equações adaptativas:

$$W_{t+1}^e = (I - \lambda)(w_t + \lambda w_{t-1} + \lambda^2 w_{t-2} + \lambda^3 w_{t-3} + \dots) \quad (2.4)$$

Onde  $w_t$  é um substituto para eletricidade assim como para o preço de GN no tempo  $t$  e,  $0 < (I - \lambda) < I$ .

O resultado do experimento sugere que modelos de equações simultâneas parecem ser mais apropriados, desde que provenham estimativas de elasticidade de preço negativas para as amostras de eletricidade por segmento; além disso, os modelos de

---

<sup>21</sup> Baseado nas proposições de Houthakker e Taylor descritas em “*Consumer demand in the United States: analyses and projections*”, publicado por Harvard University Press.

fluxo ajustado predisseram, em alguns casos, estimativas positivas de elasticidade de preço, suscitando a dúvida se estes falham na identificação de suas próprias provisões e que podem estar influenciando preços.

Com relação aos métodos de análise de série temporais, percebe-se ainda seu emprego na análise de preços de equilíbrio de demanda e oferta, como em KRICHENE (2005), PANAGIOTIDISA e RUTLEDGE (2007) e ASCHE *et al.* (2001), nos quais a demanda do GN é vista pelo enfoque de *commodity*, conforme discutido em 2.1.2 e 2.2.1.

### **2.5.2 – PREVISÃO DE DEMANDA DE GN EMPREGANDO AM**

O AM é uma área da Inteligência Artificial que estuda métodos computacionais para aquisição de novos conhecimentos, bem como meios de organizar o conhecimento já existente. É orientado aos dados e pressupõe o aprendizado automático a partir de grandes volumes de dados, gerando hipóteses a partir deste.

A idéia central da aprendizagem é que as percepções sejam usadas não só para agir, mas também para melhorar a habilidade do algoritmo de aprendizado. Com isto o elemento de aprendizagem aperfeiçoa a base de conhecimento, e o elemento de desempenho utiliza a base de conhecimento para executar a sua tarefa.

Recomenda-se ainda ao leitor no desejo de aprofundamento específico a consulta as obras de MITCHELL (1997), DUDA e HART (2000)<sup>22</sup> e HAYKIN (2001)<sup>23</sup>.

Com relação aos métodos de AM destaca-se o emprego de RN e SVM aplicadas à previsão da demanda de energia, principalmente no planejamento de curto prazo. Observa-se que seu emprego atende várias das etapas envolvidas no processo de previsão, transitando desde a definição dos parâmetros até a questão da acuracidade.

O primeiro exemplo destacado é o emprego de RN na identificação de parâmetros de modelagem aplicado pela Agência Internacional de Energia (AIE) aos mercados de GN (WERBOS, 1988). Nesta proposição empregou-se as RN para a análise de sensibilidade, na qual o *feedback* dinâmico, componente central da *backpropagation*, foi utilizado para definir o conjunto de derivadas que formaram o instrumento de diagnóstico da avaliação do modelo.

---

<sup>22</sup> Duda, Richard O. Hart, Peter E. Stork, David G. Pattern classification (2ª ed.). John Wiley & Sons, New York, p.637, 2001. ISBN: 0-471-05669-3.

<sup>23</sup> Haykin, Simon. Redes Neurais: princípios e prática. Bookman, Porto Alegre, RS, p.900, 2001. ISBN: 8-573-07718-2.



O experimento de WERBOS (1988) aponta para a importância da utilização de novas funções de perda ao invés das regressões e, que esta nova visão tende a exigir o *feedback* dinâmico para implementações eficientes e confiáveis.

Uma aplicação típica empregando AM é o trabalho de BAKIRTZIS *et al.* (1996) que propõe uma RN com treinamento por *back-propagation* para prever a carga utilizando dados horários, dados diários da temperatura máxima e mínima e a temperatura projetada para os dias que se deseja a previsão. Este experimento apresentou resultados satisfatórios para dias comuns, mas obteve altos erros para feriados, quando implementado em um período correspondente. Como solução, BAKIRTZIS *et al.* (1996) empregaram a RN básica com dois componentes: os do dia normal e com ajustamento específico para os feriados.

YOO e PIMMEL (1999) propuseram um previsor para o STLF com o intervalo de uma hora até um dia à frente, com emprego de dados de corrente, temperaturas prévias e cargas de energia. Nesta proposição empregou-se uma RN adaptável auto-supervisionada. Este algoritmo é um auto-organizador, ou algoritmo de aprendizado por topologia de vizinhança.

Na previsão horária os resultados do experimento de YOO e PIMMEL (1999) foram similares aos outros métodos; entretanto, na previsão de um dia à frente os resultados foram superiores aos outros modelos empregados, mesmo com estruturas e algoritmos mais complexos.

Os autores destacam que, além de apresentar a vantagem quanto ao treinamento, visto a quantidade de *layers* em redes supervisionadas, a proposta apresentou vantagem relacionada com a adaptabilidade, pois a RN auto-supervisionada é re-treinada para cada previsão, enquanto outras abordagens demandam extensas sessões de treinamento.

KHOTANZAD *et al.* (2000) propõem um sistema de previsão de GN com duas fases. Na primeira, dois previsores trabalham em paralelo e produzem previsões independentes de despacho diário de GN, que em uma segunda fase são combinadas para gerar a previsão final.

Os autores justificam esta escolha por entenderem que os sistemas baseados em regressões lineares prevêm a demanda como uma função linear dos dados recentes de consumo e ambiente. Porém, devido à natureza não linear do problema, principalmente pelas questões ambientais, tais métodos não capturam esta relação adequadamente. Além disso, consideram que as técnicas tradicionais com emprego de um único previsor podem não ser a melhor escolha para características particulares dos dados.

Na verdade KHOTANZAD *et al.* (1998) empregaram a abordagem de combinações no que foi intitulada terceira geração de RN orientadas ao problema de carga de curto prazo, desenvolvido pela *Southern Methodist University* em parceria com *Electric Power Research Institute*. Neste experimento, o modelo consistia de três módulos, com duas RN para prever e uma terceira para combinar as duas previsões e gerar a previsão final.

No trabalho posterior (KHOTANZAD *et al.*, 2000) a primeira fase continha um previsor com arquitetura *multilayer feedforward* e, um segundo com arquitetura por *link* funcional com alguns parâmetros oriundos de funções não lineares. Ambos são treinados com algoritmo BP para erro. Já na segunda fase, de combinação, as duas previsões são misturadas considerando estratégias alternativas, por exemplo: média das previsões, mínimos quadrados, lógica *fuzzy*, entre outros..

Com base nos resultados os autores concluíram que a combinação das previsões resulta em melhoria da precisão da previsão; e que a combinação baseada em uma única ANN apresentou desempenho superior as demais.

Em outra aplicação de previsão para STLF, YANG *al.* (2001) propuseram uma RN BP GA, empregando algoritmos genéticos no treino das conexões de peso, antes que o erro da aprendizagem tendesse à estabilidade. Isto aumentou a velocidade de treinamento, bem como reduziu a possibilidade de convergência para um mínimo local.

Em um ensaio mais recente para prever carga de EE no curto prazo, LAURET *et al.* (2008) utilizou abordagem *Bayesiana* à estimação dos parâmetros da RN, pontualmente pela aplicação de funções de densidade de probabilidade sobre o intervalo dos pesos, com o objetivo de minimizar o problema de *overfitting* a partir do controle da complexidade do modelo.

Tal ensaio mostrou que, independente das técnicas de RN tradicionais, o método é hábil para lidar com a complexidade modelo e, desta forma com o problema de *overfitting*; e ainda contribui com a seleção das variáveis mais importantes do modelo.

Ainda em 2007, SEETHA e SARAVANAN (2007) combinaram RN com *Fuzzy back-propagation* à previsão de carga, com a idéia central de aplicar *Fuzzy* para modelar o conhecimento sobre o sistema e seus parâmetros de entrada e, a RN para capturar os relacionamentos de entradas e saídas *fuzzy*. Esta aproximação híbrida permitiu explorar as propriedades das RN em conjunto com as dos sistemas *fuzzy*, e os experimentos demonstraram que os resíduos foram consideravelmente menores para a BP *Fuzzy* quando comparadas com RN *multilayer perceptron*.

Como últimas citações sobre o emprego de RN ao problema de previsão, Mishra e Patra, respectivamente em (MISHRA e PATRA, 2008a) e (MISHRA e PATRA, 2008b) propuseram processar os sinais de resposta, em camadas próprias, contendo neurônios como se fossem camadas de *feedforward* com emprego do *Particle Swarm Optimization*<sup>24</sup>. No primeiro caso a RN foi estruturada conforme a figura 2.10.

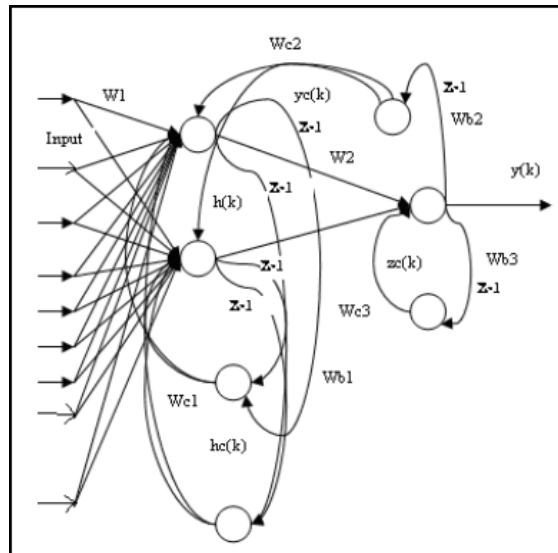


Figura 2.10 – Estrutura de Camadas Proposta em MISHRA e PATRA (2008a).

Esta estrutura, além das camadas *feedforward*, possui duas camadas de avaliação local e uma de avaliação global, às quais se aplicam as somas ponderadas dos atrasos de saída das camadas escondidas para ativar funções, como nos neurônios da camada *feedforward*.

Esta proposição, que emprega um número menor de neurônios, apresentou acuracidade superior a outras proposições, bem como as relações temporais providas através de neurônios e não só pelos caminhos de resposta, aumentaram a capacidade de mapeamento não-linear.

Na segunda proposição os autores compararam o desempenho das RN treinadas com GA e com *Particle Swarm Optimization*. Constatou-se com isto que na primeira abordagem, por GA, houve uma previsão melhor, porém com mais tempo de treinamento, enquanto na segunda aplicação o treinamento é mais rápido, mas com resíduos maiores.

No campo da utilização de SVM, LIU *et al.* (2004a) salientam que este método ofereceu um melhor desempenho de generalização por procurar a minimização do risco estrutural, particularmente quando a informação disponível é limitada.

<sup>24</sup> Baseado em comportamento social humano, modela o problema como um jogo de  $n$  partículas, que representam as  $n$  dimensões do espaço de soluções. Estas partículas se movem no espaço de soluções à procura da solução ótima.

O experimento de LIU *et al.* (2004a) foi orientado à previsão de carga de GN no curto prazo e apresentou resultado satisfatório, permitindo aos autores concluir que a solução não sofre das limitações de regularização da RN, o que pode levar-lhes a mínimos locais.

Além desta limitação, ao contrário da RN, o SVM não exige que a arquitetura seja definida a priori, bem como ela própria auxilia na identificação das observações úteis contribuindo assim para formar o conjunto de informações importantes.

Em LIU *et al.* (2004b) os autores reproduziram o modelo anterior aplicando mínimos quadrados na modificação da função de otimização<sup>25</sup>, para qual os resultados foram tão satisfatórios quanto os anteriores e a principal diferença para a SVM padrão foram as restrições de igualdade que simplificaram o problema.

Outro artigo que introduz o emprego de SVM na atividade de previsão é o trabalho de PAI e HONG (2005) sobre a aplicação de *Recurrent Support Vector Machines with Genetic Algorithms* (RSVGM) na previsão de carga de curto prazo; no qual se compara o desempenho da RSVGM com outros métodos tradicionais de previsão empregados no STLF.

Os autores observam que os métodos de séries temporais têm a limitação de ignorar a influência de fatores externos e, que os de inteligência artificial consomem grande tempo no treinamento, a despeito de apresentarem melhor desempenho do que as técnicas de séries temporais e métodos de aceleração de convergência. Segundo PAI e HONG (2005), a acuracidade da previsão do modelo RSVGM superou as outras aplicações pelas seguintes razões:

- A RSVGM tem capacidade de mapear relações não lineares e desta forma identificar com mais facilidade os padrões existentes nos dados de carga;
- Os GA podem determinar de melhor forma os parâmetros para a previsão reduzindo os riscos de *overfitting* ou *underfitting* no modelo SVM; e,
- O RSVGM minimiza o risco estrutural ao invés de minimizar os erros de treinamento.

Percebe-se que a aplicação de AM à previsão é uma via em franco desenvolvimento; entretanto, com clara orientação ao STLF que, mesmo apresentando similaridades com o curto prazo do GN, como salientado por LIU *et al.* (2004a), não é

---

<sup>25</sup> Baseada na proposição de Suykens e Vandewalle intitulada “Least Squares Support Machine Classifiers”, *Neural Processing Letter*, 1999, 284-288 V01.9, N, p.293-300.

tão pesquisada. No horizonte da previsão de médio e longo prazo, nos quais os pressupostos sócio-econômicos e de estruturação da economia têm maiores apelo e influência, constata-se que a contribuição do AM ainda é inexpressiva, tanto na questão da energia, quanto do GN.

### 2.5.3 - A ANÁLISE DE DADOS ESPACIAIS

O processo de análise espacial pode ser dividido em (ANSELIN, 1999b):

- Seleção - métodos de consulta e apresentação de dados, seja em mapas, tabelas ou diretamente ao BDE;
- Manipulação - processos que criam outros dados espaciais, tais como: agregação, interpolação, sobreposição (*overlay*), entre outras;
- Análise Exploratória - trata da descrição, visualização e análise das associações espaciais (aglomerados, autocorrelações) e sua instabilidade (heterogeneidade e estacionariedade).
- Análise Confirmatória - lida com a estimação e a validação necessária à realização de análises espaciais multivariadas, nas quais se empregam métodos de regressões espaciais.

De interesse a esta tese encontra-se a análise exploratória e a confirmatória.

O dado para emprego em métodos de análise espacial segue três características a observar:

- Descrição - aborda a forma mais tradicional na qual os dados são encontrados, ou seja, discreto ou contínuo;
- Contexto - lida com o significado do dado, ou melhor, o tipo de fenômeno que ele representa, que pode ser ambiental ou sócio-econômico; e,
- Forma de representação - trata de como encontramos o dado representado espacialmente, ou melhor, de forma pontual, por áreas ou em forma de matriz.

Na verdade não existe uma padronização sobre o tema. Por exemplo, CÂMARA *et al.* (2004a) cita a existência de dois dados associados ao problema de análise espacial: os ambientais, em geral representados por valores distribuídos de forma regular ou irregular no espaço e os sócio-econômicos, usualmente representados por polígonos, como abordado no tópico 2.4.2.1.

CRASSIE (1993) considera a existência de dados: Geoestatísticos, *Lattice Data* e Pontos. A abordagem de dados *Lattice* permite criar um modelo de dependência espacial entre amostras vizinhas. Sua disposição no espaço assume a forma de uma

matriz, com reticulado regular ou irregular. Desta forma atendem tanto a abordagens pontuais quanto para áreas.

LLOYD (2006) apresenta uma configuração que considera os dados em forma de *Grid*, Áreas, Geoestatísticos e padrões pontuais. Os dados do tipo *Grid* são dados matriciais. Embora existam diferenças técnicas entre *Lattice* e *Grid data*, estes são empregados para descrever um formato que armazena a localização em linhas e colunas.

Finalmente, BAILEY e GATRELL (1995) propõem a divisão mais usual entre os geocientistas: dados pontuais, por áreas, superfícies contínuas e interações espaciais.

Os dados do tipo *Grid*, *raster* ou matricial, abordados em LLOYD (2006), estão incorporados nas representações contínuas e, a interação espacial (link ou dados de fluxo), consiste de medidas associadas aos pares de localizações representando pontos ou áreas.

As técnicas de análise espacial possibilitam produzir informações, ou novos dados, com base nas características dos elementos contidos num SIG. Nesta tese adotar-se-á a tipologia de dados proposta em BAILEY e GATRELL (1995); entretanto, ressalta-se que FISHER (2006) provê uma interessante descrição, apresentada na tabela 2.1, na qual associa os tipos de dados com os respectivos métodos de análise, exploratória e confirmatória, adotando a linha de BAILEY e GATRELL (1995).

Tabela 2.1 – Análises Espaciais - Tipos de Dados e Modelos - Extraída de FISHER (2006).

| <b>Tipo de Dados</b>               | <b>Análise Exploratória</b>   | <b>Modelos de Análise de dados Espaciais</b>   |
|------------------------------------|---|--|
| <b>Padrões de Pontos</b>           | Métodos Quadráticos<br>Estimação de Densidade Kernel<br>Vizinhos mais próximos<br>Análise de função K                     | Modelos Poisson de processos homogêneo e heterogêneo e extensões multivariadas                                     |
| <b>Padrões de Áreas</b>            | Medidas globais de associação espacial:<br>Moran's I e Geary's c  | Modelos de Regressão espacial  |
|                                    | Medidas locais de associação espacial:<br>Gi e Gi* e Correlação Moran's   | Modelos de regressão espacial com resíduos autocorrelacionados   |
| <b>Contínuos</b>                   | Variograma e covariograma   | Modelos de superfície de tendência   |
|                                    | Estimação de Densidade Kernel   | Predição espacial e krigagem   |
|                                    | Poligonação de Thiessen   | Modelagem linear geral espacial  |
| <b>Interação espacial de dados</b> | Técnicas exploratórias para representação de dados  | Modelos de interação espacial  |
|                                    | Técnicas para descobrir evidências de estrutura hierárquica dos dados como teoria dos grafos e técnicas de regionalização | Modelos de distribuição local<br>Escolha de espaço e modelos de procura<br>Modelagem de Caminhos e fluxos de redes |

Cabe lembrar que os conceitos chave aplicado à análise espacial são a dependência espacial e a heterogeneidade espacial, abordadas no tópico 2.4.1; entretanto, ressalta-se que as análises acima citadas são *cross section*, seguindo pressupostos de efeitos globais e locais, a partir de medidas tomadas em um determinado instante de tempo  $t = 0$ .

Há de se tomar cuidado em não confundir o significado do termo predição. No senso espacial, este tem o sentido de reconstruir uma superfície a partir de técnicas para estimar valores para um fenômeno, em posições não amostradas, empregando as amostras do fenômeno proveniente da vizinhança, com métodos determinísticos ou geoestatísticos.

Os métodos determinísticos empregam relações de medidas diretas entre os eventos, destacando-se os procedimentos baseados em: triangulações, poligonação de *Thiessen*, IDW (*inverse distance weighting*), *trend surface* e interpolação polinomial.

Já os métodos geoestatísticos assumem a falta de conhecimento relativo a como as propriedades variam no espaço, sendo assim as variações espaciais são modeladas compreendendo um componente determinístico (tendência) e um componente estocástico (aleatório). Destaca-se o emprego da Krigagem<sup>26</sup>, simples e ordinária, e da co-krigagem.

Como o conjunto de dados desta tese é composto por dados pontuais de consumo e dados sócio-econômicos representados em polígonos, não serão aplicados métodos de interpolação para gerar de superfícies, ou métodos de análise de interação espacial. Sugere-se para aprofundamento nestes métodos a leitura de CRASSIE (1993), BAILEY e GATRELL (1995) e LLOYD (2006).

### **2.5.3.1 – AS ANÁLISES DE EVENTOS PONTUAIS**

Um padrão de distribuição de pontos compreende um conjunto de localizações pontuais,  $x_i$ ,  $i = 1, 2, \dots, n$ , com coordenadas espaciais  $x$  e  $y$ , caracterizando os eventos contidos em uma AE (LLOYD, 2006).

O interesse primário na análise deste tipo de padrão de distribuição é determinar se os eventos exibem padrão sistemático, ao invés de estar distribuídos aleatoriamente; nesta análise busca-se detectar a existência de padrões espaciais pela constatação de um número de casos excessivamente próximos acima do esperado.

---

<sup>26</sup> Na literatura empregam-se os termos Krigagem ou Krigeagem, porém ambos tratam do mesmo conjunto de técnicas, e são derivados do nome de Daniel G. Krige.

A partir de uma distribuição estocástica, constrói-se a hipótese de aleatoriedade. Caso o padrão da distribuição pontual desvie do padrão gerado a partir da distribuição de Poisson, é provável que ocorram aglomerados espaciais. Como em qualquer processo estocástico, o processo pode ser descrito em termos de primeira e segunda ordem.

Os efeitos de primeira ordem, considerados globais, correspondem a variações no valor médio do processo. Neste caso o interesse é na intensidade do processo (número de eventos / área). Os efeitos de segunda ordem, denominados locais, representam a dependência espacial no processo. Neste caso procura-se estimar o relacionamento entre pares de eventos, por unidade de área, no espaço.

BAILEY e GATRELL (1995) citam quatro grandes métodos aplicados à análise de padrões pontuais: métodos quadráticos, estimador de intensidade, vizinho mais próximo e a função K.

A análise estatística de distribuição de pontos requer um modelo teórico, base para desenvolver os métodos que verifiquem a significância dos resultados exploratórios. O modelo mais simples é conhecido como aleatoriedade espacial completa ou *complete spatial randomness* (CSR).

No padrão CSR os eventos seguem uma distribuição de Poisson sobre a área  $R$  em estudo. Este modelo divide  $R$  em subáreas  $A$  e modela a distribuição dos eventos como um processo aleatório  $\{y(A_i), A \in R\}$  (BAILEY, e GATRELL 1995). Isto implica que  $y(A_i)$  e  $y(A_j)$  são independentes para qualquer troca de  $A_i$  e  $A_j$ , e que a probabilidade de distribuição de  $y(A)$  é uma distribuição de Poisson com valor de significância  $\lambda S$ , onde  $S$  é área de  $A$ . Deste modo a distribuição de probabilidade de  $y(A)$  é:

$$f_{y(A)}(y) = \frac{(\lambda A)^y}{y!} e^{-\lambda S} \quad (2.5)$$

Onde  $\lambda$  é uma constante (intensidade média de eventos por unidade de área) e que pode-se considerar que a posição dos eventos é independente e que estes têm igual probabilidade de ocorrência em toda a região  $R$ . Esta formulação permite estabelecer uma base de comparação entre uma distribuição totalmente aleatória, gerada pelo processo Poisson, e as observações de campo.

Baseados nos conceitos de CSR pode-se buscar caracterizar os efeitos de segunda ordem em distribuição de pontos, utilizando os métodos do vizinho mais próximo e da função K. Em BAILEY e GATRELL (1995) encontra-se uma completa descrição da formulação destes quatro métodos citados.



### 2.5.3.2 – AS ANÁLISES DE ÁREAS

Tal qual o caso das distribuições pontuais, o objetivo da análise da distribuição de eventos em áreas é identificar padrões espaciais nos valores observados. Análoga a localização por pontos, são estabelecidas relações de proximidade espacial que definem a contiguidade entre as áreas englobadas na AE, porém neste caso as relações são estabelecidas com base nos centróides das mesmas (BAILEY e GATRELL, 1995).

O conceito consiste em empregar uma matriz  $W$  de proximidade, ou contiguidade espacial, na qual cada elemento da matriz  $W_{ij}$ , representa o relacionamento entre as áreas  $A_i$  e  $A_j$ . Para estabelecer estas relações emprega-se uma estratégia, usualmente baseada em topologia ou distância, através da qual o aspecto de dependência espacial aparecerá, como por exemplo:

$$w_{ij} = \begin{cases} 1 & \text{centróide de } A_j \text{ é um dos } k \text{ centróides mais próximos de } A_i \\ 0 & \text{de outra forma} \end{cases}$$

$$w_{ij} = \begin{cases} 1 & \text{centróide de } A_j \text{ esta dentro de uma determinada distância de } A_i \\ 0 & \text{de outra forma} \end{cases}$$

$$w_{ij} = \begin{cases} 1 & A_j \text{ possui uma fronteira com } A_i \\ 0 & \text{de outra forma} \end{cases}$$

$$w_{ij} = \begin{cases} d_{ij}^\Phi & \text{se a distância entre centróides } d_{ij} < y \text{ (} y > 0; \Phi < 0 \text{)} \\ 0 & \text{de outra forma} \end{cases}$$

$$w_{ij} = \begin{cases} \frac{l_{ij}}{L_i} & \text{onde } l_{ij} \text{ é o comprimento de fronteira comum entre } A_j \text{ e } A_i \\ l_{ij} & \text{e } l_{ij} \text{ é o perímetro de } A_i \end{cases}$$

Sobre isto, GETIS (2009) aponta pelo menos três formas de estabelecer as matrizes para emprego em estudos de autocorrelação espacial:

- Teórica - a matriz é exógena ao sistema, com base em uma estrutura preconcebida, normalmente uma distância pré-definida. Por exemplo, uma função de declínio fixo, como  $d_{ij}^2$ , onde  $d_{ij}$  é a distância entre as feições espaciais então os elementos da matriz de  $W$  são  $w_{ij} = 1/d_{ij}^2$ , com  $w_{ij}=0$ ;
- Empírica - tem-se a idéia de que um sistema flexível de pesos pode realçar características que se acredita ser mais importante em uma AE; e,
- Topológica - surge da necessidade de descrever de forma mais real as relações entre as feições contidas na AE. Nesta visão, a matriz pode ser especificada

por número de vizinhos, lado comum (*Rook*), vértice comum (*Bishop*), vértices e lados comuns (*Queen*), entre outros.

Outra questão sobre as formas das matrizes é a ordem dos atrasos. Análogo ao atraso temporal, no senso espacial o atraso é “medido” pela relação entre os pontos, ou entre os centróides das áreas, de forma similar às estratégias de topologia ou distância.

Nas análises por áreas, onde as formas das feições são variadas e, por conseguinte com medidas de centróides diferentes, a precisão para definir as relações de ordem é maior nas matrizes baseadas em topologia, do que nas por distâncias, visto ser praticamente impossível determinar uma distância que englobe somente vizinhos de uma mesma ordem.

Quanto à distribuição espacial do padrão, BAILEY e GATRELL (1995), utiliza a matriz de contiguidade  $W_{ij}$  para definir pesos entre as áreas vizinhas, e explora como o valor da média  $\mu_i$  dos atributos de interesse varia através da região de estudo, estabelecendo o que é conhecido como média móvel espacial, que é dada por:

$$\hat{\mu}_i = \frac{\sum_{j=1}^n w_{ij} y_j}{\sum_{j=1}^n w_{ij}} \quad (2.6)$$

O denominador é desnecessário se  $W_{ij}$  for “padronizado” sendo a soma das unidades das colunas, o que acontece na maioria dos casos.

De forma similar ao padrão de pontos é possível empregar o estimador de intensidade (*kernel estimator*) para fornecer uma visão aproximada da distribuição espacial do padrão. Neste caso os valores observados representam a média das observações e emprega-se o estimador para calcular o valor médio do atributo por unidade de área.

A média espacial móvel e o estimador de intensidade são úteis para mostrar padrões e tendências espaciais; porém, para explorar a dependência espacial, mostrando como os valores estão correlacionados no espaço, torna-se necessário medir a autocorrelação, para qual se empregam métodos globais e locais.

Os métodos globais são empregados para avaliar a interação entre os dados e podem ser facilmente visualizados pelo significado de um variograma espacial, uma série de autocorrelações medidas para diferentes ordens de contiguidade (FISCHER, 2006).

Existem várias técnicas, mas as de maior emprego estimam a correlação espacial com base na medida da covariância, onde se destacam os índices:

a) Moran's I – onde a correlação espacial dos atributos  $y_i$  é estimada por:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\left( \sum_{i=1}^n (y_i - \bar{y})^2 \right) \sum_{i \neq j} w_{ij}} \quad (2.7)$$

b) Geary's C – onde a correlação espacial dos atributos  $y_i$  é estimada por:

$$I = \frac{(n-1) \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})^2}{2 \left( \sum_{i=1}^n (y_i - \bar{y})^2 \right) \left( \sum_{i \neq j} w_{ij} \right)} \quad (2.8)$$

Onde,

$n$  - número de áreas;

$y_i$  - valor do atributo considerado na área  $i$ ;

$\bar{y}$  - valor médio do atributo na região de estudo; e,

$w_{ij}$  - pesos atribuídos conforme a conexão entre as áreas  $i$  e  $j$ .

O método global fornece um valor único como medida da associação espacial para todo o conjunto de dados; porém às vezes é preciso examinar padrões numa escala maior, visto a possibilidade de existir dependência espacial em um subconjunto da AE. Para tal são utilizados indicadores locais de associação espacial que possam ser associados a diferentes localizações de uma variável distribuída espacialmente.

O indicador local fornece um valor específico por objeto, e permite identificar agrupamentos com valores de atributos semelhantes (clusters), objetos anômalos (*outliers*) e de mais de um regime espacial. Especialmente destacam-se os índices (ANSELIN, 1995):

a) Local de Moran – o índice local do objeto  $I_i$  é multiplicado pelo desvio correspondente  $z_i$  e pela média dos desvios de seus vizinhos, sendo expresso por:

$$I = \frac{z_i \sum_j w_{ij} z_j}{\sum_{i=1}^n z_i^2} \quad (2.9)$$

b) Estatísticas  $G_i$  e  $G_i^*$  - o numerador é a soma de todos os valores de todos os vizinhos dentro de uma distância  $d$  do ponto considerado. A estatística  $G_i^*$  inclui a localização visitada em seu cálculo.

$$G_i(d) = \frac{\sum_{j=1}^n w_{ij}(d)x_j}{\sum_{j=1}^n x_j}, j \neq i \quad (2.10)$$

$$G_i^*(d) = \frac{\sum_{j=1}^n w_{ij}(d)x_j}{\sum_{j=1}^n x_j} \quad (2.11)$$

### 2.5.3.3 – OS MODELOS DE REGRESSÃO ESPACIAL – *CROSS SECTION*

Os métodos descritos anteriormente são empregados usualmente com o objetivo de investigar os relacionamentos entre os valores dos atributos  $y_i$ , relativos ao arranjo espacial  $A_i$ , e, a possibilidade de empregar os valores de outros atributos,  $x_i^T = (x_{i1}, \dots, x_{ip})$ , associados a cada  $A_i$  (BAILEY e GATRELL, 1995).

Para abordagens mais formais de modelagem do problema espacial, aplicam-se os Modelos de Regressão Espacial, baseados em regressões não espaciais, na forma de  $y = X\beta + \varepsilon$ <sup>27</sup> e em regressões ordinárias de mínimos quadrados, (*Ordinary Least Squares* - OLS) modificadas para relaxar a suposição de independência, inerente ao modelo de regressão simples na forma de  $y = X\beta + u$ <sup>28</sup>.

Para dados em formato de áreas existem algumas dificuldades em admitir que os efeitos de segunda ordem sejam estacionários e podem ser modelados como uma função ajustada pela distância entre os pontos, assim como no caso dos dados contínuos. Mesmo que existisse um processo com espaço contínuo estacionário, o uso de valores agregados e observações indiretas sobre áreas irregulares resultariam em variâncias e covariâncias que não seriam iguais para todas as áreas; sendo assim não se teria uma medida simples de distância entre as áreas (BAILEY e GATRELL, 1995).

Embora seja possível modelar a estrutura de covariância para áreas com base no correlograma residual, é preferível assumir que os valores dos dados acontecem nos

---

<sup>27</sup> Onde  $y$  é vetor ( $n \times 1$ ) de variáveis de interesse, ou mais estritamente  $y(A_i)$  no caso de área  $A_i$ ,  $X$  é a matriz ( $n \times k$ ) de variáveis explicativas, em cada área, com colunas de vetores  $x_i^T$ ,  $i = 1, \dots, n$ ; e  $\varepsilon$  é o vetor ( $n \times 1$ ) de média zero de variáveis randômicas,  $\varepsilon_i$ , representando as flutuações das tendências ou valor de média  $\mu_i = x_i^T \beta$  na área  $A_i$ .

<sup>28</sup> Onde  $u$  é vetor ( $n \times 1$ ) de média zero de erros com matriz de covariância  $C$  ( $E(U) = 0$  e  $E(UU^T) = C$ ), então os mínimos quadrados estimados para  $\beta$  e correspondendo aos erros padrões torna-se:  $\hat{\beta} = (X^T C^{-1} X)^{-1} X^T C^{-1} y$  e  $VAR(\hat{\beta}) = (X^T C^{-1} X)^{-1}$ . Os mínimos quadrados ponderados são um caso espacial onde  $C$  é uma matriz diagonal e pode ser apropriada no simples caso em que os erros,  $u_i$ , são heterocedásticos.

centróides das áreas como uma forma indireta de especificar o modelo de covariância. Isto é razoável para explicar as variações de segunda ordem observadas pela demonstração de aceitáveis graus de independência espacial nos resíduos do modelo.

Ou seja, ao invés de tentar especificar a matriz de covariância  $C$  diretamente por OLS, faz-se indiretamente por um “esquema de interações”, que inclui no modelo os relacionamentos entre as variáveis e seus valores vizinhos, os quais vão especificar indiretamente a forma de  $C$ .

Como a interação das variáveis nos modelos não precisa assumir estacionariedade para os componentes de segunda ordem, não estão restritas à estrutura de covariância e, sendo assim são ajustadas por funções de distância euclidiana (BAILEY e GATRELL, 1995). Um modelo simples de interação de variáveis é escrito na forma:

$$\begin{aligned} y &= X\beta + u \\ u &= pWu + \varepsilon \end{aligned} \quad (2.12)$$

Sendo,  $W_{ij}$  padronizada com a soma das linhas igual a 1 e  $\varepsilon$  o vetor de erros aleatórios independentes com variância  $\sigma^2$  constante,  $E(\varepsilon) = 0$  e  $E(\varepsilon\varepsilon^T) = \sigma^2 I$ ; ajustando-se 2.12 pela interação do parâmetro  $p$  com relação à  $y$  e  $X\beta$ , obtém-se outra forma desta classe de modelos, conhecido como modelo de erros autocorrelacionados (*Spatial Error Model* – SEM) e pode ser escrito na forma de (BAILEY e GATRELL, 1995) e (LESAGE, 1999):

$$y = X\beta + pWu + \varepsilon \quad (2.13)$$

Na qual  $y$  é expresso como uma resposta às influências  $y_i$  na área  $A_i$ , que depende dos valores vizinhos  $y_j (j \neq i)$ , através do termo  $pWu$  e, da direção geral da relação de regressão  $X\beta$ . Este modelo pode ser generalizado apresentando parâmetros de interação e matrizes  $W_{ij}$  para diferentes atrasos, conforme:

$$\begin{aligned} y &= X\beta + u \\ u &= p_1 W^{(1)}u + p_2 W^{(2)}u + \dots + \varepsilon \end{aligned} \quad (2.14)$$

Entretanto, os problemas de estimação numérica associados com as interações em (2.14) podem ser grandes. De forma alternativa pode-se simplificar o modelo em (2.13) reescrevendo-o formalmente como:

$$\begin{aligned} y &= X\beta + pWu + \varepsilon \therefore \\ y &= X\beta + pW(y - X\beta) \therefore \\ y &= X\beta + pWy - pWX\beta + \varepsilon \end{aligned} \quad (2.15)$$

Nesta forma às influências  $y_i$  na área  $A_i$  passam a depender dos valores vizinhos  $y_j(j \neq i)$ , através dos termos  $pWy$  e  $pWX\beta$ . Este modelo é conhecido como *Spatial Durbin Model* (SDM) em analogia a sugestão de Durbin para o caso de um modelo de séries temporais com autocorrelação residual (ANSELIN, 1988).

Alternativamente a supressão do termo  $pWX\beta$  faz com que as variações espaciais estejam associadas somente a  $y$ , ou seja, por um efeito puramente autoregressivo. Isto resulta em uma estimativa de máxima verossimilhança, a qual é computada de forma mais simples que no modelo de erros autocorrelacionados; desta forma tem-se:

$$y = X\beta + pWy + \varepsilon \quad (2.16)$$

Este modelo é intitulado Modelo Autoregressivo-Regressivo Espacial (*Spatial Autoregressive-Regressive Model* - SAR) (LESAGE, 1999), que seguido da supressão do termo regressivo  $X\beta$ , resulta no modelo puramente autoregressivo, intitulado espacial de primeira ordem (*First Order Spatial AR* - FAR):

$$y = pWy + \varepsilon \quad (2.17)$$

Assumindo que todas as variações no processo são capturadas pelas variações de ordem espacial, não existe dependência média em  $X$ , o que não é verdade para processos com variações de ordem temporal; apesar da estimação de parâmetros por máxima verossimilhança ser relativamente correta. Neste caso, uma alternativa aos modelos puramente autoregressivos é o modelo de média móvel, na forma de  $Y = (I + pW)\varepsilon$ .

Segundo BAILEY e GATRELL (1995) existe uma variedade de estruturas baseadas em autoregressão condicional. Em LESAGE (1999) encontra-se a proposição de uma estrutura geral que considera as condições autoregressivas espaciais, incluindo os atrasos espaciais e a estrutura de erros correlacionados espacialmente, como segue:

$$\begin{aligned} y &= pW_1y + X\beta + u \\ u &= \lambda W_2u + \varepsilon \\ \varepsilon &\sim N(0, \sigma_\varepsilon^2) \end{aligned} \quad (2.18)$$

LESAGE (1999) ressalta que ocorrendo dependência espacial na estrutura de erros de um modelo autoregressivo espacial, os modelos baseados em condições podem representar uma abordagem apropriada para modelar a dependência destes erros.

Estes modelos também podem ser empregados para simular processos de decisão com base em localização, nos quais, por exemplo, emprega-se a matriz  $W_2$  como o inverso da distância entre pontos.

Existem outras abordagens espaço regressivas na literatura. Em especial cita-se a proposição de CRASSIE (1993) sobre a re-colocação da parcela autoregressiva nos modelos de média móvel, obtendo-se assim os modelos espaciais autoregressivos e médias móvel (*Space-time Autoregressive*

*Moving Average - STARMA*) e o emprego da abordagem bayesiana, para qual o distúrbio  $\varepsilon$  apresenta variação não constante com diferentes valores para todas as observações. Sendo assim a matriz  $V$ , da distribuição normal dos erros, é formada por parâmetros a serem estimados ( $v_i, i = 1, \dots, n$ ). Com isto o modelo geral proposto em LESAGE (1999) assume a forma de:

$$\begin{aligned} y &= \rho W_1 y + X\beta + u \\ u &= \lambda W_2 u + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 V) \\ V &= \text{diag}(v_1, v_2, \dots, v_n) \end{aligned} \quad (2.19)$$

Para aprofundamento no tema recomenda-se a leitura de BAILEY e GATRELL (1995), LESAGE (1999) e, particularmente por uma abordagem de *Lattice Data*, a leitura de CRASSIE (1993).

#### 2.5.3.4 – O MODELO DE REGRESSÃO ESPACIAL PONDERADA

Intitulado GWR (*Geographically Weighted Regression*) ou regressão geograficamente ponderada, baseia-se na ponderação dos dados de acordo com a distância entre as suas localizações, seja por pontos ou centróides de áreas. A principal contribuição metodológica é o uso de subconjuntos de dados ponderados pela distância para proceder às regressões localmente lineares para estimar pontos em todo o espaço.

As decisões na GWR estão centradas na escolha da função de ponderação, no formato e na largura do Kernel. Usualmente emprega-se um Kernel com largura de faixa única para todas as localizações, porém é possível empregar uma largura de faixa adaptada localmente (LLOYD, 2006).

A GWR baseia-se em uma regressão para duas variáveis independentes, do tipo:

$$z(x_i) = \beta_0(x_i) + \beta_1(x_i)y_1 + \beta_2(x_i)y_2 + \varepsilon \quad (2.20)$$

Na qual  $x_i$  é a localização para quais os parâmetros são estimados; com isto é possível estimar valores para localizações onde não existem dados coletados.

Os parâmetros para a GWR podem ser estimados resolvendo:

$$\beta(x_i) = (Y^T W(x_i) Y)^{-1} Y^T W(x_i) z \quad (2.21)$$

Na qual  $W(x_i)$  é uma matriz quadrada  $n \times n$  e os elementos da diagonal principal são as ponderações geográficas ao redor do ponto  $i$ .

Assim como no caso das matrizes  $W_{ij}$ , para estabelecer o relacionamento no espaço é possível aplicar diferentes funções de ponderação; por exemplo, a função Gaussiana  $w_i(x) = \exp[-.5(d/\tau)^2]$ , onde  $d$  é a distância euclidiana entre a observação  $i$  e a localização  $x$  e  $\tau$  é a largura da faixa do Kernel; ou a abordagem de vizinhos mais próximos, na qual a largura da faixa é inversa à concentração de observações por área. Neste caso os pesos seriam providos por:

$$w_i(x) \begin{cases} [1-(d_i/\tau)^2]^2 & \text{se } d_i < \tau \\ \mathbf{0} & \text{caso contrário} \end{cases} \quad (2.22)$$

Assim como para as regressões espaciais, existe uma abordagem Bayesiana para emprego no método GWR. De forma similar à primeira, o desvio  $\varepsilon$  apresenta variação não constante com valores diferentes para todas as observações; sendo assim a matriz  $V$ , da distribuição normal dos erros, é formada por parâmetros a estimar (LESAGE, 1999).

### 2.5.3.5 – OS MODELOS DE REGRESSÃO ESPAÇO-TEMPORAL - *PANEL DATA*

No tópico anterior observou-se a modelagem de regressão em um senso estritamente pontual em relação ao tempo. Na teoria geral econômica tal abordagem se denomina *cross section*, para qual, no caso espacial, os modelos citados têm somente a capacidade de investigar os relacionamentos entre os valores dos atributos  $y_i$ , e as respectivas influências espaciais, para um dado momento no tempo.

Para investigar os eventos ao longo do tempo faz-se necessário empregar séries históricas de eventos para a variável de interesse ou, dados em painel (*Panel Data*), quando as observações acontecem para mais de um indivíduo ao longo do tempo.

A taxonomia típica aplicada a dados em painel apresenta estrutura na forma de um modelo de regressão linear simples, conforme:

$$y_{it} = \beta_{0it} + \beta_{1it}x_{1it} + \dots + \beta_{nit}x_{nit} + \varepsilon_{it} \quad (2.23)$$

Onde,

$y$  - variável dependente;

$x_k$  - variáveis explicativas do modelo;

$\beta_0$  - intercepto;

$\beta_k$  - parâmetros a serem estimados;

$i$  - referem-se aos diferentes indivíduos, localizações, empresas; e,

$t$  - refere-se ao período de tempo que está sendo analisado.



Segundo BALTAGI (2001) sua estrutura difere das séries temporais e das regressões *cross section*, por subscrever duplamente as variáveis do modelo. Ainda segundo o autor existem vários benefícios na aplicação de dados em painel, entre eles:

- Controlar a heterogeneidade individual, pois ao contrário das séries temporais e dos dados *cross section*, os dados em painel enxergam os eventos individuais de forma heterogênea reduzindo o risco de resultados com viés;
- Fornecer mais informação, variabilidade, mais graus de liberdade, eficiência e menos colinearidade;
- Ser mais hábil para estudar a dinâmica econômica, uma vez que os dados *cross section* tem um olhar relativamente mais estável quando se avalia a amplitude de acontecimentos no tempo;
- Ter melhor capacidade para identificar e medir efeitos que não são detectáveis em dados *cross section* e séries temporais puras; sendo assim permitem construir e testar modelos para comportamentos mais específicos; e,
- Permitir o emprego de dados em unidade de descrição menores, o que melhora a acuracidade da medida das variáveis mitigando os vieses, quando da agregação dos resultados em diferentes indivíduos, localizações, empresas.

Entretanto, ELHORST (2001; 2003) observa a ocorrência de certas dificuldades a partir da incorporação da componente espacial pela estrutura de dados em painel, sendo as principais: a dependência e a heterogeneidade espacial.

A primeira refere-se à possibilidade de existir, em diferentes momentos do tempo, dependência espacial entre as observações da mesma variável de interesse. O fato de a distância afetar o comportamento econômico é uma das principais razões para que observações com relacionamento espacial tenham dependência entre si, e isto pode alterar com o passar do tempo.

E a segunda refere-se à instabilidade estrutural, que faz com que os dados em painel tenham uma componente locacional, na qual os parâmetros não são homogêneos sobre o espaço, mas variam preferencialmente sobre as diferentes posições no espaço.

Tais colocações vão ao encontro da análise de WOOLDRIDGE (2002) sobre a influência da dependência espacial, ocorrência típica em dados *cross section*, visto que a coleta de dados amostrais relacionados à população pode seguir vários tipos de estratificações: cidades, estados, províncias. Neste caso, os resultados das unidades vizinhas provavelmente são correlacionados e se a correlação acontece principalmente

com as variáveis explicativas (ao contrário das não observáveis), na prática, nada precisa ser feito (embora a análise assintótica possa ser complicada).

De fato, algumas vezes as co-variáveis para uma região, aparecem como variáveis explicativas na equação para unidades vizinhas, como uma forma de capturar os efeitos associados.

Observando o interesse espacial sobre os dados em painel a equação em (2.23) pode ser re-escrita inicialmente de forma agrupada conforme:

$$Y_{it} = \beta X + \varepsilon_{it} \quad (2.24)$$

Na qual as observações podem ser estaqueadas como sucessivas *cross sections* para  $t = 1, \dots, T$ , referidas como  $Y_t$  (vetor  $N \times 1$  observações *cross sections* para  $t = 1, \dots, T$ ),  $X_t$  (matriz  $N \times K$  observações *cross sections* para as variáveis explicativas em  $t$ ) e  $\varepsilon_t$  (vetor  $N \times 1$  *cross sections* de erros para  $t$ ) e escritas na forma de (ANSELIN e LOZANO-GRACIA, 2008):

$$Y_t = \beta X_t + \varepsilon_t \quad (2.25)$$

Onde,  $Y_t = (Y_{1t}, \dots, Y_{Nt})'$ ,  $X_t = (X'_{1t}, \dots, X'_{Nt})'$  e  $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$ .

Entretanto, conforme ELHORST (2003) a principal objeção para este modelo é que não apropria a heterogeneidade espacial. As unidades espaciais podem diferir em suas variáveis de base, que são geralmente específicas ao espaço e sem variação no tempo, mas afetam a variável dependente e são difíceis de medir ou de obter.

Para tal uma solução é introduzir uma variável intercepto  $\mu_i$  representando os efeitos não observáveis que são particulares a cada unidade espacial considerada. Desta forma a equação em (2.25) pode ser acrescida desta variável e, assume a forma de,

$$Y_t = X_t \beta + \mu + \varepsilon_t \quad (2.26)$$

Onde,  $Y_t = (Y_{1t}, \dots, Y_{Nt})'$ ,  $X_t = (X'_{1t}, \dots, X'_{Nt})'$ ,  $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Nt})'$  e  $\mu = (\mu_1, \dots, \mu_N)'$ .

Na verdade a solução apontada por ELHORST (2003) segue a orientação básica pra tratar efeitos não observáveis em dados em painel.

Observando WOOLDRIDGE (2002) verifica-se que o termo proposto em (2.26) recebe uma série de outras determinações, quando da classificação dos efeitos não observáveis: componente não observável, variável latente e heterogeneidade não observável.

Ou seja, a proposição de ELHORST (2003) mantém a estrutura (2.26) em linha com as possíveis formas para estimar os parâmetros de um modelo de dados em painel:

a partir de método *pooled* OLS, métodos de efeitos randômicos e métodos de efeitos fixos.

Neste caso, para os efeitos fixos introduz-se uma variável do tipo *dummy* para cada unidade espacial como medida à variável intercepto; e nos efeitos aleatórios o intercepto é tratado como uma distribuição i.i.d com  $\mu = 0$  e variância  $\sigma_\mu^2$ . Além disto, assume-se que as variáveis randômicas  $\mu_i$  e  $\varepsilon_t$  são independentes.

Conforme ELHORST (2003), o modelo com um intercepto acomoda a heterogeneidade espacial até certo ponto; o problema permanece se os dados em tal modelo estão corretamente aninhados. Quando a heterogeneidade espacial não é capturada completamente pelo termo intercepto, a generalização permite que os parâmetros das regressões variem. Neste caso, eles podem igualmente ser considerados fixos ou aleatórios, e distribuídos entre unidades espaciais.

Se os parâmetros são fixos, mas diferem através das unidades espaciais, cada uma pode ser tratada de forma separada; ou seja, caso a equação  $Y_i = X_i\beta_i + \varepsilon_i$  seja uma dentre  $N$  equações empilhadas pela unidade espacial no tempo, a única forma de relacioná-las é assumindo correlação entre os termos de erro destas equações. Este fenômeno é conhecido como correlação contemporânea do erro, e o conjunto de  $N$  equações pode ser representado como:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \quad (2.27)$$

Na qual,  $E(\varepsilon_i) = 0$ ,  $E(\varepsilon_i \varepsilon'_j) = \sigma_{ij}^2 I_i$  e  $i, j = (1, \dots, N)$ .

Este modelo é conhecido como modelo de regressões aparentemente relacionadas (*Seemingly Unrelated Regressions Model* – SUR).

Para maior detalhamento sobre a aplicação dos modelos de efeitos fixos e aleatórios, bem como o modelo SUR na solução de regressões para dados em painéis, sugere-se a leitura de ANSELIN (1988), BALTAGI *et al.* (2009a) e WOOLDRIDGE (2002).

Com relação ao modelo em (2.26), cabe ressaltar que este contempla somente efeitos espaciais específicos, fixos ou aleatórios; não aborda os efeitos das interações espaciais, mais particularmente os efeitos espaciais relacionados ao processo de resíduos espacialmente correlacionados, o *spatial error*, e o relacionado com o atraso espacial da variável de interesse, o *spatial lag*.

Na verdade, mesmo com uma literatura pouco documentada, os modelos econométricos para dados em painel, com processos de *spatial error* e *spatial lag*, têm apresentado um crescente número de desenvolvimentos nos últimos anos.

Além das contribuições citadas de BALTAGI (2001), ELHORST (2001, 2003), ANSELIN e LOZANO-GRACIA (2008), outros textos relevantes sobre o tema são: KAPPOR *et al.* (2007), BALTAGI *et al.* (2009a e 2009b), BALTAGI e PIROTTE (2010), FLINGETON e LE GALLO (2008) e FLINGETON (2008); porém a contribuição mais utilizada como base para a econometria espacial é ANSELIN (1988).

Entretanto, esta documentação usual é pouco amigável, pois contém siglas e representações matemáticas não padronizadas e que dificultam o entendimento; por isto, nesta tese adotar-se-á a estruturação dos modelos seguindo a proposição de ELHORST (2010), por ser mais aderente às práticas econométricas aplicadas a dados em painel e, por apresentar o desenvolvimento de *tool box* que permite sua utilização.

Quando se especificam as interações entre as unidades espaciais (UE), o modelo pode apresentar uma variável dependente atrasada espacialmente ou um processo autoregressivo espacial nos resíduos, conhecidos respectivamente como modelo de atraso espacial e modelo de erros espaciais.

Cabe ressaltar que, a AE é a área total onde o fenômeno será estudado, as UGR são as possíveis formas de representar a AE; ou seja, os municípios com seus bairros e setores censitários e, a UE é um objeto (área) dentro da UGR, por exemplo, um bairro.

No modelo de atraso espacial a variável de interesse depende das observações da própria variável nas unidades vizinhas e de um conjunto de características locais observadas através das variáveis explicativas. Assim como os modelos para dados *cross section* o modelo de atrasos espaciais é denominado *SAR Panel*.

No modelo espacial de erros correlacionados, a variável de interesse depende de um conjunto de características observadas localmente e os termos de erro estão correlacionados através do espaço. Do mesmo modo que os modelos para dados *cross section*, o modelo de erros correlacionados é denominado *SEM Panel*.

A extensão do modelo em (2.26) para as variantes com inter-relação espacial passa classicamente pela inclusão do termo de associação espacial contendo a matriz  $W_{ij}$ , que no caso espacialmente atrasado, o *SAR Panel*, assume a forma de:

$$Y_t = pWY_t + X_t\beta + \mu + \varepsilon_t, \quad (2.28)$$

E no caso do erro espacialmente correlacionado, o SEM *Panel*, é reescrito como:

$$Y_t = X_t\beta + \mu + \varphi_t, \quad \varphi_t = pW\varphi_t + \varepsilon_t \quad (2.29)$$

Na qual  $E(\varepsilon_t) = 0$ ,  $E(\varepsilon_t\varepsilon_t') = \sigma^2 I_N$ .

No SAR *Panel* o termo de atraso espacial tem a seguinte composição:

$$pWY_t = p \sum_{j=1}^N W_{ij} y_{jt} \quad (2.30)$$

No SEM *Panel* o termo correlacionado espacialmente tem a seguinte forma:

$$pW\varphi_t = p \sum_{j=1}^N W_{ij} \varphi_{jt} \quad (2.31)$$

Na consulta a ANSELIN *et al.* (2008) e BALTAGI *et al.* (2009b) percebe-se que o modelo de erros correlacionados suporta a aplicação de uma estrutura de médias móveis. Neste caso o modelo é conhecido como modelo espacial de erros por média móvel (*Spatial Moving Average Error Model – SMA*); entretanto, não se encontrou desenvolvimento de algoritmos de resolução para este modelo.

#### 2.5.3.6 – EMPREGO DE ANÁLISE ESPACIAL NA PREVISÃO DE DEMANDA DE GN

Quanto ao problema de previsão de demanda de GN, não se encontrou aplicação com emprego das abordagens descritas neste tópico. Também não foram observados exemplos de desenvolvimentos com uma linha baseada em dependência espacial e modelos autoregressivos com o foco na previsão de demanda de energia no senso geral.

Neste caso, observou-se de forma comum o seu emprego na previsão de carga de energia através da redução da AE em pequenas áreas, para coleta de informações de consumo e carga e, para limitar a área de operação de equipamentos (WILLIS, 2002).

Outro ponto abordado nos modelos para previsões de carga com cunho espacial é o uso do solo; para o qual verifica-se como as regiões se desenvolvem, pelo prisma da industrialização, adensamento imobiliário, entre outros (WILLIS, 2002). De certo tal situação ocorre em decorrência da redução da AE, pois não prever o comportamento de uma determinada célula, ou área, pode inviabilizar a previsão apresentada.

Tome-se como exemplo, uma região tipicamente residencial horizontal, que no tempo  $t_0$ , absorve contornos de verticalização. O crescimento nesta área pode se tornar muito maior do que o projetado uma vez que espacialmente ocorre o efeito de adensamento. Entretanto, o que se observa sobre o assunto é que os exemplos citados não atendem aos conceitos dispostos neste tópico relativos à dependência espacial.

## 2.6 – CONSIDERAÇÕES DA FUNDAMENTAÇÃO TEÓRICA E DIRETRIZES À TESE

O objetivo desta tese é o desenvolvimento de um modelo de previsão de demanda de GN, aplicado à atividade de distribuição, tendo como abordagem a previsão pela perspectiva espaço-temporal.

Nesta via, o cerne da proposição é o emprego da informação espacial na caracterização das relações entre os fenômenos no tempo e no espaço; porém, antes de apresentar a metodologia aplicada ao problema, realizou-se a definição da estrutura do modelo empregado nesta tese.

Com relação a tal questão, fruto da revisão bibliográfica, tem-se o seguinte cenário acerca dos métodos de previsão aplicados ao consumo de energéticos:

- Os métodos baseados em séries temporais são os mais empregados em previsão, independente do setor, do segmento e do período de previsão;
- O AM tem maior direcionamento às questões de curto prazo e à substituição dos métodos de séries temporais em problemas típicos da operação;
- A econometria espacial enfatiza a análise da dependência e heterogeneidade espacial, bem como a caracterização dos relacionamentos espaciais, sejam estes em uma visão *cross section*, ou para dados em painel;
- Na prática, entretanto, a dimensão espacial tem sido empregada na apresentação dos resultados, ou seja, para acompanhar e representar as variáveis ou resultados no tempo, com sua respectiva projeção no espaço;
- Fruto disto, problemas de ordem temporal e resolução espacial, entre outros, são subtraídos à análise e composição da solução do problema;
- Tal situação estende-se aos métodos de análise espacial, nos quais a tônica é o emprego em uma visão estática, ou seja, que apresenta e analisa dados em uma visão unicamente espacial desintegrada do tempo;
- Como exemplo, cita-se a previsão de carga de EE, na qual seu emprego busca mitigar questões associadas ao uso do solo, e não no sentido de avaliar o consumo e a estruturação da distribuição ao longo do tempo; e,
- Apesar das possíveis combinações com outros métodos, como a taxonomia levantada por FISCHER (2006), especificamente não se encontrou a aplicação da dimensão espacial voltada à previsão de demanda do setor de EE e GN observando a análise da dependência espacial, ou ainda pelo emprego de modelos de regressões espaciais, base da econometria espacial.

Com isto concluiu-se que a questão espacial pode ser aplicada em problemas de previsão de forma mais estruturada, nos quais sua utilização na modelagem de demanda tangencie questões de resolução, escala, dependência espacial, nível de representação e agregação, entre outros.

Entretanto, para seu emprego, identifica-se na revisão bibliográfica duas questões principais quanto à econometria espacial e a dependência e a heterogeneidade espacial:

- A natureza dos dados aplicados ao problema; e,
- O termo residual dos modelos espaço-temporais.

O primeiro está relacionado com a disponibilidade, forma e conteúdo de dados para emprego na caracterização do fenômeno visando os processos de estimação e previsão. Já o segundo está relacionado com a forma com que são apropriados ao modelo os efeitos dos processos espaciais, na forma de um atraso espacial da variável dependente ou como erros espacialmente autocorrelacionado.

#### **2.6.1 – A NATUREZA DOS DADOS APLICADOS AO PROBLEMA**

Um ponto inicial sobre a distribuição da rede de GN é que esta ocorre essencialmente no espaço urbano. Com isto, espera-se que a apresentação dos resultados de um processo de previsão esteja em uma UGR adequada ao estudo da distribuição da rede dentro de um espaço qualquer de interesse.

Outro ponto, é que a natureza da previsão prevê o emprego de variáveis sócio-econômicas, que em sua maioria, encontram-se em séries temporais apresentadas em unidades de análise delimitadas por polígonos fechados, ou seja, em áreas.

Na verdade estes dois pontos são os principais direcionadores para a seleção e a coleta dos dados candidatos ao emprego no desenvolvimento, que em síntese apresentam a seguinte composição, visto as características estruturais dos também modelos candidatos à solução do problema, citados na revisão bibliográfica:

- Feições espaciais, por ponto ou polígono, para representar no espaço, o posicionamento e o relacionamento dos objetos de interesse; e,
- O conjunto de dados relativo à variável dependente, objeto do estudo; e de variáveis explicativas que representam as características da atividade econômica, demografia, consumo e do preço de energéticos.

Especificamente, no formato pontual encontra-se a representação do consumo de GN e, no formato de polígonos, a representação das variáveis explicativas em nível municipal.

A análise desta composição indica que a variável dependente não se encontra no mesmo nível geográfico de discretização das variáveis explicativas, independente do intervalo temporal da série de dados. A solução deste inconveniente passa por realizar um dos seguintes procedimentos:

1. As variáveis explicativas, com representação espacial poligonal, devem ser desagregadas em unidades de representação do tipo pontual via processo de interpolação; ou,
2. A variável dependente deve ser associada aos seus respectivos pontos no espaço, para permitir sua correta contabilização em uma unidade de área de maior escala.

Com relação à proposição 1 acima, cabe ressaltar que apesar de existir a possibilidade de gerar dados na mesma escala de resolução da variável dependente a partir do emprego de métodos de interpolação geoestatística ou por ponderação geográfica, como por exemplo, a GWR, tal opção não pode ser adotada pelos seguintes motivos:

- 1a) Como as variáveis explicativas encontram-se representadas em áreas, para gerar uma superfície seus valores são primeiramente atribuídos ao centróide da feição, o que não representa o correto balanço de massa da ocorrência da variável ao longo do espaço. Por isto tal estratégia é empregada para gerar dados no mínimo na mesma escala de resolução e não para escalas maiores;
- 1b) Para a correta apropriação dos dados representados em áreas, é necessário um conjunto amostral em formato de grade de pontos com a ocorrência da variável principal, ou de outra variável com alta correlação, para servir de base para a interpolação; ainda assim, tal grade deve ter amostras temporais, para permitir a correta alocação da variável de interesse também nesta perspectiva;
- 1c) Além do exposto, algumas variáveis econômicas ou de consumo, por exemplo, os dados de preço de combustíveis, não apresentam variações ao longo do espaço porque são melhor descritas pela existência ou não da concorrência de outro energético; e,
- 1d) Especialmente no caso dos dados econômicos coletados para esta proposição, apesar de alguns encontrarem-se em frequência mensal, são representativos no nível de país e da unidade da federação, o que produziria uma interpolação com vieses, visto a representação espacial global destes eventos, contra uma representação espacial local do consumo do GN.



Com relação à proposição 2 é possível realizar a contabilização dos dados pontuais em dados de área. Porém, seu nível de discretização deve estar abaixo do nível municipal, preferencialmente em um nível que permita representar os resultados ao longo do espaço delimitado pela UGR de maior representação, por exemplo, em bairros, regiões administrativas, setores censitários ou outra composição.

Em decorrência do que foi discutido neste tópico têm-se as seguintes orientações à escolha do modelo:

(A) **Modelos Multivariados** – visto à característica da previsão de demanda de energéticos empregar variáveis sócio-econômicas;

(B) **Causalidade de Granger**<sup>29</sup> – visto a impossibilidade de reduzir o universo amostral das variáveis explicativas à mesma UGR da variável dependente, não é possível avaliar o mapeamento de  $x \rightarrow y$  e vice-versa; sendo assim não é possível avaliar a causalidade de Granger e adotar modelos que utilizam realimentação;

(C) **Matrizes de Contiguidade Espacial** – ainda em decorrência da impossibilidade de compatibilizar a discretização dos universos amostrais das variáveis, explicativas e dependente, também não é possível gerar matrizes  $W_{ij}$  com a mesma estrutura espacial. Isto reduz a capacidade de mapeamento e comparação dos valores da dependência espacial entre variáveis de naturezas distintas.

(D) **Modelos Lineares** – os modelos apresentados em 2.5.3.3 e 2.5.3.5 seguem prioritariamente a realização de mapeamentos lineares, baseados em modelos de regressão linear baseados em estimadores do tipo OLS, para estimar os parâmetros do modelo. Com isto devem-se relaxar as hipóteses de:

- Linearidade nos parâmetros e regressores não estocásticos;
- A expectativa de média zero dos distúrbios;
- Homocedasticidade;
- Ausência de autocorrelação dos resíduos no tempo;
- Ausência de correlação entre regressores e os distúrbios; e,
- Existência de um  $n^\circ$  de observações maior do que o  $n^\circ$  de regressores.

---

<sup>29</sup> A causalidade de Granger é um conceito estatístico da causalidade que é baseada na predição. De acordo com a causalidade de Granger, se um sinal  $x_t$ ; "Granger-causes" (ou "G-causes") um sinal  $y_t$ , então valores do passado de  $x_t$  devem conter informações que ajudem a prever  $y_t$ , além dos valores passados contidos unicamente em  $y_t$ . Sua formulação matemática é baseada na modelagem da regressão linear de processos estocásticos (Granger 1969).

### 2.6.2 – O TERMO RESIDUAL

Por uma questão de normatização, a parcela  $\varepsilon_i$  será intitulada resíduo ou erro do modelo. Sendo assim não se empregará o termo distúrbio a partir deste ponto. Tal colocação se faz necessária para mitigar o risco da confusão de entendimento entre o variado emprego da palavra erro nas literaturas temporal e espacial.

Um ponto importante na discussão sobre os resíduos é que estes podem conter, em um modelo puramente temporal, parcelas relacionadas com a influência espacial da distribuição das variáveis, explicativas e dependentes, empregadas no modelo.

Entretanto, conforme abordado por DUBIN (2004), de um ponto de vista teórico, um modelo econométrico pode conter ambos os resíduos: provenientes do atraso espacial da variável dependente e de erros espacialmente correlacionado.

DUBIN (2004) ressalta que na prática tais modelos são raramente usados, visto a dificuldade de estimá-los, uma vez que a matriz  $W_{ij}$  empregada em ambos os termos de erro, *spatial lag* e *spatial error*, geralmente é a mesma.

De certa forma, seria possível conceber que um modelo econométrico espaço-temporal, como os modelos da família autoregressiva apresentados na revisão bibliográfica, teriam uma estrutura de resíduos com uma composição próxima de:

$$\varepsilon_i = \varepsilon_T + \varepsilon_E, \quad (2.32)$$

Na qual,

$\varepsilon_T$  - parcela de resíduos relativos efeitos temporais; e,

$\varepsilon_E$  - parcela de resíduos relativos efeitos espaciais.

Esta separação seria uma forma muito clara de perceber o problema caso realmente fosse possível tratar as estruturas dos modelos espaciais assim como são tratadas as estruturas dos modelos temporais, como por exemplo, na resolução de questões como a heterocedasticidade e a autocorrelação. Na prática, os testes para a especificação de modelos espaciais tomaram um caráter não-acoplado; ou seja, o pesquisador tenta determinar que modelo de erro, *spatial lag* e *spatial error*, melhor ajusta os dados.

Ou seja, em modelos onde os resíduos apresentam dependência espacial, existem abordagens para testar se a correlação espacial está presente nos resíduos, ou o quanto a inclusão do termo *spatial lag* elimina a dependência espacial dos resíduos do modelo.

No caso desta tese a questão do resíduo é amplificada pelo fato de que as variáveis, explicativas e dependente, estão em UGR com níveis de resolução espacial

diferentes, impossibilitando a construção de matrizes  $W_{ij}$  que permitam avaliar o atraso espacial do conjunto de variáveis explicativas.

Em função das colocações introduzidas neste tópico, são estabelecidas as seguintes orientações complementares às introduzidas no tópico 2.6.1:

(E) – **Dependência Espacial das Variáveis Explicativas** – com o intuito de reduzir os efeitos espaciais das variáveis explicativas no modelo, recomenda-se analisar a sua dependência espacial antes de efetuar a escolha; privilegia-se assim, no aspecto temporal, as de maior correlação com a variável dependente e no aspecto espacial, as de menor dependência espacial; com isto reduz-se o risco de introduzir variáveis explicativas que adicionem dependência espacial ao modelo.

(F) – **Modelo Espacial** – na seleção do modelo deve-se atentar para a incapacidade, imposta por (C), de mapear, nos resíduos, os efeitos espaciais relacionados com o conjunto de variáveis explicativas.

### 2.6.3 – ESCOLHA DO MODELO

A partir da análise das proposições (A), (B), (C), (D), (E) e (F) apresentadas nos tópicos anteriores, foi possível avaliar os seguintes modelos / família de modelos que não atendem às condições impostas por tais proposições:

- Os modelos espaciais com estrutura do tipo média móvel (SMA), combinados ou não com estruturas autoregressivas (SARMA, STARMA), não podem ser empregados pela impossibilidade gerada pela proposição (C). Esta limitação impede a avaliação dos resíduos com o objetivo de estabelecer se os atrasos estão associados a efeitos espaciais ou a efeitos temporais; e desta forma gerariam uma componente de atraso tendenciosa;
- Os modelos espaciais com estruturas de vetores: autoregressivos, ou *Spatial Vector Autoregressive* (SpVAR), bem como os modelos baseados em autômatos celulares, não podem ser empregados pela impossibilidade gerada pela proposição (B), visto a necessidade da re-alimentação  $y \rightarrow x$ ;
- No tocante aos modelos espaciais dinâmicos, com base em autômatos celulares, entende-se que estes evoluem satisfatoriamente em processos de saturação, como por exemplo, o uso do solo; para isto seria necessário definir os limites de saturação de consumo de GN em cada UE de interesse, de forma a identificar como o processo evolui e, quais são as características que determinam uma mudança de estado dentro e, entre as UE.

- Os modelos espaciais de estrutura puramente autoregressiva (FAR), baseados na forma (2.17), em um primeiro momento não devem ser empregados porque a natureza dos modelos de previsão de demanda de energéticos não está baseada no mapeamento exclusivo da variável dependente. Condicionante introduzida pela proposição (A);
- Apesar de serem modelos da família autoregressiva espacial com estrutura que permite mapeamento de variáveis exógenas, os modelos SEM e SDM, respectivamente (2.13) e (2.15), não podem ser empregados, devido à incapacidade de gerar matrizes  $W_{ij}$  para as variáveis explicativas na mesma resolução da matriz para a variável dependente. Incapacidade esta gerada pela proposição (C);

Face à análise e às condicionantes apresentadas, propõe-se aplicar nesta tese o modelo de estrutura SAR *Panel*, visto a característica de previsão temporal com base em séries históricas de consumo de GN e, sua estrutura apresentar melhor consolidação frente às condicionantes impostas pelas proposições (A) até (F), ou seja:

(A) – é um modelo multivariado, que permite privilegiar a seleção e estimação a partir de variáveis explicativas com melhor correlação com a variável dependente, no tocante a questão temporal. Sua estrutura é similar aos modelos de estrutura ARX, autoregressivos com adição de um termo onde são incluídas as variáveis explicativas.

(B) – não necessita re-alimentação;

(C) – aplica uma única matriz  $W_{ij}$  relacionada com a variável dependente;

(D) – admite mapeamento linear;

(E) – por sua característica estrutural é possível incluir no termo regressivo  $X\beta$ , variáveis explicativas que apresentem menor dependência espacial, mitigando assim do vetor dos resíduos a parcela relativa à autocorrelação espacial da variável explicativa; e,

(F) – mitiga do vetor dos resíduos a parcela relativa à autocorrelação espacial da variável dependente, pois maximiza parte dos atrasos espaciais, *spatial lag*, na estimação do modelo (estimação de  $p$ ).

### 3 - METODOLOGIA

Neste capítulo descreve-se detalhadamente o método empregado na solução do problema identificado; em especial, aprofunda-se a análise da dependência espacial para constituição do modelo, bem como o desenvolvimento e a previsão, a partir do modelo SAR *Panel*. Para tal, encontra-se dividida em cinco etapas:

- Estruturação e análise de dados;
- Diretrizes para a modelagem;
- Desenvolvimento e previsão do modelo SAR *Panel*;
- Testes do modelo e análises dos resultados; e,
- Comparação com outras estruturas de previsão.

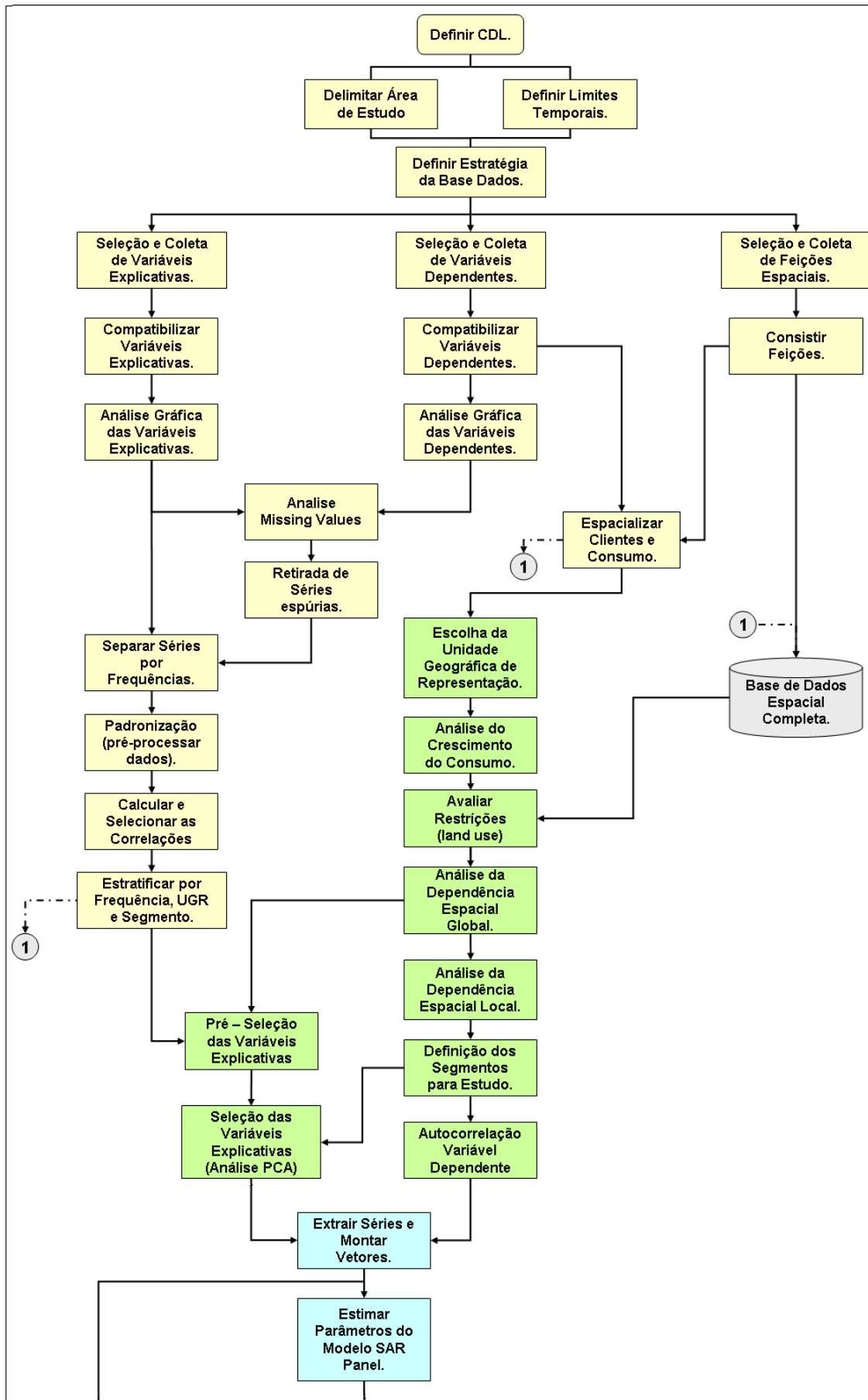
A primeira etapa engloba as atividades relacionadas com a coleta e compatibilização dos dados referentes às variáveis dependentes, independentes e espaciais. Nesta fase foram realizadas, entre outras, a compatibilização, tratamento e organização dos dados, a retirada de séries de dados espúrias e a montagem do banco de dados.

Nas diretrizes para a modelagem, à luz dos dados alfanuméricos e espaciais obtidos, foram realizadas as análises de crescimento espacial do consumo e dos segmentos de negócio que apresentam nítida influência espacial, a análise do uso do solo e a redução da dimensionalidade do conjunto original de variáveis independentes.

A terceira parte trata da metodologia para o desenvolvimento do modelo propriamente dito, que englobam toda a manipulação associada à confecção deste até a análise dos resultados obtidos com a proposição. Por último, descrevem-se as atividades relacionadas com os testes do modelo, a análise dos resultados e a comparação com outros modelos e previsão de caráter linear e não-linear.

Na figura 3.1 apresenta-se, para uma visão completa do processo, o fluxo da metodologia executada nesta tese. Neste fluxo, para identificar as cinco etapas citadas, aplicou-se um padrão diferenciado de cores para o conjunto de atividades associadas em cada uma das etapas da metodologia. Sendo assim:

- Amarelo - Estruturação e análise de dados;
- Verde - Diretrizes para a Modelagem;
- Azul - Desenvolvimento e previsão do modelo SAR *Panel*;
- Branco - Teste e análise dos resultados; e,
- Ouro - Comparação com outras estruturas de previsão.



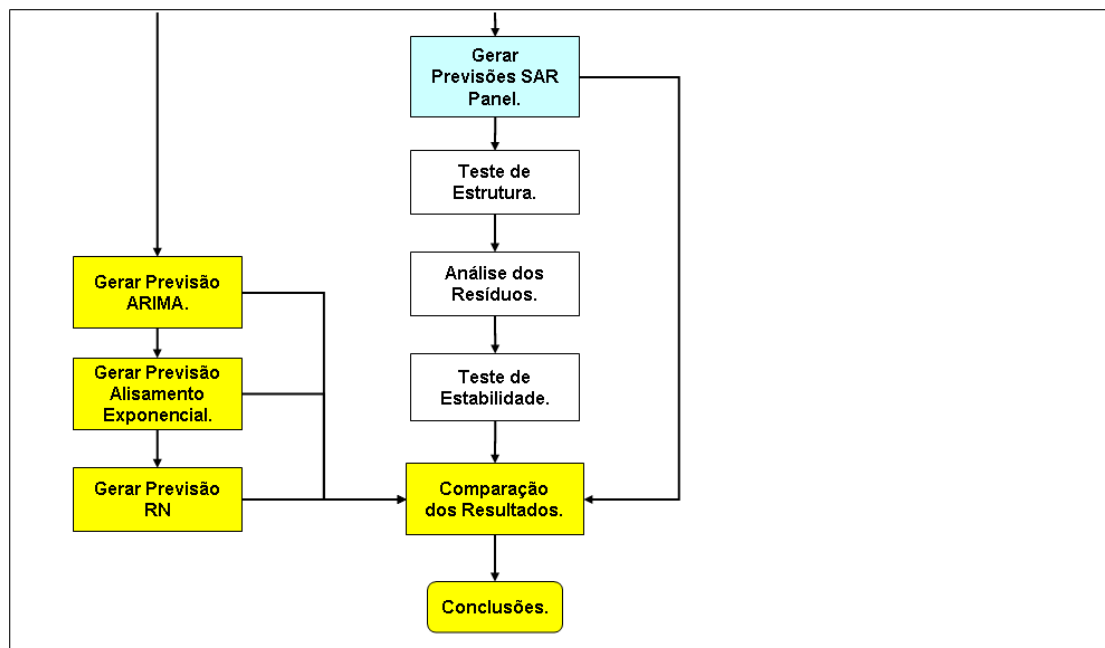


Figura 3.1 – Fluxograma da Metodologia Proposta.

### 3.1 - ESTRUTURAÇÃO E ANÁLISE DE DADOS

A tese inicia-se pela definição da CDL e os seus respectivos limites temporal e espacial. O primeiro relacionado com a frequência e o período de dados disponíveis à variável de interesse, neste caso o consumo de GN. O segundo relacionado com o espaço geográfico, onde se encontram os limites da AE.

Os critérios para escolha da CDL são a análise do consumo de GN, a partir dos dados dos clientes e, a existência de um Mobiliário Urbano Básico<sup>30</sup> (MUB), que permita a espacialização dos clientes no espaço.

O consumo de GN foi tabulado pelo volume total anual de GN consumido, por município da área de concessão da CDL. Após a sua apropriação, delimitou-se a AE com base na avaliação conjunta da:

- Sobreposição entre a área de concessão e as regiões economicamente contiguas; e,
- Representatividade econômica do município atendido dentro desta área contigua a partir do percentual do PIB individual de cada município.

Em seguida definiu-se a construção do BDE por meio da estratégia dual de relacionamento. Nesta abordagem as feições no terreno, ou objetos, são definidas pela sua localização e formato, bem como por uma série de características não geométricas.

<sup>30</sup> MUB é a base cartográfica contendo as informações do cadastro dos equipamentos urbanos. Sua estrutura varia de acordo com as necessidades de aplicação dentro das esferas da gestão pública. É a base para espacializar os clientes; condição necessária para a correta agregação dos valores de consumo de GN em cada CDL.

Estas características são associadas à geometria dos objetos por meio de identificadores (*object identifier*) que associam os dados temáticos aos geométricos. O BDE contemplou duas dimensões de dados:

- Espaciais – relativos às feições que identificam a AE, seus municípios e UGR de menor resolução, bairros e setores censitários, a posição dos clientes de GN e a infra-estrutura de distribuição da CDL; e,
- Alfanuméricos – variáveis, dependente e explicativas, que descrevem o consumo de energéticos, a atividade econômica e a demografia, no tempo.

Após definida a estratégia do BDE, iniciam-se as atividades de seleção e coleta de dados, alfanuméricos e espaciais. As feições espaciais foram coletadas e consistidas quanto à sua qualidade topológica e referencial, para garantir a correta indexação e espacialização dos clientes, reduzindo assim os possíveis problemas de precisão; também foi realizada a correta indexação dos dados alfanuméricos.

Especificamente na consistência espacial da UGR município, empregada para indexar as variáveis explicativas, atentou-se para as variações na forma ao longo do tempo e, por conseguinte às possíveis alterações nas relações espaços-temporais. Conforme abordado no tópico 2.4.3.3, as alterações ocorridas ao longo do tempo, trazem conseqüências à representação dos fenômenos espaciais e ao relacionamento espacial entre as séries de variáveis.

As atividades de validação espacial foram realizadas empregando o ArcGIS, ferramenta SIG da empresa ESRI (*Environmental Systems Research Institute*).

Os dados de clientes foram espacializados a partir de seus endereços. Para isto utilizou-se o procedimento denominado geocodificação, que determina a localização espacial a partir de um endereço, conforme descrito no Apêndice B.

O BDE foi desenvolvido empregando o software ArcGIS, para armazenar, na forma de tabelas, os atributos alfanuméricos dos objetos geográficos e arquivos para guardar as representações geométricas destes objetos.

Quanto aos modelos de previsão de demanda de energia, que em geral, empregam variáveis de natureza: operacional, preço e econômicas; foram coletadas as seguintes séries de dados com frequências mensal e anual:

- Variável dependente - consumo de GN, originado dos registros históricos de consumo das respectivas CDL's; e,
- Variáveis explicativas - consumo e preço dos energéticos (energia elétrica, álcool e derivados de petróleo), atividade econômica e demografia; obtidos



prioritariamente em fontes públicas, com exceção do consumo de energia elétrica, originado dos históricos de consumo da Copel e da Escelsa.

Ressalta-se que as variáveis climáticas foram descartadas pela não utilização, no Brasil, do GN em aquecimento, ou em refrigeração, visto a indisponibilidade no mercado nacional de equipamentos movidos a GN para tal finalidade.

As séries de dados coletadas foram compatibilizadas em variáveis individuais; bem como foram geradas variáveis totalizadoras por: município, frequência e AE. Nesta coleta, as séries de preço e demografia foram interpoladas, a partir do emprego de regressões de forma a compatibilizar a frequência anual e a mensal. Após a compatibilização retiraram-se as variáveis que continham valores nulos ou faltantes (*missing values*), reduzindo assim o conjunto final de variáveis.

As variáveis foram catalogadas conforme a AE e a frequência. Após esta organização, com base no número de eventos disponíveis por série de dados, efetuou-se a normalização das mesmas a partir da subtração da média e divisão pela variância.

Para identificar as variáveis explicativas que melhor explicam o consumo, após a normalização estabeleceu-se a matriz de correlação cruzada empregando a ferramenta *Statistics / Basic Statistics / Correlation Matrices*, do software *Statistica 7.0*, na qual para cada par de variáveis determinou-se a correlação, a determinante, a potência da correlação e o número de eventos úteis considerados.

Com base nas correlações determinou-se a frequência a ser empregada no desenvolvimento do modelo, neste caso mensal, pois com esta configuração as variáveis explicativas apresentaram as maiores séries de dados disponíveis na mesma frequência da variável de interesse.

## **3.2 - DIRETRIZES PARA A MODELAGEM**

Após a determinação da frequência da série de dados a ser utilizada na tese procedem-se nesta ordem: a definição da UGR que será empregada para caracterizar o consumo, a análise da dependência espacial, a seleção das áreas para aplicação do modelo e a definição das variáveis explicativas empregadas no modelo.

### **3.2.1 - DEFINIÇÃO DA REPRESENTAÇÃO ESPACIAL DO CONSUMO**

A definição da UGR para representar o consumo se dá pela identificação do padrão espacial da distribuição dos clientes na AE (aleatório, agrupado ou regular). Para tal, com o uso das ferramentas de classificação e consulta do ArcGIS construíram-se

mapas com a posição espacial dos clientes por tipo de segmento de consumo, e gráficos com a evolução do número de UE com clientes de GN.

Com isto é possível identificar as características básicas do processo espacial na AE, o que auxilia na definição da UGR que será empregada para avaliar a dependência espacial do consumo de GN; ou seja, avaliar na relação entre o consumo de GN e as UGR, os arranjos que podem apresentar autocorrelação espacial. Além disso, permite reavaliar o próprio limite territorial da AE.

Por exemplo, se o número de clientes por área  $C$ , do segmento  $S$ , cresce nas áreas  $i$  e  $j$ , sendo  $i \neq j$ , entre os tempos  $t_0, t_1, \dots, t_n$ , pode-se assumir que existe um crescimento da base de clientes ao longo de todo o espaço, e que este crescimento pode ter algum nível de relação entre si (dependência espacial). Porém, se isto acontece somente em uma das áreas,  $i$  ou  $j$ , ou ainda em nenhuma delas, pode-se assumir que o fenômeno é espacialmente clusterizado, e nesta região o crescimento tende a ser avaliado potencialmente por técnicas econométricas tradicionais.

Para lembrança a AE é a área total onde o fenômeno será estudado, as UGR são as possíveis formas de representar a AE, ou seja, os municípios com seus bairros e setores censitários e, a UE é um objeto (área) dentro da UGR, por exemplo, um bairro.

As UE têm relação direta com a matriz  $W_{ij}$ , pois representam cada uma das  $n$  áreas  $i \neq j$  do estudo, segundo a UGR adotada para determinar as relações espaciais; com isto, permitem a associação e emprego dos dados dos clientes no modelo.

Como já salientado, para definir a representação espacial do consumo, foram realizadas duas análises: do crescimento do número de UE com consumo de GN, e da disposição dos clientes versus as UGR definidas.

A análise do crescimento do número de UE com consumo de GN foi realizada por gráficos com a evolução do número de UE, por bairros e setores, ao longo do tempo. Nestes gráficos foram estipuladas três curvas contabilizando o número de UE contendo pelo menos 1, 2 ou 3 clientes. Com isto foi possível avaliar se o crescimento se deu também em número de clientes de GN por UE.

A análise da disposição foi realizada a partir de mapas temáticos com a representação da distribuição espacial dos clientes por UE, conforme a UGR e o segmento de consumo. Isto permite definir a malha de polígonos de UE, que representa de forma mais homogênea a distribuição dos clientes sobre o espaço ao longo do tempo; reduzindo a análise da dependência espacial somente à influência de UE relacionadas topologicamente, ou dentro de uma determinada distância.

Ainda com o software ArcGIS avaliaram-se as restrições espaciais referentes ao uso do solo e à área definida para a previsão, que podem ser de caráter:

- Ambiental - florestas, manguezais, lagos e lagoas, entre outros; e,
- Planejamento – diretrizes de planos urbanos, constituição de áreas de uso restrito, zonas industriais, entre outros.

### **3.2.2 - ANÁLISE DE DEPENDÊNCIA ESPACIAL GLOBAL E PRÉ-SELEÇÃO DE VARIÁVEIS EXPLICATIVAS.**

Após a avaliação da representação espacial do consumo, investigou-se a dependência espacial global para as UGR e segmentos caracterizados como propensos a tal dependência, por meio do índice de Moran Global conforme (2.7).

Esta análise, classificada como exploratória, tem o objetivo de entender as associações e instabilidades espaciais, sendo diretamente relacionada com a dependência e a heterogeneidade espacial.

Quanto à instabilidade estrutural, os índices globais não conseguem identificar as provenientes dos regimes espaciais presentes no conjunto total de objetos analisados. Ele é um indicador de existência de autocorrelação espacial, mas a forma com que esta instabilidade se apresenta só é possível de ser detectada com as análises locais.

Ressalta-se que o emprego de dados *cross section* conferem certeza às medidas de instabilidade, porém, no caso de dados em painel, podem ocorrer diferentes regimes espaciais ou mudança nestes ao longo do tempo (ELHORST, 2010). Por consequência é necessário avaliar o comportamento e a estacionariedade do índice ao longo do tempo, visto que a dependência espacial pode se alterar; tanto globalmente, pela pura mudança do índice, quanto localmente, pela mudança de regimes.

Cabe ressaltar que a condição necessária e suficiente para emprego de um modelo espaço-temporal é que a variável de interesse apresente dependência espacial estacionada. Com isto a abordagem de testar a posteriori a existência de dependência espacial nos resíduos, ou se a inclusão de um termo atrasado elimina a dependência espacial dos resíduos do modelo, conforme citado em 2.6.2, passa a ser complementar.

Apesar de um modelo econométrico espaço-temporal, como os modelos da família autoregressiva, apresentar resíduos que admitem em sua composição a presença de duas parcelas, relativas a efeitos temporais e espaciais; entende-se que, a manipulação nos dados, a modificação da estrutura do modelo ou mesmo a mudança do tipo do estimador, altera esta composição e conseqüentemente os resíduos esperados.

Sendo assim, a motivação inicial para empregar um modelo espacial é existir dependência espacial sobre o evento de interesse e não a apuração de uma possível dependência espacial após estimar um modelo no senso temporal.

O cálculo dos índices globais de autocorrelação espacial foi realizado com base no consumo de GN por segmento, e com a adoção de matrizes  $W_{ij}$  de vizinhança de 1ª ordem e por distância. Para efetuar os cálculos aplicaram-se o GEODA<sup>31</sup> e o ArcGIS, pelas seguintes razões:

- O GEODA consegue capturar, nas matrizes  $W_{ij}$  de 1ª ordem, as áreas que apresentam somente 1 nó em contiguidade. Esta matriz é intitulada *Queen*;
- O ARCGIS não consegue capturar este tipo de relacionamento. A matriz de 1ª ordem considera os polígonos que tenham face-nó-face em contiguidade; desta forma, os polígonos contíguos somente pelo nó não são interpretados como vizinhos. Este modelo de matriz é conhecido como *Rook*; e,
- Opcionalmente o ArcGIS fornece uma avaliação estatística do índice conforme sua variação em relação ao desvio padrão do valor esperado e do valor calculado, sendo útil para complementar a análise dos dados.

Nas matrizes baseadas em distância aplicou-se a medida euclidiana e o conceito de zona de indiferença, ou seja, a partir de uma distância  $d$  entre centróides, a UE não é mais considerada vizinha, independente de ser contígua. Com isto as matrizes por distância foram construídas considerando um valor mínimo que garanta ao menos 1 vizinho para cada UE.

Os resultados foram apresentados em gráficos com o comportamento do índice global ao longo do tempo. O índice foi calculado pelo GEODA, e para os casos em que os resultados obtidos não permitiram uma avaliação direta, complementou-se a avaliação com as análises qualitativas fornecidos pelo ArcGIS.

O caminho natural da tese seria proceder à análise da dependência espacial local, porém introduziu-se uma etapa intermediária referente a pré-seleção das variáveis explicativas, pelos seguintes motivos:

- Na escolha da UGR de representação, verificou-se a impossibilidade de aferir a dependência global para as variáveis explicativas, que associada à decisão de empregar exclusivamente dados em frequência mensal, reduziu o universo total de variáveis explicativas disponíveis; e,

---

<sup>31</sup> *Software free* desenvolvido pela equipe de Luc Anselin, no laboratório de análises espaciais da Universidade de Illinois, EUA.

- Como o volume de trabalho na análise de dependência espacial local é grande, preferiu-se verificar antes se existem variáveis explicativas para empregar no desenvolvimento do modelo, para os segmentos indicados a esta análise.

Com isto, foram selecionadas do conjunto total de correlações aquelas que relacionam exclusivamente a variável de interesse com as explicativas, com frequência mensal e, pertençam aos segmentos com propensão à existência de dependência, conforme a análise global de dependência espacial.

Neste conjunto expurgaram-se as variáveis com índice entre  $\pm 0,8$  de correlação temporal, mantendo-se assim as correlações fortes ou extremamente fortes e o conjunto de variáveis resultante foi priorizado segundo critérios qualitativos e quantitativos.

Para esta priorização as variáveis foram separadas em três blocos, por segmento de mercado e AE, nos quais os dois primeiros contêm variáveis macroeconômicas, no nível do país e da UF e o terceiro as variáveis de consumo e preço estratificados no nível municipal. Ressalta-se que estes grupos foram definidos em função da existência de dados específicos para alguns níveis de discretização da administração pública.

A qualificação seguiu prioritariamente a linha de análise econômica, buscando encontrar critérios que refletissem os possíveis impactos / significados da variável analisada com relação ao consumo de GN. Assim, adotou-se aleatoriamente um número de quinze variáveis em cada segmento, distribuídas preferencialmente de forma igualitária entre os três conjuntos; ou seja, cinco variáveis econômicas nos níveis do e da UF e, cinco variáveis de consumo e preço no nível municipal.

Este número de quinze variáveis é empírico, condicionado puramente à organização da tese, pois em cada grupo de variáveis resultante da aplicação dos critérios de priorização, existem pelo menos mais de cem variáveis explicativas disponíveis. Como exemplo, na tabela 3.1, apresenta-se a classificação e quantificação realizada para o consumo de GN em um segmento qualquer, onde na 1ª coluna encontra-se a variável de interesse e na 2ª a variável explicativa correlacionada.

Tabela 3.1 – Exemplo de priorização de Variáveis Explicativas.

| Correlação Consumo GN Comercial Curitiba x Variáveis Explicativas - UGR Brasil |        |                |        |    |           |   |  |
|--|--------|----------------|--------|----|-----------|---|--|
| Variável Explicativa   | r(X,Y) | r <sup>2</sup> | p      | N  | Priorizar | Motivo  |  |
| % Operações de Crédito Pessoa Física/PIB- BR (M)                               | 0,9595 | 0,9207         | 0,0000 | 79 | Sim       | Nível de crédito ao Consumo por Pessoa Física     |  |
| % Operações de Crédito Privado/PIB- BR (M)                                     | 0,9443 | 0,8917         | 0,0000 | 79 | Sim       | Nível de crédito ao Consumidor privado            |  |
| Nível Pessoal Empregado Indústria - BR (M)                                     | 0,9410 | 0,8854         | 0,0000 | 79 | Sim       | Aumento da produção                               |  |
| % Operações de Crédito ao Comércio / PIB - BR (M)                              | 0,9347 | 0,8736         | 0,0000 | 79 | Sim       | Crédito ao comércio (abastecimento e crescimento) |  |
| Nível Pessoal Empregado Indústria (Dess) - BR (M)                              | 0,9334 | 0,8712         | 0,0000 | 79 | Sim       | Aumento da produção                               |  |
| Consumo de EE Total - BR (M)   | 0,9261 | 0,8576         | 0,0000 | 79 | Não       | Acima da QTD de 5 Variáveis                       |  |

Para a UGR Brasil, a variável percentual de operações de crédito de pessoas físicas sobre o PIB foi selecionada como a principal neste grupo, por apresentar uma correlação de 0,95953 com a variável de interesse e também pelo entendimento de que reflete o nível de crédito concedido ao consumidor final.

Já a variável de consumo de EE total no Brasil não foi selecionada por exceder a quantidade de cinco variáveis priorizadas por conjunto, para cada segmento.

Após a priorização retiraram-se as variáveis explicativas redundantes, ou seja, aquelas que apresentam alta correlação entre si.

### **3.2.3 - ANÁLISE DE DEPENDÊNCIA ESPACIAL LOCAL E DEFINIÇÃO DAS UE PARA EMPREGO DO MODELO SAR *PANEL***

Uma vez pré-selecionadas as variáveis explicativas para o desenvolvimento do modelo, procedeu-se à análise dependência local para a variável de interesse. Esta investigação tem o intuito de identificar agrupamento de objetos, neste caso UE, com valores de atributos semelhantes (*clusters*) ou objetos anômalos (*outliers*).

Conforme apresentado em 2.5.3.2 existem dois índices que podem ser empregados na realização desta tarefa: o Local de Moran e as Estatísticas Getis e Ord -  $G_i$  e  $G_i^*$ . Nesta tese adotou-se o Índice Getis e Ord Star, conhecido como  $G_i^*$ , por incluir em sua formulação o valor da AE observada, conforme (2.10).

Esta opção se deu pelo entendimento de que o modelo SAR *Panel* tem a premissa de contemplar o evento na UE de interesse, bem como deve permitir identificar os possíveis regimes espaciais, conforme observado em ELHORST (2003), o que na prática, pode gerar vários modelos locais para uma mesma AE.

O índice de Getis Ord interpreta como os dados encontram-se distribuídos no espaço de forma mais direta do que o de Moran. Este último, apesar de sua utilidade, apresenta problemas na interpretação, visto sua distribuição probabilística não ser conhecida perfeitamente e ter que ser estimada por amostragem. O indicador Getis Ord permite o teste de hipóteses sobre a concentração espacial da soma dos valores associados com a vizinhança da UE de interesse, sendo estas estatísticas computadas para as variáveis normalizadas.

Em condições normais, a distribuição das permutações simuladas de  $G_i$  e  $G_i^*$  aproxima-se da normal, o que facilita a aplicação de testes de significância. Ou seja, a hipótese nula é a de que os valores  $x_i$  dentro de uma distância estabelecida sejam uma amostra aleatória independente de sua posição geográfica.

Supondo que os resultados assumam a forma de uma distribuição i.i.d, são construídos intervalos de confiança que indicam valores extremos correspondentes a regiões com regimes espaciais próprios. Como resultados identificam-se aglomerados espaciais de valores altos e baixos, nos quais valores padronizados significativos ( $p$ -valor  $< 0,05$ ) e positivos indicam uma aglomeração espacial de valores altos.

Já valores padronizados significantes ( $p$ -valor  $< 0,05$ ) e negativos apontam para uma aglomeração espacial de valores baixos. A existência de polígonos com valores altos ou baixos de forma isolada significa a presença de um *outlier*.

Os índices de Getis e Ord são estimados a partir de grupos de vizinhos dentro de uma distância crítica  $d$  de cada área  $i$  obtida a partir da matriz  $W_{ij}$ , na qual cada elemento encontra-se dentro da distância crítica, ou  $W_{ij}(d)$ .

Ressalta-se que a assunção sobre a distância não é clara. Nas bibliografias pesquisadas, como GETIS (2009), BAILEY e GRATEL (1995) e ANSELIN (1988), não citam-se limitações para as regras de  $W_{ij}(d)$  na formação da matriz de contiguidade. Por exemplo, o conceito de zona de indiferença, atende à premissa de distância crítica, bem como o próprio conceito de declínio fixo também atende tal premissa.

No limite, a extrapolação do conceito teórico, observado por GETIS (2009), no qual a matriz é exógena ao sistema, ou seja, é fornecida ao sistema baseando-se em uma estrutura preconcebida, permite adotar vários arranjos que atendam  $W_{ij}(d)$ . Na verdade, este conceito prioriza a relação de acoplamento entre as UE, ou as relações de vizinhança, em suas possíveis formas, e não somente baseada na distância.

Na contramão, abdicar desta possibilidade é determinar que regras de vizinhança baseadas em topologia e que geram diferentes distâncias físicas, como as matrizes do tipo *Rook*, *Queen* ou *Bishop*, não podem ser utilizadas para avaliar os efeitos locais da dependência espacial.

Assumindo que a matriz é exógena, não existe empecilho matemático em conceituar as relações da matriz baseadas em várias distâncias críticas, ou melhor, várias distâncias críticas concêntricas. Neste caso, pode-se utilizar a vizinhança baseada em topologia, que apresenta como vantagens sobre as matrizes de distância fixa:

- Não comparar vizinhos situados entre regiões com vazio ocupacional; e,
- Não estabelecer relações contendo efeitos de ordens secundárias que influenciem no resultado do índice.

Com base no exposto, nesta tese a dependência espacial local, assim como no caso da dependência global, também foi observada à luz de matrizes de distância fixa e

topologia baseada em vizinhança de 1ª ordem. Quanto à estrutura das matrizes não houve alteração, uma vez que não ocorreu alteração da estrutura espacial. Ressalta-se que os valores aplicados à matriz por distância garantem pelo menos 1 vizinho para a feição mais afastada do conjunto.

Os índices foram calculados empregando o ArcGIS, considerando uma série de 24 eventos mensais, entre janeiro de 2007 e dezembro de 2008. Após calculados avalia-se a significância, com base em uma normalização do tipo *Z Score*, na qual valores padronizados acima de 1,96 Z apresentam significância com 95% de confiança.

No tocante a esta tese, estes índices permitem validar a condição necessária para aplicar o modelo SAR *Panel*, ou seja, existir dependência espacial em um grau de predominância global ou local, bem como a condição suficiente para determinar o emprego em caráter temporal, ou seja, estacionariedade da dependência espacial.

De forma prática, depois do cálculo da dependência espacial local é possível definir onde aplicar o modelo SAR *Panel* validando o segmento por AE, qual a UGR de representação e a extensão do modelo, ou seja, englobando todas ou algumas UE contidas na UGR de representação.

Tal etapa é intitulada de definição das áreas para emprego do modelo espacial e predominantemente investiga a estacionariedade temporal do índice local de dependência espacial, e qual o seu comportamento ao longo do espaço. Esta investigação emprega mapas temáticos contendo a significância do índice local, e tem o objetivo de verificar e quantificar a existência de regiões (regimes espaciais) sustentáveis ao longo do tempo.

Com relação a isto cabe uma consideração. A instabilidade estrutural, ocasionada pela dispersão dos eventos ao longo do espaço, afeta a formação das matrizes espaciais, e, por conseguinte o cálculo da dependência espacial, pois:

- Geram matrizes que comparam objetos em regiões não contíguas; e,
- Nas matrizes por distância o valor que garante ao menos 1 vizinho para o objeto de interesse acarreta no mapeamento de vizinhos de ordens enésimas.

Na prática, observa-se a formação de diferentes regimes espaciais. Por exemplo, nas figuras 3.2 e 3.3, apresenta-se uma porção territorial de um mapa temático com a significância  $G_i^*$  para o mês de julho de 2008 no segmento residencial de Curitiba, considerando respectivamente, entre as UE, a vizinhança de 1ª ordem e a por distância.



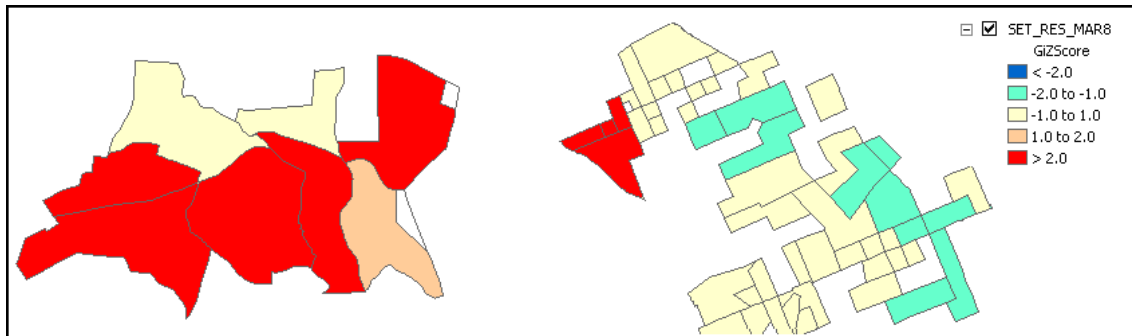


Figura 3.2 – Significância  $G_i^*$  (07/2008) - Curitiba Residencial - Vizinhos de 1ª ordem.

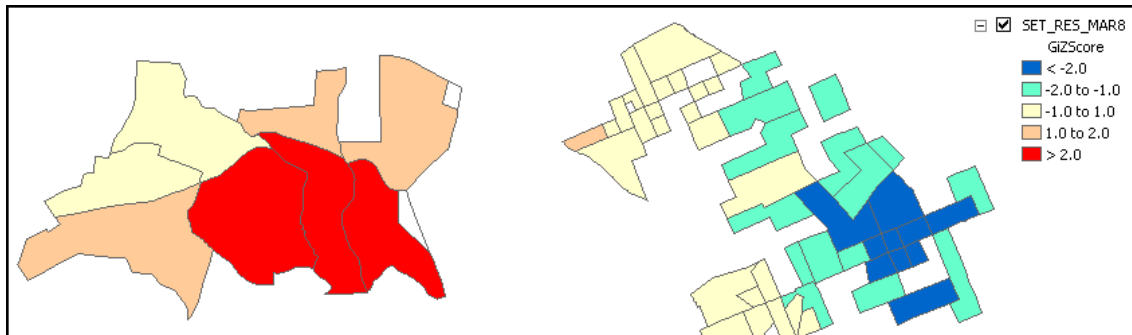


Figura 3.3 – Significância  $G_i^*$  (07/2008) - Curitiba Residencial - Vizinhos por distância.

Comparando as figuras nota-se a presença de diferentes regimes espaciais, tanto em número quanto em posição e valor. Na figura 3.3 existe um cluster de valor negativo com uma zona de transição entre este regime e a indiferença; já na figura 3.2 somente existem agrupamentos de UE com valores positivos.

Esta particularidade também é observada no senso temporal, ou seja, existem as mesmas variações de significância quando se observa a dependência espacial à luz de duas matrizes de estrutura diferente. Tal característica impacta a estacionariedade, pois o valor da autocorrelação espacial em uma determinada UE hora tem significância, hora pode não tê-la mais. Diante disto, nesta tese, definiram-se as seguintes regras para quantificar a estacionariedade temporal do índice de dependência espacial local:

- Para mitigar a instabilidade estrutural, foram considerados os índices calculados empregando as matrizes  $W_{ij}$  de 1ª ordem; e,
- Para considerar a dependência espacial “estacionada”, as UE deveriam ter a mesma classificação da significância ao longo de toda série, admitindo pequenas flutuações (máximo 2) entre os períodos.

Com base nos mapas temáticos do índice, atribuiu-se no tempo o valor 1 para UE com significância alta, positiva ou negativa e, 0 para as demais. Estes valores foram tabulados em uma matriz  $m \times n$ , onde  $m$  identifica a UE e  $n$  o tempo, e sobre esta identificaram-se as UE que apresentam significância alta ao longo do tempo. Como

exemplo, na figura 3.4, após a atribuir os valores e construir a matriz, as UE B a D, seriam as elencadas com índice de autocorrelação espacial significativa e estacionado.

| Eventos | UE Vizinhas |      |      |      |      |      |     |      |
|---------|-------------|------|------|------|------|------|-----|------|
|         | UE A        | UE B | UE C | UE D | UE E | UE F | ... | UE n |
| jan/07  | 1           | 1    | 1    | 1    | 0    | 0    | ... | 0    |
| fev/07  | 1           | 1    | 1    | 1    | 0    | 0    | ... | 0    |
| mar/07  | 1           | 1    | 1    | 1    | 0    | 0    | ... | 0    |
| abr/07  | 0           | 1    | 1    | 0    | 0    | 0    | ... | 0    |
| mai/07  | 0           | 1    | 1    | 1    | 0    | 0    | ... | 0    |
| ⋮       | ⋮           | ⋮    | ⋮    | ⋮    | ⋮    | ⋮    | ... | ⋮    |
| jun/08  | 0           | 1    | 1    | 1    | 1    | 1    | ... | 0    |
| jul/08  | 0           | 1    | 0    | 1    | 1    | 1    | ... | 0    |
| ago/08  | 1           | 1    | 1    | 1    | 0    | 0    | ... | 0    |
| set/08  | 1           | 1    | 1    | 1    | 0    | 0    | ... | 0    |
| out/08  | 0           | 1    | 1    | 0    | 0    | 0    | ... | 0    |
| nov/08  | 0           | 1    | 1    | 1    | 0    | 0    | ... | 0    |
| dez/08  | 1           | 1    | 1    | 1    | 0    | 0    | ... | 0    |

Figura 3.4 – Dependência Espacial Local - Exemplo da Matriz de Estacionariedade.

Na prática, estas UE seriam as de interesse ao emprego do modelo SAR *Panel* dentro da AE, pois apresentam as condições necessária e suficiente.

### 3.2.4 – SELEÇÃO DAS VARIÁVEIS PARA ESTIMAÇÃO DO MODELO SAR *PANEL*

Após definidas as UE para o modelo SAR *Panel* e conseqüentemente os segmentos e AE relacionados, executou-se a seleção das variáveis para estimar o modelo. Nesta etapa foram selecionadas variáveis explicativas do conjunto das variáveis pré-selecionadas, e de forma complementar analisou-se a contribuição da própria variável dependente, uma vez que no contexto temporal pode informar ao modelo particularidades da série de interesse.

A seleção das variáveis explicativas teve como ponto de partida as transformações por análise de componentes principais (PCA), com o objetivo de definir a melhor estratégia de emprego destas variáveis, conferindo ao modelo um conjunto de dados reduzidos, com capacidade de previsão e representatividade.

Dois pontos principais motivaram a realização da análise PCA:

- Dimensionalidade - a quantidade de variáveis pré-selecionadas, cerca de quinze, por segmento e área de estudo, poderia tornar o modelo de difícil determinação e manutenção; e,
- Robustez - o modelo deveria ter a capacidade de prever um  $\Delta t$  à frente, sem a necessidade de realizar procedimentos adicionais, bem como empregar variáveis que melhor expliquem o fenômeno.

Empregando os *toolboxes* STATISTIXL, do Excel e o algoritmo NIPALS do *Advanced Non Linear / Linear Models* do STATISTICAL 7, realizaram-se três análises em comparação com a variável dependente:

- PCA das variáveis explicativas em fase ( $k = 0$ ) com a variável dependente,
- PCA das variáveis explicativas defasada da variável dependente; e,
- Variáveis explicativas defasadas da variável dependente.

Em todos os casos as componentes principais (CP) foram calculadas com base em matrizes de correlação com extração de todas as componentes. Apesar de ser possível determinar um número mínimo de CP ou um limite mínimo de *eigenvalues*, optou-se por extrair todas as CP de forma a identificar as mais significantes.

Especificamente o *toolbox* do STATISTICAL fornece a possibilidade de definir o percentual mínimo de casos válidos por variável e vice-versa, bem como a definição dos critérios de *cross-validation*. Como isto não está disponível no STATISTIXL, optou-se pela especificação padrão do *toolbox*: *fitting methods* por número de CP para validação cruzada e método de validação cruzada de V-Fold.

Foi extraído para todas as CP o seguinte conjunto de informações: estatísticas descritivas, matriz de correlação, *eigenvalues* (variância total e acumulada), *components loading* (correlação entre a variável principal e as CP), *eigenvector* (coeficiente dos componentes) e *casewise score* PCA (dados transformados).

A PCA em fase com a variável dependente teve como foco as AE, nas quais as UE apresentaram dependência espacial estacionada, onde realizaram-se a:

- Avaliação da correlação cruzada entre os grupos espaciais de variáveis explicativas (Brasil, UF e município) oriundos da pré-seleção;
- Determinação da principal componente (PC) com posterior transformação das variáveis explicativas; e,
- Análise da correlação entre a variável dependente e as transformadas pela PC.

A análise da correlação cruzada entre as variáveis explicativas contemplou estritamente o número de eventos úteis da variável dependente correspondente. Isto difere da análise das correlações descrita em 3.1, pois naquela as variáveis explicativas foram comparadas empregando o número total de eventos úteis entre si.

Com base nesta nova matriz de correlação foi estabelecido um fator de corte das variáveis pré-selecionadas empregando como limites os valores acima de 0,97 e abaixo de 0,85, para retirar variáveis redundantes e de baixa significância respectivamente.

Após o corte, realizou-se a PCA para determinar a PC, transformou-se o conjunto de dados por esta última e determinou-se a correlação entre as variáveis transformadas e a respectiva variável de interesse.

Uma vez avaliado os resultados da PCA observou-se a possibilidade de retirar outras variáveis com objetivo de reduzir a dimensionalidade sem afetar a contribuição dos valores transformados em relação à variável dependente. Desta foram, este ciclo foi repetido, de forma que a redução do conjunto acontecesse em sintonia com os resultados obtidos na transformação pela PC.

Para os subconjuntos de variáveis explicativas de melhor desempenho após a PCA em  $k = 0$ , repetiu-se a PCA com até 12 atrasos. Por último, determinou-se a correlação entre as variáveis explicativas e a de interesse, para cada segmento e AE, também considerando um atraso de doze épocas. Cabe ressaltar que para este caso foram empregados os mesmos subconjuntos de variáveis utilizados na determinação da PCA.

A seleção das variáveis para o modelo realizou-se através da comparação do resultado da correlação do conjunto de dados transformado pela PC com a variável de interesse e, esta última contra as variáveis explicativas de forma individual; privilegiando a redução de dimensionalidade e a robustez, bem como os desvios relativos entre as correlações. No tocante às variáveis explicativas, considerou-se para comparação as cinco variáveis que melhor explicaram a variável de interesse, em  $k = 0$ .

Estas correlações foram dispostas em tabelas contendo na primeira linha, a correlação entre a variável de interesse e os dados transformados pela PC, conforme o atraso; e nas demais linhas, também conforme o atraso, as diferenças entre os valores das correlações da variável de interesse com os dados transformado pela PC e com as variáveis explicativas de melhor correlação com a variável de interesse em  $k = 0$ .

Em complemento à definição do conjunto de variáveis explicativas procedeu-se a análise da possível contribuição, no senso temporal, da própria variável de interesse. Isto é importante uma vez que os dados socioeconômicos, representados nas variáveis explicativas, encontram-se em um nível de resolução espacial que não permite a diferenciação das contribuições das variáveis explicativas por UE.

Assumindo que o modelo SAR *Panel* é um empilhamento no tempo de SARs *cross section*, a aplicação desta variável no termo regressivo  $X\beta$ , no senso espacial apresentará valor igual para qualquer UE, e isto não é a realidade do mercado.

Matematicamente, empregar o mesmo valor para a variável explicativa, nas  $n$  diferentes UE, não compromete sua determinação; entretanto implica que todo o resíduo

individual dos eventos de uma série de dados em painel, esteja unicamente associada ao termo autoregressivo espacial  $pWy$  da estrutura SAR. Sendo assim, a restrição de representação espacial das variáveis explicativas compromete a distinção das contribuições do termo  $X\beta$  para o modelo; a opção é extrair da variável de interesse uma contribuição individualizada sobre o fenômeno.

Conforme a estrutura SAR *Panel* em (2.28) o emprego da variável dependente no termo  $pWy$  é ponderado pela matriz  $W_{ij}$ ; bem como se dá em fase com a série temporal, mesmo nesta estrutura. Ou seja, aplicar a variável de interesse no referido termo contribui tão somente com a mitigação da autocorrelação no espaço. Desta forma, é possível empregá-la em  $X\beta$ , capturando a componente temporal por UE a partir da análise da autocorrelação parcial do consumo de GN.

Este procedimento considerou o consumo nas UE de interesse e em seus respectivos vizinhos, determinados por  $W_{ij}$ , observando a questão da dimensionalidade na determinação do vetor de cálculo da autocorrelação.

Conforme 2.5.3.5, o modelo resultante do empilhamento de SARs *cross section* no tempo pode ser resolvido aplicando uma estrutura SUR, ou seja, através de um conjunto de  $N$  equações para  $n$ . Entretanto, para tal é necessário igualdade dimensional, no tempo e no espaço, entre os termos e as variáveis do modelo. No senso espacial não existe restrição uma vez que já avaliou-se no tempo a estabilidade estrutural, e a estacionariedade da significância da autocorrelação espacial do consumo de GN.

Contudo, na análise do consumo de GN de forma individualizada, por UE, percebe-se que isto ocorreu de forma diferenciada, ou melhor, com momentos iniciais distintos. Por exemplo, na figura 3.5, onde definiu-se a UE 207 para modelagem, tem-se as seguintes UE vizinhas conforme a matriz  $W_{ij}$ : 112, 212, 214, 277, 285 e 334.



Figura 3.5 – Vizinhança da UE 207 definida para Modelagem.

Neste caso, conforme as UE vizinhas observadas têm-se os diferentes  $t$  apresentados na tabela 3.2, em virtude das diferentes datas de início do consumo individual de GN por UE.

Tabela 3.2 – Comprimento do Vetor de dados de Consumo de GN – UE 207 e Vizinhança.

| Identificador do Polígono $n_i$ | Série Histórica de $y$ em $n_i$ | $t$ para $n_i$ |
|---------------------------------|---------------------------------|----------------|
| 207                             | Set/03 a Dez/08                 | 64             |
| 112                             | Mai/06 a Dez/08                 | 32             |
| 212                             | Jun/03 a Dez/08                 | 67             |
| 214                             | Fev/06 a Dez/08                 | 35             |
| 277                             | Fev/07 a Dez/08                 | 23             |
| 285                             | Abr/04 a Dez/08                 | 57             |
| 334                             | Abr/05 a Dez/08                 | 45             |

Percebe-se que os vetores da variável dependente têm comprimentos diferentes, bem como diferem do vetor de consumo total neste segmento, no qual  $t$  é igual a 72 eventos. Com base nestas particularidades definiram-se as seguintes premissas para definição das séries para o cálculo da autocorrelação parcial:

- Todas as séries foram pré-processadas empregando o mesmo procedimento de normalização apresentado no item 3.1;
- Não houve interpolação de dados, ou seja, considerou-se o valor zero para os meses em que não houve consumo. Desta forma, a série de consumo de cada setor ficou condizente com a realidade;
- As UE vizinhas à de interesse, sem consumo de GN no período estabelecido, não devem ser contempladas na estrutura da matriz  $W_{ij}$ ; e,
- A série para cada setor teve comprimento compatível com a série de dados de consumo para o segmento.

As análises das autocorrelações consideraram 15 atrasos e empregou-se o *toolbox Time Series / Forecasting* do módulo *Advanced Non Linear / Linear Models* do STATISTICAL 7. Com base nos resultados selecionou-se, individualmente por segmento e AE, o atraso que pode conferir mais informação ao modelo. Ao final desta etapa encontram-se definidos (as) para desenvolvimento do modelo:

- As UE por segmento e AE, para aplicar o modelo SAR *Panel*, à luz da dependência espacial;
- A dimensão de  $n$  e as respectivas matrizes  $W_{ij}$ ; e,
- As séries de variáveis para emprego nos termos  $X\beta$  e  $pWy$ .

### 3.3 – DESENVOLVIMENTO DO MODELO E PREVISÃO DO CONSUMO DE GN

Uma vez definidas as diretrizes para aplicação do modelo SAR *Panel*, procedem-se na seqüência: a estimação dos parâmetros do modelo e a previsão propriamente dita. Na primeira parte ressalta-se a escolha da forma funcional do modelo SAR *Panel* e da estratégia de entrada de dados. Entretanto antes de tangenciar estes tópicos cabe o esclarecimento do mecanismo de estimação do modelo.

#### 3.3.1 – MECANISMO DE ESTIMAÇÃO DO MODELO

Assumindo que o modelo SAR *Panel* é um empilhamento no tempo de SARs *cross section*, faz-se necessário primeiro entender seu mecanismo de estimação, onde, observando (2.16), segue a seguinte ordem de execução:

I. Estimar  $\beta_0$  a partir de estimador OLS, para o modelo  $y = X\beta_0 + \varepsilon_0$ , (3.1)

II. Estimar  $\beta_L$ , também empregando OLS, para o modelo  $Wy = X\beta_L + \varepsilon_L$ , (3.2)

III. Calcular os resíduos  $\varepsilon_0 = y - X\hat{\beta}_0$  e  $\varepsilon_L = Wy - X\hat{\beta}_L$ , (3.3 e 3.4)

IV. Dado  $e_0$  e  $e_L$ , estimar o escalar  $p$  que maximize a função concentrada de verossimilhança:

•  $L_n(L) = -(n/2) \ln(\pi) - (n/2) \ln(1/n) (\varepsilon_0 - p\varepsilon_L)'(\varepsilon_0 - p\varepsilon_L) + \ln|I - pw|$ , (3.5)

onde o operador  $|\bullet|$  é o determinante;

V. Dado  $\hat{p}$  que maximize  $L_C$ , calcular:

•  $\hat{\beta} = (\hat{\beta}_0 - p\hat{\beta}_L)$ ; e, (3.6)

•  $\hat{\sigma}_\varepsilon^2 = (1/n) (\varepsilon_0 - p\varepsilon_L)'(\varepsilon_0 - p\varepsilon_L)$ . (3.7)

O processo inicia-se com a estimação dos parâmetros das regressões nas etapas (I) e (II), seguido de uma otimização univariada de parâmetros, na função de probabilidade concentrada, com relação ao parâmetro  $p$  na etapa (IV).

A estimação de parâmetros baseados em OLS, para as regressões em (I) e (II), assume a condição de que este estimador é o melhor estimador linear sem tendência (*best linear unbiased estimator* – BLUE), em conjunto com um estimador de máxima verossimilhança para a função em (IV). Tal combinação torna-se necessária, pois o estimador OLS produz uma estimativa tendenciosa para o parâmetro  $p$ , uma vez que sobre a assunção de dependência espacial, não é possível admitir que  $Wy$  seja fixo em repetidas amostragens.

Esta situação é similar à dos modelos autoregressivos de séries temporais, nos quais as observações são geradas sequencialmente e conseqüentemente não são independentes.

Uma vez que as estimativas de (I) e (II), realizadas sobre a hipótese de BLUE do estimador OLS, mitigaram as influências temporais das variáveis explicativas sobre a variável dependente (3.3 e 3.4), na etapa IV tem-se a expectativa de maximizar o valor de  $p$ , e com isto ampliar a capacidade de mapear os resíduos do modelo.

Como alternativa à deficiência demonstrada para o emprego de um estimador OLS na estimação de  $p$  pode ser aplicado o método de momentos ou o de máxima verossimilhança. Como exemplo da aplicação dos métodos de momentos, para estimação de  $p$  em modelos SAR, citam-se ARNOLD (2007) e LEE (2007).

Nesta proposta assumiu-se que o emprego de uma abordagem por máxima verossimilhança venha conferir maior robustez ao entendimento das distribuições, considerando que, após a análise e tratamento dos dados coletados, o conjunto de variáveis e o tamanho das séries históricas é pequeno

Ou seja, uma vez que a distribuição de  $y$  é condicionada a um número de variáveis  $x$  conhecidas e a um pequeno número de coeficientes  $\varepsilon$  desconhecidos, pode-se empregar a abordagem de máxima verossimilhança para estimar estes parâmetros desconhecidos, escolhendo-os de forma que correspondam à melhor distribuição resultante possível, de maneira que possam ser definidos mais precisamente, considerando os dados observados.

Para exemplificar, tome-se como exemplo um modelo de regressão simples que apresenta resíduos com distribuição normal, ou seja,  $y_i = \beta x_i + u_i$ ; onde  $u_i \sim N(0, \sigma^2)$ , e a função de densidade do erro  $u_i$  é dada por:

$$f(u_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u_i^2}{2\sigma^2}\right) \quad \text{para } i = 1, 2, \dots, n \quad (3.8)$$

Como o erro apresenta distribuição normal com média zero e variância  $\sigma^2$ ,  $y_i$  também será normal e, sendo assim apresenta uma função de densidade na forma de:

$$f(y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2\right] \quad \text{para } i = 1, 2, \dots, n \quad (3.9)$$

No caso espacial, a probabilidade limite (*plim*) do termo  $y'W\varepsilon$  não pode ser zero, pois quando  $p = 0$  não existe dependência espacial na amostra de dados. Desta forma, para o modelo SAR proposto, o estimador de máxima verossimilhança será empregado



para encontrar o valor de  $p$  que maximize a função de densidade, levando-se em consideração o resultado da estimação OLS em (3.1 e 3.2) e o cálculo dos resíduos  $\varepsilon_0$  (3.3) e  $\varepsilon_L$  (3.4).

Neste caso, a função de densidade tem a forma de:

$$L(\varepsilon / p, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} |I_n - pW| \exp\left[-\frac{1}{2\sigma^2} (\varepsilon_0 - p\varepsilon_L)^2\right] \quad (3.10)$$

Para simplificar o problema da maximização, obtém-se o logaritmo natural da função concentrada de verossimilhança com base na eliminação dos distúrbios da variância  $\sigma^2$ , através da substituição de:

$$\hat{\sigma}^2 = (1/n)(e_0 - pe_L)'(e_0 - pe_L), \quad (3.11)$$

em (3.10), e deste modo obtém-se a função concentrada de verossimilhança proposta em (3.5). Após o cálculo de  $\hat{p}$  que maximize  $L_n(L)$ , deve-se calcular (3.6 e 3.7).

Na mesma linha, o modelo SAR *Panel* com a extensão de atrasos espaciais, conforme (2.28), também pode ser estimado por máxima verossimilhança a partir de dados em *cross section*; a diferença é que os dados são expandidos da visão por secção de  $N$  observações, para um painel de  $N \times T$  observações (ANSELIN e HUDAK, 1992).

Inicialmente as observações *cross section* para  $t = 1, \dots, T$  são empilhadas para obter  $(NT, 1)$  vetores de  $Y^*$  e  $(I_T \otimes W)Y^*$ , e uma matriz  $(NT, K)$   $X^*$  de variáveis degradadas. Ressalta-se que este processo deve ser realizado uma única vez e que a  $(NT, NT)$  matriz diagonal  $(I_T \otimes W)$  não necessita ser armazenada, pois aumentaria consideravelmente o processamento computacional do estimador por máxima verossimilhança para séries de dados com muitos eventos.

Uma vez que  $b_0$  e  $b_L$  representem o estimador OLS das sucessivas regressões  $Y^*$  e  $(I_T \otimes W)Y^*$  em  $X^*$ , e  $e_0^*$  e  $e_L^*$  correspondem aos respectivos resíduos, então o estimador por máxima verossimilhança de  $p$  é obtido pela maximização da função concentrada de verossimilhança conforme:

$$Ln(L) = C - \frac{NT}{2} \ln[(\varepsilon_0^* - p\varepsilon_L^*)'(\varepsilon_0^* - p\varepsilon_L^*)] + T \ln|I_N - pW| \quad (3.12)$$

Onde,  $C$  é uma constante independente de  $p$ .

Este problema de maximização só pode ser resolvido via métodos numéricos, uma vez que não existe uma forma definida para estimar  $p$ . Entretanto, desde que a função de máxima verossimilhança seja côncava em  $p$ , a solução numérica é única.

De certo, no caso de modelos onde as matrizes  $W_{ij}$  têm tamanho elevado, os métodos de obtenção da função log-determinante exigem elevado esforço computacional; desta forma é necessário aplicar métodos computacionais para otimizar este processo. Para tal, PACE e BARRY (1997, 1998 e 1999) propuseram métodos para o cálculo da função  $\ln|I - pw|$ , onde gera-se um vetor predeterminado com os possíveis valores de  $p$ , a partir de um grid no qual os valores variam entre  $1/\omega_{min}$  a  $1$  considerando  $W_{ij}$  normalizada.

Com base nestes valores predeterminados, pode-se testar a função concentrada de log-verossimilhança com todos os valores admissíveis de  $p$  e determinar qual o valor maximiza a referida função. Por último, computam-se os estimadores de  $\beta$  e  $\sigma^2$ , dado a estimativa numérica de  $p$ , conforme:

$$\beta = b_0 - pb_L = (X^{*'} X^*)^{-1} X^{*'} [Y^* - p(I_T \otimes W)Y^*] \quad (3.13)$$

$$\sigma^2 = \frac{1}{NT} (\varepsilon_0^* - p\varepsilon_L^*)' (\varepsilon_0^* - p\varepsilon_L^*) \quad (3.14)$$

Ressalta-se que o operador asterisco relaciona-se com o procedimento de eliminação dos efeitos fixos espaciais introduzidos pela variável intercepto  $\mu_i$  na regressão em (2.26), a partir da degradação de  $y$  e  $x$  com base nas seguintes transformações,

$$y_{it}^* = y_{it} - \frac{1}{T} \sum_{t=1}^T y_{it} \quad (3.15)$$

$$x_{it}^* = x_{it} - \frac{1}{T} \sum_{t=1}^T x_{it} \quad (3.16)$$

O processo de estimação dos parâmetros foi realizado empregando o *Tool Box* de econometria espacial desenvolvido para aplicação no *software* MATLAB. Os códigos foram desenvolvidos originalmente por James P. LeSage em 2000 e o modelo original SAR *Panel* especificamente, escrito por Elhorst em 2003, foi desenvolvido com base na estrutura da função SAR original escrita por LeSage em 2000.

### 3.3.2 – FORMA FUNCIONAL E MÉTODOS DE OTIMIZAÇÃO DO MODELO

Para aplicação da estrutura SAR *Panel*, o primeiro passo é definir a forma funcional, que pode ser agrupada, de efeitos fixos, temporais e / ou espaciais, ou de efeitos aleatórios. A escolha passa pela capacidade do modelo estimar os parâmetros de forma consistente, e relaciona-se com a quantidade de eventos, temporais e espaciais, disponíveis à estimação. Sobre isto, ELHORST (2003) estabelece algumas diretrizes:

- Modelos de efeitos fixos:
  - Em painéis pequenos, onde  $T$  é fixo e  $N \rightarrow \infty$ , os efeitos fixos temporais podem ser estimados de forma consistente;
  - Em painéis grandes, onde  $T \rightarrow \infty$  e  $N$  é fixo, os efeitos fixos espaciais podem ser estimados de forma consistente; e,
  - Quando  $N$  e  $T$  são compatíveis em tamanho, os efeitos fixos espaciais e temporais podem ser estimados de forma consistente.
- Modelos de efeitos aleatórios:
  - Os parâmetros podem ser consistentemente estimados quando  $N \rightarrow \infty$ ,  $T \rightarrow \infty$  ou  $N$  e  $T \rightarrow \infty$ .

Entretanto, ELHORST (2003) salienta que o modelo de efeitos aleatórios pode não ser apropriado quando são empregadas observações em UE irregulares. Este caso é similar à distribuição do consumo de GN ao longo do território, objeto desta tese.

Com relação à questão do atraso espacial (*spatial lag*), ELHORST (2003) propôs uma estratégia de estimação baseada em regimes espaciais, a partir da aplicação de modelos intitulados de coeficientes, fixos ou aleatórios. Esta proposição tem a desvantagem do grande número de parâmetros a serem estimados, pois são gerados valores individualizados para as diferentes UE, e seu emprego seria indicado para análises onde  $N$  é pequeno.

Cabe ressaltar que não foi encontrada citação ou aplicações do modelo de coeficientes, fixos ou aleatórios, proposto em ELHORST (2003); bem como, na revisão em ELHORST (2010) tal estratégia de modelagem não é mais citada.

Independente do modelo, a estratégia de estimação por regimes espaciais é uma saída para redução do número de parâmetros estimados, uma vez que geram-se parâmetros observando condições específicas de atrasos espaciais. Na verdade, a prévia análise da autocorrelação local reduz o universo de UE a um conjunto que possui o mesmo regime de dependência espacial; face a isto decidiu-se estimar os parâmetros do modelo nas formas funcionais agrupada (PSAR) e de efeitos espaciais fixos (SARSFE).

A forma funcional agrupada foi escolhida por ser a forma padrão para estimar dados em painel e a de efeitos espaciais fixos pelo entendimento de que o conjunto de UE definido após a análise de dependência espacial apresenta um efeito espacial fixo no tempo; ou seja, os índices apresentam estacionariedade.

A diferença entre as formas PSAR e SARSFE é que, a primeira desconsidera a eliminação dos efeitos fixos espaciais pela degradação via as transformações (3.15) e (3.16), procedimento este automático no código computacional do modelo SARSFE.

Quanto à estimação de  $p$ , ressalta-se que os modelos espaciais estimados por máxima verossimilhança, requerem o cálculo da função log-determinante da matriz positiva definida na forma de  $\ln|I - pw|$ . Com relação a isto, nesta tese, foram testados os três métodos de determinação propostos por Pace e Barry; entretanto, a aplicação da proposição em PACE e BARRY (1998), que realiza a estimação para um grid com  $0 < p < 1$ , apresentou erros durante o processo de estimação.

Sendo assim, empregou-se a proposta de PACE e BARRY (1997), que é a forma original para determinar o grid de valores, intitulada método de determinação completo (*Indet full*) e a proposta de PACE e BARRY (1999), na qual otimiza-se a estimação da determinante via algoritmo de Monte Carlo (*Indet MC*). Do mesmo modo que para o modelo SAR *Panel*, os métodos de determinação para o cálculo da função  $\ln|I - pw|$  estão disponíveis no toolbox de Econometria espacial para emprego no MATLAB.

### 3.3.3 – ENTRADA DE DADOS PARA ESTIMAÇÃO DO MODELO

Durante a revisão bibliográfica não se encontrou citação ou exemplo de como proceder com a entrada de dados e suas possíveis formas, uma vez que a estrutura do modelo é flexível no termo  $X\beta$ , para estimação do modelo. Com isto o primeiro ponto a entender no processo de estimação é como os dados são fornecidos para o modelo.

A ordem de entrada descrita no algoritmo proposto por ELHORST (2003) prevê o fornecimento dos dados primeiro na ordem espacial e depois pela ordem temporal; ou seja, primeiro são fornecidos os dados das diversas UE no instante 1, posteriormente os dados das mesmas UE, na mesma ordem, no tempo 2, e assim sucessivamente. De forma esquemática o modelo a ser estimado apresenta a seguinte configuração,

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_T = \begin{bmatrix} x_{a1} & x_{b1} & \cdots & x_{s1} \\ x_{a2} & x_{b2} & \cdots & x_{s2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{an} & x_{bn} & \cdots & x_{sn} \end{bmatrix}_T \begin{bmatrix} \beta_a \\ \beta_b \\ \vdots \\ \beta_s \end{bmatrix} + p \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \end{bmatrix}_T \begin{bmatrix} y_a \\ y_b \\ \vdots \\ y_n \end{bmatrix}_T + \begin{bmatrix} \varepsilon_a \\ \varepsilon_b \\ \vdots \\ \varepsilon_n \end{bmatrix}_T \quad (3.17)$$

No qual,

$n$  – são as UE,

$s$  – índice das diversas *variáveis explicativas*  $x$ ;  $e$ ,

$T$  – tempo.

As variáveis de interesse ( $y$ ) são fornecidas em formato de vetor coluna e as variáveis explicativas em formato de matriz; sendo o número de colunas igual à quantidade de variáveis explicativas e o número de linhas é o produto da quantidade de UE pelo número de eventos no tempo.

A série de dados disponível para a tese vai até dezembro de 2008, e deste modo previu-se valores individuais mensais para 12 eventos a partir de janeiro de 2008. Com isto o último mês de dados conhecido é dezembro de 2007.

Para lembrança, o termo  $X\beta$  é composto pela representação do conjunto de variáveis explicativas transformadas pela PC com atraso de ordem 12 ( $PCA12$ ), e a própria variável dependente com atraso de ordem 1 ( $y_{(k-1)}$ ).

Em  $pWy$  empregaram-se as mesmas matrizes  $W_{ij}$ , por topologia, utilizadas na análise de dependência espacial, o escalar  $p$  e o vetor de variáveis dependentes, simplesmente intitulado  $y$ .

Tendo dezembro de 2007, como último mês de dados conhecidos e os atrasos das variáveis explicativas têm-se, conforme tabela 3.3, o comprimento útil para os vetores de estimação e conforme a tabela 3.4, as seguintes configurações das séries históricas.

Tabela 3.3 – Casamento Dimensional ( $T$ ) do Conjunto de Variáveis

| Série   | Residencial de Curitiba | Residencial de Vitória | Comercial de Vitória |
|---|-------------------------|------------------------|----------------------|
| Consumo Total   | 76                      | 69                     | 72                   |
| Ordem Atraso da PCA                                     | 12                      | 12                     | 12                   |
| <i>PCA 12</i>   | 64                      | 57                     | 60                   |
| Meses a prever  | 12                      | 12                     | 12                   |
| <i>PCA12</i> útil / $y$ útil / $y_{(k-1)}$ útil ( $T$ ) | 52                      | 45                     | 48                   |

Tabela 3.4 – Relação de UE de Interesses e Respectivas Séries Históricas por Variáveis.

| Segmento                | UE de Interesse | Termo    | Variável     | Descrição da série |
|-------------------------|-----------------|----------|--------------|--------------------|
| Residencial de Curitiba | 624 e 638       | $X\beta$ | <i>PCA12</i> | Set/02 a Dez/06    |
|                         |                 |          | $y_{(k-1)}$  | Ago/03 a Nov/07    |
|                         |                 | $pWy$    | $y$          | Set/03 a Dez/07    |
| Residencial de Vitória  | 247 e 258       | $X\beta$ | <i>PCA12</i> | Abr/03 a Dez/06    |
|                         |                 |          | $y_{(k-1)}$  | Mar/04 a Nov/07    |
|                         |                 | $pWy$    | $y$          | Abr/04 a Dez/07    |
| Comercial de Vitória    | 207             | $X\beta$ | <i>PCA12</i> | Jan/03 a Dez/06    |
|                         |                 |          | $y_{(k-1)}$  | Dez/03 a Nov/07    |
|                         |                 | $pWy$    | $y$          | Jan/04 a Dez/07    |

Cabe ressaltar que os valores conhecidos de janeiro a dezembro de 2008 servirão como conjunto de teste para os valores previstos e, desta forma não foram incluídos na

estimação. Além disso, o comprimento dos vetores foi estabelecido após a avaliação do casamento dimensional entre as séries de dados disponíveis.

Com relação ao formato de entrada de dados em  $X\beta$ , adotaram-se três opções de distribuição das variáveis na matriz, respeitando a estrutura em (3.17). A primeira opção, intitulada (A), relacionou na primeira coluna a variável  $PCAI2$ , com o mesmo valor para todas as UE e, na segunda coluna a variável  $y_{(k-1)}$ , individualizada por UE.

$$\begin{bmatrix} PCAI2_1 & y_{(k-1)_1} \\ PCAI2_2 & y_{(k-1)_2} \\ \vdots & \vdots \\ PCAI2_n & y_{(k-1)_n} \end{bmatrix}_T \quad (3.18)$$

Ressalta-se que o valor de  $n$  é referente ao número da UE de interesse e de sua vizinhança e a quantidade de eventos temporais é igual ao tamanho útil das séries após análise de casamento dimensional, conforme tabela 3.3.

As segunda e terceira opções de entrada de dados, respectivamente intituladas opções (B) e (C), foram propostas com o intuito de encapsular os efeitos temporais da variável dependente individualmente por UE. Desta forma o processo de estimação por OLS observa cada UE individualmente.

Na opção (B) introduziu-se uma coluna individual, para cada  $y_{(k-1)}$ , conforme UE correspondente. A distribuição dos dados na matriz de entrada assumiu a forma de:

$$\begin{bmatrix} PCAI2_1 & y_{(k-1)_1} & 0 & \cdots & 0 \\ PCAI2_2 & 0 & y_{(k-1)_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ PCAI2_n & 0 & 0 & \cdots & y_{(k-1)_n} \end{bmatrix}_T \quad (3.19)$$

A última opção, a (C), fornece  $y_{(k-1)}$ , somente para a UE de interesse, reduzindo sua participação no termo  $X\beta$ . Por exemplo, em um modelo onde a segunda UE é a de interesse, a distribuição dos dados na matriz de entrada assume a forma de:

$$\begin{bmatrix} PCAI2_1 & 0 \\ PCAI2_2 & y_{(k-1)_2} \\ \vdots & \vdots \\ PCAI2_n & 0 \end{bmatrix}_T \quad (3.20)$$

Ao todo serão realizados doze experimentos para cada UE de interesse, fruto da combinação entre as formas funcionais (PSAR e SARSFE), métodos de determinação de  $\ln|I - pW|$  (*Indet full* e MC) e opções de entrada de dados (A), (B) e (C).

### 3.3.4 – PREVISÃO

Inicialmente ressalta-se que, assim como no caso da entrada de dados, não se encontrou citação ou exemplo aplicado de como proceder com o emprego dos parâmetros estimados para gerar uma previsão real. Deste modo, devem ser realizados entendimentos adicionais quanto ao equilíbrio entre os termos, do lado esquerdo e do lado direito do modelo.

No senso temporal, a primeira condição para uma previsão é que o vetor de variáveis dependentes, do lado esquerdo do modelo, apresente defasagem com o lado direito. Porém, isto não é condição necessária no senso espacial, pois como o atraso é na dimensão do espaço, ou melhor, na relação entre vizinhos no espaço, não existe defasagem temporal formal entre  $pWy$  e o  $y$  de interesse (lado esquerdo do modelo).

Na verdade, na pós-multiplicação o emprego do valor de  $y$  da UE de interesse não influencia o resultado projetado, pois o produto de  $y$  pela sua posição na matriz  $W_{ij}$  é igual a zero. Sendo assim, o termo  $pWy$  é influenciado somente pelos vizinhos relacionados na matriz  $W_{ij}$  empregada, e com isto tem a função exclusiva de incorporar ao modelo a contribuição dos  $n$  vizinhos, relacionados com o  $y$  de interesse, a partir do critério de relacionamento entre as UE: ordem, distância, topologia, etc...

Tal situação atesta a estratégia de manter a matriz  $W_{ij}$  empregada na determinação da dependência espacial, pois caso contrário empregar-se-iam relações de vizinhança diferentes das utilizadas na determinação da dependência espacial, e isto comprometeria o resultado da estimação.

Este entendimento torna razoável a colocação de LESAGE (1999) de que, em um modelo de atrasos espaciais (*spatial lag*), é possível verificar somente o quanto a adição da parcela espacial mitiga os resíduos.

No processo de estimação, os valores dos  $y$  "independentes" das diversas UE em  $Wy$  estão em fase com os  $y$  dependentes; com isto a defasagem temporal do modelo acontece pela aplicação das variáveis atrasadas  $PCAI2$  e  $y_{(k-1)}$  em  $X\beta$ , assumindo assim a característica de autoregressivo com exógenas (ARX).

Uma segunda condição chave é a questão de como prever o próximo  $y$ . Mesmo que a parcela  $Wy$  não tenha função temporal em  $y$ , e que o escalar  $p$  além de atemporal, equaliza os dois termos do modelo; como os dados estão em painel, a otimização de  $p$  empregará dados históricos, neste caso as séries de  $y$  dos vizinhos.

Observando o mecanismo de estimação, percebe-se que o passo 1 considera somente  $x$  e  $y$  conhecidos, sem o emprego de  $Wy$ . Sua estimação desta forma considera

atrasos espaciais embutidos no resíduo, fato este que será mitigado à frente, quando da determinação de  $\varepsilon_L = Wy - X\hat{\beta}_L$  e da função concentrada de verossimilhança.

Uma vez definido  $p$ , que maximiza  $x$  em  $y$  sem os efeitos espaciais, calcula-se  $\hat{\beta} = (\hat{\beta}_0 - p\hat{\beta}_L)$ ; e com isto conclui-se que não é correto reinjetar a parcela  $pWy$  para determinar o próximo passo de  $X\beta$  pelos seguintes motivos:

- Como  $pWy$  refere-se à influência dos vizinhos e não tem função temporal, reinjetá-lo em  $X\beta$  seria estipular um passo temporal para os efeitos espaciais com base no otimizador do termo regressivo temporal; e,
- Como o atraso espacial se dá individualmente para cada evento do painel de dados, reinjetá-lo em  $X\beta$  criaria um efeito em cascata.

O terceiro e último ponto de atenção refere-se ao avanço de  $Wy$ . Apesar do modelo contemplar a influência dos eventos vizinhos, não há abordagem sobre como tratar, para efeito de previsão, os efeitos do tempo nestes vizinhos. A princípio tais valores também deveriam ser previstos para compor o conjunto de valores de teste.

Entretanto, qual deve ser o modelo empregado para cada previsão individualizada por UE vizinha? Está dúvida se fundamenta no fato de que os modelos locais investigam a dependência espacial em um conjunto de UE mais restrito.

Nesta tese, por exemplo, a análise de autocorrelação espacial determinou um conjunto de UE que apresentava dependência espacial “estacionada” no tempo, a partir de uma investigação de caráter local, com base numa vizinhança 1ª ordem.

Na realidade, a UE de interesse pode ter vizinhos de 1ª ordem que, isoladamente, podem apresentar dependência espacial com outras UE de ordem  $n$ , não vizinhas a esta.

Para o caso em que os vizinhos da UE de interesse não apresentam dependência espacial com outras UE de ordem  $n$  qualquer, estas últimas podem ter seus valores previstos por qualquer procedimento puramente temporal. Porém, no caso contrário, estimar modelos para os vizinhos de 1ª ordem de forma isolada demandaria determinar novas matrizes  $W_{ij}$  individuais e análise de estacionariedade; além disso, acarretaria um efeito cascata, pois para prever os vizinhos de ordem  $n$  seria necessário prever os  $i$  vizinhos de ordem  $n$  à frente.

Entretanto, como o objetivo principal é avaliar o resultado de um processo de previsão espacial, definiu-se que os valores empregados na etapa de previsão, como entrada de  $y$  das UE vizinhas, seriam os valores conhecidos. Tal decisão permitiu que o impacto de  $p$  fosse única e exclusivamente pelo senso espacial, pois na previsão do



termo  $pWy$  foram empregados os valores reais, ou seja, sem viés temporal. Desta forma, empregou-se em  $Wy$  os  $y$  conhecidos das UE vizinhas de janeiro de 2008 a dezembro de 2008, em fase com o  $y$  dependente.

Observando as considerações relativas ao equilíbrio do modelo, para uma dada UE qualquer de interesse, o processo de previsão segue o seguinte roteiro:

1. Previsão de  $X\beta$ :

$$\left[ \begin{array}{cc} PCA12_{(Jan/07)} & PCA12(bhat) \\ PCA12_{(Fev/07)} & PCA12(bhat) \\ \vdots & \vdots \\ PCA12_{(Dez/07)} & PCA12(bhat) \end{array} \right] x + \left[ \begin{array}{cc} y_{(k-1)}UC_{n(dez/07)} & y_{(k-1)}(bhat) \\ Xb_1^n & y_{(k-1)}(bhat) \\ \vdots & \vdots \\ Xb_{11}^n & y_{(k-1)}(bhat) \end{array} \right] = \left[ \begin{array}{c} Xb_1 \\ Xb_2 \\ \vdots \\ Xb_{12} \end{array} \right]^n \quad (3.21)$$

2. Previsão de  $pWy$ :

$$\left( \begin{array}{ccc} yUE_{1(Jan/08)} & \cdots & yUE_{1(Dez/08)} \\ yUE_{2(Jan/08)} & \cdots & yUE_{2(Dez/08)} \\ \vdots & \cdots & \vdots \\ yUE_{j(Jan/08)} & \cdots & yUE_{j(Dez/08)} \end{array} \right) x [W_{ij}] x [p] = \left[ \begin{array}{c} pWy_1 \\ pWy_2 \\ \vdots \\ pWy_{12} \end{array} \right]^n \quad (3.22)$$

3. Previsão de  $X\beta + pWy$ :

$$\left[ \begin{array}{c} PREV_{Jan/08} \\ PREV_{Fev/08} \\ \vdots \\ PREV_{Dez/08} \end{array} \right]^n = \left[ \begin{array}{c} Xb_1 \\ Xb_2 \\ \vdots \\ Xb_{12} \end{array} \right]^n + \left[ \begin{array}{c} pWy_1 \\ pWy_2 \\ \vdots \\ pWy_{12} \end{array} \right]^n \quad (3.23)$$

Como o processo de previsão requer a execução dos termos  $X\beta$  e  $pWy$  separados, e a sua composição segue regras particulares de casamento dimensional, utilizou-se para isto o *Microsoft Excel*, pela sua flexibilidade para trabalhar com dados dispostos em matrizes de ordens diferentes, bem como a disponibilidade de facilidades de BD, gráficos e de cálculos matemáticos na mesma ferramenta.

### 3.4 – TESTES DO MODELO E ANÁLISES DOS RESULTADOS

Este tópico trata das atividades realizadas para validar o emprego da estrutura SAR *Panel* ao problema em questão. Neste sentido realizaram-se as análises de estrutura, resultados das previsões e estabilidade do modelo.

A avaliação da estrutura tem a função de averiguar o quanto a estrutura SAR mitigou os efeitos espaciais associados. Em continuidade, a verificação dos resíduos objetiva avaliar os desvios da previsão, contra a massa de teste. Por último, os testes de

estabilidade avaliam a robustez do modelo, ou seja, a capacidade de repetir os resultados da previsão frente à deterioração dos dados.

#### **3.4.1 – TESTES DE ESTRUTURA SAR PANEL**

Inicialmente, cabe lembrar que de acordo com DUBIN (2004) um modelo econométrico pode conter os resíduos provenientes do atraso espacial da variável dependente (*spatial lag*) e de erros espacialmente correlacionado (*spatial error*). Na prática, a situação de atrasos conjuntos, raramente é investigada, visto a dificuldade em estimá-los, porque a matriz  $W_{ij}$  empregada em ambos os erros geralmente é a mesma.

Nesta tese os testes de estrutura do modelo SAR *Panel*, são baseados na estratégia concebida por LESAGE (1999) que busca identificar se a inclusão do termo de atraso espacial mitiga a dependência espacial nos resíduos do modelo. Esta linha difere dos testes aplicados em modelos baseados em erros, pois permite avaliar a influência do atraso espacial da variável dependente no modelo. O teste para dependência de espaço é condicionado à existência de um parâmetro  $\rho$  diferente de zero.

Especificamente quanto aos modelos *Panel*, ELHORST (2003) assevera que as estruturas de atrasos podem ser comparadas às de erros correlacionados, a partir de testes de multiplicadores de *Lagrange* ou *F-test*. Nesta tese foram empregadas a forma comum, intitulada *LM test* e a forma robusta, intitulada *Robust LM Test*

Estes testes são realizados a posteriori da estimação, entretanto, de forma diferente da análise de correlação, os testes de estrutura empregam a matriz de entrada de dados em  $X\beta$ , a matriz  $W_{ij}$  e o vetor de variáveis dependente  $y$ , além dos resíduos. O teste de estrutura foi realizado empregando o código *LMSARSEM\_PANEL*, do *Tool Box* de econometria espacial desenvolvido para aplicação no *software* MATLAB. Assim como no caso da estimação, este código foi descrito por ELHORST (2003) e foi desenvolvido com base na estrutura do teste SAR original escrita por LeSage em 2000.

#### **3.4.2 – ANÁLISE DOS RESULTADOS DA PREVISÃO SAR PANEL**

A análise dos resultados das previsões teve como base os desvios entre os valores previstos e os valores da massa de teste, a partir dos quais foram elaborados os gráficos de:

- Erro relativo mensal – percentual mensal, positivo ou negativo, do desvio relativo entre o valor previsto e o real;

- Erro acumulado total – posição mensal, em valores percentuais, do desvio acumulado entre as somas dos valores previstos e reais; e,
- Erro absoluto mensal - percentual mensal do desvio absoluto entre o valor previsto e o valor real.

O primeiro gráfico tem o objetivo de verificar o comportamento da previsão no tempo com relação ao centro da meta, ou os valores reais da massa de teste. Tal gráfico demonstra a capacidade que o modelo apresentou em corrigir os valores previstos, reduzindo o erro frente aos dados de teste.

O segundo gráfico permite acompanhar o quanto a previsão acumulada desvia dos valores reais acumulados. Neste caso o valor do mês de dezembro demonstra o erro acumulado total entre o consumo previsto e o real no período da previsão.

Esta análise é importante, pois a contratação de GN junto à transportadora é por volume, e com isto uma boa previsão total individual local por UE permitirá uma boa previsão global. Ou seja, mesmo que a previsão não apresente um bom ajuste mensal, pode ter um bom desempenho total, subsidiando as estratégias de contratação.

O terceiro gráfico apresenta o erro mensal absoluto, base para determinar o erro percentual médio absoluto (*Mean Absolute Percent Error* – MAPE) que permite avaliar o quanto as previsões individuais mensais se distanciaram do real. Tal avaliação é importante, pois permite determinar o grau das variações locais por UE no escopo da distribuição espacial do consumo, fator preponderante para permitir uma melhor espacialização da rede, uma vez que frente às incertezas da previsão, opta-se por distribuir a rede pelas UE que apresentaram o menor MAPE.

Ainda empregando o MAPE será realizado o teste de hipótese com objetivo de avaliar o quanto as previsões individuais por UE são distintas. Neste caso a hipótese nula  $H_0$  significa que os desvios entre a previsão e os valores reais têm a mesma significância e a hipótese alternativa  $H_1$ , significa que os desvios têm significâncias diferentes e neste caso, pode-se optar pelas previsões com o menor MAPE.

Empregou-se na análise da hipótese o Teste  $t$  de *Student* combinado para duas amostras, conforme MONTGOMERY e RUNGER (2009), com distribuição bi-caudal e  $\alpha$  igual a 0,05. Tanto os gráficos quanto os índices desta etapa foram realizados empregando o *Microsoft Excel*.

### 3.4.3 – TESTES DE ESTABILIDADE DO MODELO

Os testes de estabilidade foram realizados para os experimentos SAR *Panel* que obtiveram as melhores previsões individuais por UE. Ao todo foram cinco pares de experimentos selecionados, um para cada UE, considerando os arranjos de formas funcionais, opção de entrada de dados e método de determinação.

O teste de estabilidade baseou-se na degradação e substituição dos valores das séries de testes das UE, a partir da geração de conjuntos de dados aleatórios. O gerador de números aleatórios empregado foi o do pacote de ferramentas estatísticas Análise de dados do *Microsoft Excel*. Dentre as possibilidades disponíveis para a geração aleatória optou-se pela forma de distribuição uniforme pelos seguintes motivos:

- Permite gerar números aleatórios com intervalos de  $-\infty$  a  $+\infty$ ;
- Permite gerar uma série de números aleatórios de igual valor ao da massa de teste, neste caso 12 eventos; e,
- Permite gerar números respeitando os intervalos individuais de cada UE.

Ao todo foram realizados vinte experimentos comparativos, sendo quatro por UE, dois para cada arranjo de forma funcional (PSAR e SARSFE) e método de determinação (*Indet full e MC*). A diferenciação entre os dois testes, intitulados 1 e 2, está no limite adotado para definir os intervalos da distribuição uniforme destes dois conjuntos de dados.

No primeiro caso, considerou-se os limites mínimos e máximos dos doze valores de  $y$  das próprias séries de teste de cada UE para gerar os números aleatórios. No segundo caso, o intervalo considerou a série completa de  $y$  para cada UE.

O processo de previsão seguiu o mesmo procedimento descrito em 3.3.4, adotando os mesmos parâmetros estimados para os experimentos originais, ocorrendo somente a substituição dos valores de  $y$  original pelos degradados no cálculo de  $pWy$ .

Esta prática permitiu degradar toda a contribuição espacial no resultado da previsão e tal situação vai ao encontro à questão abordada sobre o avanço de  $Wy$ . Em condições normais estes valores também deveriam ser previstos para compor a previsão; entretanto, pelos motivos abordados naquele item, definiu-se que os valores utilizados como entrada para as UE vizinhas seriam os valores conhecidos.

Sendo assim, a degradação gera uma massa de  $y$  desconhecidos, diferente da amostra original e desta forma, cumpre o papel de fornecer ao modelo um conjunto de dados que, por exemplo, poderia ser fruto da previsão individual de cada UE vizinha.

Para comparar os valores da previsão original com a degradada, empregou-se o *Microsoft Excel*, e foram elaborados os mesmos gráficos, bem como o cálculo do MAPE e o teste de hipótese.

### **3.5 – COMPARAÇÃO COM OUTRAS ESTRUTURAS DE PREVISÃO**

Para avaliar os resultados da previsão do modelo SAR *Panel* frente a outros modelos empregados ao tema, estabeleceram-se três previsores: dois lineares, empregando estrutura ARIMA e método de alisamento exponencial e, um não linear empregando RN *feedforward back propagation*.

Para estes previsores também foram realizados os gráficos de erro e os testes de hipóteses para comparar seus resultados, bem como aos dos experimentos SAR *Panel*.

#### **3.5.1 – SIMULAÇÃO DOS PREVISORES LINEARES**

O previsor ARIMA foi estimado a partir do *Time Series Analysis Tool Box*, opção *ARIMA and Autocorrelation Functions* do STATISTICAL, forma funcional *Single Series ARIMA*. Para cada UE determinou-se a previsão baseada em quatro experimentos, considerando doze atrasos de sazonalidade e a estimação de constante:

- ARIMA (1,0,0) - atraso de ordem 1, sem transformação;
- ARIMA (2,0,0) - atraso de ordem 2, sem transformação;
- ARIMA (1,2,0) - atraso de ordem 1, com transformação de variáveis por diferença com atraso 1 e dois passos; e,
- ARIMA (2,2,0) - atraso de ordem 2, com transformação de variáveis por diferença com atraso 1 e dois passos.

Cabe ressaltar que para cada experimento devem ser realizados o teste de raiz unitária para identificar se o modelo é não-estacionário ou se a estrutura não permite a inversão das matrizes. Para estes casos, procedem-se os experimentos sem a determinação da constante  $e$  / ou sem os efeitos de reinjeção da parcela média móvel.

O segundo previsor linear estimado foi o de alisamento exponencial, realizado com o STATISTICAL 7.0, a partir do *Time Series Analysis Tool Box*, opção *Exponential Smoothing & Forecasting*, forma funcional *Seasonal and Non-Seasonal Exponential Smoothing*. Neste caso, para cada UE foram determinadas a previsão baseada em dois experimentos:

- Alisamento simples sem componentes de tendência e sazonal; e,

- Alisamento simples com componente de tendência linear, *trend holt e*, sem componente sazonal.

Não foi possível incluir a componente sazonal nos experimentos porque o comprimento total da série não é suficiente para atender o requisito de que existam pelo menos cinco ciclos sazonais completos e três observações para cada ciclo. Sendo assim, procederam-se aos experimentos sem a análise da sazonalidade.

Para a comparação com os resultados *SAR Panel*, estes valores foram tabulados e apresentados em gráficos com a posterior determinação do MAPE e dos testes de hipóteses, com base no Teste *t* de *Student* combinado para duas amostras. Os gráficos, o cálculo do MAPE e o teste de hipótese foram realizados empregando o *Microsoft Excel*, do mesmo que na elaboração para o modelo *SAR Panel*.

### 3.5.2 – SIMULAÇÃO DO PREVISOR NÃO-LINEAR

O previsor não-linear privilegiou a avaliação da capacidade das RN em mapear o problema espacial, a partir do fornecimento do consumo da UE de interesse e de seus vizinhos, sem a apresentação das relações de contiguidade (matriz  $W_{ij}$ ).

A partir do *Neural Network Tool* do MATLAB 7.0, estabeleceu-se para a UE, na qual o previsor SAR apresentou a pior previsão, uma RN com uma camada intermediária com 5 neurônios com função de ativação por tangente hiperbólica e uma camada de saída com 1 neurônio com função de ativação linear, conforme a figura 3.6.

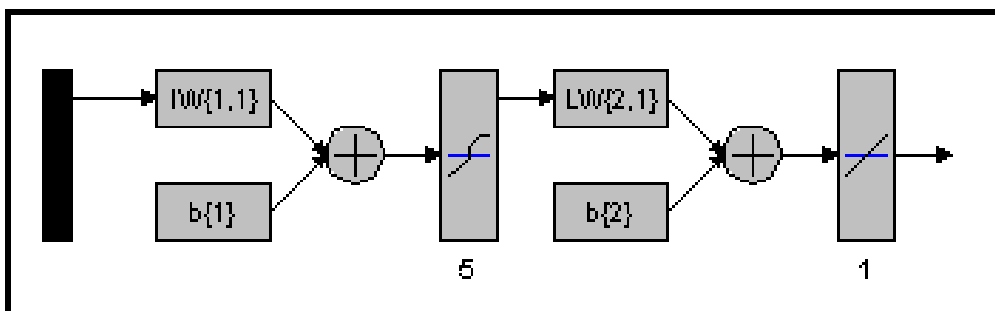


Figura 3.6 – Estrutura da Rede Neural.

Inicialmente mapeou-se esta UE observando as possíveis combinações dentre um conjunto de métodos de treinamento e funções de adaptação de parâmetros de treinamento (deste ponto em diante intitulada funções de adaptação)

No conjunto de treinamento da RN foram elencados os métodos de otimização de *Levenberg-Marquardt* (TRAINLM), gradiente descendente com taxa adaptativa de aprendizado (TRAINGDM), otimização de *Levenberg-Marquardt* com regularização



A determinação das matrizes de dados para o treinamento e simulação da RN seguiu o mesmo casamento dimensional avaliado em 3.3.3, conforme tabela 3.4 e 3.5.

Cabe lembrar que para o treinamento da RN não se forneceram as relações de vizinhança conforme a matriz  $W_{ij}$ , ou seja, espera-se que a instabilidade estrutural espacial, tal como a dependência espacial sejam “capturadas” pela RN, à luz dos dados fornecidos. Esta particularidade faz com que o previsor neural tenha caráter espacial global, pois consideram-se no treinamento todos os atrasos espaciais possíveis, independente de uma regra de contiguidade (distância e topologia).

Ainda sobre os dados de entrada ressalta-se que não foram fornecidos dados de UE vizinhas que apresentassem consumo zero no período estipulado para o treinamento. Tal procedimento seguiu a mesma regra adotada na previsão SAR *Panel*, onde após a determinação das UE que apresentaram estacionariedade da dependência espacial local, retiraram-se as unidades vizinhas que não apresentaram consumo, no período estipulado para determinação dos coeficientes.

Com base nestas premissas realizou-se o treinamento das RN, seguido da geração do conjunto de gráficos das previsões e o cálculo do MAPE, para cada experimento realizado. Para os casos onde o treino convergiu rapidamente, o experimento foi refeito do zero, estabelecendo uma nova RN com as mesmas características. Sendo assim, não se empregou a opção de reverter ou re-inicializar os pesos para um novo treinamento.

O mapeamento inicial foi expandido para mais duas UE, respectivamente segunda e terceira piores performances com o SAR *Panel*. Nesta replicação considerou-se o experimento de melhor desempenho, conforme a melhor condição obtida no treinamento<sup>32</sup> da primeira UE.

O formato de entrada de dados para estes dois novos testes considerou a mesma estrutura empregada em (3.24) a (3.26), respeitando a quantidade de UE vizinhas e o casamento dimensional da série de dados.

Assim como no caso anterior foram gerados o conjunto de gráficos das previsões e o cálculo do MAPE para cada experimento realizado, complementados pelos testes de hipótese de significância do experimento RN contra o SAR *Panel*.

---

<sup>32</sup> Entenda-se como melhor condição de treinamento as estruturas da RN e suas respectivas funções de treinamento e adaptação de treinamento, bem como os parâmetros de: inclinações, épocas de treinamento, inicialização de  $\mu$ , decréscimo / incremento de  $\mu$ , meta, variação do gradiente, tempo, dentre as outras condições finais do treino.



### 3.5.2.1 – SIMULAÇÃO DOS PREVISORES POR RN SEM A VIZINHANÇA

Até aqui os experimentos de RN tem o objetivo de investigar a capacidade de mapear o problema espacial da UE de interesse e de seus vizinhos, com o fornecimento dos dados de consumo; não avalia, entretanto, a influência dos dados da vizinhança nos resultados, ou seja, o mote espacial no experimento não linear com base em RN.

Sendo assim, coube investigar o quanto estes experimentos alcançaram a condição de mapeamento, a partir da apresentação dos dados de consumo das UE vizinhas a de interesse, ou são frutos unicamente da capacidade de mapeamento das RN.

Para tal replicaram-se os experimentos do tópico anterior com visão puramente temporal, alterando-se somente a entrada de dados pela retirada da informação da vizinhança. Com isto as dimensões das matrizes ficaram conforme a tabela 3.5.

Tabela 3.5 – Matrizes de Entrada de Dados para RN puramente Temporais.

| Entrada ( $m \times n$ )  |          | Target ( $m \times n$ )  |          | Teste ( $m \times n$ )    |         |
|---------------------------|----------|--------------------------|----------|---------------------------|---------|
| Variáveis                 | Eventos  | Variáveis                | Eventos  | Variáveis                 | Eventos |
| PCA ( $k-12$ )            | 1        |                          |          | PCA ( $k-12$ )            | 1       |
| UE <sub>a</sub> ( $k-2$ ) | 1 52     | UE <sub>a</sub> ( $k0$ ) | 1 52     | UE <sub>a</sub> ( $k-2$ ) | 1 1     |
| UE <sub>a</sub> ( $k-1$ ) | 1        |                          |          | UE <sub>a</sub> ( $k-1$ ) | 1       |
| Total                     | (3 × 52) | Total                    | (1 × 52) | Total                     | (3 × 1) |
| PCA ( $k-12$ )            | 1        |                          |          | PCA ( $k-12$ )            | 1       |
| UE <sub>b</sub> ( $k-2$ ) | 1 48     | UE <sub>b</sub> ( $k0$ ) | 1 48     | UE <sub>b</sub> ( $k-2$ ) | 1 1     |
| UE <sub>b</sub> ( $k-1$ ) | 1        |                          |          | UE <sub>b</sub> ( $k-1$ ) | 1       |
| Total                     | (3 × 48) | Total                    | (1 × 48) | Total                     | (3 × 1) |
| PCA ( $k-12$ )            | 1        |                          |          | PCA ( $k-12$ )            | 1       |
| UE <sub>c</sub> ( $k-2$ ) | 1 45     | UE <sub>c</sub> ( $k0$ ) | 1 45     | UE <sub>c</sub> ( $k-2$ ) | 1 1     |
| UE <sub>c</sub> ( $k-1$ ) | 1        |                          |          | UE <sub>c</sub> ( $k-1$ ) | 1       |
| Total                     | (3 × 45) | Total                    | (1 × 45) | Total                     | (3 × 1) |

Ao final foram gerados o conjunto de gráficos das previsões e o cálculo do MAPE, para os testes das RN sem os valores da vizinhança; seguido da execução dos testes de hipótese de significância do experimento RN sem os vizinhos contra, respectivamente as RN com vizinhança e o SAR *Panel*.

O último passo da tese é emitir as considerações finais sob a forma de conclusões indicando quais os possíveis desenvolvimentos ou melhorias que se seguem a esta.

## 4 – DESENVOLVIMENTO DA TESE

Conforme estabelecido na metodologia o desenvolvimento divide-se em cinco partes:

- 4.1 – Estruturação e Análise de dados;
- 4.2 - Diretrizes para a Modelagem;
- 4.3 – Desenvolvimento do Modelo;
- 4.4 – Análise de Resultados; e,
- 4.5 – Comparações com Modelos Lineares e Não-lineares.

### 4.1 – ESTRUTURAÇÃO E ANÁLISE DE DADOS

Neste tópico descrevem-se as atividades de coleta e tratamento das variáveis dependente, explicativas e espaciais, onde são detalhadas a compatibilização, a organização, a retirada de séries de dados espúrias e a montagem do BD.

#### 4.1.1 – ESCOLHA DA CDL E DEFINIÇÃO DOS LIMITES TEMPORAL E ESPACIAL

Com base na tabela 4.1, que apresenta a quantidade de clientes por segmento em abril de 2008, elencou-se como possíveis CDL de interesse, as que pertencem as UF do Rio de Janeiro, São Paulo, Espírito Santo e Paraná. Após as visitas técnicas para avaliar o (a): cadastro de clientes e de logradouro, MUB georreferenciado e espacialização de clientes, priorizaram-se à tese as duas últimas.

Tabela 4.1 – Quantidade de Clientes, por Segmento e por CDL – Fonte: Petrobras.

| Distribuidora           | Municípios atendidos | Clientes    |             |            |                             |            |               |           | Extensão da rede (Km) |
|-------------------------|----------------------|-------------|-------------|------------|-----------------------------|------------|---------------|-----------|-----------------------|
|                         |                      | Industriais | Automotivos | Comerciais | Residenciais <sup>(1)</sup> | Co-geração | Termelétricas | Total     |                       |
| Algas                   | 12                   | 18          | 36          | 271        | 12.724                      | -          | -             | 13.049    | 216                   |
| BahiaGas                | 13                   | 95          | 50          | 63         | 1.732                       | -          | 1             | 1.941     | 530                   |
| Cegas <sup>(2)</sup>    | 16                   | 109         | 75          | 34         | 8                           | 6          | 2             | 234       | 260                   |
| BR                      | 6                    | 26          | 25          | 109        | 8.004                       | -          | -             | 8.164     | 123                   |
| Gasnig <sup>(3)</sup>   | 22                   | 173         | 92          | -          | -                           | -          | 2             | 267       | 366                   |
| Compagas <sup>(4)</sup> | 7                    | 94          | 26          | 167        | 2.773                       | 2          | 1             | 3.063     | 475                   |
| CEG                     | 16                   | 339         | 384         | 10562      | 713498                      | 0          | 3             | 724786    | 3704                  |
| CEG Rio                 | 17                   | 77          | 77          | 224        | 19.389                      | -          | 3             | 19.770    | 825                   |
| SCGAS                   | 42                   | 136         | 94          | 102        | 174                         | -          | -             | 506       | 726                   |
| COMGAS <sup>(5)</sup>   | 66                   | 989         | 384         | 8.563      | 683.692                     | 16         | 2             | 693.646   | 5.255                 |
| GAS Brasileiro          |                      |             |             |            |                             |            |               | 0         |                       |
| Gas Natural             |                      |             |             |            |                             |            |               | 0         |                       |
|                         | 217                  | 2.056       | 1.243       | 20.095     | 1.441.994                   | 24         | 14            | 1.465.426 |                       |

(1) Quantidade de unidades habitacionais

(2) Quantidade de clientes residenciais = quantidade de condomínios

(3) O total de 173 clientes industriais é composto de: Industrial = 90; GNC Industrial = 6; Uso Geral = 69; GNL = 8

(4) O total de 94 clientes industriais é composto de: Industrial = 93; Matéria-prima = 1

(5) Relatório de Administração 2007 - <http://www.comgas.com.br/investidores/media/pdf/1151.08%20RAComgas%202007.pdf>

A Compagas, que atua no Paraná, iniciou as atividades 1998, distribuindo gás de refinaria às indústrias de Curitiba, Araucária e Campo Largo. A partir de 2000, com a inauguração do ramal sul do GASBOL, passou a distribuir GN proveniente da Bolívia. Sua rede de distribuição até abril de 2008 atendia os municípios relacionados na tabela

4.2, com atuação direta nos segmentos industrial, comercial, GNV e residencial, exclusivamente na cidade de Curitiba, além de fornecer GN para co-geração, matéria-prima e para geração elétrica na Usina Termoelétrica de Araucária.

Na BR, a Petrobras Distribuidora S.A., subsidiária da Petróleo Brasileiro S.A., obteve em 1993 a concessão para distribuição de GN na UF do Espírito Santo, nos segmentos Residencial, Comercial, Industrial, Construtoras<sup>33</sup> e GNV.

#### 4.1.1.1 - LIMITE TEMPORAL E DA ÁREA DE CONCESSÃO DAS CDLS

Com relação ao limite temporal é possível definir a partir das tabelas 4.2 e 4.3, a seguir, que a série de dados disponíveis para a Compagas compreende os anos de 2000 a 2008 e, de 1999 a 2008 para a BR. Cabe ressaltar que os dados apresentados nas duas tabelas encontram-se em frequência anual, porém a série original encontra-se em base mensal como poderá ser visto mais adiante no item referente à análise dos dados.

Com base no consumo apresentado nestas tabelas delimitou-se o seguinte conjunto de municípios como de interesse para definir a AE:

- Compagas – Araucária, Balsa Nova, Campo Largo, Curitiba, Palmeira, Ponta Grossa e São José dos Pinhais.
- BR – Aracruz, Cariacica, Linhares, Serra, Viana, Vila Velha e Vitória.

Tabela 4.2 – Consumo Total Anual de GN em MM<sup>3</sup> – Fonte: Compagas.

| Município  | Segmento    | 1999  | 2000  | 2001  | 2002  | 2003  | 2004  | 2005  | 2006  | 2007  | 2008  |
|------------|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Aracruz    | VEICULAR    | 0,0   | 0,0   | 0,0   | 0,2   | 1,0   | 0,9   | 1,0   | 1,0   | 1,1   | 0,8   |
|            | INDUSTRIAL  | 63,1  | 64,5  | 64,6  | 83,6  | 93,4  | 94,9  | 101,0 | 105,7 | 106,9 | 112,9 |
| Cariacica  | VEICULAR    | 0,0   | 2,1   | 10,5  | 14,4  | 16,0  | 13,7  | 13,4  | 13,1  | 12,1  | 9,7   |
|            | INDUSTRIAL  | 22,5  | 23,0  | 20,1  | 19,8  | 18,8  | 21,1  | 19,1  | 21,2  | 22,0  | 20,0  |
| Colatina   | GNV         | 0,0   | 0,0   | 0,0   | 0,0   | 0,0   | 0,0   | 0,0   | 0,5   | 1,0   | 0,8   |
| Linhares   | VEICULAR    | 0,0   | 0,0   | 0,0   | 0,0   | 0,0   | 0,0   | 0,0   | 0,7   | 1,6   | 1,2   |
| Serra      | VEICULAR    | 0,0   | 5,0   | 6,1   | 6,9   | 9,7   | 9,9   | 10,5  | 12,1  | 13,1  | 11,8  |
|            | INDUSTRIAL  | 25,2  | 26,2  | 28,9  | 33,8  | 41,6  | 44,3  | 44,0  | 48,3  | 36,9  | 46,0  |
| Viana      | VEICULAR    | 0,0   | 0,0   | 0,0   | 0,0   | 0,0   | 0,0   | 0,0   | 1,2   | 2,9   | 2,3   |
|            | INDUSTRIAL  | 1,2   | 4,0   | 4,2   | 4,8   | 6,7   | 7,9   | 8,6   | 8,4   | 7,2   | 9,9   |
| Vila Velha | VEICULAR    | 0,0   | 0,0   | 0,0   | 0,0   | 0,0   | 4,1   | 7,5   | 11,0  | 12,4  | 10,6  |
|            | INDUSTRIAL  | 0,0   | 0,0   | 0,0   | 0,0   | 0,0   | 0,0   | 0,0   | 0,1   | 2,5   | 2,6   |
| Vitória    | COMERCIAL   | 0,0   | 0,0   | 0,0   | 0,0   | 0,1   | 0,3   | 1,0   | 1,2   | 1,5   | 1,6   |
|            | VEICULAR    | 0,0   | 0,0   | 0,0   | 4,1   | 8,2   | 8,7   | 8,6   | 8,8   | 8,8   | 9,6   |
|            | INDUSTRIAL  | 105,6 | 140,0 | 199,0 | 179,9 | 198,3 | 203,6 | 170,5 | 168,8 | 210,2 | 430,6 |
|            | RESIDENCIAL | 0,0   | 0,0   | 0,0   | 0,0   | 0,0   | 0,1   | 0,3   | 0,6   | 0,7   | 1,0   |

<sup>33</sup> Este segmento é um modelo de negócio voltado ao segmento de mercado residencial. Seu foco é viabilizar o GN ao empreendimento já no ato de sua construção, ou seja, uma vez o empreendimento concluído já possui a infraestrutura necessária para disponibilização do GN as unidades residenciais. Em outras CDL isto tem sido tratado na regulação, pela adoção de códigos de obras que obrigam a existência de infra-estruturas para distribuir GLP e GN.

Tabela 4.3 – Consumo Total Anual de GN em MM<sup>3</sup> – Fonte: BR Distribuidora.

| Município          | Segmento    | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 |
|--------------------|-------------|------|------|------|------|------|------|------|------|------|
| Araucária          | COMERCIAL   | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  |
|                    | INDUSTRIAL  | 3,2  | 24,9 | 29,8 | 34,2 | 46,0 | 37,9 | 42,1 | 47,5 | 54,9 |
|                    | VEICULAR    | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  | 0,4  | 0,2  | 0,0  |
| Balsa Nova         | INDUSTRIAL  | 8,4  | 18,7 | 17,1 | 0,0  | 0,0  | 0,0  | 0,0  | 0,2  | 0,2  |
| Campo Largo        | COMERCIAL   | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  |
|                    | INDUSTRIAL  | 1,6  | 23,5 | 24,9 | 24,9 | 26,7 | 22,6 | 23,7 | 28,4 | 31,6 |
|                    | VEICULAR    | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  | 0,4  | 1,3  | 1,4  |
| Curitiba           | COMERCIAL   | 0,0  | 0,0  | 0,1  | 0,4  | 0,6  | 0,9  | 2,0  | 2,5  | 3,0  |
|                    | INDUSTRIAL  | 2,4  | 11,0 | 20,1 | 29,1 | 37,9 | 35,9 | 38,3 | 42,4 | 41,9 |
|                    | RESIDENCIAL | 0,0  | 0,0  | 0,0  | 0,1  | 0,1  | 0,3  | 0,6  | 0,8  | 1,3  |
|                    | VEICULAR    | 0,0  | 0,2  | 2,3  | 3,5  | 4,3  | 14,2 | 27,1 | 29,4 | 29,6 |
| Palmeira           | INDUSTRIAL  | 1,1  | 2,9  | 3,2  | 2,0  | 1,9  | 3,6  | 5,3  | 5,6  | 6,2  |
| Ponta Grossa       | INDUSTRIAL  | 2,5  | 16,1 | 17,6 | 13,8 | 9,9  | 8,7  | 10,6 | 10,5 | 11,3 |
| São J. dos Pinhais | COMERCIAL   | 0,0  | 0,0  | 0,1  | 0,1  | 0,1  | 0,1  | 0,1  | 0,1  | 0,1  |
|                    | INDUSTRIAL  | 0,1  | 20,7 | 22,2 | 21,3 | 24,9 | 27,6 | 25,3 | 27,9 | 29,4 |
|                    | VEICULAR    | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  | 0,0  | 0,6  | 2,4  | 2,6  |

#### 4.1.1.2 - LIMITE ESPACIAL DA AE DA COMPAGAS E DA BR

Observando o conjunto de municípios relacionados no tópico anterior, percebe-se que existe uma região economicamente contígua às duas capitais contidas no conjunto: Curitiba e Vitória. Esta região é denominada Região Metropolitana (RM).

A RM original de Curitiba era formada por: Curitiba, Alm. Tamandaré, Araucária, Bocaiúva do Sul, Campo Largo, Colombo, Contenda, Piraquara, S. José dos Pinhais, Rio Branco do Sul, Campina Grande do Sul, Quatro Barras, Mandirituba e Balsa Nova.

Com o tempo agregaram-se os municípios de: Adrianópolis, Agudos do Sul, Fazenda Rio Grande, Dr. Ulysses, Campo Magro, Cerro Azul, Itaperuçu, Lapa, Pinhais, Quitandinha, Tijucas do Sul e Tunas do Paraná, totalizando 26 municípios apresentados na figura 4.1 a seguir.



Figura 4.1 – RM – Fonte: IBGE.

Na figura 4.2, apresentam-se respectivamente os municípios atendidos pela Compagas, classificados conforme o volume de GN comercializado em 2008 e a sobreposição da RM de Curitiba.

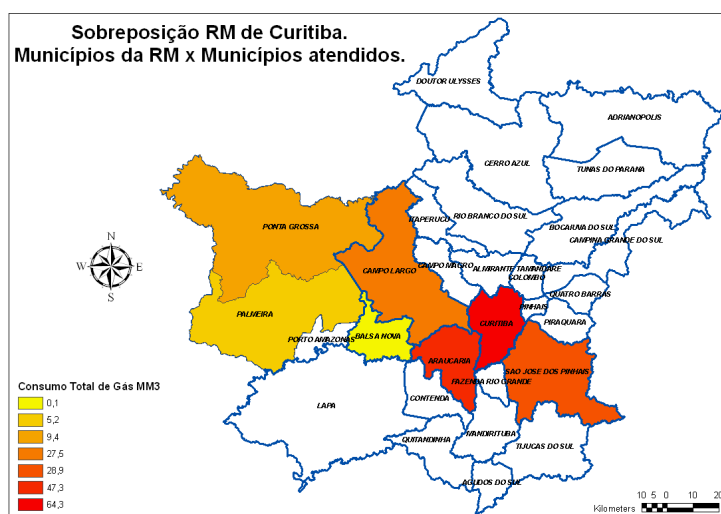


Figura 4.2 – Consumo Total de GN Municipal x Municípios da RM – Fonte: Compagas.

A partir da análise da sobreposição são extraídos inicialmente da AE, os municípios atendidos de Ponta Grossa e Palmeira, pois são externos à RM. Ressalta-se que este critério atende à questão da distância do centro de gravidade de consumo de GN que fica entre Curitiba e Araucária.

Para analisar a contiguidade espacial e a influência econômica dos municípios vizinhos, adotou-se como linha de corte a representatividade econômica, representada pelo percentual de PIB individual de cada município, fornecido pelo Instituto de Pesquisa e Planejamento Urbano de Curitiba (IPPUC) e que pode ser visto na figura 4.3.

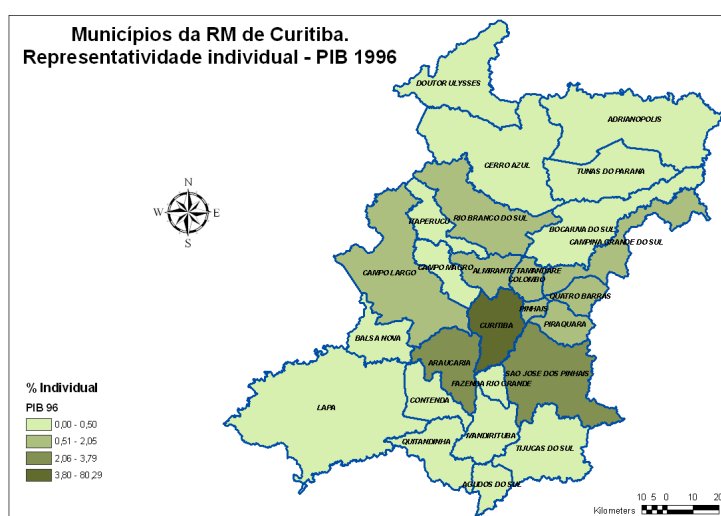


Figura 4.3 – Classificação Municipal - Representatividade % do PIB – Fonte: IPPUC.

A classe mais alta de representação tem um único elemento, Curitiba; por isto incluíram-se os municípios constantes das duas classes centrais de representatividade



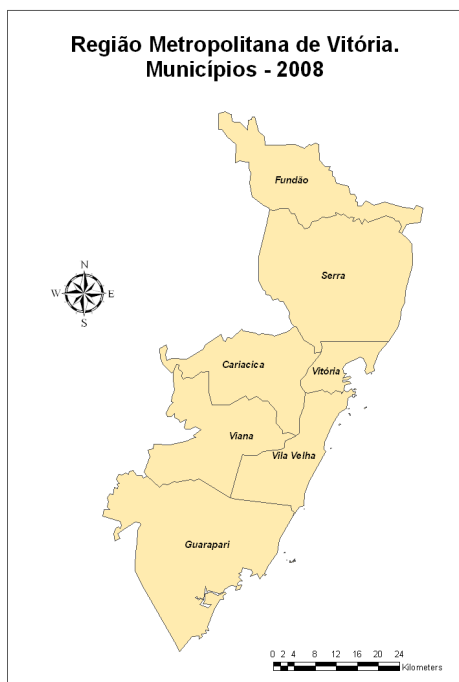


Figura 4.5 – RM de Vitória – Fonte: IBGE.

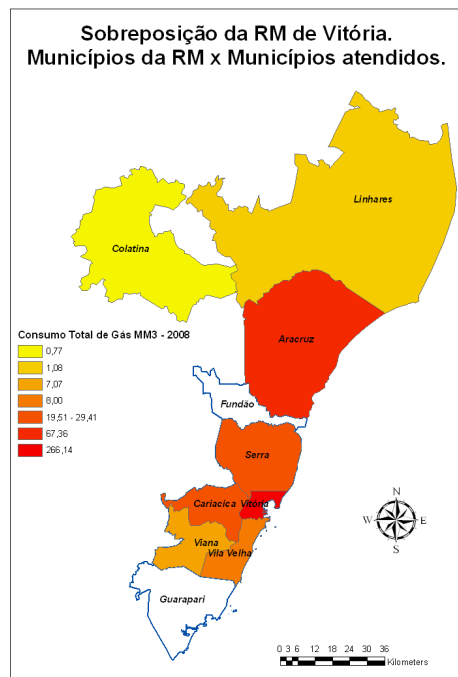


Figura 4.6 – Consumo Municipal de GN x Municípios da RM – Fonte: BR Distribuidora.

Analogamente à CDL anterior, a partir da sobreposição extraem-se inicialmente da AE os municípios atendidos de Aracruz, Colatina e Linhares; e a seleção dos municípios com possível influência espacial baseou-se na participação econômica do município na RM, a partir do valor agregado por municípios, fornecido pelo Instituto de pesquisa Jones Santos Neves (IJSN), do Estado do Espírito Santo, para qual a classificação pode ser vista na figura 4.7.

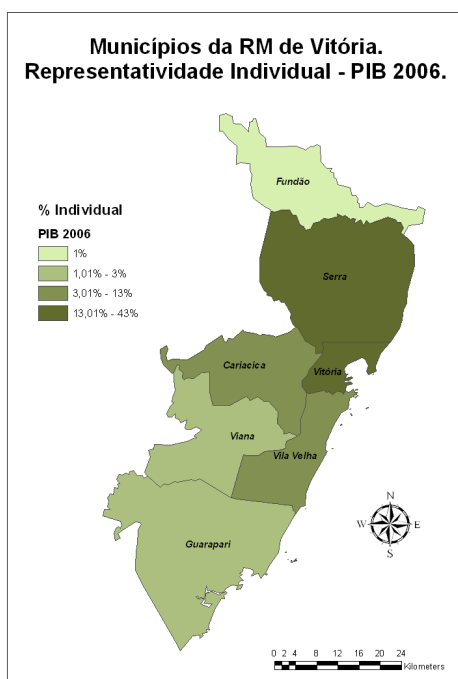


Figura 4.7 – Classificação da Representatividade.% do PIB - Fonte: IJSN.

A classe mais alta de representação teve dois elementos, Vitória e Serra. A classe imediatamente inferior com municípios com representatividade entre 3,01% e 13% também apresenta dois elementos, Cariacica e Vila Velha

Vitória representa cerca de 43% do PIB na RM. Para garantir que todas as suas dependências espaciais fossem incorporadas, a AE contemplou os municípios imediatamente contíguos, ou seja, Cariacica, Serra e Vila Velha. Este conjunto representa cerca de 95% do PIB da RM, e o seu contorno é apresentado na figura 4.8.

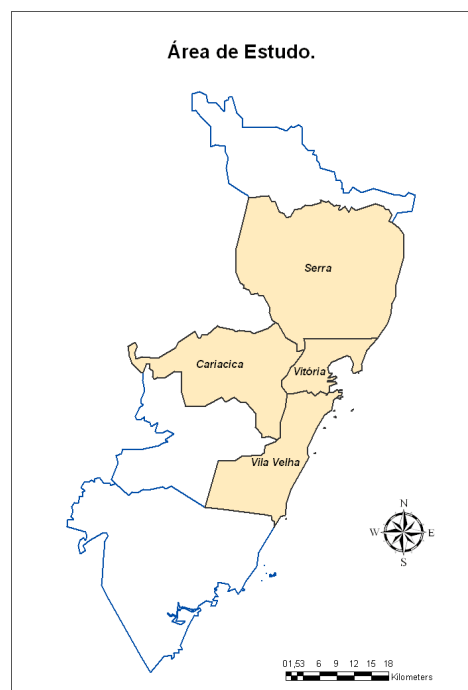


Figura 4.8 – AE BR

Ressalta-se que Aracruz, com cerca de 23,6% do consumo de GN entre 1999 e 2008 (tabela 4.3), apresenta a mesma quantidade de clientes desde 2002 (tabela 4.5), ou seja, só houve acréscimo do volume de GN distribuído. Esta característica, aliada à descontinuidade espacial, faz crer que o consumo de GN em Aracruz não apresenta dependência espacial, ou seja, seu consumo não tem influência do espaço.

Tabela 4.5 – Quantidade Total de Clientes – Base Dez – Fonte: BR Distribuidora.

| Município  | Segmento    | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008  |
|------------|-------------|------|------|------|------|------|------|------|------|------|-------|
| Aracruz    | VEICULAR    | 0    | 0    | 0    | 1    | 1    | 1    | 1    | 1    | 1    | 1     |
|            | INDUSTRIAL  | 2    | 2    | 2    | 2    | 2    | 2    | 2    | 2    | 2    | 2     |
| Cariacica  | VEICULAR    | 0    | 1    | 4    | 4    | 4    | 4    | 4    | 4    | 4    | 4     |
|            | INDUSTRIAL  | 4    | 4    | 4    | 3    | 3    | 4    | 4    | 4    | 4    | 4     |
| Colatina   | GNV         | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 1    | 1    | 1     |
| Linhares   | VEICULAR    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 1    | 1    | 1     |
| Serra      | VEICULAR    | 0    | 1    | 1    | 3    | 4    | 6    | 6    | 7    | 7    | 7     |
|            | INDUSTRIAL  | 7    | 7    | 6    | 7    | 8    | 8    | 8    | 11   | 11   | 11    |
| Viana      | VEICULAR    | 0    | 0    | 0    | 0    | 0    | 0    | 0    | 1    | 1    | 1     |
|            | INDUSTRIAL  | 3    | 3    | 3    | 3    | 3    | 4    | 4    | 5    | 5    | 5     |
| Vila Velha | VEICULAR    | 0    | 0    | 0    | 0    | 0    | 1    | 3    | 4    | 5    | 5     |
|            | INDUSTRIAL  | 0    | 0    | 0    | 0    | 0    | 0    | 1    | 2    | 2    | 2     |
| Vitória    | COMERCIAL   | 0    | 0    | 0    | 0    | 21   | 29   | 47   | 72   | 109  | 143   |
|            | VEICULAR    | 0    | 0    | 0    | 2    | 3    | 4    | 4    | 4    | 5    | 6     |
|            | INDUSTRIAL  | 4    | 4    | 4    | 5    | 5    | 5    | 5    | 5    | 7    | 7     |
|            | RESIDENCIAL | 0    | 0    | 0    | 0    | 357  | 1665 | 2819 | 4835 | 7350 | 12415 |



#### 4.1.2 – SELEÇÃO, COLETA E FORMAÇÃO DA BASE DE DADOS

As atividades de seleção, coleta e formação da BDE, tiveram como base duas dimensões de dados: espacial e alfanumérica. A primeira contém as feições espaciais e a segunda os dados de consumo, preço e socioeconômicos.

##### 4.1.2.1 - DADOS ESPACIAIS

Para caracterizar a distribuição do GN, foram coletadas para cada AE as seguintes feições para a composição do BDE, conforme as tabelas 4.6 e 4.7.

Tabela 4.6 - Configuração das feições da Base de dados espacial à RM de Curitiba.

| Curitiba  | Município   | Bairro   | Setor censitário | Logradouros | Clientes | Rede     |
|-----------|-------------|----------|------------------|-------------|----------|----------|
| Fonte     | IBGE        | IPPUC    | IBGE             | IPPUC       | Compagas | Compagas |
| Escala    | 1:2.500.000 | 1:10.000 | 1:250.000        | 1:10.000    | -        | -        |
| Data      | 2007        | 2007     | 2007             | 2007        | -        | -        |
| Primitiva | Polígono    | Polígono | Polígono         | Linha       | Ponto    | Linha    |

Tabela 4.7 - Configuração das feições da Base de dados espacial à RM de Vitória.

| Vitória   | Município   | Bairro   | Setor censitário | Logradouros | Clientes | Rede  |
|-----------|-------------|----------|------------------|-------------|----------|-------|
| Fonte     | IBGE        | BR       | IBGE             | Promaps     | BR       | BR    |
| Escala    | 1:2.500.000 | -        | 1:250.000        | -           | -        | -     |
| Data      | 2007        | 2009     | 2007             | -           | -        | -     |
| Primitiva | Polígono    | Polígono | Polígono         | Linha       | Ponto    | Linha |

O BDE espacial contemplou as seis camadas citadas e o posicionamento dos clientes foi obtido via processo de espacialização, a partir de seus endereços. Em Curitiba, 70 clientes não foram posicionados de forma automática. Para tal corrigiu-se a grafia e foi feita a localização das coordenadas via *Google Earth*, com posterior inserção destes 70 clientes no BDE. Em Vitória, 650 clientes foram posicionados de forma direta, 50 após a correção do endereço e 30 com auxílio do *Google Earth*.

Cabe ressaltar que o processo de interpolação para referenciamento dos clientes a partir do endereço é uma função linear que não considera as irregularidades da numeração ao longo das vias. Com isto, o processo fornece uma localização aproximada do cliente; entretanto, para esta tese considera-se esta aproximação suficiente, pois permite avaliar a dispersão e a posição dos mesmos.

##### 4.1.2.2 – VARIÁVEL DE INTERESSE

Para caracterizar o consumo de GN empregou-se a quantidade de clientes e o volume consumido em m<sup>3</sup>, por segmento de mercado, na maior série histórica possível, em frequência mensal. No caso da Compagas, o período disponível de dados vai de

Agosto de 2000 até Dezembro de 2008 e, na BR a série histórica contemplou o período de Janeiro de 1999 até Dezembro de 2008.

Ressalta-se que, em função da estrutura particular de gestão dos dados de clientes na CDL BR houve a necessidade de realizar um *clean-up* para gerar uma série de dados que contemple um único código de clientes; mitigando assim o risco de violações de normatização na formação do BDE para a tese.

#### **4.1.2.3 - VARIÁVEL EXPLICATIVA - ÁLCOOL E DERIVADOS DE PETRÓLEO**

Obteve-se na Agência Nacional do Petróleo (ANP) as séries históricas anuais de 1990 a 2006, por municípios, do consumo de Álcool Hidratado, Gasolina automotiva e de Aviação, GLP, Óleos Combustível e Diesel, Querosene de Aviação e Iluminante. Os dados estavam classificados por segmentos de uso, destacando-se a (o):

- Indústria (não metais, cerâmica, cimento, vidros, metalúrgica, papel, química, têxtil, produtos alimentícios, bebidas e extração mineral);
- Revenda (doméstica, comercial, entidades públicas, forças armadas); e,
- Geração de EE.

Deste conjunto foram relacionados para a tese, por AE, o álcool hidratado, a gasolina automotiva, GLP, óleos combustível e Diesel.

#### **4.1.2.4 - VARIÁVEL EXPLICATIVA - CONSUMO DE EE**

Inicialmente o objetivo era obter os dados para caracterizar o comportamento do consumo de EE nas capitais. Para tal contactou-se o Operador Nacional do Sistema (ONS) que somente dispõe de supervisão dos fluxos ativo e reativo dos transformadores com tensão de primário igual ou superior a 230 kV.

No caso da RM de Curitiba, seriam os transformadores de Umbará (2 x 230/69 kV), Uberaba (2 x 230/69 kV) e Cidade Industrial (2 x 230/69 kV e 2 x 230/13,8 kV). Já em Vitória, seriam os transformadores de Vitória (5 x 345/138 kV) e Viana (2 x 345/138 kV), sendo que estas subestações atendem não apenas Vitória, mas cerca de 80% das cargas do Estado do Espírito Santo.

Em ambas as capitais existem as redes de subtransmissão, com tensões de 138 kV ou menores que interligam a capital ao interior, fechando anéis com outros pontos de suprimento ao estado e conectando com outras centrais que não são despachadas pelo ONS. Em função da configuração desta rede de subtransmissão e do despacho destas outras usinas, a região do interior do estado, ora pode ser importadora, ora exportadora

de energia para a capital. Portanto, o somatório dos fluxos nestes transformadores não representa adequadamente o consumo de energia elétrica nestas capitais.

Além disso, o ONS só dispõe de histórico destas medidas a partir de novembro de 2001 e não tem a visão de distribuição espacial (o que vai para cada município), somente a visão por barras. O ONS sugeriu que as melhores fontes de informação para o estudo seriam os agentes de distribuição ELCELSA e COPEL.

Assim como o caso do consumo de GN, solicitaram-se às Companhias de Distribuição de Energia Elétrica, os dados referentes às séries históricas individuais mensais dos municípios, conforme variáveis abaixo descritas:

- Total mensal da carga elétrica de potência ativa nos barramentos / subestações, para os anos de 1990 até 2008;
- Consumo total em MWh por classe de consumo (domiciliar, comercial e industrial), base mensal, ou em outra frequência no máximo anual, de 1990 até 2008, ou maior período disponível; e,
- Quantidade de clientes total por classe de consumo (domiciliar, comercial e industrial), base mensal, ou em outra frequência no máximo anual, de 1990 até 2008, ou maior período disponível.

Na COPEL (Companhia Paranaense de Energia Elétrica) conseguiram-se os últimos dois dados com base anual de 1980 a 2007 e com frequência mensal para o ano de 2007 e 2008. Na ELCELSA (Centrais Elétricas do Espírito Santo) conseguiram-se os últimos dois dados com base mensal de janeiro de 2000 até junho de 2008, também segmentado por classe de consumo.

Em complemento, para a RM de Vitória obteve-se no IJSN os dados de consumo e número de consumidores de EE com base anual de 1994 a 2007.

#### **4.1.2.5 – VARIÁVEIS EXPLICATIVAS - PREÇO DOS ENERGÉTICOS**

A ANP forneceu os preços mínimo, médio e máximo mensais de revenda e de distribuição para os derivados do petróleo e, o Álcool, em R\$, entre julho de 2001 e dezembro de 2008.

Para os óleos combustíveis A1, A2 e B1, a ANP forneceu os preços médios semanais ponderados praticados pelos produtores, entre junho de 2004 e dezembro de 2008, para as regiões sul e sudeste. Neste caso, para equiparação na frequência mensal, foi necessário realizar um procedimento de interpolação descrito no Apêndice C.

Para a EE, em virtude da diferença de políticas de preços por segmento, foram fornecidas as tarifas médias mensais líquidas (livres de ICMS), por segmento (residencial, comercial e industrial), em R\$ / Kwh. No caso da Escelsa de janeiro de 1994 a dezembro de 2008 e, no caso da COPEL de janeiro de 2004 a dezembro de 2008.

No GN a política de preço admite a diferenciação de tarifas por segmento e faixa de consumo. Este último impossibilita adotar um preço específico por segmento, pois os clientes mudam de faixa de consumo no tempo. Sendo assim, adotou-se a média em R\$/m<sup>3</sup> por segmento com frequência mensal. Na Compagas a série contempla janeiro de 2002 até dezembro de 2008 e, na BR vai de janeiro de 1999 até dezembro de 2008.

#### **4.1.2.6 – VARIÁVEIS EXPLICATIVAS – ECONOMIA E DEMOGRAFIA**

Para a caracterização da atividade econômica e da demografia pesquisou-se no Ministério do Trabalho e Emprego (MTE), IBGE, IPEA, BACEN e nas secretarias estaduais de governo, quatro conjuntos de dados ao temas proposto: emprego e número de empresas por segmento da atividade econômica, PIB segmentado por tipo de conta, indicadores econômicos e financeiros e dados censitários.

(A) DADOS DE EMPREGO E EMPRESAS - O MTE coleta dados para elaborar estatísticas de trabalho via a RAIS<sup>34</sup> e o CAGED<sup>35</sup>. Em ambos os casos o empregador é obrigado a informar, por estabelecimento, qualquer movimentação no quadro de empregados (admissão, desligamento ou transferência).

Os dados são públicos, com acesso por *link* no site do MTE (<http://www.mte.gov.br/geral/estatisticas>), no qual encontra-se uma visão pela RAIS e outra por estabelecimentos. A série histórica é anual de 1985 a 2007, e os dados são classificados por estabelecimentos, município, número de empregados e segmento da economia, tanto na visão do IBGE quanto do código nacional de atividades<sup>36</sup>.

Para esta tese foram recuperados os dados do número de estabelecimentos, segundo a classificação do IBGE, ou seja, contemplando 26 segmentos da economia, para todos os 17 municípios das duas AE no período 1985 até 2007 totalizando assim 374 tabelas (22 anos x 17 municípios).

---

<sup>34</sup> A RAIS (Relação Anual de Informações Sociais) tem o objetivo de suprir às necessidades de controle da atividade trabalhista no País; bem como, prover dados para elaborar estatísticas do trabalho às entidades governamentais.

<sup>35</sup> O CAGED (Cadastro Geral de Empregados e Desempregados) registra as admissões e dispensas de empregados, sob o regime da CLT. É a base para elaborar estudos, pesquisas e programas ligados ao mercado de trabalho.

<sup>36</sup> Tem o objetivo de padronizar os códigos de atividade econômica e os critérios de enquadramento empregados pelos diversos órgãos da Administração Tributária do país.

(B) PRODUTO INTERNO BRUTO - O documento Sistemas de Contas Nacionais (SCN), do IBGE, considera o PIB um elemento agregado a partir do SCN. Por definição, consideram-se como agregado do SCN os indicadores de síntese e variáveis chaves para a análise macroeconômica e para comparações no espaço e no tempo.

A continuidade do SCN é o desdobramento das contas nacionais em contas regionais e municipais, obtidas a partir de estimativas da própria SCN. Entretanto, a revisão das contas regionais, apresentou descontinuidade nas séries históricas do PIB. No site do IBGE existiam duas séries municipais de PIB, uma de 1999 a 2004, com metodologia antiga, e uma de 2002 a 2005, com os resultados da nova<sup>37</sup> série do SCN.

Conforme informações obtidas no IBGE, na Coordenação de Contas Nacionais, estas séries não conversam e é um erro empregá-las em conjunto por que:

“A referência da nova SCN é o ano 2000, com maior detalhamento de atividades e produtos para as Tabelas de Recursos e Usos e de setores para as Contas Econômicas Integradas. Para antes de 2000 fez-se a retroposição até 1995, dentro da disponibilidade de dados para o período, estimando-se uma nova série encadeada com a nova série no ano 2000.”

Em face destas inflexões na medição do PIB municipal, bem como o PIB ser calculado e não apurado, este item econômico não foi empregado na tese.

(C) - INDICADORES ECONÔMICOS E FINANCEIROS - Os dados econômicos foram complementados com variáveis, de fontes públicas, para caracterizar o fluxo de capital e o nível da atividade econômica, a partir dos seguintes catálogos de dados: IPEADATA<sup>38</sup>, FGV DADOS<sup>39</sup>, SIDRA<sup>40</sup> e SGS<sup>41</sup>.

---

<sup>37</sup> Em 2007 foram publicados os resultados da nova série do SCN, tendo como referência o ano 2000. Nesta foram introduzidos novos conceitos e, uma nova BD, com níveis geográficos mais desagregados. A atualização da série da SCN, usualmente referida como mudança de base, é a atualização do peso das atividades econômicas para o cálculo do PIB e de seus componentes a preços constantes de um determinado ano. (IBGE, 2008).

<sup>38</sup> É uma BD com cerca de 7.000 séries históricas de fontes estatísticas brasileiras e internacionais, atualizadas e padronizadas no tempo e no espaço, com valores nominais e reais expressos em uma mesma moeda (R\$). Temas incluídos: população, emprego, salário e renda, produção, consumo e vendas, contas nacionais, finanças públicas, preços, juros, moeda e crédito, balanço de pagamentos e economia internacional. <http://www.ipeadata.gov.br>

<sup>39</sup> O BD de indicadores econômicos do Instituto Brasileiro de Economia da Fundação Getúlio Vargas, contempla as séries históricas estatística de índices de preços, preços recebidos e pagos pelos produtores agrícolas, sondagens industriais e, variado conjunto de indicadores, produzidos por outras instituições. Disponibiliza mais de 2.500 séries de cerca de 50 diferentes fontes <http://www.fgvdados.fgv.br>.

<sup>40</sup> O SIDRA é disponibilizado pelo IBGE para acesso aos dados agregados de suas pesquisas. Um dado agregado pode ser obtido, por exemplo, através do somatório dos valores de quesitos contidos em um questionário de pesquisa, que está associado às unidades de um nível territorial (unidade da federação, município, etc.), a um período de tempo e, muitas vezes, a um conjunto de classificações que o qualificam. <http://www.sidra.ibge.gov.br/>

Excetuando o SIDRA, que contém dados oriundos dos censos demográficos e das pesquisas do IBGE, e parte do FGVDADOS que não é gratuita, quase que na totalidade, o conteúdo disponibilizado nestes acessos são similares em fontes e naturezas, variando a forma de apresentação e a consulta. Em função disto optou-se pelo SGS por apresentar um desempenho melhor na recuperação das variáveis.

Com base nestas fontes foram coletadas as seguintes variáveis:

1. Frequência anual e granulometria país:
  - a) Venda de veículos automotores – ANFAVEA;
2. Frequência trimestral e granulometria país:
  - a) Índice de Confiança do Empresário Industrial - CNI;
3. Frequência mensal e granulometria país:
  - a) Venda de veículos automotores – ANFAVEA;
  - b) Indicadores de pessoal, produção e faturamento da indústria - CNI;
  - c) Indicadores de Inflação;
  - d) Taxas de Juros – BACEN;
  - e) Indicadores de produção de bens - BACEN;
  - f) Consumo de energia Elétrica – MME;
  - g) Vendas no varejo – IBGE;
  - h) Operações de crédito do sistema financeiro – BACEN;
  - i) Câmbio, balança comercial e ações – BACEN;
4. Frequência mensal e granulometria unidade da federação:
  - a) *Espírito Santo:*
    - i. Exportações e importações – BACEN;
    - ii. Receita de estados e municípios – Ministério da Fazenda;
    - iii. Arrecadação de ICMS - Ministério da Fazenda;
    - iv. Emprego formal – MTE;
    - v. Vendas no varejo – IBGE;
    - vi. Operações de Crédito do sistema financeiro – BACEN;
  - b) *Paraná:*
    - vii. Exportações e importações – BACEN;
    - viii. Receita de estados e municípios – Ministério da Fazenda;
    - ix. Arrecadação de ICMS - Ministério da Fazenda;
    - x. Emprego formal – MTE;
    - xi. Índice Nacional de Preços ao Consumidor – IBGE;
    - xii. Vendas, pessoal e horas empregadas na indústria – FIEP;
    - xiii. Vendas no varejo – IBGE; e,
    - xiv. Operações de Crédito do sistema financeiro – BACEN.

---

<sup>41</sup> O Sistema Gerenciador de Séries Temporais do BACEN contém dados desagregados em séries históricas, sobre economia regional, finanças públicas, indicadores de moeda e crédito e economia internacional. <http://www.bcb.gov.br/?SERIETEMP>

(D) DADOS DEMOGRÁFICOS - os dados demográficos são fornecidos pelo IBGE ou pelas secretarias estaduais de governo. Nesta tese, os dados de população e número de domicílios, foram solicitados diretamente ao IBGE, na Coordenação de População e Indicadores Sociais.

Foram fornecidas respectivamente a população dos censos entre 1940 e 2000 e a contagem da população de 1996 e 2007, para Curitiba e Vitória e o número de domicílios particulares permanentes dos censos de 1940 a 2000. A série de domicílios foi complementada com a pesquisa ao site do IBGE com relação aos domicílios da contagem de 1996 e 2007. Para equiparação na frequência anual com a mensal, estas variáveis foram equacionadas a partir de aplicação de regressões.

#### 4.1.3 – ANÁLISE GRÁFICA E DE CORRELAÇÃO CRUZADA DOS DADOS

Na coleta de dados foram levantadas ao total 2338 séries de variáveis brutas que após a compatibilização e consistência segundo as premissas descritas em 3.1, gerou a estratificação por tema, CDL e nível regional, apresentada na tabela 4.8

Tabela 4.8 – Característica das Variáveis Brutas.

| Característica Variáveis Brutas                  | BR         | Compagás    | Brasil    |
|--|------------|-------------|-----------|
| Dados da RM (Consumo combustíveis, Energia, MTE) | 478        | 1347        |           |
| Dados de Preço de combustíveis.                  | 59         | 76          |           |
| Dados de Consumo GN                              | 102        | 102         |           |
| Indicadores macro-econômicos                     | 24         | 30          | 78        |
| Demográficos                                     | 12         | 30          |           |
| <b>Total</b>                                     | <b>675</b> | <b>1585</b> | <b>78</b> |

Com base nos dados consistidos construíram-se 71 gráficos no âmbito da BR e 107 para a Compagás que, seguida pela redução de dimensionalidade em função da análise de *missing values* e séries espúrias, resultou em um conjunto de 1441 séries de variáveis caracterizadas conforme a tabela 4.9.

Tabela 4.9 – Característica das Variáveis Brutas Extratificadas.

| Característica das Variáveis para Pré-processamento | BR         |            | Compagás   |            | Brasil    |       |
|---|------------|------------|------------|------------|-----------|-------|
|   | Mensal     | Anual      | Mensal     | Anual      | Mensal    | Anual |
| Dados da RM (Consumo combustíveis, Energia, MTE)    | 50         | 211        | 130        | 593        |           |       |
| Dados de Preço de combustíveis.                     | 65         |            | 88         |            |           |       |
| Dados de Consumo GN                                 | 38         | 38         | 44         | 44         |           |       |
| Indicadores macro-econômicos                        | 24         |            | 30         |            | 62        |       |
| Demográficos  |            | 12         |            | 12         |           |       |
| <b>Total</b>  | <b>177</b> | <b>261</b> | <b>292</b> | <b>649</b> | <b>62</b> |       |

Conforme previsto em 3.1, as séries foram normalizadas antes da análise da correlação, que adotou relacionamento pareado entre as variáveis, ou seja, independente das diferenças de tamanhos entre as séries, o sistema considerou o intervalo útil de eventos para cálculo, analisando as sobreposições individuais entre estes pares.

Ao todo foram geradas, conforme apresentado na tabela 4.10, 321936 correlações, divididas pela frequência e nível regional, correspondendo ao cruzamento de todas as variáveis, dependentes e independentes entre si.

Tabela 4.10 – Quantidade de Correlações.

| Quantidade de Correlações Cruzadas | BR ES        | Compagas      | Brasil      |
|------------------------------------|--------------|---------------|-------------|
| Frequência Mensal                  | 26550        | 60230         | 1891        |
| Frequência Anual                   | 29890        | 203375        |             |
| <b>Total</b>                       | <b>56440</b> | <b>263605</b> | <b>1891</b> |

Este universo final de correlações foi estratificado pelos municípios que apresentam efetivo consumo de GN, conforme citado no item relativo ao limite temporal, e, organizados por frequência, UGR e segmento.

#### 4.1.4 – CONSIDERAÇÕES A CERCA DA BASE DE DADOS

Os dados espaciais coletados não apresentaram inconsistência de formato e identificação permitindo a correta execução do processo de espacialização dos clientes, bem como, na análise gráfica da série histórica da variável de interesse pode-se verificar a inexistência de descontinuidades ao longo do tempo.

Na Compagas, o segmento industrial foi o primeiro a ser atendimento, estratégia comum nas CDLs. Na figura 4.9 verifica-se que, a despeito da entrada em operação de outros segmentos, o industrial representa o grande consumidor de GN na CDL.

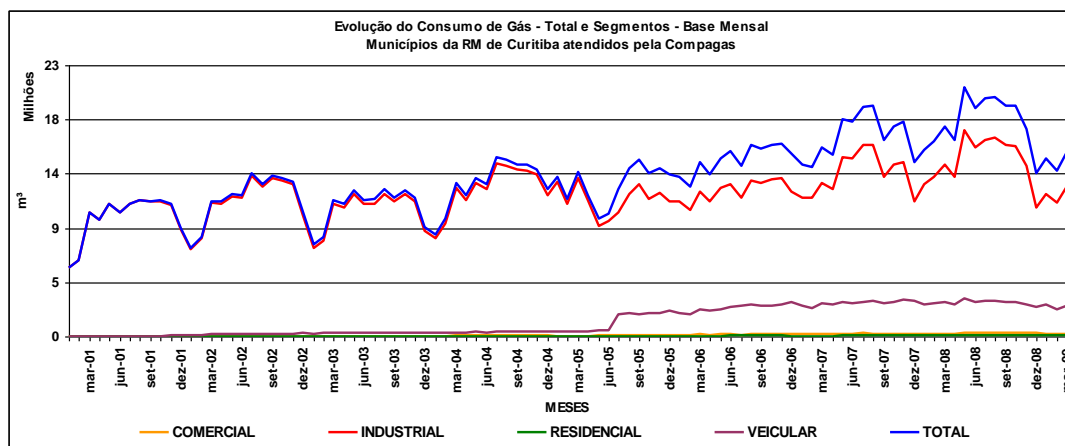


Figura 4.9 – Consumo Mensal de Gás MM<sup>3</sup> por Segmentos – Fonte: Compagas.



Até meados de 2005 o consumo é quase que totalmente industrial com ciclos de frequência aproximada de 12 meses.

A partir de 2005 o consumo industrial se mantém aparentemente estável até o final de 2007, quando apresenta uma evolução rápida até outubro de 2008 com retrocesso imediato nos meses subsequentes. Neste período, 2006 até 2008, o acréscimo de consumo se dá prioritariamente pelo segmento veicular.

Já no número de clientes, conforme figura 4.10, o histórico é inverso. Até o início de 2005 não existe crescimento nesta base. O crescimento se dá a partir de meados de 2005, num primeiro momento nos segmentos comercial e residencial, e logo depois, início de 2006 quase que exclusivamente neste último segmento.

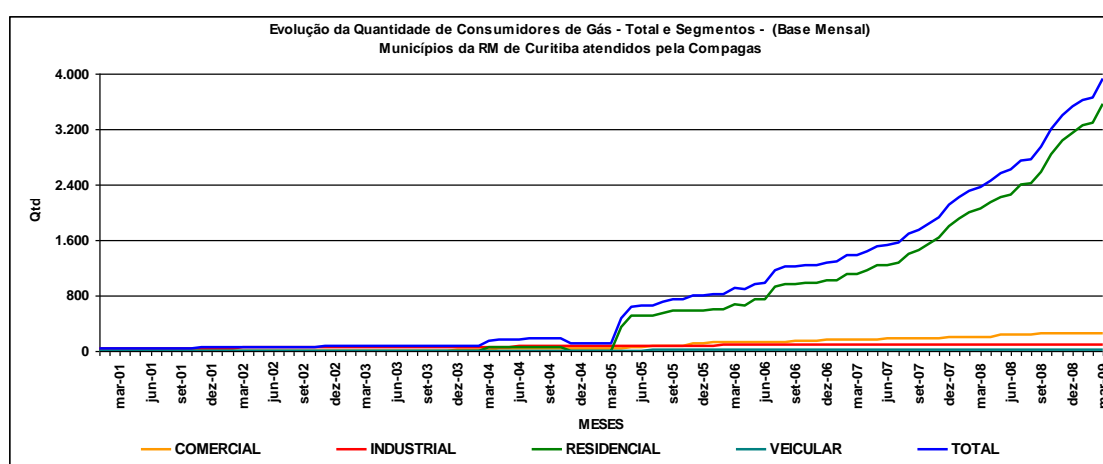


Figura 4.10 – Evolução Mensal do N<sup>o</sup> de Clientes de GN por Segmentos – Fonte: Compagas.

Similar à Compagas, na BR o atendimento também se iniciou pelo segmento industrial. Na figura 4.11, verifica-se que o crescimento no consumo se dá de forma mais uniforme ao longo do tempo, independente da entrada em operação de outros segmentos de negócio, e que a série original aparenta não possuir ciclos na frequência.

No primeiro ano o consumo é exclusivamente industrial e, a partir do ano 2000 desloca-se de patamar pela entrada do GNV.

A partir de 2003 até dezembro de 2007, tanto o consumo industrial, quanto o veicular, se mantém estável, quando apresenta uma evolução rápida até outubro de 2008 com retrocesso imediato nos meses subsequentes. Esta rápida evolução se dá exclusivamente no segmento industrial.

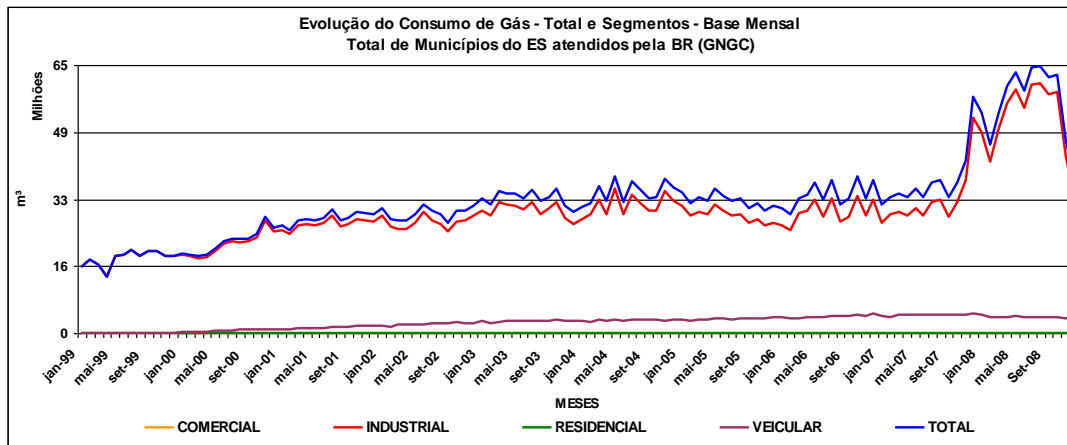


Figura 4.11 – Consumo Mensal de Gás MM<sup>3</sup> por Segmentos – Fonte: BR.

No caso da quantidade de clientes, conforme figura 4.12, verifica-se que até meados de 2003 não existe um crescimento considerável da mesma.

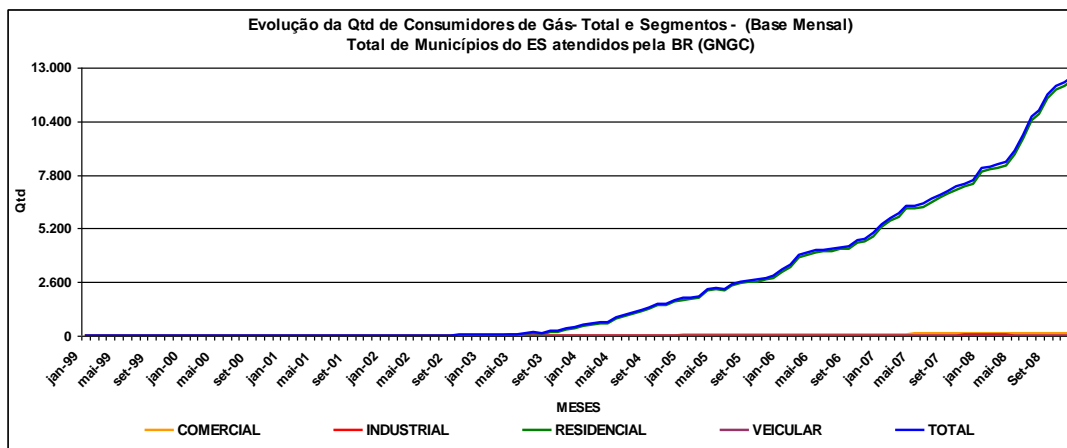


Figura 4.12 – Qtd. Mensal de Clientes GN por Segmentos – Fonte: BR Distribuidora.

Na figura 4.13, em escala logarítmica, nota-se que após 2003 existe um crescimento acelerado no segmento residencial, seguido de um pequeno crescimento no segmento comercial a partir de 2005. Os outros segmentos não apresentam crescimento.

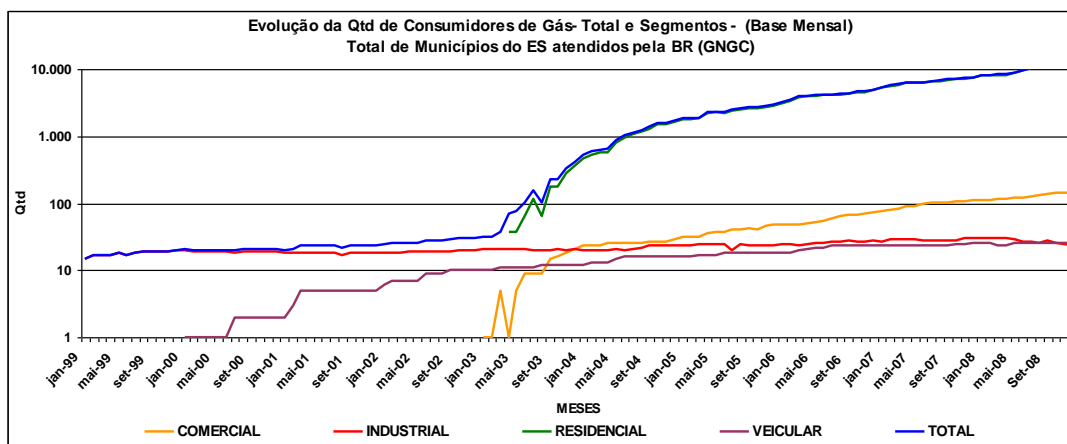


Figura 4.13 – Qtd. Mensal de Clientes GN por Segmentos (Log) – Fonte: BR Distribuidora.

Com relação às variáveis explicativas coletadas percebeu-se um comportamento bem distinto entre as de frequência anual e as mensais. Na primeira houve acentuado número de séries expurgadas pela presença de *missing values*, principalmente quanto às descontinuidades e valores zerados nos dados de consumo de combustíveis da ANP.

As correlações apresentaram baixa quantidade de pares de eventos úteis entre a variável de interesse e as explicativas, oscilando entre sete e nove períodos para AE de Curitiba e, entre oito e dez para a de Vitória. Na análise de cruzamento entre as próprias variáveis explicativas esta sobreposição chegou a trinta e oito períodos.

As correlações entre as variáveis explicativas e a de interesse na frequência anual, apresentaram altos coeficientes de correlação, direta e indireta, acima de 90%, associado a altos valores de significância apresentados nos valores da potência. Apesar desta capacidade de explicar a variável de interesse, as séries anuais não foram empregadas, por apresentarem correlações com pequeno número de eventos úteis entre as séries.

Por outro lado, a correlação entre as variáveis explicativas de frequência mensal apresentou intervalos com até 204 eventos para análise. Especialmente na correlação com a variável de interesse, apresentou sobreposições superiores a trinta e seis eventos.

Em face disto, somado aos fatos de se encontrarem na mesma frequência da variável de interesse e apresentarem séries de consumo de EE, de preço e de energéticos substitutos, optou-se pelo uso das variáveis explicativas na frequência mensal.

Ressalta-se que o número de correlações acima de 90%, reduz-se de forma significativa. Por exemplo, para o consumo de GN no segmento comercial de Curitiba, existiam 212 correlações acima de 90%, na frequência anual, contra 12 na mensal.

## **4.2 – DIRETRIZES PARA MODELAGEM**

Neste tópico descrevem-se as análises desenvolvidas com relação às premissas do desenvolvimento do modelo, tais como: a análise do crescimento espacial do consumo, os segmentos de negócio que apresentam dependência espacial, a análise do uso do solo e a redução da dimensionalidade do conjunto de variáveis explicativas.

### **4.2.1 – ESCOLHA DA UGR PARA A VARIÁVEL DEPENDENTE.**

Nas figuras 4.14 e 4.15 apresenta-se, com base nos dados de clientes de dezembro de 2008, respectivamente para a Compagas e para a BR, o resultado dos processos de espacialização de clientes, para as duas AE definidas no item 4.1.1.2.

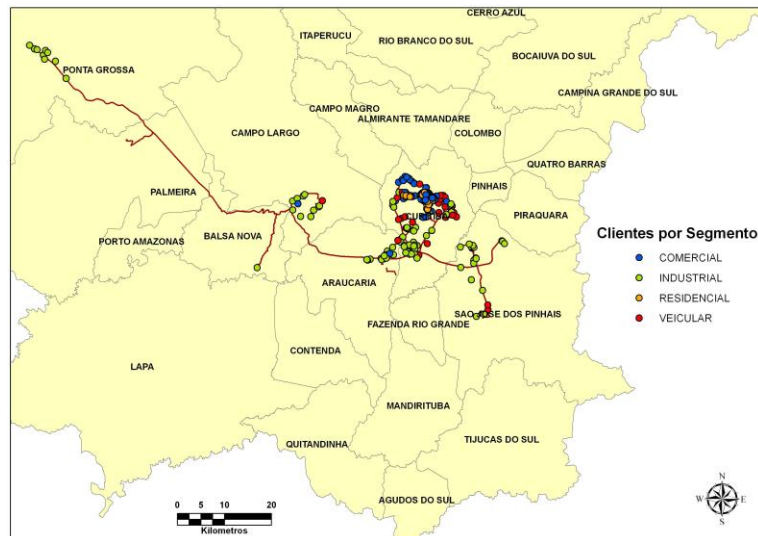


Figura 4.14 – Distribuição Espacial dos Clientes por Segmento – Fonte: Compagas.

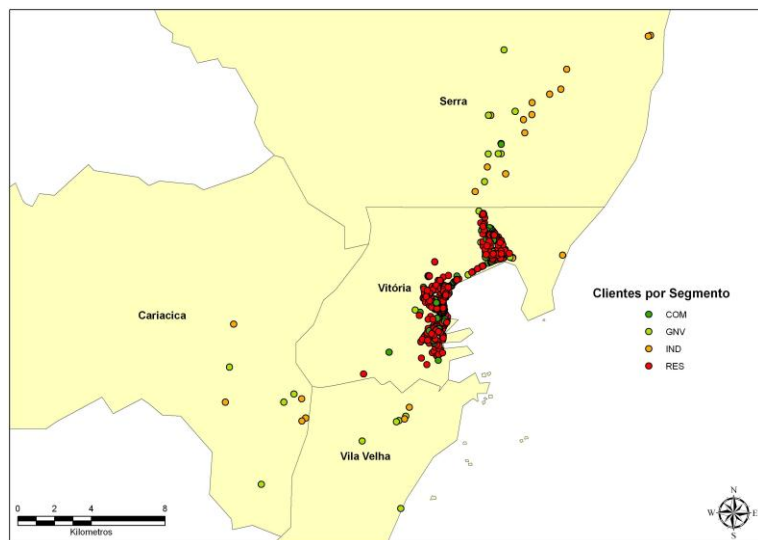


Figura 4.15 – Distribuição Espacial dos Clientes por Segmento – Fonte: BR Distribuidora.

Observa-se nas figuras a aglomeração de clientes nas capitais, principalmente para o segmento residencial e comercial, nos quais a distribuição é quase que exclusiva nesta região; em função disto, redefiniu-se a AE aos limites geográficos das capitais.

Cabe ressaltar que isto causa impacto na escolha das variáveis explicativas. Apesar de ser possível determinar a correlação temporal entre o consumo de GN dos clientes de um segmento na capital, com uma variável explicativa em outro município B qualquer, não é possível avaliar o impacto da mesma variável explicativa no consumo de GN deste próprio segmento na região B, pois não existe a ocorrência do fenômeno.

Com isto, mesmo existindo alta correlação temporal entre a variável de interesse na capital e as explicativas nas diferentes regiões vizinhas, no senso espacial não é possível avaliar a ocorrência da dependência espacial global e / ou local entre estas, pois não é possível avaliar a autocorrelação espacial de um evento dado uma única região.

Na prática a escolha de variáveis explicativas com menor autocorrelação espacial individual, conforme proposto em 2.6.2, não se aplica nesta tese, pois:

- Não é possível determinar se a variável explicativa que influencia o consumo na capital, também influencia o consumo nos municípios vizinhos; e,
- Como a UGR para o modelo deve apresentar um nível de discretização menor do que o delimitado na AE (neste caso o polígono que representa a capital), as variáveis explicativas assumem o mesmo valor para qualquer área  $n$  envolvida na modelagem, na estrutura do modelo SAR *Panel*.

Desta forma, estabeleceram-se como possíveis UGR para desenvolvimento do modelo, as duas representações espaciais poligonais (dados de áreas) imediatamente inferiores em escala e disponíveis na base de dados: bairros e setores censitários.

#### 4.2.1.1 – ANÁLISE DA DISPOSIÇÃO DOS CLIENTES NAS UGR DEFINIDAS

Após ajustes na espacialização dos clientes ao longo dos logradouros permitindo assim refinar a avaliação do grau de associação entre os polígonos e os possíveis aglomerados espaciais, foram construídos os gráficos de crescimento de números de bairros e setores com consumo, por segmento e, produzidos os mapas temáticos contendo a disposição dos clientes versus os bairros e os setores censitários.

No Apêndice D encontra-se disponível, por AE, o conjunto de gráficos por segmento e UGR, similar aos apresentados nas figuras 4.16 e 4.17, que demonstram respectivamente o crescimento da quantidade de setores com clientes no segmento veicular e comercial de Curitiba.

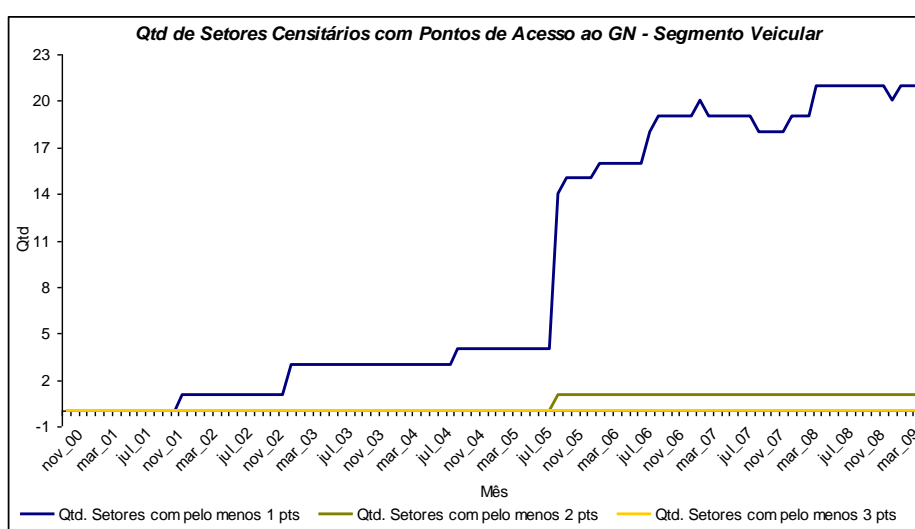


Figura 4.16 – Segmento Veicular x Setor Censitário – Curitiba.

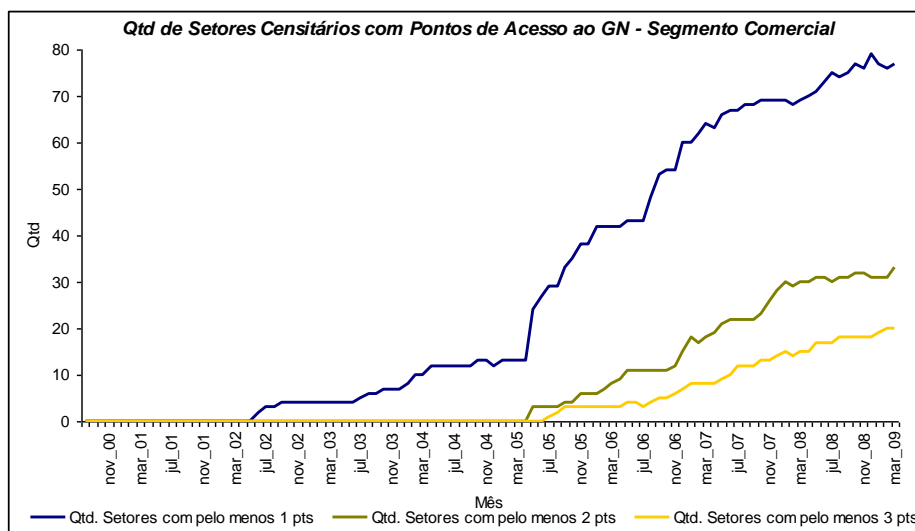


Figura 4.17 – Segmento Comercial x Setor Censitário – Curitiba.

Nota-se que no segmento veicular apesar do crescimento de polígonos com consumo, este não se dá em número de clientes. Já para o segmento comercial o crescimento é em ambos. Isto ocorre, pois o segmento comercial apresenta uma maior densidade de clientes na AE, bem como um crescimento da própria área ocupada, ou seja, o consumo está ampliando espacialmente.

Ao analisar os gráficos realizados percebe-se que quando separados por tipo de segmento, a quantidade de clientes é pequena; em alguns casos houve feições com um único cliente, além de que o segmento industrial não apresenta um comportamento de aumento da ocupação espacial ao longo do tempo.

Com base na observação dos gráficos e dos mapas são as seguintes diretrizes para emprego na análise da dependência espacial:

- (A) Os conjuntos de feições (bairros e setores) apresentam condição de relacionar no espaço todo o conjunto de clientes de cada AE, bem como permite estabelecer as matrizes de  $W_{ij}$ ;
- (B) O segmento industrial de ambas as AE e o segmento veicular de Vitória não apresentam as condições necessárias, contiguidade espacial e crescimento territorial, para existir dependência. Sendo assim devem ser objeto de modelagem de previsão puramente temporal;
- (C) O segmento veicular de Curitiba, apesar do crescimento territorial, não apresenta elevação do número de clientes por UGR; ou seja, atende a necessidades localizadas de GN, e não fruto de um processo de adensamento por influência do consumo nas regiões vizinhas. Sendo assim deve ser objeto de modelagem de previsão puramente temporal;

- (D) A dispersão dos clientes tem impacto nos relacionamentos de contiguidade entre setores, pois apesar da menor extensão territorial, quando comparados aos bairros, observa-se uma maior quantidade de interstícios;
- (E) Por outro lado a quantidade de clientes por bairros, bem como sua extensão territorial, tem impacto nas matrizes  $W_{ij}$ , pois “adoçam” os relacionamentos baseadas em distância (todos passam a ser vizinhos de todos); e,
- (F) A avaliação da dependência espacial deve contemplar as UGR bairro e setores, para os segmentos residencial e comercial de Vitória e Curitiba e o segmento veicular deste último.

Nesta etapa geraram-se também as séries históricas da variável dependente para as UE individuais dentro das UGR de bairro e setor, por segmento. Estas são as séries que serão empregadas nas análises de autocorrelação espacial e na montagem do modelo auto-regressivo espacial.

#### **4.2.1.2 – ANÁLISE DAS RESTRIÇÕES (LAND USE).**

Observando o conjunto de setores e bairros empregados na espacialização do consumo de GN, conclui-se que não existem, dentro das respectivas áreas de estudo, regiões que sofram impacto de questões ambientais, áreas de proteção ou exclusivas. Ou seja, os fatores apresentados não influenciam na análise de dependência espacial e sendo assim podem ser empregados na constituição das matrizes  $W_{ij}$  sem que seja necessário realizar qualquer tipo de edição cartográfica.

#### **4.2.2 – ÍNDICE DE AUTOCORRELAÇÃO ESPACIAL GLOBAL (MORAN GLOBAL).**

O resultado das análises realizadas no item 4.2.1 indicou a possibilidade de empregar o modelo SAR *Panel* para os segmentos residencial e comercial de Curitiba e Vitória, e o veicular deste último. Entretanto, tal suposição baseou-se exclusivamente na avaliação da dispersão dos clientes ao longo do espaço.

Para avaliar a condição necessária à aplicação do modelo espaço-temporal definido realizou-se primeiramente o cálculo do índice de autocorrelação espacial global, conforme o detalhamento apresentado em 3.2.2.

Sobre o desenvolvimento, ressalta-se que, para as matrizes baseadas em distância, observou-se que os valores máximos de distância entre centróides, que garantiam ao menos 1 vizinho por feição, diferiram entre o GEODA e o ArcGIS, conforme apresenta-se na tabela 4.11.

Tabela 4.11 – Distâncias Euclidianas empregadas nas matrizes de proximidade.

| <b>Matrizes <math>W_{ij}</math> por distância - Distâncias euclidianas</b> |               |              |               |
|--|---------------|--------------|---------------|
| <b>AE Curitiba</b>   |               |              |               |
| <b>Segmento</b>  | <b>Feição</b> | <b>GEODA</b> | <b>ARCGIS</b> |
| Residencial  | Setor         | 805          | 800           |
|  | Bairro        | 2300         | 2084          |
| Veicular   | Setor         | 4300         | 4318          |
|  | Bairro        | 7800         | 7305          |
| Comercial  | Setor         | 1500         | 1754          |
|  | Bairro        | 3300         | 4625          |
| <b>AE Vitória</b>  |               |              |               |
| <b>Segmento</b>  | <b>Feição</b> | <b>GEODA</b> | <b>ARCGIS</b> |
| Residencial  | Setor         | 786          | 1072          |
|  | Bairro        | 1300         | 1475          |
| Comercial  | Setor         | 1121         | 1131          |
|  | Bairro        | 1500         | 1475          |

Não foi possível determinar o porquê da diferença de valores; as suspeitas recaem sobre a forma como os dois aplicativos determinam o centróide da feição, e seu impacto na avaliação das descontinuidades espaciais (“vazios”). Independente disto, este valor é o que a aplicação entende como necessário para garantir a regra de pelo menos um vizinho; sendo assim, não cabem correções ou ajustes sobre o mesmo.

Com relação à autocorrelação espacial global observa-se que no segmento residencial de Curitiba, a análise dos gráficos de Moran indica que, mesmo em estágio de redução, existe dependência entre os setores. Para os bairros a dependência é pequena na análise por distância e alta e em decréscimo, para vizinhança de 1ª ordem; o que é corroborado pela indicação qualitativa extraída do ArcGIS.

No caso do segmento comercial de Curitiba, apesar de uma dispersão menor, a análise dos gráficos Moran indica propensão ao emprego de estrutura por bairros. Como detalhe ressalta-se que a indicação qualitativa do ArcGIS sugere uma dependência associada a um padrão clusterizado resultado de um processo aleatório.

Em Vitória, os clientes apresentam dispersão mais homogênea do que em Curitiba, desta forma os “vazios” ocupacionais são menores, em distância e número, não apresentando indícios de comprometimento na formação das matrizes, principalmente a por distância.

Nota-se em Vitória, tanto para o segmento residencial quanto para o comercial, que o comportamento do índice aponta para o emprego da feição setor. Porém, em nenhum dos casos identifica-se uma variação maior que 0,25, positivo ou negativo, nas



faixas de cada uma das matrizes testadas; o que em parte pode explicar a maioria das análises qualitativas apontarem que o padrão não é clusterizado nem disperso.

De qualquer forma, no caso de Vitória há uma maior concordância entre o conjunto de análises: trabalhar por setores. Contudo, somente com a análise global não se pode extrair conclusão; sendo assim sugerem-se, com base nas matrizes de 1ª ordem e distância construídas, avaliar a dependência espacial local, por setores e bairros, para os segmentos residencial e comercial de ambas as AE.

Os valores dos índices, calculados pelos dois softwares, estão disponíveis nos gráficos do Apêndice E.

#### 4.2.3 – PRÉ-SELEÇÃO DAS VARIÁVEIS EXPLICATIVAS

Antes da pré-seleção nos segmentos priorizados, existiam as seguintes quantidades de correlações entre as variáveis explicativas e o consumo de GN na frequência mensal: 353 correlações para os segmentos residencial e comercial em Curitiba e 238 correlações para os segmentos residencial e comercial em Vitória.

O limite de  $\pm 0,8$  na correlação reduziu este universo aos valores apresentados na tabela 4.12, com seguinte distribuição por UGR.

Tabela 4.12 – Distribuição de Variáveis Explicativas por UGR – Frequência Mensal.

| Segmento e AE        | Visão Macroeconômica |        | Visão Consumo e preço | Total de Variáveis |
|----------------------|----------------------|--------|-----------------------|--------------------|
|                      | UGR Brasil           | UGR UF | UGR Município         |                    |
| Residencial Curitiba | 17                   | 10     | 59                    | 86                 |
| Comercial Curitiba   | 20                   | 12     | 56                    | 88                 |
| Residencial Vitória  | 20                   | 16     | 83                    | 119                |
| Comercial Vitória    | 17                   | 15     | 88                    | 120                |

O conjunto final de variáveis candidatas, mesmo contendo uma quantidade inicial razoável, conforme a tabela 4.12, é funcionalmente menor, pois muitas destas são descartadas pelos seguintes motivos:

- Os clientes ou consumo não são relacionados com o segmento de interesse;
- A base de clientes ou consumo não é relacionada com a AE de interesse; e,
- O energético não é empregado no segmento.

Após a priorização, selecionou-se dos conjuntos, no limite de quinze variáveis, as principais candidatas respeitando o equilíbrio entre os três conjuntos, seguido pela retirada das variáveis redundantes. Nas tabelas 4.13 a 4.16 estão apresentados os conjuntos finais de variáveis pré-selecionadas para ensaio do modelo, respectivamente para os segmentos residencial e comercial em Curitiba e Vitória, conforme os grupos de variáveis econômicas para o Brasil e para a UF e de consumo e preço para o município.

Tabela 4.13 – Variáveis Explicativas Pré-selecionadas para o Modelo – Curitiba Residencial.

| Consumo de GN - Seg. Residencial Curitiba x Var. Explicativas |   |        |                |        |    |
|---|---|--------|----------------|--------|----|
| UGR   | Variável Explicativa                                      | r(X,Y) | r <sup>2</sup> | p      | N  |
| BRASIL  | Operações de Crédito do SFN - Setor privado serviços /PIB | 0,9505 | 0,9035         | 0,0000 | 76 |
|   | Operações de Crédito do SFN - Setor comercial/PIB         | 0,9493 | 0,9012         | 0,0000 | 76 |
|   | Total Mensal Pessoal Empregado Indústria                  | 0,9345 | 0,8732         | 0,0000 | 76 |
|   | Operações de Crédito do SFN - Pessoas físicas/PIB         | 0,9329 | 0,8703         | 0,0000 | 76 |
| UF  | Operações de Crédito do SFN - Pessoas físicas             | 0,9429 | 0,8891         | 0,0000 | 60 |
|   | Nível do emprego formal                                   | 0,9422 | 0,8877         | 0,0000 | 76 |
|   | Arrecadação ICMS - Receita estados e municípios           | 0,8973 | 0,8051         | 0,0000 | 76 |
|   | Preço Médio Mensal GN Seg. Residencial                    | 0,8575 | 0,7353         | 0,0000 | 76 |
|   | Arrecadação de ICMS - Setor secundário                    | 0,8532 | 0,7280         | 0,0000 | 76 |
|   | Vol. vendas no varejo - Automóveis, motocicletas          | 0,8491 | 0,7209         | 0,0000 | 76 |
| Município   | QTD Mensal de Clientes GN Seg. Comercial Curitiba         | 0,9439 | 0,8909         | 0,0000 | 76 |
|   | QTD Mensal de Clientes EE Seg. Comercial Curitiba         | 0,8911 | 0,7941         | 0,0000 | 24 |
|   | Consumo Mensal de EE Seg. Residencial Curitiba            | 0,7925 | 0,6281         | 0,0000 | 24 |

Tabela 4.14 – Variáveis Explicativas Pré-selecionadas para o Modelo – Curitiba Comercial.

| Consumo de GN - Seg. Comercial Curitiba x Var. Explicativas |  |        |                |        |    |
|---|--|--------|----------------|--------|----|
| UGR   | Var2   | r(X,Y) | r <sup>2</sup> | p      | N  |
| BRASIL  | Operações de Crédito do SFN - Pessoas físicas/PIB    | 0,9595 | 0,9207         | 0,0000 | 79 |
|   | Operações de Crédito do SFN - Setor privado/PIB      | 0,9443 | 0,8917         | 0,0000 | 79 |
|   | Total Mensal Pessoal Empregado Indústria             | 0,9410 | 0,8854         | 0,0000 | 79 |
|   | Operações de Crédito do SFN - Setor comercial/PIB    | 0,9347 | 0,8736         | 0,0000 | 79 |
| UF  | Nível do emprego formal                              | 0,9530 | 0,9083         | 0,0000 | 79 |
|   | Operações de Crédito do SFN - Pessoas físicas        | 0,9262 | 0,8578         | 0,0000 | 60 |
|   | Arrecadação ICMS - Receita estados e municípios      | 0,8922 | 0,7961         | 0,0000 | 79 |
|   | Arrecadação ICMS - Setor terciário                   | 0,8737 | 0,7634         | 0,0000 | 79 |
|   | Vol. vendas no varejo - Automóveis, motocicletas     | 0,8651 | 0,7484         | 0,0000 | 79 |
|   | Arrecadação ICMS - Setor secundário                  | 0,8364 | 0,6996         | 0,0000 | 79 |
| Município   | Consumo Mensal de GN Seg. Residencial Curitiba       | 0,9402 | 0,8840         | 0,0000 | 76 |
|   | Consumo Mensal de EE Total Curitiba                  | 0,8413 | 0,7078         | 0,0000 | 24 |
|   | Consumo Mensal de EE Seg. Residencial Curitiba       | 0,8153 | 0,6647         | 0,0000 | 24 |
|   | QTD Mensal de Clientes de EE Seg. Comercial Curitiba | 0,8090 | 0,6545         | 0,0000 | 24 |

Tabela 4.15 – Variáveis Explicativas Pré-selecionadas para o Modelo – Vitória Residencial.

| Consumo de GN - Seg. Residencial Vitória x Var. Explicativas |  |        |                |        |    |
|--|--|--------|----------------|--------|----|
| UGR  | Var2   | r(X,Y) | r <sup>2</sup> | p      | N  |
| BRASIL   | Operações de Crédito do SFN - Setor comercial/PIB        | 0,9690 | 0,9390         | 0,0000 | 69 |
|  | Operações de Crédito do SFN - Pessoas físicas/PIB        | 0,9626 | 0,9265         | 0,0000 | 69 |
|  | Total Mensal Pessoal Empregado Indústria                 | 0,9545 | 0,9111         | 0,0000 | 69 |
|  | Operações de Crédito do SFN - Setor privado serviços/PIB | 0,9364 | 0,8769         | 0,0000 | 69 |
| UF   | Nível do emprego formal                                  | 0,9700 | 0,9408         | 0,0000 | 69 |
|  | Operações de Crédito do SFN - Pessoas físicas            | 0,9505 | 0,9034         | 0,0000 | 60 |
|  | Arrecadação ICMS - Receita estados e municípios          | 0,9444 | 0,8920         | 0,0000 | 69 |
|  | Vol. vendas no varejo - Automóveis, motocicletas         | 0,9327 | 0,8699         | 0,0000 | 69 |
|  | Arrecadação ICMS - Setor secundário                      | 0,8991 | 0,8083         | 0,0000 | 69 |
|  | Arrecadação ICMS - Setor terciário                       | 0,8907 | 0,7933         | 0,0000 | 69 |
| Município  | QTD Mensal de Clientes de EE Seg. Residencial Vitória    | 0,9661 | 0,9334         | 0,0000 | 63 |
|  | QTD Mensal de Clientes de EE Seg. Comercial Vitória      | 0,9603 | 0,9222         | 0,0000 | 63 |
|  | QTD Mensal de Clientes de GN Seg. Comercial Vitória      | 0,9552 | 0,9124         | 0,0000 | 69 |
|  | Preço Mínimo Mensal do GLP em Vitória                    | 0,8812 | 0,7765         | 0,0000 | 69 |

Tabela 4.16 – Variáveis Explicativas Pré-selecionadas para o Modelo – Vitória Comercial.

| Consumo de GN - Seg. Comercial Vitória x Var. Explicativas |  |        |                |        |    |
|--|--|--------|----------------|--------|----|
| UGR  | Variável Explicativa   | r(X,Y) | r <sup>2</sup> | p      | N  |
| BRASIL   | Operações de Crédito do SFN - Pessoas físicas/PIB                | 0,9644 | 0,9301         | 0,0000 | 72 |
|  | Total Mensal Pessoal Empregado Indústria (Série dessazonalizada) | 0,9317 | 0,8681         | 0,0000 | 72 |
|  | Operações de Crédito do SFN - Setor comercial/PIB                | 0,9162 | 0,8394         | 0,0000 | 72 |
|  | Operações de Crédito do SFN - Setor privado/PIB                  | 0,9017 | 0,8130         | 0,0000 | 72 |
| UF   | Nível do emprego formal  | 0,9673 | 0,9357         | 0,0000 | 72 |
|  | Arrecadação ICMS - Receita estados e municípios                  | 0,9259 | 0,8572         | 0,0000 | 72 |
|  | Vol. vendas no varejo - Automóveis, motocicletas                 | 0,9011 | 0,8120         | 0,0000 | 72 |
|  | Arrecadação ICMS - Setor terciário                               | 0,8988 | 0,8079         | 0,0000 | 72 |
|  | Vol. vendas no varejo - Hiper/supermercado, Prod. Alimentícios   | 0,8832 | 0,7800         | 0,0000 | 72 |
| Município  | QTD Mensal de Clientes de EE Seg. Comercial Vitória              | 0,9642 | 0,9296         | 0,0000 | 66 |
|  | QTD Mensal de Clientes de EE Total Vitória                       | 0,9496 | 0,9018         | 0,0000 | 66 |
|  | Consumo Mensal de GN Seg. Residencial Vitória                    | 0,9185 | 0,8437         | 0,0000 | 69 |
|  | Preço Mínimo Mensal do GLP em Vitória                            | 0,8918 | 0,7953         | 0,0000 | 72 |

De forma geral, apesar do grande número de variáveis coletadas e analisadas por variável de interesse, observa-se que funcionalmente são poucas as variáveis explicativas que podem contribuir de forma significativa na explicação da primeira.

Variáveis de caráter de consumo, como por exemplo, preço do GLP ou o consumo de EE, das quais se esperava maior grau de correlação, apresentam coeficientes abaixo de 0,8, e por isto foram descartadas. Por outro lado, as variáveis econômicas, como as de arrecadação e crédito, apresentam a priori uma maior capacidade explicativa.

Tal característica é representativa na medida em que o GN vem sendo ao longo do tempo tratado como energo-intensivo e desta forma, geralmente encontra-se mais relacionado com orientações macroeconômicas do que com orientações de livre mercado, como por exemplo, a decisão de substituição.

Nos conjuntos de variáveis pré-selecionadas, nota-se que as de caráter macroeconômico apresentam, principalmente em Curitiba, melhores índices de correlação com o consumo de GN. O que pode ser um reflexo na forma da CDL atuar no mercado, onde mesmo para segmentos não energo-intensivos, aplica-se diretrizes comuns à indústria, ou seja, com visão de atacado e não de varejo.

#### 4.2.4 – ÍNDICE DE AUTOCORRELAÇÃO ESPACIAL LOCAL (GETIS E ORD\*)

Os resultados dos índices de autocorrelação local são apresentados em forma de mapas temáticos com a representação da significância. Como exemplo, na figura 4.18, encontra-se o mapa referente ao segmento residencial de Vitória, para o mês de dezembro de 2008, empregando a matriz de vizinhança de 1ª ordem.

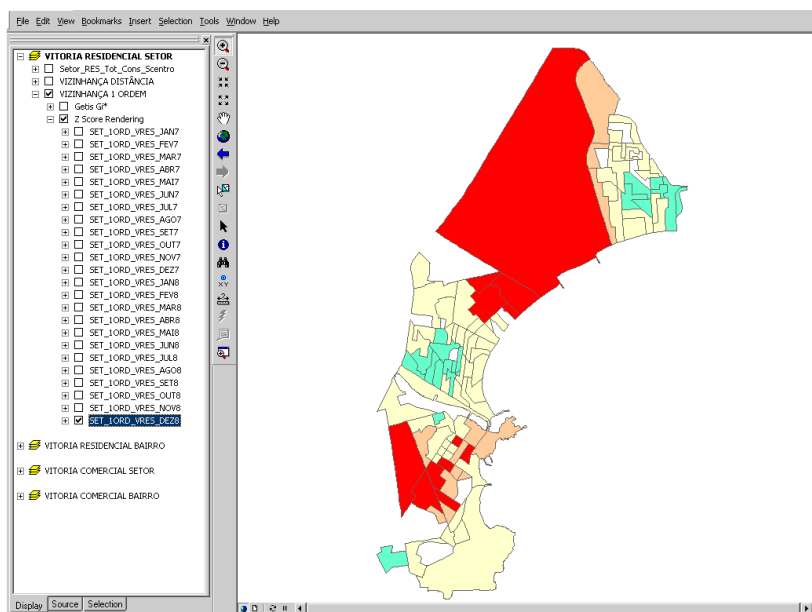


Figura 4.18 – Autocorrelação Local Dez/08 - Vizinhança de 1ª ordem – Vitória Residencial.

Os resultados estão estratificados em cinco faixas de significância:

- Alta e positiva – cor vermelha ( $> + 1,96$  Z Score);
- Alta e negativa – cor azul escuro ( $< -1,96$  Z Score);
- Baixa e positiva – cor laranja ( $> 1,1$  e  $< 1,96$  Z Score);
- Baixa e negativa – cor azul claro ( $< - 1,1$  e  $> -1,96$  Z Score); e,
- Sem significância ( $< 1,1$  a  $> - 1,1$  Z Score).

O exemplo da figura 4.18 indica que no mês de dezembro de 2008, o consumo de GN no segmento residencial, apresentou dependência espacial localizada em duas regiões com alta significância. Foram realizados ao todo 24 mapas temáticos, classificados conforme a significância, para as duas áreas de estudo, conforme os segmentos, as matrizes por distância e por vizinhança e a UGR setor e bairro.

Com base nos mapas temáticos são os seguintes os comentários por AE:

#### A. Curitiba – ArcGIS:

| SEGMENTO | FEIÇÃO | COMENTÁRIO – BASE ARCGIS   |
|----------|--------|--|
| RES      | Setor  | Existência de <i>clusters</i> e <i>outliers</i> significantes. O principal <i>cluster</i> é de valor positivo, mantêm-se ao longo de todo período de análise; entretanto, com flutuação na quantidade de UE. O segundo é de valor negativo, só ocorre com matriz por distância e desaparece nos seis últimos meses da série. |
|          | Bairro | Sem indicação de existência de <i>clusters</i> ou <i>outliers</i> , com qualquer matriz.   |
| COM      | Setor  | Com matriz de vizinhança não foi observada a existência de <i>clusters</i> , somente <i>outliers</i> provenientes da instabilidade estrutural. Já com matriz baseada em distância, ao longo do tempo identifica-se a existência de um <i>cluster</i> , com 3 objetos, de valor negativo em um único mês da série.            |
|          | Bairro | Sem indicação de existência de <i>clusters</i> ou <i>outliers</i> , com qualquer matriz.   |

## B. Vitória – ArcGIS:

| SEGMENTO | FEIÇÃO | COMÉNTARIO – BASE ARCGIS   |
|----------|--------|--|
| RES      | Setor  | Existência de <i>clusters</i> e de <i>outliers</i> significantes. Na matriz por distância o <i>cluster</i> principal é de valor negativo e com cerca de nove UE. Com a matriz de vizinhança existem dois <i>clusters</i> de valores positivos. Percebe-se que os <i>clusters</i> positivos se mantêm ao longo do tempo.  |
|          | Bairro | Sem indicação de existência de <i>clusters</i> ou <i>outliers</i> , com qualquer matriz.   |
| COM      | Setor  | Existência de <i>clusters</i> e de <i>outliers</i> significantes. Na matriz por distância o <i>cluster</i> principal é de valor negativo e com pequena quantidade de UE. O segundo é de valor positivo. Com a matriz de vizinhança existe um agrupamento de valor positivo. Os três apresentam ao longo do tempo pequena quantidade de UE e flutuação na ocorrência. |
|          | Bairro | Sem indicação de existência de <i>clusters</i> ou <i>outliers</i> , com qualquer matriz.   |

Os resultados da análise local, bem como a instabilidade apresentada ao longo do tempo, justificam as flutuações e os baixos índices apurados na análise global. Com isto, em ambas as AE é possível, observando as análises de dependência global e local, afirmar que não existe motivador para a adoção de um modelo espacial global.

No caso da UGR bairros, independente da AE e do segmento, a análise local indicou a não existência de *clusters* ou *outliers*, ou seja, não existe motivador para emprego desta UGR na formatação do modelo espacial. Já no caso dos setores observou-se a formação de *clusters* e *outliers*, mesmo que em pequenas quantidades.

Sendo assim adotou-se para desenvolvimento do modelo a UGR setores e a estratégia de desenvolvimento de modelos locais.

### 4.2.5 – DEFINIÇÃO DAS ÁREAS PARA EMPREGO DO MODELO ESPACIAL.

Os índices de autocorrelação foram tabulados em uma matriz  $m \times n$ , onde  $m$  identifica o tempo e  $n$  a UE, a partir da qual se avaliou a estacionariedade do índice ao longo do tempo.

No segmento residencial de Curitiba, com UGR por setores, o agrupamento positivo mantém-se forte ao longo de todo o período de análise caracterizando a existência de dependência espacial local estacionada. Já no segmento comercial não se identificou a existência de dependência espacial estacionada.

Nos segmentos de Vitória, mesmo com algumas flutuações entre os períodos, os *clusters* se mantiveram ao longo do tempo, caracterizando assim a existência de dependência espacial local estacionada, e como tal, justificam o emprego do modelo espacial para estes segmentos.

A partir disto foram determinadas, conforme as figuras 4.19, 4.20 e 4.21 as seguintes UE para ensaio do modelo.

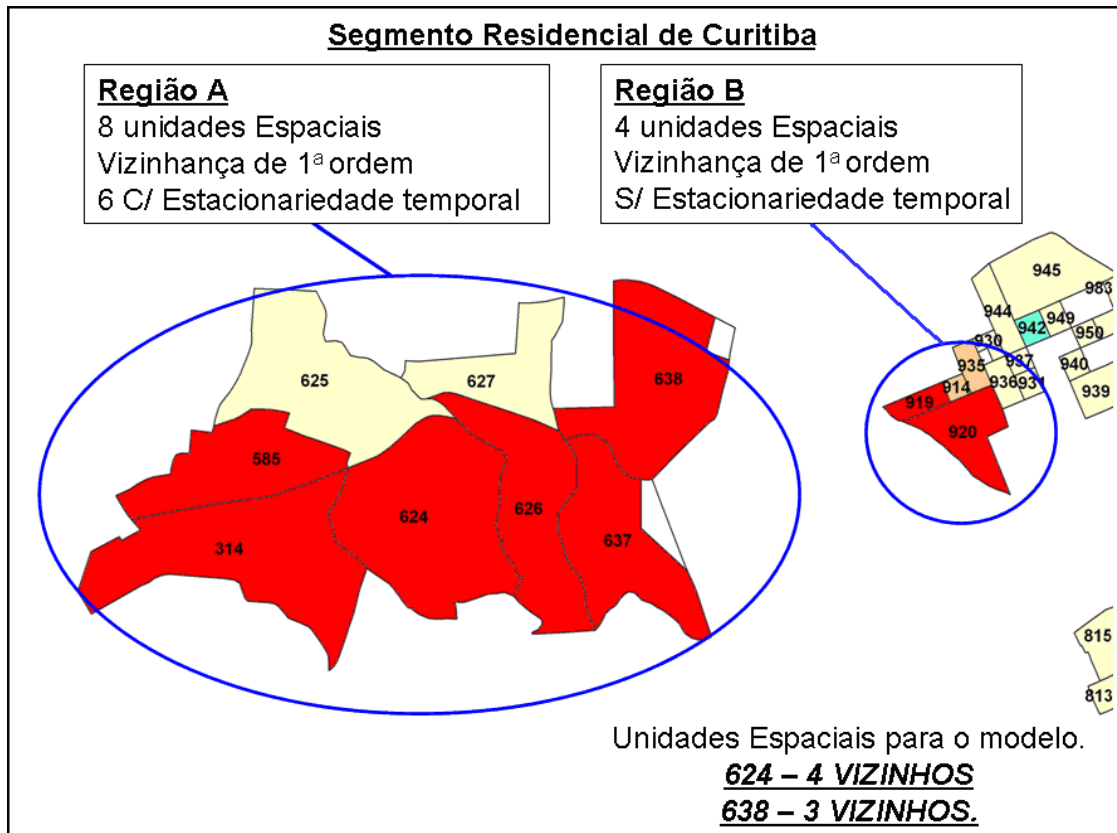


Figura 4.19 – Regiões de Significância e UE com Estacionariedade - Curitiba Residencial.

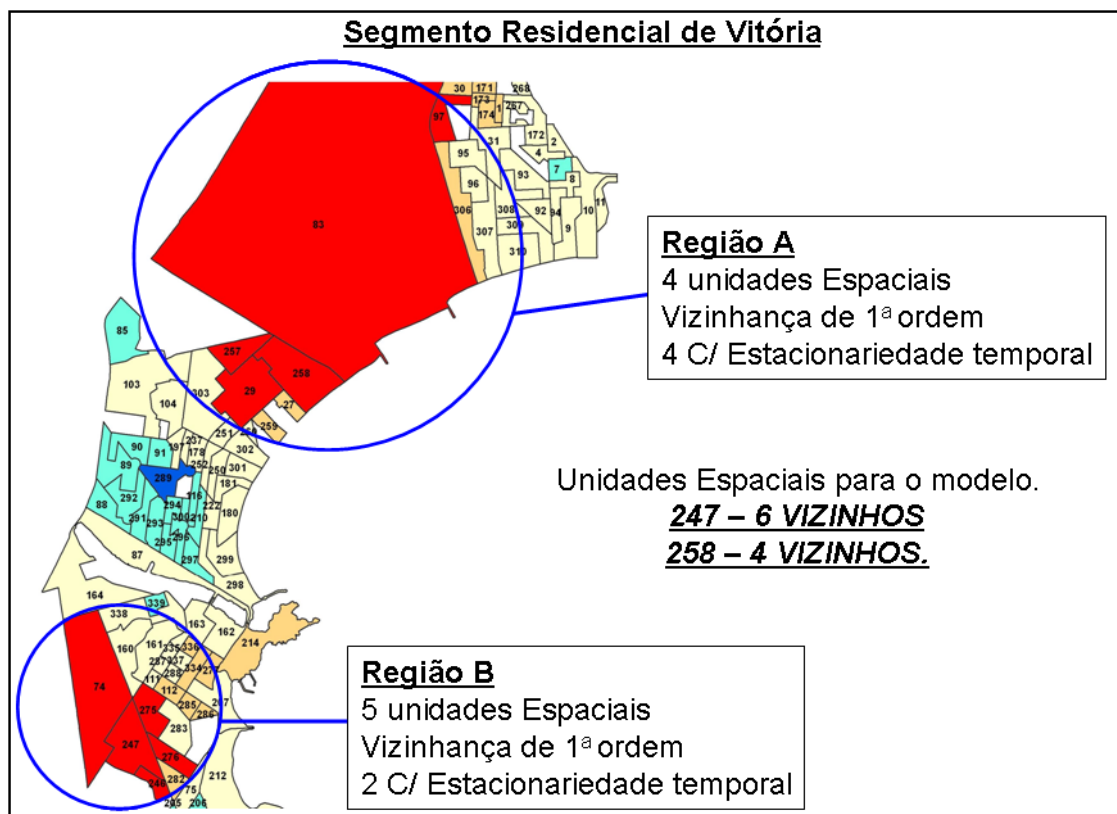


Figura 4.20 – Regiões de Significância e UE com Estacionariedade - Vitória Residencial.

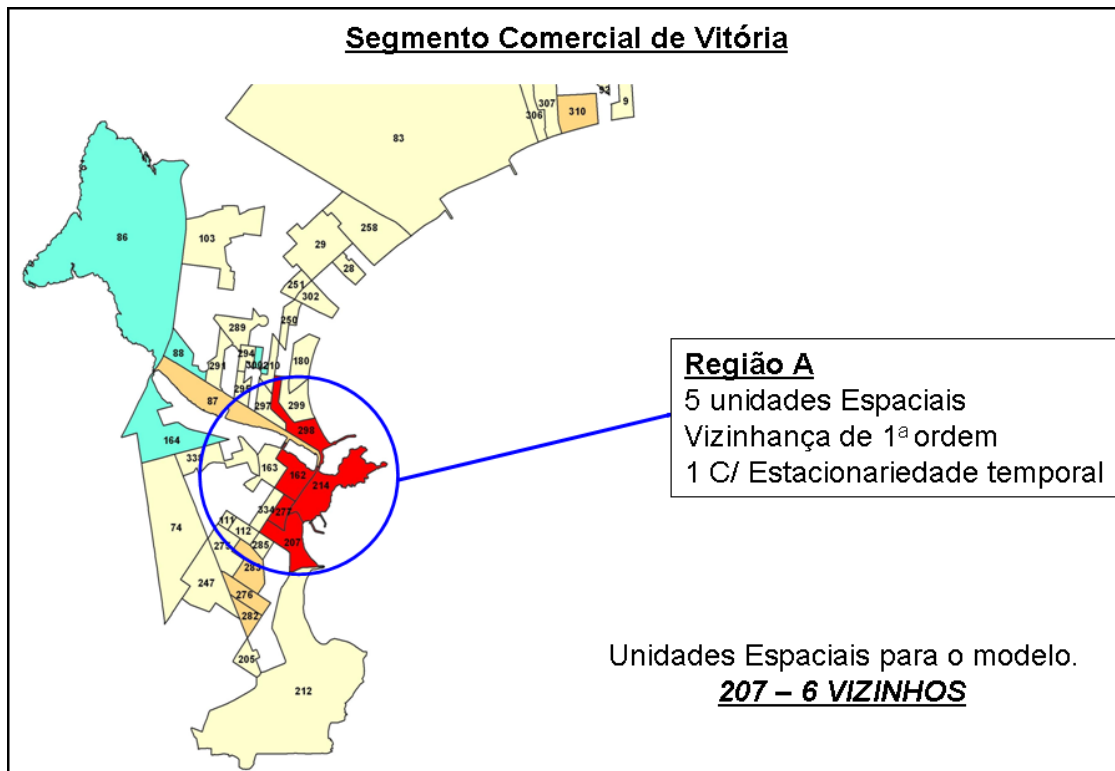


Figura 4.21 – Regiões de Significância e UE com Estacionariedade - Vitória Comercial.

No segmento comercial de Curitiba só foram identificados *outliers* frutos da instabilidade estrutural; por isto o segmento não tem a condição necessária para aplicar o modelo espaço-temporal.

Como existe mais de uma UE candidata à aplicação do modelo, estabeleceram alguns critérios de seleção, para mitigar os efeitos da instabilidade e da não estacionariedade dos índices, sendo eles:

- Selecionar UE com mais de um vizinho e que nesta vizinhança existam, ao mesmo tempo, UE com e sem dependência espacial estacionária; e,
- Desde que possível UE em regiões entre regimes de transição.

Com isto foram selecionadas respectivamente para os segmentos:

- Residencial de Curitiba – duas UE em uma região com regime espacial próprio;
- Residencial de Vitória – duas UE, uma em cada região separada por um regime de transição;
- Comercial de Vitória – uma UE em região com unidades vizinhas com e sem dependência espacial estacionária.

Na tabela 4.17 a seguir, apresenta-se o conteúdo final das UE de interesse para emprego do Modelo SAR *Panel*, com sua respectiva vizinhança e quantidade de UE ( $n$ ).



Tabela 4.17 – Relação Final de UE e Respectiva Vizinhaça.

| Segmento                | UE de Interesse | Vizinhaça                         | Tamanho de <i>n</i> |
|-------------------------|-----------------|-----------------------------------|---------------------|
| Residencial de Curitiba | 624             | 314 / 585 / 626                   | 4                   |
|                         | 638             | 626 / 627 / 637                   | 4                   |
| Residencial de Vitória  | 247             | 74 / 160 / 246 / 275 / 276 / 282  | 7                   |
|                         | 258             | 27 / 29 / 83 / 257                | 5                   |
| Comercial de Vitória    | 207             | 112 / 212 / 214 / 277 / 285 / 334 | 7                   |

#### 4.2.6 – SELEÇÃO DAS VARIÁVEIS EXPLICATIVAS

A seleção final das variáveis explicativas teve início com a realização da PCA em fase com a variável de interesse, empregando as variáveis explicativas pré-selecionadas descritas no item 4.2.3. Uma vez determinadas a matriz de correlação e as CP, procederam-se às seguintes reduções no conjunto de variáveis pré-selecionadas:

- Curitiba Residencial – das treze variáveis pré-selecionadas (tabela 4.13) retiraram-se: Vol. vendas no varejo - Automóveis, motocicletas, QTD Mensal de Clientes EE Seg. Comercial Curitiba, Consumo Mensal de EE Seg. Residencial Curitiba e Operações de Crédito do SFN - Pessoas físicas; reduzindo de vinte e quatro para quatro correlações  $< 0,85$  e de dez para seis correlações  $> 0,97$ . O conjunto final ficou com nove variáveis.
- Vitória Residencial – na primeira redução excluiu-se a variável: Operações de Crédito do SFN - Pessoas físicas. Em seguida retirou-se da pré-seleção (tabela 4.15): Operações de Crédito do SFN - Pessoas físicas/PIB e Nível do emprego formal por redundância. Com isto houve uma redução de dezesseis para cinco correlações acima de 0,97. O conjunto final ficou com onze variáveis.
- Vitória Comercial – inicialmente contemplou todo o conjunto pré-selecionado (tabela 4.16), apresentando vinte correlações acima de 0,97. Na redução foram retiradas: Operações de Crédito do SFN - Pessoas físicas/PIB e Operações de Crédito do SFN - Setor privado/PIB por redundância. Com isto houve uma redução de vinte para sete correlações acima de 0,97. O conjunto final ficou com onze variáveis.

Ao todo foram realizadas cinco PCA, para as quais os resultados finais estão disponíveis no Apêndice F, sendo uma para o segmento residencial de Curitiba, e duas para cada segmento de Vitória. Na tabela 4.18 apresentam-se para cada redução e PCA: o número de variáveis originais e priorizadas, a quantidade de CP, a correlação entre a PC e a variável de interesse, o *Test t* e a significância (*p*).



Tabela 4.18 – Tabela com os Resultados das PCAs para as Variáveis Explicativas.

| Segmento            | AE       | Fonte Dados Transformados | Nº de Variáveis |          | Nº CP | Correlação PC | Test t  | Signific. (p) |
|---------------------|----------|---------------------------|-----------------|----------|-------|---------------|---------|---------------|
|                     |          |                           | Pré-sel.        | Prioriz. |       |               |         |               |
| Residencial Opção 1 | Curitiba | STATISTIXL                | 13              | 9        | 3     | 0,9566        | 28,2435 | <0,0000       |
|                     |          | STATISTICAL               | 13              | 9        | 3     | 0,9566        | 28,2434 | <0,0000       |
| Residencial Opção 1 | Vitória  | STATISTIXL                | 14              | 13       | 5     | 0,9720        | 33,8789 | <0,0000       |
|                     |          | STATISTICAL               | 14              | 13       | 5     | 0,9720        | 33,8789 | <0,0000       |
| Residencial Opção 2 | Vitória  | STATISTIXL                | 14              | 11       | 4     | 0,9708        | 33,1596 | <0,0000       |
|                     |          | STATISTICAL               | 14              | 11       | 4     | 0,9708        | 33,1596 | <0,0000       |
| Comercial Opção 1   | Vitória  | STATISTIXL                | 13              | 13       | 5     | 0,9546        | 26,8386 | <0,0000       |
|                     |          | STATISTICAL               | 13              | 13       | 5     | 0,9546        | 26,8386 | <0,0000       |
| Comercial Opção 2   | Vitória  | STATISTIXL                | 13              | 11       | 4     | 0,9553        | 27,0551 | <0,0000       |
|                     |          | STATISTICAL               | 13              | 11       | 4     | 0,9553        | 27,0551 | <0,0000       |

Citam-se como pontos importantes os seguintes resultado na PCA em  $k = 0$ :

- Há similaridade dos resultados, independente da redução da dimensionalidade e da ferramenta empregada, o que confere robustez aos dados transformados;
- Todos os conjuntos de dados transformados apresentam correlação superior a 95% com a respectiva variável dependente; e,
- As CP significativas explicam mais de 95% da variância acumulada e, nas cinco análises, a PC explica mais de 92% da variância acumulada.

Na figura 4.22, a seguir, apresentam-se os gráficos da variância das CP com base nos resultados apresentados via STATISTIXL.

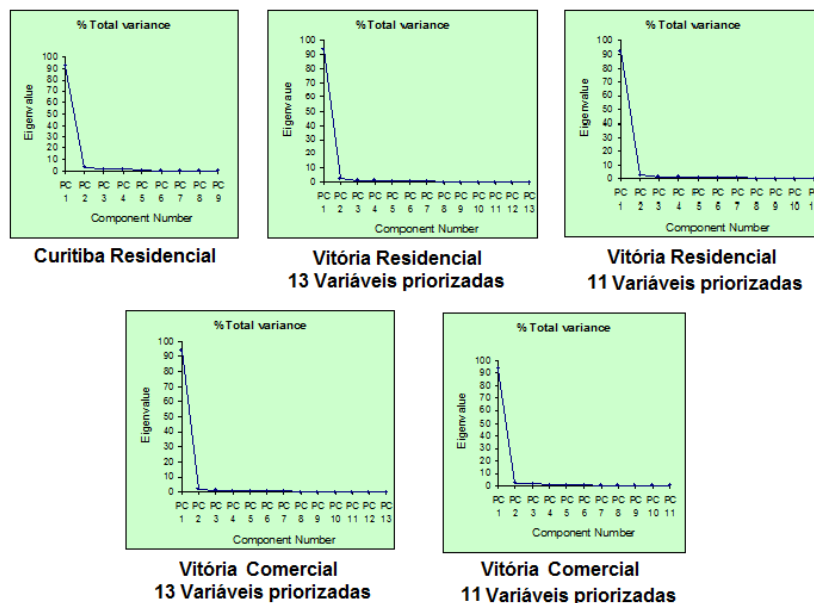


Figura 4.22 – Comportamento da Variância Individual das Principais Componentes.

Destaca-se que na maioria dos casos, a correlação entre os valores transformados e a variável de interesse é maior do que a correlação entre esta e as variáveis explicativas pré-selecionadas (ver tabelas 4.13 a 4.16); ou seja, os valores transformados apresentam melhor ajustamento com a variável dependente.

#### 4.2.6.1 - PCA DEFASADA COM A VARIÁVEL DEPENDENTE

Em complemento à PCA em  $k = 0$ , procedeu-se a determinação das PCA atrasadas até  $12 \ k = 12$ , empregando os subconjuntos de dados Residencial opção 1 Curitiba, Residencial e Comercial opção 2 Vitória. No caso de Vitória, a escolha baseou-se na redução da dimensionalidade mantendo os valores de correlação e do *Test t*.

Nas tabelas 4.19 a 4.21 são apresentados respectivamente, para o segmento residencial de Curitiba, o segmentos residencial e o comercial de Vitória os valores da correlação entre a variável de interesse e a série transformada pela PC, na fase e nos atrasos, do *Test t*, dos autovalores e do % da variância da PC.

Tabela 4.19 – Correlação PC c/ a Variável de Interesse ( $K = 0, \dots, 12$ ) – Curitiba Residencial.

| Atrasos | N  | Correlação | T-test  | Eigenvalue | % Var   |
|---------|----|------------|---------|------------|---------|
| Fase    | 76 | 0,95661    | 28,2435 | 8,29694    | 92,1882 |
| 1       | 75 | 0,95171    | 26,4869 | 8,29786    | 92,1984 |
| 2       | 74 | 0,94567    | 24,6808 | 8,30790    | 92,3099 |
| 3       | 73 | 0,93935    | 23,0797 | 8,27346    | 91,9273 |
| 4       | 72 | 0,93809    | 22,6582 | 8,25358    | 91,7064 |
| 5       | 71 | 0,94165    | 23,2389 | 8,22188    | 91,3541 |
| 6       | 70 | 0,94380    | 23,5472 | 8,17964    | 90,8848 |
| 7       | 69 | 0,94640    | 23,9839 | 8,13561    | 90,3956 |
| 8       | 68 | 0,94868    | 24,3719 | 8,08026    | 89,7806 |
| 9       | 67 | 0,94850    | 24,1400 | 8,03810    | 89,3122 |
| 10      | 66 | 0,95028    | 24,4128 | 8,03114    | 89,2349 |
| 11      | 65 | 0,94794    | 23,6264 | 8,00965    | 88,9960 |
| 12      | 64 | 0,94219    | 22,1405 | 8,03607    | 89,2896 |

Tabela 4.20 – Correlação PC c/ a Variável de Interesse ( $K = 0, \dots, 12$ ) – Vitória Residencial.

| Atrasos | N  | Correlação | T-test  | Eigenvalue | % Var   |
|---------|----|------------|---------|------------|---------|
| Fase    | 69 | 0,97086    | 33,1596 | 10,17331   | 92,4846 |
| 1       | 68 | 0,96529    | 30,0256 | 10,23397   | 93,0360 |
| 2       | 67 | 0,96101    | 28,0203 | 10,23985   | 93,0895 |
| 3       | 66 | 0,95817    | 26,7838 | 10,21810   | 92,8918 |
| 4       | 65 | 0,95639    | 25,9892 | 10,19124   | 92,6476 |
| 5       | 64 | 0,95498    | 25,3472 | 10,16563   | 92,4148 |
| 6       | 63 | 0,95857    | 26,2833 | 10,14778   | 92,2525 |
| 7       | 62 | 0,96182    | 27,2220 | 10,11933   | 91,9939 |
| 8       | 61 | 0,96472    | 28,1443 | 10,09824   | 91,8021 |
| 9       | 60 | 0,96376    | 27,5128 | 10,08420   | 91,6745 |
| 10      | 59 | 0,96249    | 26,7837 | 10,06497   | 91,4997 |
| 11      | 58 | 0,96241    | 26,5152 | 10,03213   | 91,2011 |
| 12      | 57 | 0,95647    | 24,3074 | 9,98355    | 90,7595 |

Tabela 4.21 – Correlação PC c/ a Variável de Interesse ( $K = 0, \dots, 12$ ) – Vitória Comercial.

| Atrasos | N  | Correlação | T-test  | Eigenvalue | % Var   |
|---------|----|------------|---------|------------|---------|
| Fase    | 72 | 0,95536    | 27,0551 | 10,27360   | 93,3963 |
| 1       | 71 | 0,95500    | 26,7438 | 10,28470   | 93,4972 |
| 2       | 70 | 0,94867    | 24,7353 | 10,29139   | 93,5580 |
| 3       | 69 | 0,94606    | 23,9014 | 10,27754   | 93,4321 |
| 4       | 68 | 0,94863    | 24,3574 | 10,27684   | 93,4258 |
| 5       | 67 | 0,94681    | 23,7214 | 10,26822   | 93,3474 |
| 6       | 66 | 0,94591    | 23,3249 | 10,27206   | 93,3823 |
| 7       | 65 | 0,94572    | 23,0973 | 10,26485   | 93,3167 |
| 8       | 64 | 0,94604    | 22,9874 | 10,24888   | 93,1716 |
| 9       | 63 | 0,94203    | 21,9284 | 10,25206   | 93,2005 |
| 10      | 62 | 0,93788    | 20,9381 | 10,23739   | 93,0671 |
| 11      | 61 | 0,93129    | 19,6372 | 10,22450   | 92,9499 |
| 12      | 60 | 0,92857    | 19,0535 | 10,20367   | 92,7606 |

Citam-se como pontos importantes os seguintes resultado na PCA defasada:

- O decaimento com as defasagens é pequeno. No segmento comercial de Vitória, este valor está na casa dos 3 %, nos outros não passa de 1,5 %;
- Independente do segmento, os conjuntos de dados transformados apresentam correlação superior a 92% com a respectiva variável de interesse; e,
- A PC explica mais de 89% da variância acumulada.

#### 4.2.6.2 – DEFASAGEM ENTRE VARIÁVEIS EXPLICATIVAS E DE INTERESSE

Nas tabelas 4.22 a 4.24 encontram-se relacionadas às variáveis explicativas e as respectivas correlações com a variável de interesse, de  $k = 0, \dots, 12$ , para o segmento residencial de Curitiba e Vitória e o comercial de Vitória. Em cada uma indicou-se as cinco variáveis explicativas que apresentam as maiores correlações com a variável de interesse em  $k = 0$ , e que serão empregadas para comparar com os resultados da PC.

Tabela 4.22 – Correlações Variáveis Explicativas x Interesse ( $K = 0, \dots, 12$ ) – Curitiba Res.

| Correlation Matrix (R)          | FASE  | ATRASO 1 | ATRASO 2 | ATRASO 3 | ATRASO 4 | ATRASO 5 | ATRASO 6 | ATRASO 7 | ATRASO 8 | ATRASO 9 | ATRASO 10 | ATRASO 11 | ATRASO 12 |
|---------------------------------|-------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| %OPER_CRED_PRIOTSSERV/PIB_BR(M) | 0,951 | 0,947    | 0,949    | 0,952    | 0,952    | 0,953    | 0,947    | 0,939    | 0,927    | 0,915    | 0,903     | 0,893     | 0,880     |
| %OPER_CRED_COM/PIB_BR(M)        | 0,949 | 0,943    | 0,944    | 0,942    | 0,944    | 0,946    | 0,947    | 0,947    | 0,945    | 0,939    | 0,933     | 0,923     | 0,912     |
| GN_CLIE_CUR_COM                 | 0,944 | 0,940    | 0,935    | 0,932    | 0,930    | 0,928    | 0,922    | 0,918    | 0,916    | 0,917    | 0,918     | 0,917     | 0,916     |
| HIV_EMP_FORM_PR                 | 0,942 | 0,934    | 0,920    | 0,903    | 0,889    | 0,881    | 0,879    | 0,883    | 0,889    | 0,900    | 0,920     | 0,940     | 0,953     |
| PES_EMP_IJD_BR(M)               | 0,934 | 0,938    | 0,926    | 0,911    | 0,897    | 0,891    | 0,891    | 0,898    | 0,909    | 0,924    | 0,941     | 0,959     | 0,967     |
| %OPER_CRED_PF/PIB_BR(M)         | 0,933 | 0,936    | 0,937    | 0,938    | 0,939    | 0,939    | 0,940    | 0,939    | 0,937    | 0,936    | 0,934     | 0,931     | 0,927     |
| REC_EST/MUN_ARREC_ICMS_PR       | 0,897 | 0,891    | 0,874    | 0,866    | 0,873    | 0,866    | 0,889    | 0,894    | 0,889    | 0,876    | 0,863     | 0,842     | 0,807     |
| GN_PRC_RES_PR                   | 0,857 | 0,857    | 0,881    | 0,868    | 0,865    | 0,854    | 0,850    | 0,841    | 0,841    | 0,821    | 0,830     | 0,833     | 0,883     |
| ARREC_ICMS_SSEC_PR              | 0,853 | 0,831    | 0,805    | 0,787    | 0,788    | 0,815    | 0,827    | 0,834    | 0,831    | 0,832    | 0,829     | 0,800     | 0,751     |

Tabela 4.23 – Correlações Variáveis Explicativas x Interesse ( $K = 0, \dots, 12$ ) – Vitória Res.

| Correlation Matrix (R)           | FASE  | ATRASSO 1 | ATRASSO 2 | ATRASSO 3 | ATRASSO 4 | ATRASSO 5 | ATRASSO 6 | ATRASSO 7 | ATRASSO 8 | ATRASSO 9 | ATRASSO 10 | ATRASSO 11 | ATRASSO 12 |
|----------------------------------|-------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|------------|------------|
| <b>%OPER_CRED_COM/PIB_BR(M)</b>  | 0,969 | 0,967     | 0,965     | 0,965     | 0,967     | 0,969     | 0,973     | 0,973     | 0,972     | 0,968     | 0,960      | 0,951      | 0,943      |
| <b>PES_EMP_IID_BR(M)</b>         | 0,955 | 0,961     | 0,955     | 0,941     | 0,924     | 0,910     | 0,903     | 0,901     | 0,909     | 0,922     | 0,939      | 0,956      | 0,967      |
| %OPER_CRED_PRIOTSSERV/PIB_BR(M)  | 0,936 | 0,936     | 0,936     | 0,939     | 0,944     | 0,943     | 0,942     | 0,939     | 0,929     | 0,917     | 0,906      | 0,893      | 0,882      |
| <b>REC_EST_MUJ_ARREC_ICMS_ES</b> | 0,944 | 0,933     | 0,923     | 0,920     | 0,920     | 0,922     | 0,937     | 0,946     | 0,957     | 0,956     | 0,952      | 0,949      | 0,935      |
| VEN_VJ_AUTOMV_ES                 | 0,933 | 0,934     | 0,945     | 0,945     | 0,938     | 0,934     | 0,931     | 0,939     | 0,949     | 0,953     | 0,959      | 0,963      | 0,957      |
| ARREC_ICMS_SSEC_ES               | 0,899 | 0,895     | 0,891     | 0,881     | 0,878     | 0,873     | 0,879     | 0,877     | 0,877     | 0,860     | 0,857      | 0,861      | 0,842      |
| ARREC_ICMS_STERC_ES              | 0,891 | 0,878     | 0,860     | 0,856     | 0,854     | 0,856     | 0,878     | 0,897     | 0,915     | 0,924     | 0,913      | 0,905      | 0,880      |
| <b>EE_CLIE_VIT_RES</b>           | 0,963 | 0,963     | 0,963     | 0,962     | 0,961     | 0,959     | 0,956     | 0,951     | 0,946     | 0,941     | 0,935      | 0,929      | 0,922      |
| EE_CLIE_VIT_COM                  | 0,940 | 0,940     | 0,940     | 0,939     | 0,938     | 0,936     | 0,936     | 0,933     | 0,929     | 0,925     | 0,922      | 0,920      | 0,914      |
| <b>GN_CLIE_VIT_COM</b>           | 0,955 | 0,951     | 0,948     | 0,945     | 0,941     | 0,938     | 0,936     | 0,935     | 0,936     | 0,937     | 0,940      | 0,944      | 0,947      |
| GLP_PRC_MIN_VIT                  | 0,881 | 0,879     | 0,869     | 0,860     | 0,854     | 0,852     | 0,852     | 0,850     | 0,845     | 0,842     | 0,840      | 0,833      | 0,827      |

Tabela 4.24 – Correlações Variáveis Explicativas x Interesse ( $K = 0, \dots, 12$ ) – Vitória Com.

| Correlation Matrix (R)           | FASE  | ATRASSO 1 | ATRASSO 2 | ATRASSO 3 | ATRASSO 4 | ATRASSO 5 | ATRASSO 6 | ATRASSO 7 | ATRASSO 8 | ATRASSO 9 | ATRASSO 10 | ATRASSO 11 | ATRASSO 12 |
|----------------------------------|-------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|------------|------------|
| <b>PES_EMP_IID_DESS_BR(M)</b>    | 0,932 | 0,932     | 0,934     | 0,938     | 0,943     | 0,945     | 0,945     | 0,944     | 0,940     | 0,934     | 0,930      | 0,926      | 0,921      |
| %OPER_CRED_COM/PIB_BR(M)         | 0,916 | 0,920     | 0,924     | 0,925     | 0,930     | 0,934     | 0,934     | 0,931     | 0,929     | 0,928     | 0,925      | 0,917      | 0,918      |
| <b>IIV_EMP_FORM_ES</b>           | 0,967 | 0,964     | 0,961     | 0,957     | 0,953     | 0,946     | 0,942     | 0,940     | 0,943     | 0,942     | 0,937      | 0,936      | 0,933      |
| <b>REC_EST_MUJ_ARREC_ICMS_ES</b> | 0,926 | 0,935     | 0,932     | 0,932     | 0,939     | 0,938     | 0,942     | 0,945     | 0,943     | 0,940     | 0,925      | 0,913      | 0,920      |
| VEN_VJ_AUTOMV_ES                 | 0,901 | 0,889     | 0,874     | 0,869     | 0,859     | 0,861     | 0,856     | 0,851     | 0,861     | 0,855     | 0,857      | 0,856      | 0,841      |
| ARREC_ICMS_STERC_ES              | 0,899 | 0,900     | 0,889     | 0,890     | 0,910     | 0,913     | 0,912     | 0,919     | 0,913     | 0,912     | 0,895      | 0,870      | 0,880      |
| VEN_VJ_SUP/ALIM_ES               | 0,883 | 0,905     | 0,889     | 0,871     | 0,894     | 0,887     | 0,884     | 0,887     | 0,891     | 0,873     | 0,876      | 0,858      | 0,845      |
| <b>EE_CLIE_VIT_COM</b>           | 0,966 | 0,961     | 0,957     | 0,954     | 0,950     | 0,945     | 0,945     | 0,946     | 0,948     | 0,952     | 0,953      | 0,952      | 0,952      |
| <b>EE_CLIE_VIT_TOT</b>           | 0,946 | 0,940     | 0,935     | 0,930     | 0,926     | 0,921     | 0,917     | 0,913     | 0,910     | 0,906     | 0,902      | 0,895      | 0,888      |
| GN_CON_VIT_RES                   | 0,924 | 0,922     | 0,914     | 0,913     | 0,908     | 0,906     | 0,916     | 0,914     | 0,912     | 0,915     | 0,915      | 0,915      | 0,916      |
| GLP_PRC_MIN_VIT                  | 0,892 | 0,888     | 0,882     | 0,878     | 0,872     | 0,863     | 0,859     | 0,855     | 0,851     | 0,841     | 0,834      | 0,834      | 0,817      |

#### 4.2.6.3 — SELEÇÃO DAS VARIÁVEIS EXPLICATIVAS

Para definir as variáveis explicativas selecionadas para emprego no modelo, os resultados de todas as correlações foram tabelados em 4.25 a 4.27, respectivamente para o para o segmento residencial de Curitiba e Vitória e o comercial de Vitória.

Estas tabelas apresentam na primeira linha, a correlações entre a variável dependente e a PC, conforme o atraso, e nas demais linhas também conforme o atraso, as diferenças entre os valores das correlações da variável de interesse: com a PC e com as variáveis explicativas de melhor correlação com a variável de interesse em  $k = 0$ .

Na prática os valores em vermelho apontam onde a variável explicativa, individualmente apresentou melhor desempenho do que a PC. Os demais valores apontam onde a PC, conforme a atraso, apresentou melhor correlação do que a variável explicativa isolada.

Tabela 4.25 – Correlações PC e Var. Explicativas x Var. de Interesse - Curitiba Residencial.

| Comparação de Correlações                                | FASE  | ATRASO 1 | ATRASO 2 | ATRASO 3 | ATRASO 4 | ATRASO 5 | ATRASO 6 | ATRASO 7 | ATRASO 8 | ATRASO 9 | ATRASO 10 | ATRASO 11 | ATRASO 12 |
|--|-------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| (A) = CORREL (PC1 X VAR DEP)                             | 0,957 | 0,952    | 0,946    | 0,939    | 0,938    | 0,942    | 0,944    | 0,946    | 0,949    | 0,948    | 0,950     | 0,948     | 0,942     |
| (A) - CORREL (%OPER_CRED_PRIOTSSERV/PIB_BR(M) x VAR DEP) | 0,006 | 0,005    | (0,003)  | (0,012)  | (0,014)  | (0,012)  | (0,003)  | 0,007    | 0,022    | 0,033    | 0,047     | 0,055     | 0,062     |
| (A) - CORREL (%OPER_CRED_COMP/PIB_BR(M) x VAR DEP)       | 0,007 | 0,008    | 0,001    | (0,003)  | (0,006)  | (0,004)  | (0,003)  | (0,001)  | 0,004    | 0,010    | 0,017     | 0,025     | 0,030     |
| (A) - CORREL (GN_CLIE_CUR_COM x VAR DEP)                 | 0,013 | 0,011    | 0,010    | 0,007    | 0,008    | 0,014    | 0,022    | 0,029    | 0,032    | 0,031    | 0,032     | 0,031     | 0,026     |
| (A) - CORREL (NIV_EMP_FORM_PR x VAR DEP)                 | 0,014 | 0,017    | 0,026    | 0,036    | 0,049    | 0,061    | 0,065    | 0,064    | 0,059    | 0,049    | 0,031     | 0,008     | (0,011)   |
| (A) - CORREL (PES_EMP_IND_BR(M) x VAR DEP)               | 0,022 | 0,014    | 0,020    | 0,028    | 0,041    | 0,051    | 0,053    | 0,048    | 0,040    | 0,024    | 0,009     | (0,011)   | (0,025)   |

Tabela 4.26 – Correlações PC e Var. Explicativas x Var. de Interesse - Vitória Residencial.

| Comparação de Correlações                          | FASE  | ATRASO 1 | ATRASO 2 | ATRASO 3 | ATRASO 4 | ATRASO 5 | ATRASO 6 | ATRASO 7 | ATRASO 8 | ATRASO 9 | ATRASO 10 | ATRASO 11 | ATRASO 12 |
|--|-------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| (A) = CORREL (PC1 X VAR DEP)                       | 0,971 | 0,965    | 0,961    | 0,958    | 0,956    | 0,955    | 0,959    | 0,962    | 0,965    | 0,964    | 0,962     | 0,962     | 0,956     |
| (A) - CORREL (%OPER_CRED_COM/PIB_BR(M) x VAR DEP)  | 0,002 | (0,002)  | (0,004)  | (0,006)  | (0,011)  | (0,014)  | (0,014)  | (0,011)  | (0,007)  | (0,004)  | 0,003     | 0,011     | 0,014     |
| (A) - CORREL (EE_CLIE_VIT_RES x VAR DEP)           | 0,008 | 0,003    | (0,002)  | (0,004)  | (0,005)  | (0,004)  | 0,003    | 0,011    | 0,019    | 0,023    | 0,027     | 0,033     | 0,034     |
| (A) - CORREL (GN_CLIE_VIT_COM x VAR DEP)           | 0,016 | 0,014    | 0,013    | 0,013    | 0,015    | 0,017    | 0,022    | 0,026    | 0,029    | 0,027    | 0,023     | 0,018     | 0,009     |
| (A) - CORREL (PES_EMP_IND_BR(M) x VAR DEP)         | 0,016 | 0,004    | 0,006    | 0,018    | 0,033    | 0,045    | 0,056    | 0,061    | 0,056    | 0,042    | 0,024     | 0,006     | (0,010)   |
| (A) - CORREL (REC_EST/MUN_ARREC_ICMS_ES x VAR DEP) | 0,026 | 0,033    | 0,038    | 0,039    | 0,036    | 0,033    | 0,022    | 0,016    | 0,007    | 0,007    | 0,010     | 0,013     | 0,022     |

Tabela 4.27 – Correlações PC e Var. Explicativas x Var. de Interesse - Vitória Comercial.

| Comparação de Correlações                          | FASE    | ATRASO 1 | ATRASO 2 | ATRASO 3 | ATRASO 4 | ATRASO 5 | ATRASO 6 | ATRASO 7 | ATRASO 8 | ATRASO 9 | ATRASO 10 | ATRASO 11 | ATRASO 12 |
|--|---------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|-----------|-----------|
| (A) = CORREL (PC1 X VAR DEP)                       | 0,955   | 0,955    | 0,949    | 0,946    | 0,949    | 0,947    | 0,946    | 0,946    | 0,946    | 0,942    | 0,938     | 0,931     | 0,929     |
| (A) - CORREL (NIV_EMP_FORM_ES x VAR DEP)           | (0,012) | (0,009)  | (0,012)  | (0,011)  | (0,004)  | 0,001    | 0,004    | 0,006    | 0,003    | 0,000    | 0,001     | (0,005)   | (0,004)   |
| (A) - CORREL (EE_CLIE_VIT_COM x VAR DEP)           | (0,011) | (0,006)  | (0,008)  | (0,008)  | (0,002)  | 0,002    | 0,001    | (0,001)  | (0,002)  | (0,010)  | (0,015)   | (0,021)   | (0,023)   |
| (A) - CORREL (EE_CLIE_VIT_TOT x VAR DEP)           | 0,009   | 0,015    | 0,014    | 0,016    | 0,023    | 0,026    | 0,029    | 0,032    | 0,036    | 0,036    | 0,036     | 0,036     | 0,040     |
| (A) - CORREL (PES_EMP_IND_DESS_BR(M) x VAR DEP)    | 0,024   | 0,023    | 0,015    | 0,008    | 0,006    | 0,001    | 0,001    | 0,002    | 0,006    | 0,008    | 0,008     | 0,006     | 0,008     |
| (A) - CORREL (REC_EST/MUN_ARREC_ICMS_ES x VAR DEP) | 0,030   | 0,020    | 0,017    | 0,014    | 0,010    | 0,008    | 0,003    | 0,001    | 0,003    | 0,002    | 0,013     | 0,018     | 0,009     |

Inicialmente, comparando os resultados em fase,  $k = 0$ , observa-se que, excetuando dois casos no segmento comercial de Vitória, a PC apresenta melhor desempenho quando comparada às variáveis explicativas individuais.

Comparando as PC, em fase e defasada, nota-se que o decaimento da correlação com as defasagens é pequeno. Conclui-se com isto, que a PC defasada apresenta maior robustez, pois propicia ao modelo a capacidade de prever um  $\Delta t$  à frente, sem a necessidade de realizar procedimentos adicionais mantendo a significância.

Na comparação entre a PC e as variáveis explicativas, com base nos desvios positivos, que demonstram o quanto a correlação da PC com a variável de interesse é maior do que a da variável explicativa, percebe-se que a PC é geralmente superior em valor e quantidade, ou seja, explicam em maior número e grau.

Observando exclusivamente o atraso em  $k = 12$ , dos quinze casos em dez a PC atrasada apresentou melhor correlação com a variável de interesse do que a variável explicativa, bem como os desvios relativos são melhores. Sendo assim, a correlação da PC atrasada degrada menos do que a das variáveis explicativas ao longo do tempo.

Desta forma, decidiu-se estimar o modelo SAR *Panel* empregando os dados transformados pela PC, conforme a opção 1 da tabela 4.18 para o segmento residencial de Curitiba e a opção 2 para o segmento residencial e comercial de Vitória, pois apresentam capacidade de conferir maior robustez ao modelo, mitigando o risco da dimensionalidade.

#### 4.2.7 – ANÁLISE DA AUTOCORRELAÇÃO DAS VARIÁVEIS DEPENDENTES

Com base nas premissas estabelecidas em 3.2.4, inicialmente definiu-se a vizinhança para as UE definidas para estimação do modelo SAR *Panel*, com base nas matrizes de contiguidade  $W_{ij}$  empregadas para cálculo da dependência espacial.

Em complemento, determinou-se o comprimento das séries de consumo de GN, considerando seu instante inicial de consumo, por segmentos e AE. Na tabela 4.28 relaciona-se as UE de interesse, sua vizinhança e o tamanho da série de consumo empregada na análise da autocorrelação da variável de interesse.

Tabela 4.28 – Relação Final de Polígonos de Interesses e Respectiva Vizinhança.

| Segmento             | UE de Interesse | Vizinhança                        | Tamanho de $n$ | Tamanho da série |
|----------------------|-----------------|-----------------------------------|----------------|------------------|
| Curitiba Residencial | 624             | 314 / 585 / 626                   | 4              | 76               |
|                      | 638             | 626 / 627 / 637                   | 4              | 76               |
| Vitória Residencial  | 247             | 74 / 160 / 246 / 275 / 276 / 282  | 7              | 69               |
|                      | 258             | 27 / 29 / 83 / 257                | 5              | 69               |
| Vitória Comercial    | 207             | 112 / 212 / 214 / 277 / 285 / 334 | 7              | 72               |

Todos os resultados indicaram a autocorrelação parcial com atraso de ordem 1 como a de maior significância. Isto vai ao encontro às premissas de dimensionalidade e

robustez aplicadas a seleção nas variáveis explicativas, pois o emprego de mais uma variável de mesma ordem no termo regressivo  $X\beta$ , confere maior capacidade de prever, sem comprometer a determinação do termo. Assim como, não comprometem o emprego das variáveis explicativas transformadas pela componente principal considerando o atraso  $k = 12$ , pois:

- A estimação de  $\beta$ , em  $X\beta$ , recebe uma informação de igual defasagem ( $k = 1$ ), de todas as UE ao mesmo tempo; e,
- Como a estimação recebe uma informação igual e de ordem  $k = 1$ , a previsão contemplará este atraso como sinal individual para cada uma das 12 previsões.

Desta forma, na estimação empregou-se para salientar o comportamento temporal do consumo de GN, as séries de dados das UE de interesse e vizinhas, com  $k = 1$ . Os resultados das autocorrelações parciais estão disponíveis no Apêndice G.

### 4.3 – ESTIMAÇÃO DO MODELO

Com base nas formas funcionais do modelo SAR *Panel*, no método de determinação de  $\ln|I - pW|$  e na estratégia de entrada de dados, definidas na metodologia, procedeu-se à estimação dos parâmetros do modelo e o cálculo das correlações entre as séries dos valores ajustados e dos valores de teste (valores reais de  $y$ ).

Tal atividade foi realizada para cada UE de interesse e os resultados são apresentados nas tabelas 4.29 a 4.33, respectivamente para as UE 624, 638, 207, 247 e 258.

Tabela 4.29 – Estimação UGR 624 - Segmento Residencial de Curitiba.

| Entrada de Dados em $X\beta$ | Forma Funcional | Determinação $\ln I - pW $ | PCA12 (b <sub>0</sub> ) | $y(k-1)$ (b <sub>0</sub> ) | PCA12 (bhat) | $y(k-1)$ (bhat) | $\rho$ (rho) | Rsqr   | Corr2  | Tstat   |         |        |
|------------------------------|-----------------|----------------------------|-------------------------|----------------------------|--------------|-----------------|--------------|--------|--------|---------|---------|--------|
| (A)                          | PSAR            | Indet full                 | 0,0245                  | 0,8590                     | -0,0145      | 0,8438          | 0,1609       | 0,8569 | 0,8305 | -0,9513 | 22,2792 | 3,9920 |
|                              |                 | Indet MC                   | 0,0245                  | 0,8590                     | -0,0123      | 0,8446          | 0,1520       | 0,8566 | 0,8317 | -0,8056 | 22,3021 | 3,7514 |
|                              | SARSFE          | Indet full                 | 0,0479                  | 0,8091                     | 0,0014       | 0,8122          | 0,1480       | 0,8592 | 0,8280 | 0,0739  | 19,8697 | 3,3602 |
|                              |                 | Indet MC                   | 0,0479                  | 0,8091                     | 0,0045       | 0,8120          | 0,1380       | 0,8590 | 0,8293 | 0,2388  | 19,8593 | 3,1150 |
| (B)                          | PSAR            | Indet full                 | 0,0200                  | 0,8690                     | -0,0263      | 0,8691          | 0,1770       | 0,8608 | 0,8304 | -1,6612 | 13,4509 | 4,4729 |
|                              |                 | Indet MC                   | 0,0200                  | 0,8690                     | -0,0250      | 0,8691          | 0,1720       | 0,8607 | 0,8311 | -1,5750 | 13,4673 | 4,3328 |
|                              | SARSFE          | Indet full                 | 0,0437                  | 0,8228                     | -0,0082      | 0,8425          | 0,1600       | 0,8619 | 0,8277 | -0,4192 | 12,5416 | 3,7256 |
|                              |                 | Indet MC                   | 0,0437                  | 0,8228                     | -0,0043      | 0,8410          | 0,1480       | 0,8616 | 0,8295 | -0,2197 | 12,5531 | 3,4178 |
| (C)                          | PSAR            | Indet full                 | 0,1943                  | 0,5303                     | 0,1427       | 0,5320          | 0,1940       | 0,5845 | 0,5690 | 6,1144  | 6,4125  | 3,3410 |
|                              |                 | Indet MC                   | 0,1943                  | 0,5303                     | 0,1461       | 0,5319          | 0,1810       | 0,5834 | 0,5682 | 6,2286  | 6,4005  | 3,0936 |
|                              | SARSFE          | Indet full                 | 0,2296                  | 0,4610                     | 0,1874       | 0,4760          | 0,1320       | 0,6514 | 0,6249 | 7,1005  | 6,1106  | 2,1576 |
|                              |                 | Indet MC                   | 0,2296                  | 0,4610                     | 0,1909       | 0,4747          | 0,1210       | 0,6509 | 0,6247 | 7,2038  | 6,0879  | 1,9674 |



Tabela 4.30 – Estimação UGR 638 - Segmento Residencial de Curitiba.

| Entrada de Dados em X <sub>8</sub> | Forma Funcional | Determinação ln  I- pW | PCA12 (b <sub>0</sub> ) | y (k-1) (b <sub>0</sub> ) | PCA12 (bhat) | y (k-1) (bhat) | p (rho) | Rsqr   | Corr2  | Tstat   |         |         |
|------------------------------------|-----------------|------------------------|-------------------------|---------------------------|--------------|----------------|---------|--------|--------|---------|---------|---------|
| (A)                                | PSAR            | Indet full             | 0,0160                  | 0,8575                    | -0,0207      | 0,7400         | 0,2930  | 0,8445 | 0,7874 | -1,6084 | 17,8755 | 6,8870  |
|                                    |                 | Indet MC               | 0,0160                  | 0,8575                    | -0,0196      | 0,7436         | 0,2840  | 0,8442 | 0,7887 | -1,5161 | 17,9569 | 6,6251  |
|                                    | SARSFE          | Indet full             | 0,0420                  | 0,8051                    | -0,0147      | 0,7222         | 0,2920  | 0,8467 | 0,7857 | -0,9135 | 16,6738 | 6,4379  |
|                                    |                 | Indet MC               | 0,0420                  | 0,8051                    | -0,0145      | 0,7224         | 0,2910  | 0,8467 | 0,7858 | -0,9009 | 16,6799 | 6,4098  |
| (B)                                | PSAR            | Indet full             | 0,0093                  | 0,9234                    | -0,0322      | 0,8431         | 0,3120  | 0,8531 | 0,7928 | -2,4552 | 15,6654 | 7,4771  |
|                                    |                 | Indet MC               | 0,0093                  | 0,9234                    | -0,0324      | 0,8426         | 0,3140  | 0,8531 | 0,7926 | -2,4771 | 15,6589 | 7,5383  |
|                                    | SARSFE          | Indet full             | 0,0354                  | 0,8738                    | -0,0287      | 0,8305         | 0,3230  | 0,8559 | 0,7896 | -1,7857 | 15,2051 | 7,2809  |
|                                    |                 | Indet MC               | 0,0354                  | 0,8738                    | -0,0285      | 0,8306         | 0,3220  | 0,8558 | 0,7898 | -1,7724 | 15,2059 | 7,2506  |
| (C)                                | PSAR            | Indet full             | 0,1551                  | 0,6386                    | 0,0519       | 0,5801         | 0,5030  | 0,6680 | 0,5379 | 2,8381  | 7,7734  | 11,0012 |
|                                    |                 | Indet MC               | 0,1551                  | 0,6386                    | 0,0523       | 0,5803         | 0,5010  | 0,6676 | 0,5376 | 2,8561  | 7,7718  | 10,9213 |
|                                    | SARSFE          | Indet full             | 0,2071                  | 0,5413                    | 0,0849       | 0,5443         | 0,4510  | 0,6991 | 0,5896 | 3,9511  | 7,5172  | 8,9609  |
|                                    |                 | Indet MC               | 0,2071                  | 0,5413                    | 0,0824       | 0,5443         | 0,4600  | 0,7005 | 0,5903 | 3,8663  | 7,5367  | 9,2630  |

Tabela 4.31 – Estimação UGR 207 - Segmento Comercial de Vitória.

| Entrada de Dados em X <sub>8</sub> | Forma Funcional | Determinação ln  I- pW | PCA12 (b <sub>0</sub> ) | y (k-1) (b <sub>0</sub> ) | PCA12 (bhat) | y (k-1) (bhat) | p (rho) | Rsqr   | Corr2  | Tstat   |         |        |
|------------------------------------|-----------------|------------------------|-------------------------|---------------------------|--------------|----------------|---------|--------|--------|---------|---------|--------|
| (A)                                | PSAR            | Indet full             | 0,0158                  | 0,8454                    | 0,0129       | 0,8443         | 0,0590  | 0,7746 | 0,7732 | 1,5809  | 27,6985 | 1,4587 |
|                                    |                 | Indet MC               | 0,0158                  | 0,8454                    | 0,0130       | 0,8444         | 0,0559  | 0,7745 | 0,7732 | 1,5997  | 27,6992 | 1,3825 |
|                                    | SARSFE          | Indet full             | 0,0256                  | 0,8192                    | 0,0215       | 0,8182         | 0,0550  | 0,7783 | 0,7541 | 2,2524  | 24,5590 | 1,2200 |
|                                    |                 | Indet MC               | 0,0256                  | 0,8192                    | 0,0222       | 0,8184         | 0,0460  | 0,7782 | 0,7542 | 2,3228  | 24,5620 | 1,0189 |
| (B)                                | PSAR            | Indet full             | 0,0096                  | 0,7407                    | 0,0082       | 0,7547         | 0,0450  | 0,7804 | 0,7796 | 0,8624  | 9,8362  | 1,0464 |
|                                    |                 | Indet MC               | 0,0096                  | 0,7407                    | 0,0083       | 0,7531         | 0,0399  | 0,7804 | 0,7796 | 0,8786  | 9,8132  | 0,9277 |
|                                    | SARSFE          | Indet full             | 0,0165                  | 0,6011                    | 0,0151       | 0,6094         | 0,0250  | 0,7878 | 0,7655 | 1,4331  | 6,3436  | 0,5162 |
|                                    |                 | Indet MC               | 0,0165                  | 0,6011                    | 0,0150       | 0,6097         | 0,0260  | 0,7878 | 0,7655 | 1,4275  | 6,3473  | 0,5373 |
| (C)                                | PSAR            | Indet full             | 0,1527                  | 1,1191                    | 0,1461       | 1,1566         | 0,1400  | 0,4192 | 0,4098 | 12,3785 | 9,6078  | 2,7580 |
|                                    |                 | Indet MC               | 0,1527                  | 1,1191                    | 0,1471       | 1,1507         | 0,1180  | 0,4180 | 0,4098 | 12,4685 | 9,5462  | 2,2975 |
|                                    | SARSFE          | Indet full             | 0,1619                  | 0,9881                    | 0,1537       | 1,0214         | 0,1140  | 0,4554 | 0,3916 | 11,5254 | 6,8151  | 2,0529 |
|                                    |                 | Indet MC               | 0,1691                  | 0,9881                    | 0,1548       | 1,0167         | 0,0980  | 0,4548 | 0,3917 | 11,6119 | 6,7793  | 1,7507 |

Tabela 4.32 – Estimação UGR 247 - Segmento Residencial de Vitória.

| Entrada de Dados em X <sub>8</sub> | Forma Funcional | Determinação ln  I- pW | PCA12 (b <sub>0</sub> ) | y (k-1) (b <sub>0</sub> ) | PCA12 (bhat) | y (k-1) (bhat) | p (rho) | Rsqr   | Corr2  | Tstat   |         |         |
|------------------------------------|-----------------|------------------------|-------------------------|---------------------------|--------------|----------------|---------|--------|--------|---------|---------|---------|
| (A)                                | PSAR            | Indet full             | 0,0072                  | 0,9313                    | 0,0572       | 0,9008         | -0,2361 | 0,9072 | 0,9172 | 5,1121  | 36,2770 | -5,4234 |
|                                    |                 | Indet MC               | 0,0072                  | 0,9313                    | -0,0034      | 0,9377         | 0,0500  | 0,9212 | 0,9198 | -0,3589 | 41,9714 | 1,3510  |
|                                    | SARSFE          | Indet full             | 0,0167                  | 0,9057                    | 0,0712       | 0,8805         | -0,2361 | 0,9164 | 0,9171 | 5,6653  | 35,0345 | -5,1017 |
|                                    |                 | Indet MC               | 0,0167                  | 0,9057                    | 0,0098       | 0,9089         | 0,0300  | 0,9250 | 0,9175 | 0,9008  | 38,9278 | 0,7316  |
| (B)                                | PSAR            | Indet full             | 0,0077                  | 0,9118                    | 0,0575       | 0,8819         | -0,2361 | 0,9084 | 0,9177 | 5,0082  | 22,0902 | -5,1820 |
|                                    |                 | Indet MC               | 0,0077                  | 0,9118                    | -0,0088      | 0,9217         | 0,0780  | 0,9233 | 0,9210 | -0,9161 | 25,5741 | 2,0170  |
|                                    | SARSFE          | Indet full             | 0,0179                  | 0,8838                    | 0,0719       | 0,8425         | -0,2631 | 0,9177 | 0,9186 | 5,6217  | 21,6546 | -4,9034 |
|                                    |                 | Indet MC               | 0,0179                  | 0,8838                    | 0,0069       | 0,8923         | 0,0480  | 0,9270 | 0,9192 | 0,6301  | 24,7332 | 1,1571  |
| (C)                                | PSAR            | Indet full             | 0,1571                  | 0,5129                    | 0,2022       | 0,4957         | -0,2361 | 0,5400 | 0,5318 | 11,5600 | 5,8646  | -3,1467 |
|                                    |                 | Indet MC               | 0,1571                  | 0,5129                    | 0,1850       | 0,5023         | -0,1460 | 0,5378 | 0,5305 | 10,7252 | 5,9339  | -1,9823 |
|                                    | SARSFE          | Indet full             | 0,1598                  | 0,5030                    | 0,2111       | 0,4690         | -0,2361 | 0,6147 | 0,5793 | 11,2622 | 5,9624  | -3,1168 |
|                                    |                 | Indet MC               | 0,1598                  | 0,5030                    | 0,1828       | 0,4877         | -0,1059 | 0,6135 | 0,5769 | 9,9891  | 6,2029  | -1,4151 |

Tabela 4.33 – Estimação UGR 258 - Segmento Residencial de Vitória.

| Entrada de Dados em X <sub>8</sub> | Forma Funcional | Determinação ln  I- pW | PCA12 (b <sub>0</sub> ) | y (k-1) (b <sub>0</sub> ) | PCA12 (bhat) | y (k-1) (bhat) | p (rho) | Rsqr   | Corr2  | Tstat   |         |         |
|------------------------------------|-----------------|------------------------|-------------------------|---------------------------|--------------|----------------|---------|--------|--------|---------|---------|---------|
| (A)                                | PSAR            | Indet full             | 0,0015                  | 0,8257                    | 0,0006       | 0,8262         | 0,0120  | 0,7365 | 0,7356 | 0,0726  | 23,1104 | 0,2437  |
|                                    |                 | Indet MC               | 0,0015                  | 0,8257                    | 0,0006       | 0,8262         | 0,0130  | 0,7365 | 0,7356 | 0,0651  | 23,1118 | 0,2637  |
|                                    | SARSFE          | Indet full             | 0,0149                  | 0,7432                    | 0,0067       | 0,7362         | 0,0990  | 0,7603 | 0,6783 | 0,6579  | 18,5988 | 1,8410  |
|                                    |                 | Indet MC               | 0,0149                  | 0,7432                    | 0,0068       | 0,7364         | 0,0970  | 0,7602 | 0,6785 | 0,6742  | 18,6032 | 1,8015  |
| (B)                                | PSAR            | Indet full             | -0,0022                 | 0,7575                    | -0,0054      | 0,7475         | 0,0450  | 0,7516 | 0,7481 | -0,6033 | 8,9231  | 0,8569  |
|                                    |                 | Indet MC               | -0,0022                 | 0,7575                    | -0,0054      | 0,7473         | 0,0460  | 0,7516 | 0,7480 | -0,6113 | 8,9204  | 0,8760  |
|                                    | SARSFE          | Indet full             | 0,0167                  | 0,6510                    | 0,0074       | 0,6334         | 0,1030  | 0,7822 | 0,7089 | 0,7245  | 7,2792  | 1,8252  |
|                                    |                 | Indet MC               | 0,0167                  | 0,6510                    | 0,0074       | 0,6334         | 0,1030  | 0,7821 | 0,7091 | 0,7594  | 7,2867  | 1,7508  |
| (C)                                | PSAR            | Indet full             | 0,0651                  | 0,6390                    | 0,0696       | 0,6568         | -0,0740 | 0,1832 | 0,1707 | 4,6611  | 4,3850  | -0,9583 |
|                                    |                 | Indet MC               | 0,0651                  | 0,6390                    | 0,0694       | 0,6561         | -0,0710 | 0,1830 | 0,1710 | 4,6491  | 4,3799  | -0,9203 |
|                                    | SARSFE          | Indet full             | 0,0819                  | 0,4623                    | 0,0712       | 0,4391         | 0,1250  | 0,4124 | 0,2307 | 4,6772  | 3,1537  | 1,7264  |
|                                    |                 | Indet MC               | 0,0819                  | 0,4623                    | 0,0719       | 0,4406         | 0,1170  | 0,4120 | 0,2302 | 4,7159  | 3,1631  | 1,6085  |



Os estimadores com índice (*bhat*) são oriundos do ajuste de  $\hat{\beta} = (\hat{\beta}_0 - p\hat{\beta}_L)$  na etapa cinco do mecanismo de estimação. Sobre este processo, observa-se que houve a correta determinação dos parâmetros em todos os experimentos, nos quais a maximização de  $p$  apresentou comportamento similar por tipo de experimento.

Observados os resultados, nota-se que casados par a par, em função da opção de entrada de dados e da forma funcional, os valores de  $p$  não diferiram muito em função do método de determinação de  $\ln|I - pW|$ . Exceção a esta colocação são os experimentos da UE 247, com forma funcional PSAR e determinação *Indet full*, que convergiram a um mesmo valor de -0,2361, independente da opção de entrada de dados.

No tocante aos resíduos, quando verificadas a determinação e a correlação entre os valores reais e os ajustados, observa-se que os resultados com a opção (C) não apresentaram bons resultados para todas as UE. Isto ocorreu pois o produto entre o parâmetro  $y_{(k-1)}$  (*bhat*) e o valor da segunda coluna nas UE, no qual o termo é nulo, é igual a zero; ou seja, o vetor de resíduos apresenta viés entre o valor real e o ajustado, corrompendo o cálculo da determinação e da correlação.

Nos demais casos têm-se o seguinte cenário:

- UE 624 – apresentam determinação alta (todas acima de 82%);
- UE 638 – apresentam boa determinação (entre de 78 e 85 %);
- UE 207 – apresentam boa determinação (entre de 75 e 79 %);
- UE 247 – apresentam determinação alta (todas acima de 91%); e,
- UE 258 - apresentam boa determinação (entre de 75 e 79 %).

Com base nestes resultados conclui-se que os modelos estimaram os parâmetros com boa qualidade, cabendo validar o emprego do modelo SAR *Panel* ao problema estudado, com base na comparação entre os valores previstos e os reais.

#### **4.3.1 – RESULTADOS DA PREVISÃO**

Observando as estruturas em (3.21), (3.22) e (3.23) procedeu-se à previsão dos valores de consumo de GN por UE. No tocante a  $pWy$ , cabe lembrar que empregaram-se os valores conhecidos de  $y$  para as diversas UE vizinhas<sup>42</sup>. Nas figuras 4.23 a 4.27 (a), (b) e (c) apresentam-se para as UE de interesse os gráficos das previsões, conforme as formas funcionais, os métodos de determinação e as opções de entrada de dados.

---

<sup>42</sup> Ressalta-se que o valor de  $y$  das diversas UE vizinhas no termo  $pWy$  poderia ser previsto por qualquer método e isto não alteraria o processo de previsão, pois seu emprego se dá mediante a utilização dos valores previstos nas posições correlatas das UE vizinhas em (3.26).

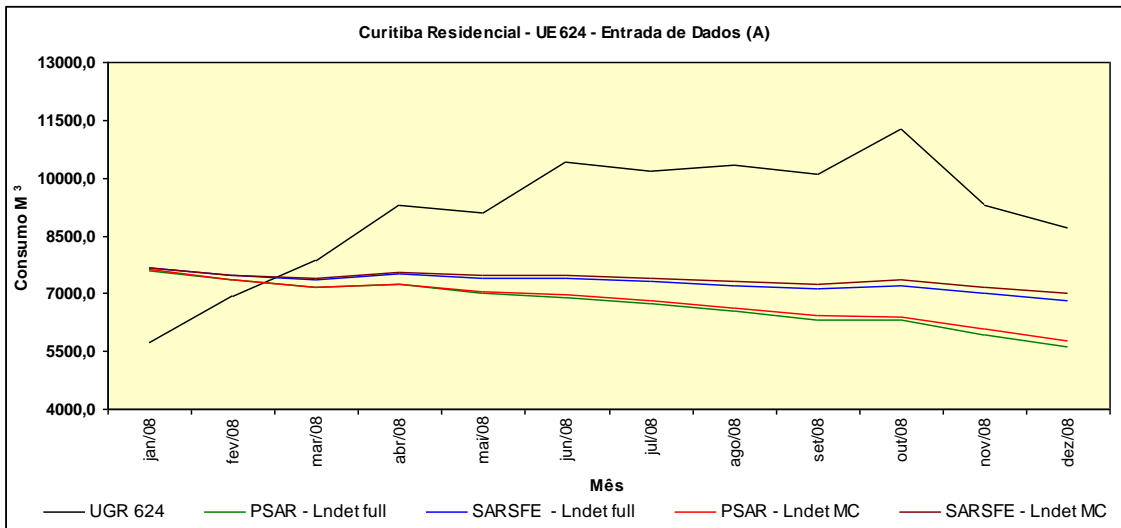


Figura 4.23 (a) – Previsão UE 624 - Segmento Residencial de Curitiba – Opção (A).

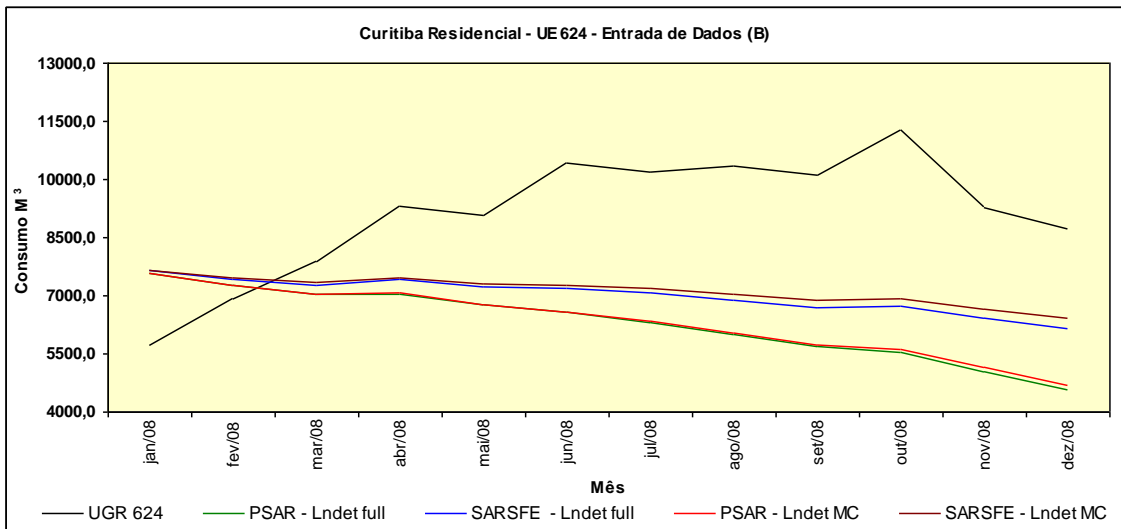


Figura 4.23 (b) – Previsão UE 624 - Segmento Residencial de Curitiba – Opção (B).

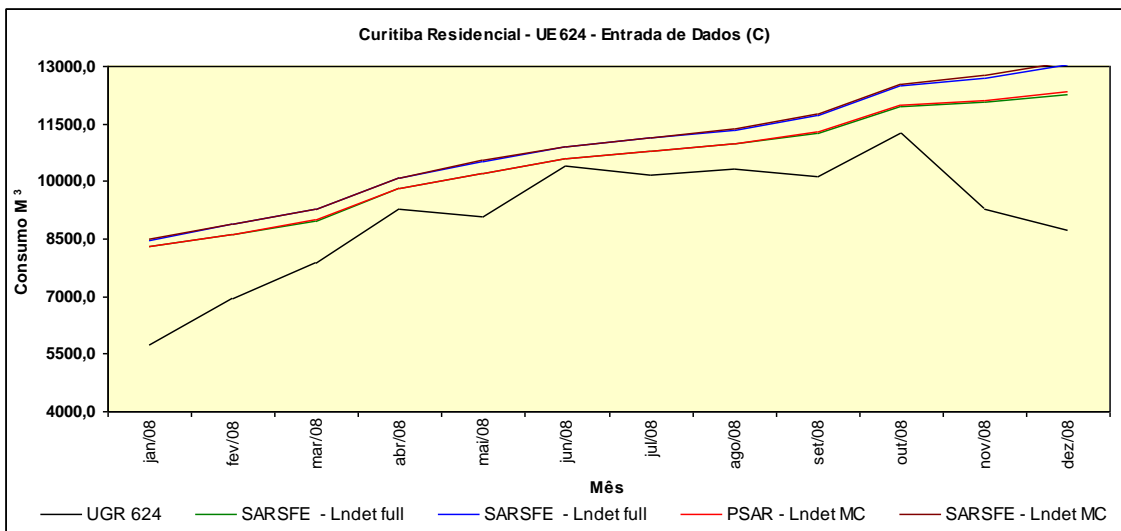


Figura 4.23 (c) – Previsão UE 624 - Segmento Residencial de Curitiba – Opção (C).

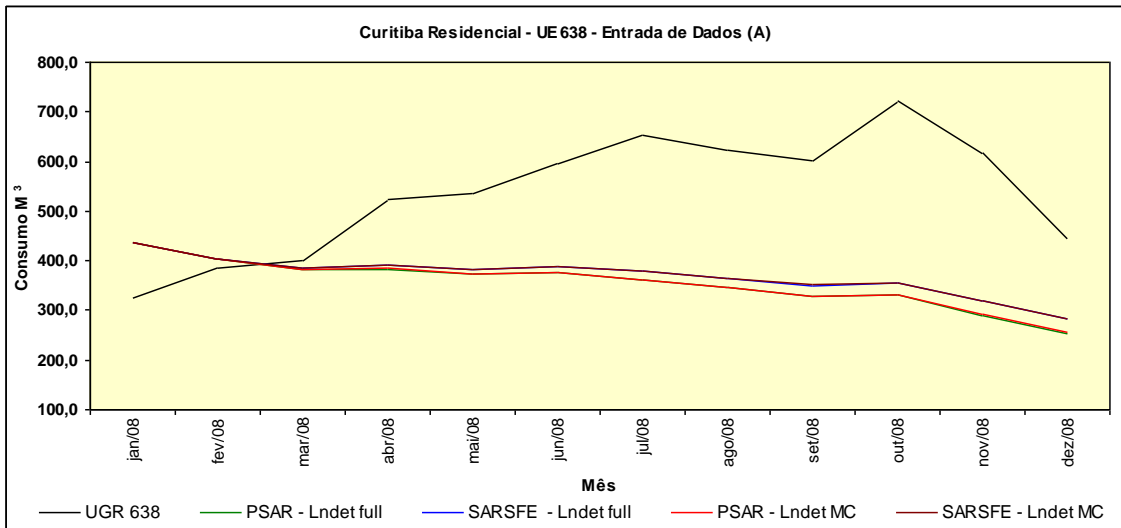


Figura 4.24 (a) – Previsão UE 638 - Segmento Residencial de Curitiba – Opção (A).

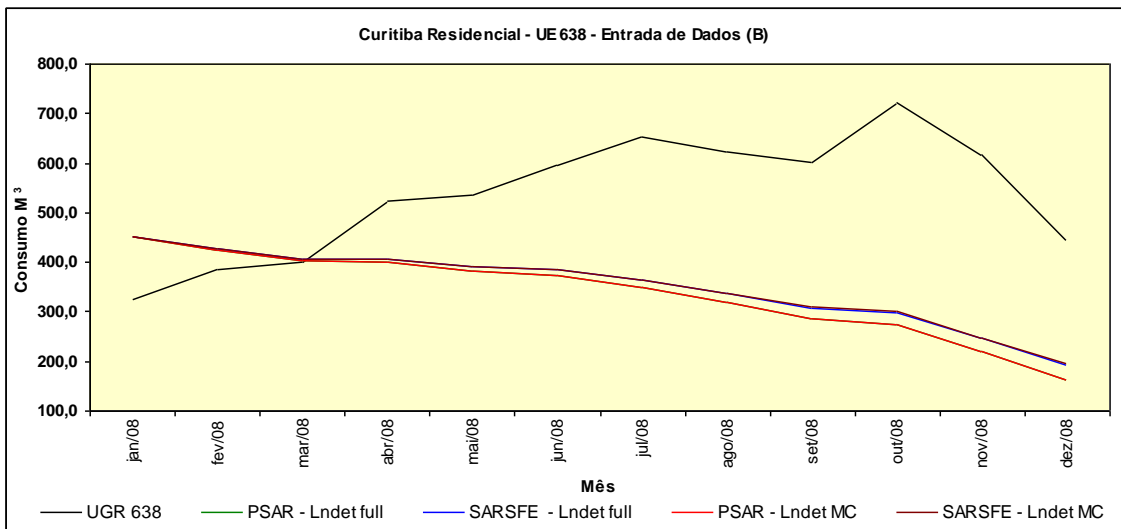


Figura 4.24 (b) – Previsão UE 638 - Segmento Residencial de Curitiba – Opção (B).

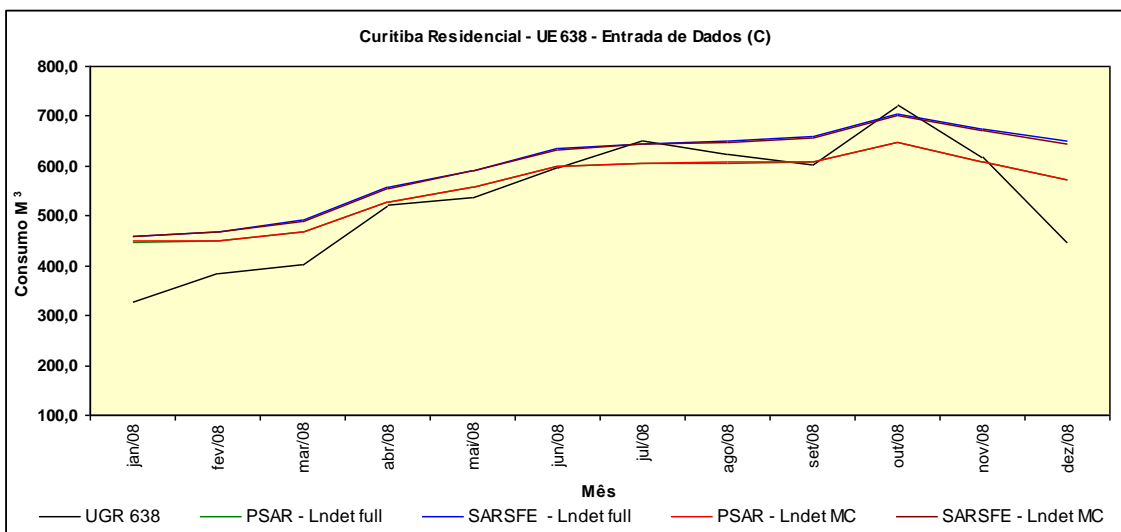


Figura 4.24 (c) – Previsão UE 638 - Segmento Residencial de Curitiba – Opção (C).

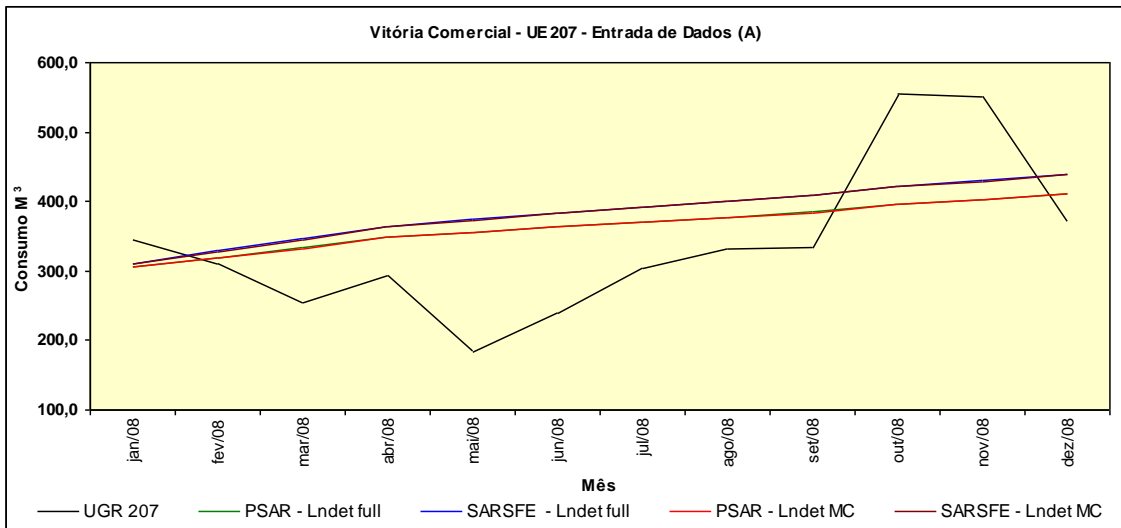


Figura 4.25 (a) – Previsão UE 207 - Segmento Comercial de Vitória – Opção (A).

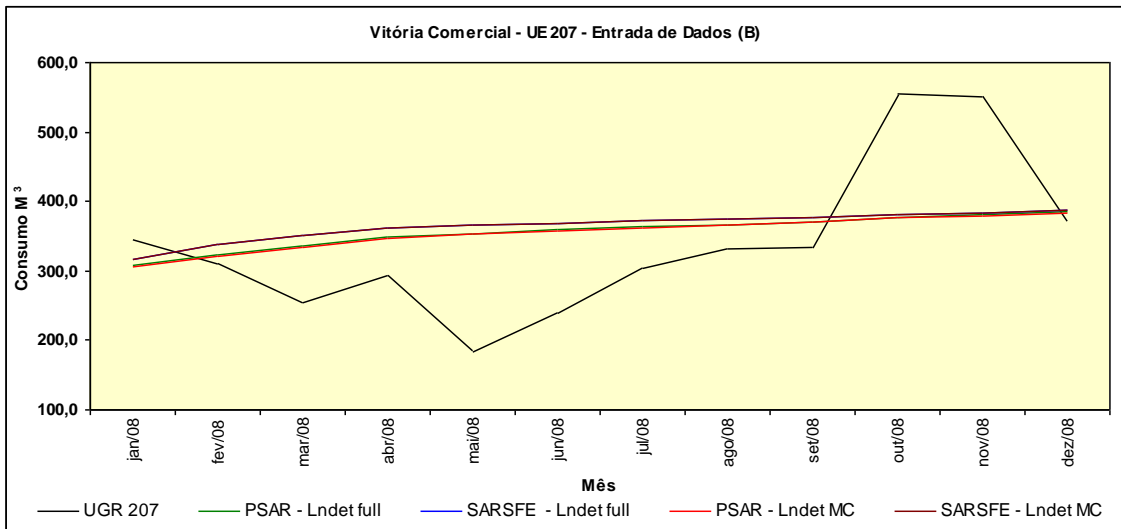


Figura 4.25 (b) – Previsão UE 207 - Segmento Comercial de Vitória – Opção (B).

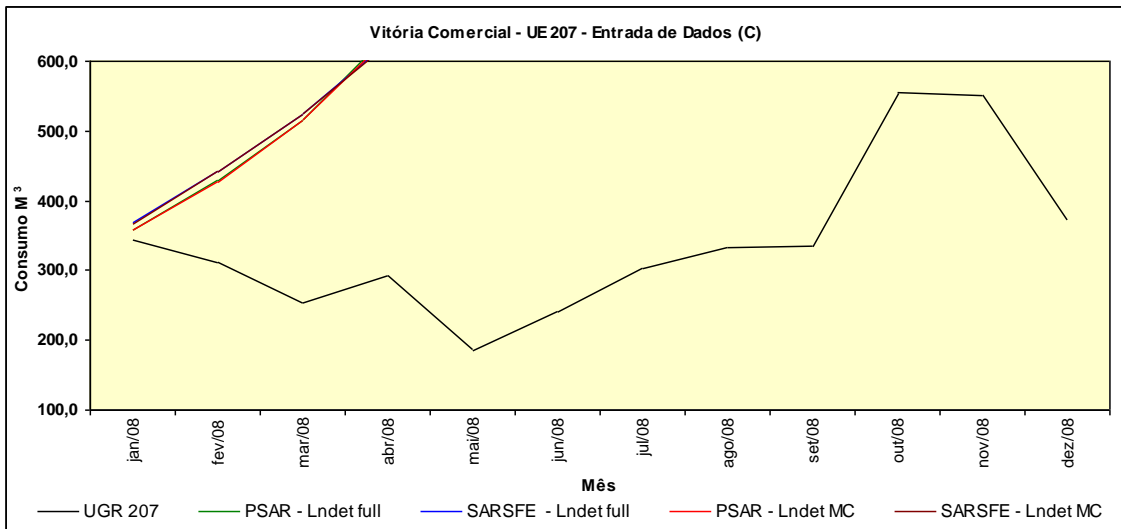


Figura 4.25 (c) – Previsão UE 207 - Segmento Comercial de Vitória – Opção (C).

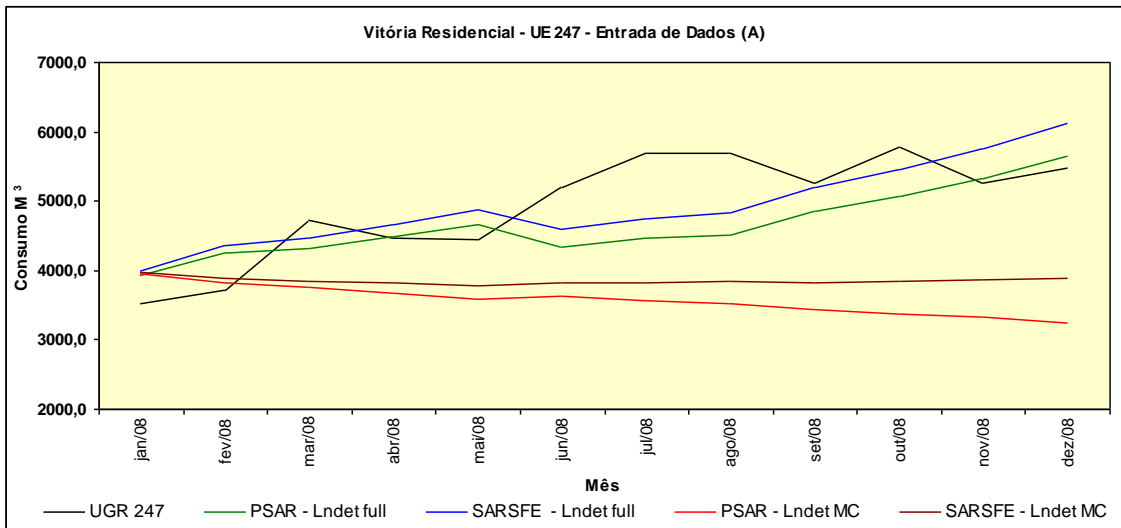


Figura 4.26 (a) – Previsão UE 247 - Segmento Residencial de Vitória – Opção (A).

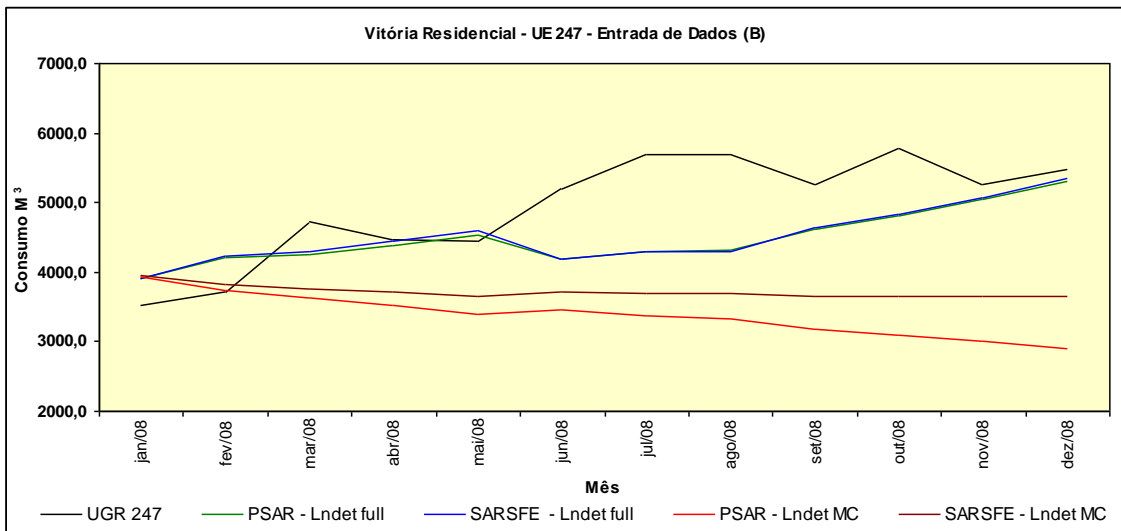


Figura 4.26 (b) – Previsão UE 247 - Segmento Residencial de Vitória – Opção (B).

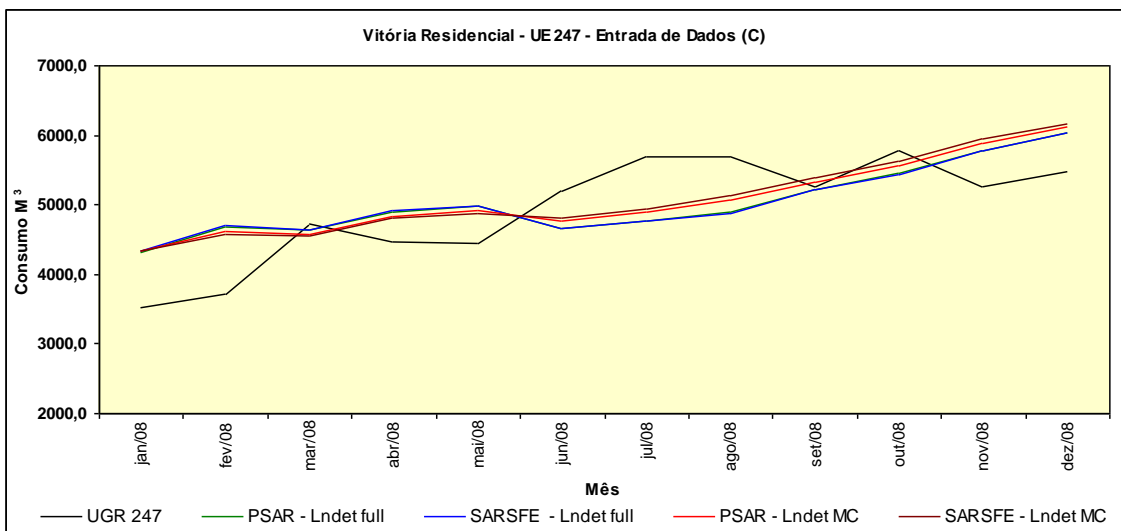


Figura 4.26 (c) – Previsão UE 247 - Segmento Residencial de Vitória – Opção (C).

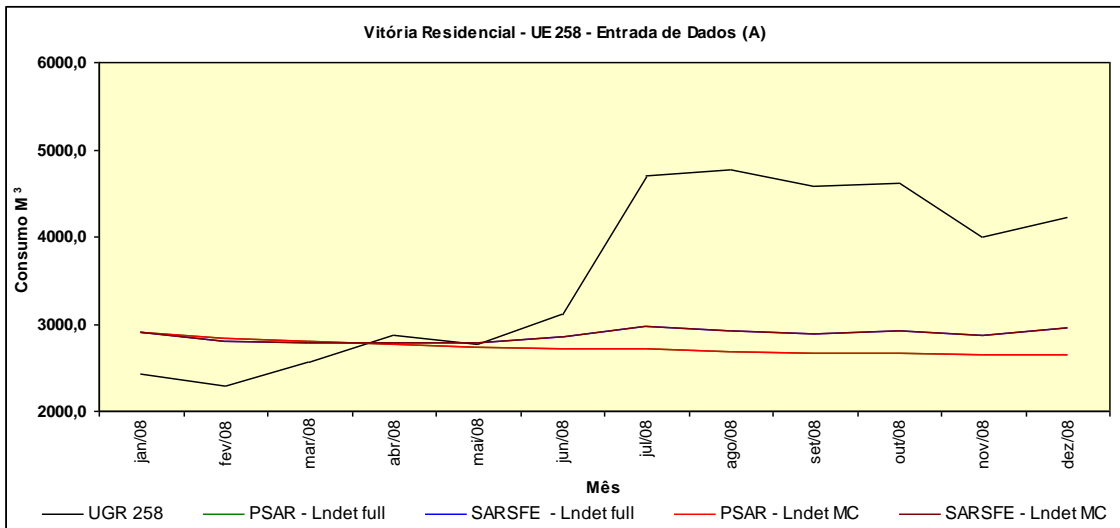


Figura 4.27 (a) – Previsão UE 258 - Segmento Residencial de Vitória – Opção (A).

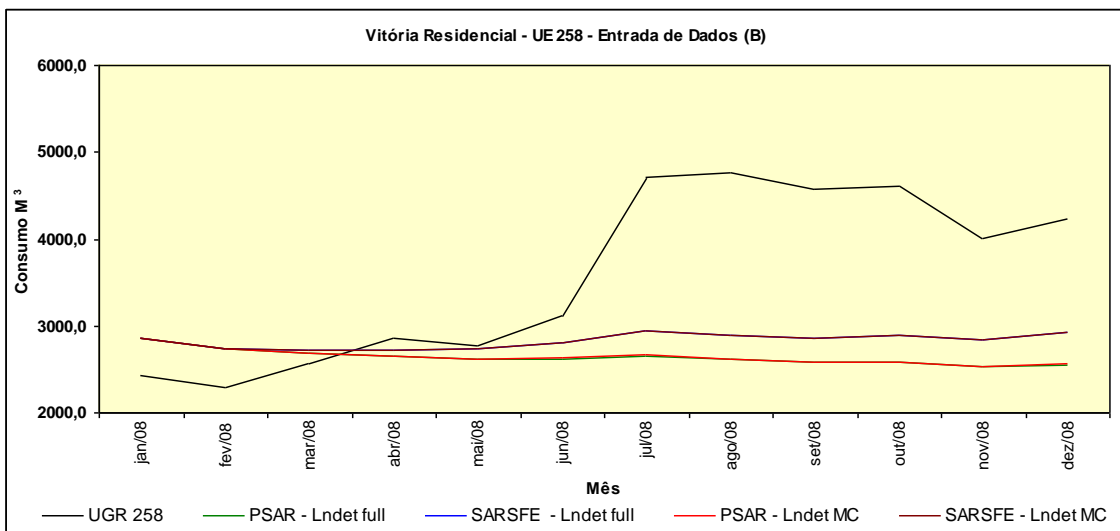


Figura 4.27 (b) – Previsão UE 258 - Segmento Residencial de Vitória – Opção (B).

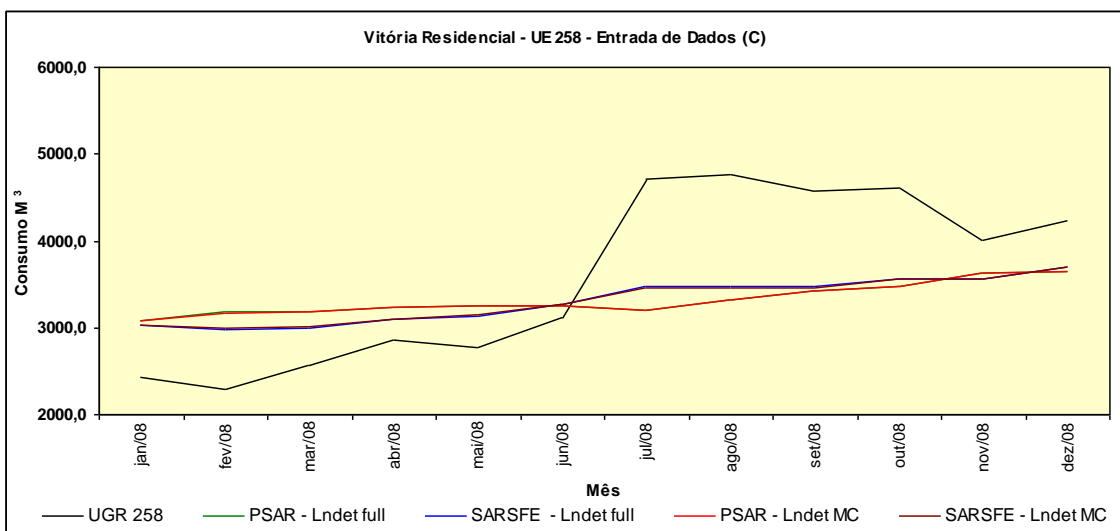


Figura 4.27 (c) – Previsão UE 258 - Segmento Residencial de Vitória – Opção (C).

Em uma observação inicial, verificou-se que em quatro dos cinco experimentos realizados, a previsão com a opção (C) de entrada de dados apresentou melhor

disposição frente aos valores de teste. Outra constatação é que excetuando o experimento da UE 247, o comportamento da previsão variou de forma pareada em função da forma funcional do modelo e não pelo método de determinação de  $\ln|I - pW|$ .

Com relação ao perfil das curvas de projeção nota-se que aparentam seguir um perfil definido por experimento: têm os pontos iniciais em Janeiro de 2008 muito próximos e, apresentam suave gradiente de defasagem ao longo do tempo.

#### 4.4 – ANÁLISE DOS RESULTADOS DAS PREVISÕES SAR

A análise dos resultados tem início pelos testes de estrutura, a partir dos quais se averiguou o quanto a estrutura SAR *Panel* mitigou os efeitos espaciais associados.

Com as proposições de entrada de dados (B) e (C), os resultados de *LM test* não foram satisfatórios ou apresentaram problemas em sua determinação. Na opção (B) o emprego de valores zerados para equilibrar a matriz propicia que  $X\beta$  se torne singular, ou mal escalada, comprometendo a determinação do gradiente e, conseqüentemente, a realização do teste. Na opção (C) ocorre um viés entre os valores reais e os previstos, conforme considerações já realizadas.

Com isto, o *LM test* só foi corretamente determinado para os valores ajustados da opção de entrada (A), conforme tabela 4.34.

Tabela 4.34 – Resultados do Teste de Estrutura – Entrada de dados opção (A).

| UE  | Forma Funcional  | Determinação $\ln  I - pW $ | LM error | LM lag  | Robust LM error | Robust LM lag |
|-----|------------------|-----------------------------|----------|---------|-----------------|---------------|
| 624 | P <sub>sar</sub> | <i>Indet full</i>           | 29,5599  | 1,0006  | 126,0941        | 97,5347       |
|     |                  | <i>Indet MC</i>             | 31,0952  | 1,5591  | 121,1988        | 91,6627       |
|     | SARSFE           | <i>Indet full</i>           | 28,5020  | 0,8647  | 130,4127        | 102,7755      |
|     |                  | <i>Indet MC</i>             | 30,2594  | 1,3734  | 127,1374        | 98,2514       |
| 638 | P <sub>sar</sub> | <i>Indet full</i>           | 24,5440  | 4,2604  | 219,3704        | 199,0868      |
|     |                  | <i>Indet MC</i>             | 26,0172  | 5,5006  | 199,1235        | 178,6069      |
|     | SARSFE           | <i>Indet full</i>           | 23,2774  | 4,9684  | 183,7804        | 165,4714      |
|     |                  | <i>Indet MC</i>             | 23,4214  | 5,0796  | 182,3829        | 164,0411      |
| 207 | P <sub>sar</sub> | <i>Indet full</i>           | 0,1165   | 0,1049  | 0,0230          | 0,0115        |
|     |                  | <i>Indet MC</i>             | 0,1560   | 0,1595  | 0,0222          | 0,0257        |
|     | SARSFE           | <i>Indet full</i>           | 0,0954   | 0,0482  | 0,0474          | 0,0002        |
|     |                  | <i>Indet MC</i>             | 0,2084   | 0,1504  | 0,0632          | 0,0051        |
| 247 | P <sub>sar</sub> | <i>Indet full</i>           | 75,0611  | 77,1527 | 10,1677         | 12,2592       |
|     |                  | <i>Indet MC</i>             | 11,1966  | 0,0346  | 20,3565         | 9,1946        |
|     | SARSFE           | <i>Indet full</i>           | 57,1795  | 39,2727 | 19,0270         | 1,1202        |
|     |                  | <i>Indet MC</i>             | 11,8205  | 0,0213  | 22,0519         | 10,2527       |
| 258 | P <sub>sar</sub> | <i>Indet full</i>           | 5,5101   | 0,0092  | 35,8796         | 30,3787       |
|     |                  | <i>Indet MC</i>             | 5,4362   | 0,0050  | 36,1048         | 30,6737       |
|     | SARSFE           | <i>Indet full</i>           | 2,6448   | 0,3503  | 9,1109          | 6,8164        |
|     |                  | <i>Indet MC</i>             | 2,7382   | 0,3932  | 9,0351          | 6,6900        |

#### 4.4.1 – CONCLUSÕES ACERCA DOS TESTES DE ESTRUTURA

##### a) Estrutura SAR

Comparando as probabilidades de não ocorrência de dependência espacial nos resíduos, ou relacionada com o atraso espacial da variável de interesse, conclui-se que a estrutura SAR mitiga da parcela dos resíduos os efeitos dos atrasos espaciais.

Fato este constatado ao verificar que, excetuando os experimentos da UE 247, com determinação de  $\ln|I - pw|$  por *Indet full* e da UE 207, com *Indet MC*, todas as probabilidades de ocorrência de efeitos espaciais associados aos atrasos foram menores do que os associados aos erros correlacionados.

##### b) Determinação de $\ln|I - pw|$

Aprofundando a colocação preliminar em 4.3.1, conclui-se que em experimentos com pequena quantidade de UE, o método de determinação de  $\ln|I - pw|$ , não impacta fortemente no resultado final da estimação de  $p$ .

Observando a tabela 4.34, nota-se que excetuando a UE 247, as diferenças individuais das probabilidades em *LM test* são baixas e em alguns casos similares. Isto se estende à forma funcional, pois a característica é comum à análise da probabilidade para não ocorrência de efeitos espaciais associados ao modelo com base em erros.

Complementarmente, ao consultar as tabelas 4.29 a 4.33, verifica-se que excetuando os doze experimentos da UE 247, para os quais ocorreu convergência a um mínimo de baixa qualidade, nos outros quarenta e oito experimentos, as diferenças individuais entre os escalares  $p$  são pequenas.

Uma vez que a dinâmica do processo de maximização é idêntica, as diferenças de valores de  $p$  ocorrem em função do conjunto de valores possíveis. Como estas diferenças são pequenas, conclui-se que os dois processos de determinação dos possíveis valores para  $p$ , que são à base de sua maximização, forneceram conjuntos de valores similares.

Cabe ressaltar que os autores consultados fazem referência à necessidade de avaliar a forma computacional mais eficiente para determinar  $\ln|I - pw|$ , quando a quantidade de UE na estimação é grande. Por outro lado, o código de determinação do MATLAB, orienta que para encontrar os valores exatos para o conjunto de possíveis  $p$ , se  $n < 500$ , deve-se escolher a forma de determinação completa, *Indet full*. Entretanto, como poderá ser visto mais adiante, no caso dos experimentos realizados nesta tese tal orientação não se confirmou.



**c) Forma funcional e Opção de entrada de dados**

Inicialmente, considerando que os testes referem-se somente à opção (A) de entrada de dados, não foi possível concluir qual a que apresenta o melhor desempenho, pois hora a probabilidade de erro referente à forma PSAR é maior do que para a SARSFE e, hora é menor. Tal situação ocorre inclusive para a probabilidade calculada em *Robust LM Test*.

**4.4.2 – ANÁLISE DOS ERROS DA PREVISÃO**

Na tabela 4.35 a seguir, são apresentados os valores percentuais do erro acumulado total (dez/08) e do MAPE, onde estão grifados em tons de amarelo para o coral, os quatros melhores resultados, sendo que o primeiro representa o melhor resultado e vice-versa.

Tabela 4.35 – Erro Acumulado Total e MAPE por UE.

| Tipo de entrada em X <sub>t</sub> | Forma Funcional do Modelo | Método de Determinação $\ln 1 - \rho W $ | UE 624               |        | UE 638               |        | UE 207               |         | UE 247               |        | UE 258               |        |
|-----------------------------------|---------------------------|--|----------------------|--------|----------------------|--------|----------------------|---------|----------------------|--------|----------------------|--------|
|                                   |                           |  | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE    | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   |
| A                                 | PSAR                      | <i>Indet full</i>                        | -26,09%              | 29,12% | -33,69%              | 35,14% | 7,22%                | 27,17%  | -5,66%               | 10,17% | -23,59%              | 25,93% |
|                                   |                           | <i>Indet MC</i>                          | -25,39%              | 28,52% | -33,50%              | 35,06% | 7,10%                | 27,08%  | -27,60%              | 27,64% | -23,59%              | 25,92% |
|                                   | SARSFE                    | <i>Indet full</i>                        | -19,98%              | 23,75% | -31,00%              | 32,66% | 12,90%               | 31,28%  | -0,28%               | 10,08% | -19,72%              | 22,47% |
|                                   |                           | <i>Indet MC</i>                          | -19,04%              | 22,94% | -30,95%              | 32,62% | 12,80%               | 31,22%  | -22,00%              | 22,80% | -19,75%              | 22,51% |
| B                                 | PSAR                      | <i>Indet full</i>                        | -30,97%              | 33,49% | -37,00%              | 39,67% | 4,73%                | 26,14%  | -9,09%               | 11,82% | -26,12%              | 26,93% |
|                                   |                           | <i>Indet MC</i>                          | -30,51%              | 33,09% | -37,04%              | 39,69% | 4,49%                | 25,99%  | -31,49%              | 30,91% | -26,09%              | 26,90% |
|                                   | SARSFE                    | <i>Indet full</i>                        | -22,98%              | 26,49% | -34,46%              | 37,42% | 7,39%                | 28,62%  | -8,60%               | 11,72% | -20,79%              | 22,49% |
|                                   |                           | <i>Indet MC</i>                          | -21,66%              | 25,34% | -34,38%              | 37,36% | 7,41%                | 28,63%  | -24,65%              | 24,99% | -20,79%              | 22,49% |
| C                                 | PSAR                      | <i>Indet full</i>                        | 15,07%               | 16,77% | 4,19%                | 10,64% | 247,83%              | 240,72% | 1,89%                | 11,59% | -7,13%               | 21,64% |
|                                   |                           | <i>Indet MC</i>                          | 15,33%               | 17,04% | 4,25%                | 10,67% | 243,70%              | 237,15% | 2,81%                | 10,87% | -7,13%               | 21,60% |
|                                   | SARSFE                    | <i>Indet full</i>                        | 19,49%               | 21,18% | 11,84%               | 15,18% | 170,83%              | 175,45% | 1,86%                | 11,69% | -7,32%               | 18,51% |
|                                   |                           | <i>Indet MC</i>                          | 19,71%               | 21,41% | 11,39%               | 14,82% | 168,81%              | 173,64% | 3,24%                | 10,65% | -7,42%               | 18,65% |

Os gráficos com os erros relativo mensal, acumulado total e absoluto mensal estão disponíveis no Apêndice H.

**4.4.2.1 – ANÁLISE DAS PREVISÕES INDIVIDUAIS POR UE**

**a) UE 624 e 638**

Observa-se nos gráficos de erros relativos mensais das UE 624 e 638 que, em ambas UE, as opções de entrada de dados (A) e (B) tiveram comportamento similar no tempo.

Por exemplo, como pode ser visto nas figuras 4.28 e 4.29, mesmo com intervalos de desvios diferentes maior na UE 638 do que na 624, os experimentos alcançaram desvio zero antes do terceiro evento previsto. Porém, após o terceiro evento previsto os desvios negativos em relação ao valor real foram constantes.

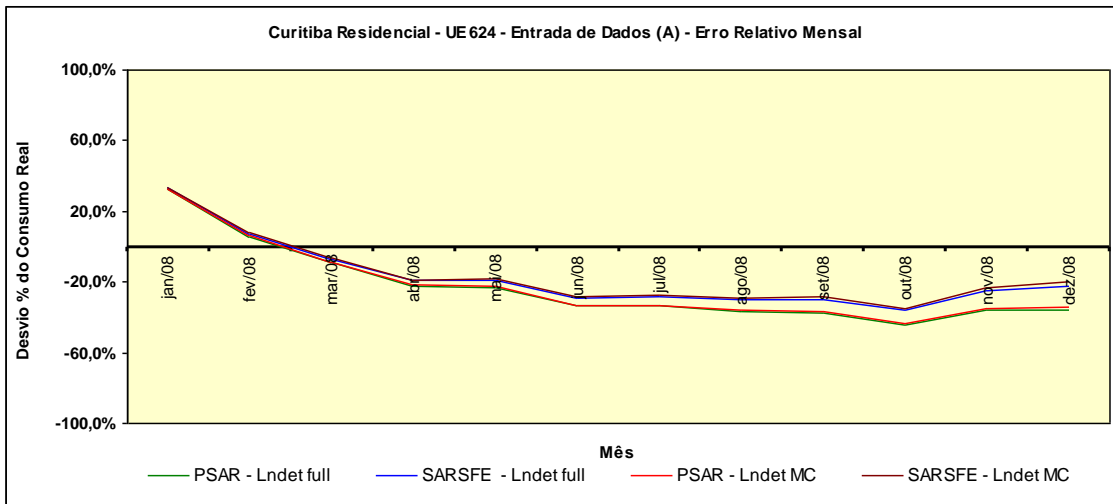


Figura 4.28 – Previsão UE 624 – Erro Relativo Mensal – Opção (A).

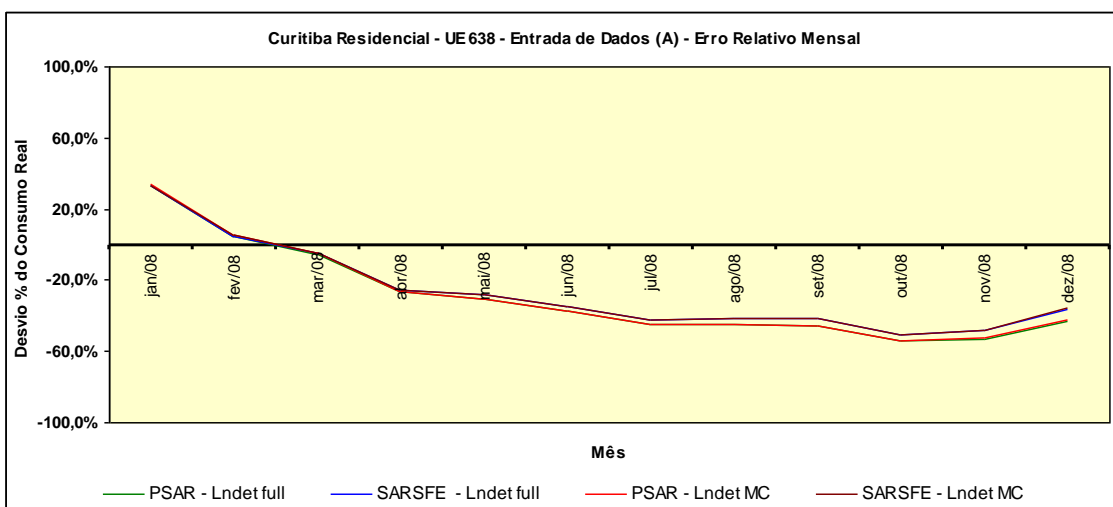


Figura 4.29 – Previsão UE 638 – Erro Relativo Mensal – Opção (A).

Já com a opção (C), nas UE 624 e 638, conforme figura 4.30 e 4.31, os valores previstos estão melhor distribuídos ao redor do centro da meta. Nestes casos o modelo corrigiu as curvas para a meta quando tendiam a um desvio maior.

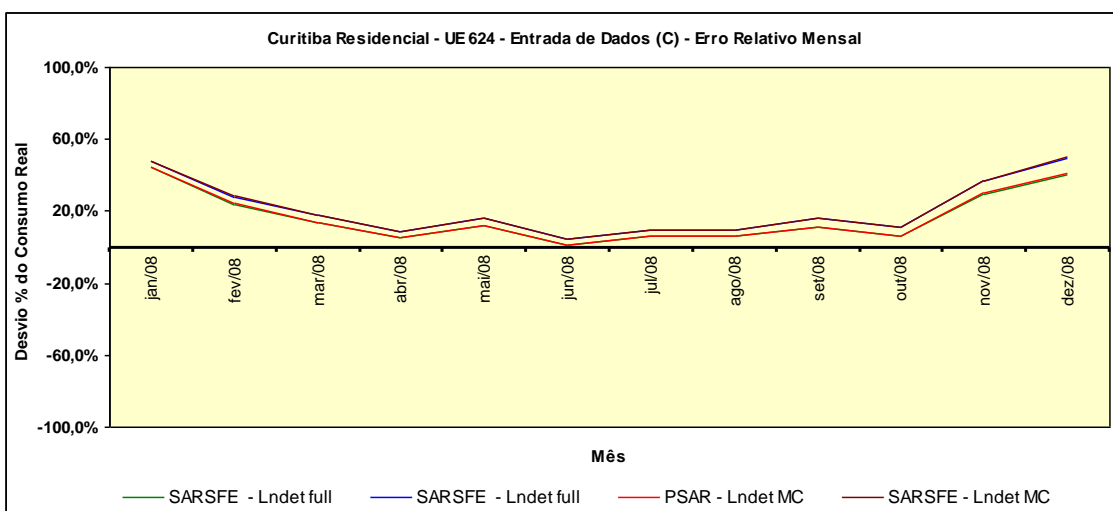


Figura 4.30 – Previsão UE 624 – Erro Relativo Mensal – Opção (C).

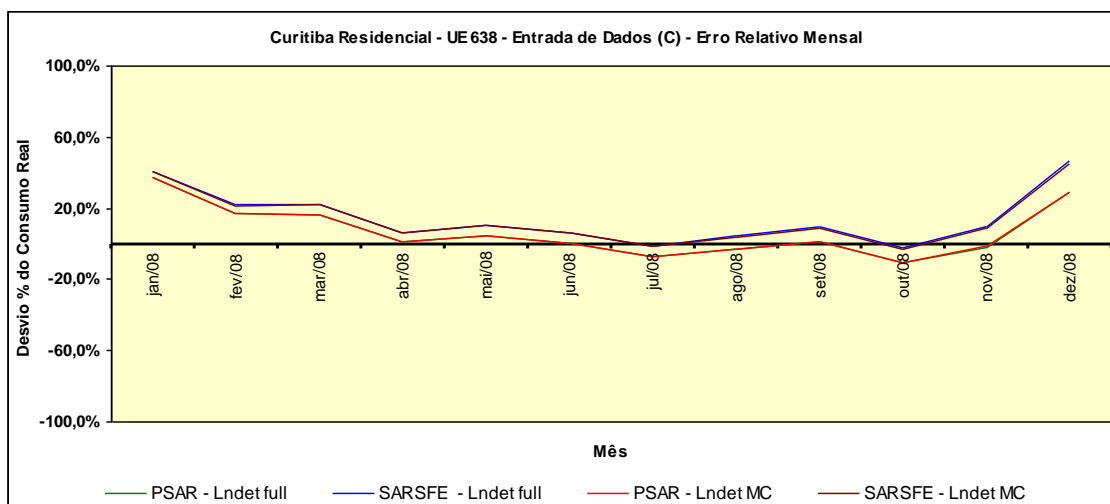


Figura 4.31 – Previsão UE 638 – Erro Relativo Mensal – Opção (C).

Nos testes de hipótese, os desvios das previsões foram comparados par a par, entre todos os experimentos. Ao total, os doze experimentos individuais por UE geraram sessenta e seis testes de hipóteses combinados par a par.

Nas UE 624 e 638, comparando inicialmente os experimentos por opção de entrada, pode-se constatar conforme tabelas 4.36 e 4.37 que a hipótese nula de igualdade, ou seja, de que os experimentos têm a mesma significância, é verdadeira em todos os casos.

Tabela 4.36 – Teste de Hipótese – UE 624 – Previsões por Opção de Entrada de Dados.

| Experimento A                  | Experimento B                  | $T_{\rho}^*$ | Graus de Liberdade | t crítico $\alpha = 0.05$ | Hipótese   |
|--------------------------------|--------------------------------|--------------|--------------------|---------------------------|------------|
| PSAR - Lndet full - Ent. (A)   | SARSFE - Lndet full - Ent. (A) | 1,0719       | 22,00              | 2,074                     | VERDADEIRO |
| PSAR - Lndet full - Ent. (A)   | PSAR - Lndet MC - Ent. (A)     | 0,1086       | 23,00              | 2,069                     | VERDADEIRO |
| PSAR - Lndet full - Ent. (A)   | SARSFE - Lndet MC - Ent. (A)   | 1,2488       | 22,00              | 2,074                     | VERDADEIRO |
| SARSFE - Lndet full - Ent. (A) | PSAR - Lndet MC - Ent. (A)     | -0,9685      | 22,00              | 2,074                     | VERDADEIRO |
| SARSFE - Lndet full - Ent. (A) | SARSFE - Lndet MC - Ent. (A)   | 0,1884       | 23,00              | 2,069                     | VERDADEIRO |
| PSAR - Lndet MC - Ent. (A)     | SARSFE - Lndet MC - Ent. (A)   | 1,1473       | 22,00              | 2,074                     | VERDADEIRO |
| PSAR - Lndet full - Ent. (B)   | SARSFE - Lndet full - Ent. (B) | 1,1811       | 21,00              | 2,080                     | VERDADEIRO |
| PSAR - Lndet full - Ent. (B)   | PSAR - Lndet MC - Ent. (B)     | 0,0592       | 23,00              | 2,069                     | VERDADEIRO |
| PSAR - Lndet full - Ent. (B)   | SARSFE - Lndet MC - Ent. (B)   | 1,3982       | 21,00              | 2,080                     | VERDADEIRO |
| SARSFE - Lndet full - Ent. (B) | PSAR - Lndet MC - Ent. (B)     | -1,1263      | 22,00              | 2,074                     | VERDADEIRO |
| SARSFE - Lndet full - Ent. (B) | SARSFE - Lndet MC - Ent. (B)   | 0,2363       | 23,00              | 2,069                     | VERDADEIRO |
| PSAR - Lndet MC - Ent. (B)     | SARSFE - Lndet MC - Ent. (B)   | 1,3451       | 21,00              | 2,080                     | VERDADEIRO |
| PSAR - Lndet full - Ent. (C)   | SARSFE - Lndet full - Ent. (C) | -0,8982      | 23,00              | 2,069                     | VERDADEIRO |
| PSAR - Lndet full - Ent. (C)   | PSAR - Lndet MC - Ent. (C)     | -0,0557      | 23,00              | 2,069                     | VERDADEIRO |
| PSAR - Lndet full - Ent. (C)   | SARSFE - Lndet MC - Ent. (C)   | -0,9362      | 23,00              | 2,069                     | VERDADEIRO |
| SARSFE - Lndet full - Ent. (C) | PSAR - Lndet MC - Ent. (C)     | 0,8388       | 23,00              | 2,069                     | VERDADEIRO |
| SARSFE - Lndet full - Ent. (C) | SARSFE - Lndet MC - Ent. (C)   | -0,0424      | 23,00              | 2,069                     | VERDADEIRO |
| PSAR - Lndet MC - Ent. (C)     | SARSFE - Lndet MC - Ent. (C)   | -0,8770      | 23,00              | 2,069                     | VERDADEIRO |

Tabela 4.37 – Teste de Hipótese – UE 638 – Previsões por Opção de Entrada de Dados.

| Experimento A                  | Experimento B                  | $T_0^*$ | Graus de Liberdade | $t_{crítico} \alpha = 0.05$ | Hipótese   |
|--------------------------------|--------------------------------|---------|--------------------|-----------------------------|------------|
| PSAR - Lndet full - Ent. (A)   | SARSFE - Lndet full - Ent. (A) | 0,3083  | 23,00              | 2,069                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (A)   | PSAR - Lndet MC - Ent. (A)     | 0,0110  | 23,00              | 2,069                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (A)   | SARSFE - Lndet MC - Ent. (A)   | 0,3135  | 23,00              | 2,069                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (A) | PSAR - Lndet MC - Ent. (A)     | -0,2973 | 23,00              | 2,069                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (A) | SARSFE - Lndet MC - Ent. (A)   | 0,0053  | 23,00              | 2,069                       | VERDADEIRO |
| PSAR - Lndet MC - Ent. (A)     | SARSFE - Lndet MC - Ent. (A)   | 0,3025  | 23,00              | 2,069                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (B)   | SARSFE - Lndet full - Ent. (B) | 0,2407  | 23,00              | 2,069                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (B)   | PSAR - Lndet MC - Ent. (B)     | -0,0022 | 23,00              | 2,069                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (B)   | SARSFE - Lndet MC - Ent. (B)   | 0,2474  | 23,00              | 2,069                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (B) | PSAR - Lndet MC - Ent. (B)     | -0,2428 | 23,00              | 2,069                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (B) | SARSFE - Lndet MC - Ent. (B)   | 0,0066  | 23,00              | 2,069                       | VERDADEIRO |
| PSAR - Lndet MC - Ent. (B)     | SARSFE - Lndet MC - Ent. (B)   | 0,2495  | 23,00              | 2,069                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (C)   | SARSFE - Lndet full - Ent. (C) | -0,9836 | 22,00              | 2,074                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (C)   | PSAR - Lndet MC - Ent. (C)     | -0,0057 | 23,00              | 2,069                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (C)   | SARSFE - Lndet MC - Ent. (C)   | -0,9195 | 22,00              | 2,074                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (C) | PSAR - Lndet MC - Ent. (C)     | 0,9769  | 22,00              | 2,074                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (C) | SARSFE - Lndet MC - Ent. (C)   | 0,0724  | 23,00              | 2,069                       | VERDADEIRO |
| PSAR - Lndet MC - Ent. (C)     | SARSFE - Lndet MC - Ent. (C)   | -0,9128 | 23,00              | 2,069                       | VERDADEIRO |

No tocante à comparação dos experimentos entre as diferentes opções de entrada, no caso da UE 638 todos os experimentos com a opção de entrada (C) diferiram das opções (A) e (B). Na tabela 4.38 são apresentados os resultados considerando a forma funcional PSAR com as duas opções de determinação e de entrada de dados opção (C).

Tabela 4.38 – Teste de Hipótese – UE 638 – PSAR c/ Opção (C) de Entrada de Dados.

| Experimento A                  | Experimento B                | $T_0^*$ | Graus de Liberdade | $t_{crítico} \alpha = 0.05$ | Hipótese |
|--------------------------------|------------------------------|---------|--------------------|-----------------------------|----------|
| PSAR - Lndet full - Ent. (A)   | PSAR - Lndet full - Ent. (C) | 4,2662  | 14,00              | 2,145                       | FALSO    |
| PSAR - Lndet full - Ent. (A)   | PSAR - Lndet MC - Ent. (C)   | 4,2609  | 14,00              | 2,145                       | FALSO    |
| SARSFE - Lndet full - Ent. (A) | PSAR - Lndet full - Ent. (C) | 4,1329  | 15,00              | 2,131                       | FALSO    |
| SARSFE - Lndet full - Ent. (A) | PSAR - Lndet MC - Ent. (C)   | 4,1271  | 15,00              | 2,131                       | FALSO    |
| PSAR - Lndet MC - Ent. (A)     | PSAR - Lndet full - Ent. (C) | 4,2615  | 14,00              | 2,145                       | FALSO    |
| PSAR - Lndet MC - Ent. (A)     | PSAR - Lndet MC - Ent. (C)   | 4,2561  | 14,00              | 2,145                       | FALSO    |
| SARSFE - Lndet MC - Ent. (A)   | PSAR - Lndet full - Ent. (C) | 4,1303  | 15,00              | 2,131                       | FALSO    |
| SARSFE - Lndet MC - Ent. (A)   | PSAR - Lndet MC - Ent. (C)   | 4,1245  | 15,00              | 2,131                       | FALSO    |
| PSAR - Lndet full - Ent. (B)   | PSAR - Lndet full - Ent. (C) | 4,2578  | 13,00              | 2,160                       | FALSO    |
| PSAR - Lndet full - Ent. (B)   | PSAR - Lndet MC - Ent. (C)   | 4,2535  | 13,00              | 2,160                       | FALSO    |
| SARSFE - Lndet full - Ent. (B) | PSAR - Lndet full - Ent. (C) | 4,2170  | 14,00              | 2,145                       | FALSO    |
| SARSFE - Lndet full - Ent. (B) | PSAR - Lndet MC - Ent. (C)   | 4,2123  | 14,00              | 2,145                       | FALSO    |
| PSAR - Lndet MC - Ent. (B)     | PSAR - Lndet full - Ent. (C) | 4,2552  | 13,00              | 2,160                       | FALSO    |
| PSAR - Lndet MC - Ent. (B)     | PSAR - Lndet MC - Ent. (C)   | 4,2509  | 13,00              | 2,160                       | FALSO    |
| SARSFE - Lndet MC - Ent. (B)   | PSAR - Lndet full - Ent. (C) | 4,2176  | 14,00              | 2,145                       | FALSO    |
| SARSFE - Lndet MC - Ent. (B)   | PSAR - Lndet MC - Ent. (C)   | 4,2129  | 14,00              | 2,145                       | FALSO    |

Já na UE 624 está razão apresentou 50 % dos experimentos com a opção de entrada (C) diferindo das opções (A) e (B). Ou seja, ao contrário da UE 638, não houve uma combinação de forma funcional, método de determinação ou opção de entrada de dados na qual a previsão fosse significativamente diferente das demais

Em complemento à tabela 4.35, observa-se que nestas duas UE, a opção (C) apresentou melhor desempenho na previsão. Nas figuras 4.32 e 4.33, são apresentados respectivamente para as UE 624 e 638, o gráfico do erro acumulado total com a referida opção (C).

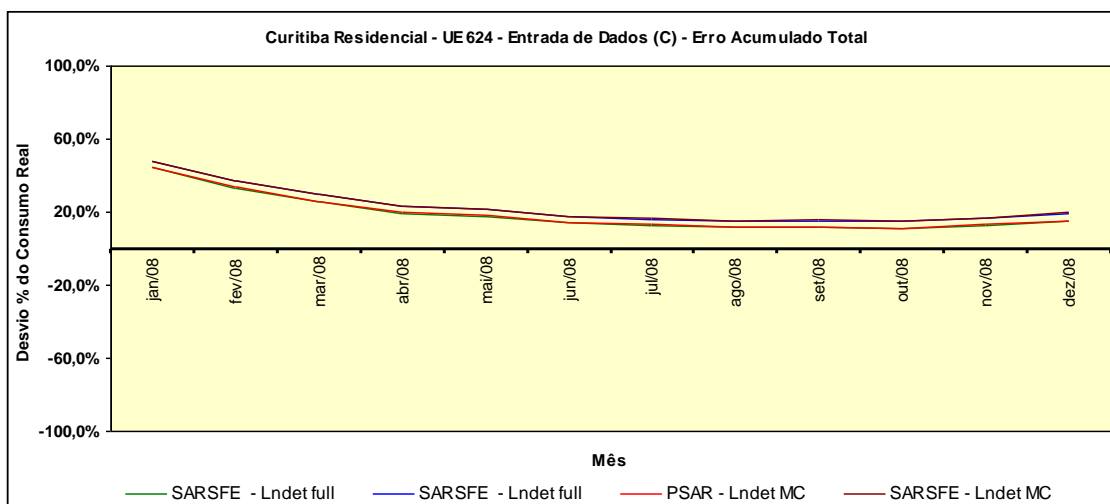


Figura 4.32 – Previsão UE 624 – Erro Acumulado Total – Opção (C).

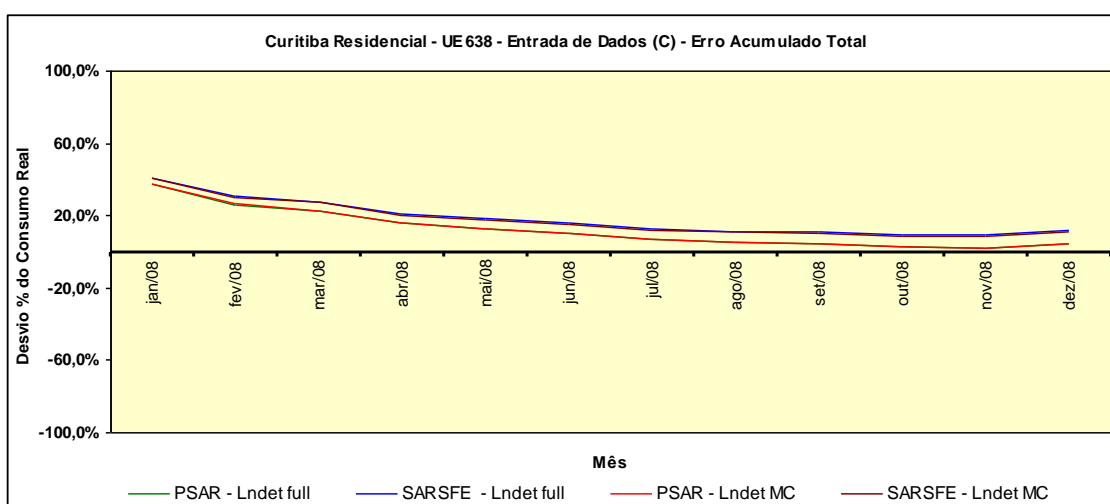


Figura 4.33 – Previsão UE 638 – Erro Acumulado Total – Opção (C).

Nestes gráficos nota-se que, fruto da capacidade do modelo corrigir as inflexões, da sexta previsão em diante, os erros acumulados totais mantêm-se estáveis.

***b) UE 207***

Comparando com os valores reais, na UE 207 a opção (C) teve o pior aproveitamento entre todos os experimentos, conforme pode-se verificar observando as figuras 4.25 (a) a (c). Neste caso, a previsão apresentou desvios positivos em todos os eventos, alcançando patamares como na forma PSAR, de cerca de 600% no último evento da série.

Nesta UE os melhores ajustes foram alcançados respectivamente com a forma funcional PSAR e opções (B) e (A), como pode ser visto na tabela 4.35. Entretanto, chama a atenção, consultando a mesma tabela, as diferenças do erro acumulado total e do MAPE entre a opção (C) e as demais. Neste último, a menor diferença é 4,5 vezes maior, enquanto que na UE 624, por exemplo, é de 1,5 vezes.

Observações corroboradas com os resultados obtidos nos testes de hipóteses, na qual nesta UE, assim como nas duas anteriores, constata-se ser verdadeira a hipótese de que estes não apresentam diferenças significativas, quando comparados os experimentos com a mesma opção de entrada, para as 3 situações.

Já quando se compara os experimentos com diferentes entradas de dados na UE 207 todas as previsões com a opção (C) diferem das demais, enquanto as de opção (A) e (B) são similares, independente da forma funcional e do método de determinação de  $\ln|I - pw|$ . Entretanto, diferente da UE 638, neste caso os melhores resultados de MAPE e erro acumulado total são com a forma funcional PSAR e opção de entrada de dados (B).

Independente disto percebe-se que em nenhum dos casos, o modelo conseguiu mapear a queda inicial dos valores da série de teste, bem como seu pico final, como pode ser visto nas figuras 4.26 (a) a (c). Sendo assim, a curva de previsão posicionou-se em uma posição central quanto aos valores reais. Observando os experimentos com opções (A) e (B), nota-se que estas curvas apresentam um comportamento típico de previsão linear.

Quanto ao erro acumulado total, conforme a figura 4.34, na opção (B) de entrada de dados, nota-se que o modelo corrige o desvio dos valores previstos a partir do quinto evento, mesmo assim somente a partir do nono evento o erro acumulado total decaiu.

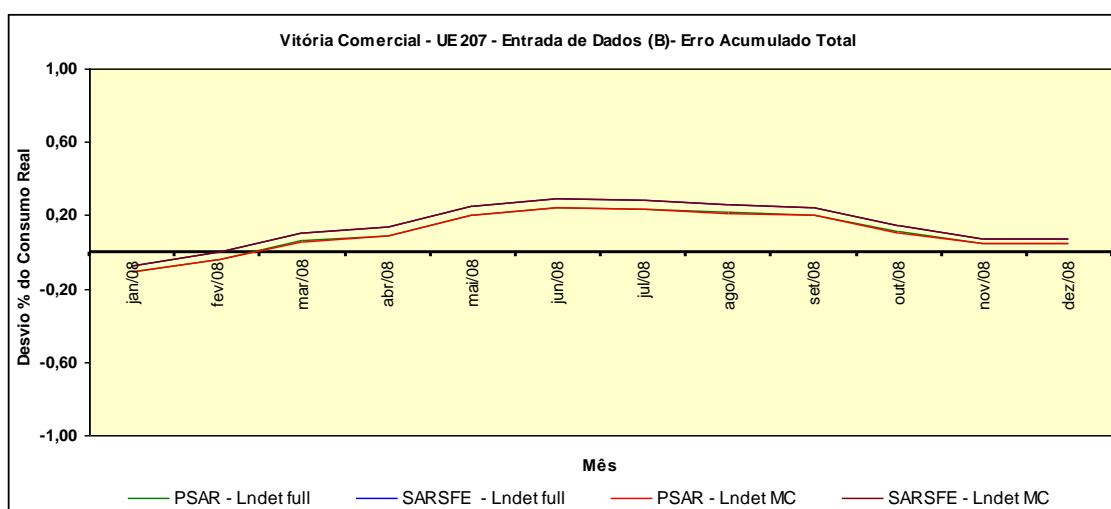


Figura 4.34 – Previsão UE 207 – Erro Acumulado Total – Opção (B).

No geral, o erro acumulado total apresentou desvio médio menor do que o MAPE e, por consequência, os resultados apontam para uma melhor capacidade de prever o valor total do consumo, do que os eventos pontuais.

c) UE 247

Diferente das outras UE de interesse a 247 apresentou os resultados mais assimétricos, bem como a maximização do escalar  $p$  apresentou convergência a um mínimo de baixa qualidade. Conforme as tabelas 4.32 e 4.35 percebe-se que nos experimentos onde  $p < 0$  o erro acumulado total e o MAPE, foram os menores. Estes experimentos apresentaram MAPE oscilando entre 10 e 12%, entretanto, somente nas previsões com a opção (C) houve regularidade.

Com relação aos resultados, nas figuras 4.26 (a) a (c), percebe-se que as previsões baseadas no método de determinação *lndet full* apresentaram bom resultados; contudo, os valores previstos com a opção (C) são mais homogêneos quando comparados aos valores de teste, conforme pode ser visto na figura 4.35.

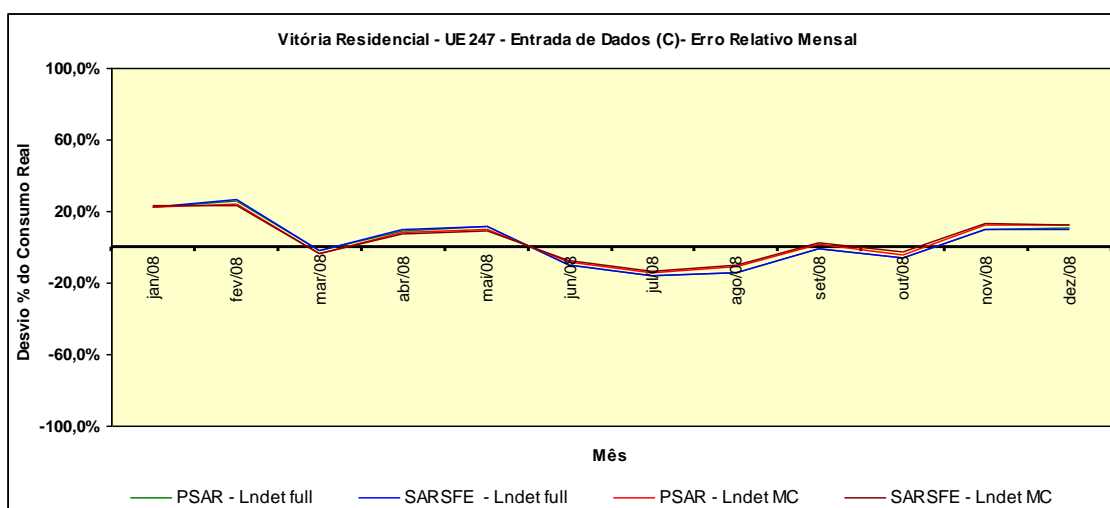


Figura 4.35 – Previsão UE 247 – Erro Relativo Mensal – Opção (C).

Situação esta corroborada nos testes de hipótese, nos quais verifica-se na tabela 4.39, que somente na opção (C) todos os resultados obtiveram a mesma significância.

Tabela 4.39 – Teste de Hipótese – UE 247.

| Experimento A                  | Experimento B                  | $T_0^*$ | Graus de Liberdade | t crítico $\alpha = 0.05$ | Hipótese   |
|--------------------------------|--------------------------------|---------|--------------------|---------------------------|------------|
| PSAR - Lndet full - Ent. (A)   | SARSFE - Lndet full - Ent. (A) | 0,1494  | 20,00              | 2,086                     | VERDADEIRO |
| PSAR - Lndet full - Ent. (A)   | PSAR - Lndet MC - Ent. (A)     | -3,6984 | 17,00              | 2,110                     | FALSO      |
| PSAR - Lndet full - Ent. (A)   | SARSFE - Lndet MC - Ent. (A)   | -3,2317 | 20,00              | 2,086                     | FALSO      |
| SARSFE - Lndet full - Ent. (A) | PSAR - Lndet MC - Ent. (A)     | -4,0283 | 13,00              | 2,160                     | FALSO      |
| SARSFE - Lndet full - Ent. (A) | SARSFE - Lndet MC - Ent. (A)   | -3,6669 | 15,00              | 2,131                     | FALSO      |
| PSAR - Lndet MC - Ent. (A)     | SARSFE - Lndet MC - Ent. (A)   | 0,9160  | 22,00              | 2,074                     | VERDADEIRO |
| PSAR - Lndet full - Ent. (B)   | SARSFE - Lndet full - Ent. (B) | 0,0345  | 23,00              | 2,069                     | VERDADEIRO |
| PSAR - Lndet full - Ent. (B)   | PSAR - Lndet MC - Ent. (B)     | -3,4904 | 17,00              | 2,110                     | FALSO      |
| PSAR - Lndet full - Ent. (B)   | SARSFE - Lndet MC - Ent. (B)   | -2,9494 | 21,00              | 2,080                     | FALSO      |
| SARSFE - Lndet full - Ent. (B) | PSAR - Lndet MC - Ent. (B)     | -3,5057 | 17,00              | 2,110                     | FALSO      |
| SARSFE - Lndet full - Ent. (B) | SARSFE - Lndet MC - Ent. (B)   | -2,9679 | 21,00              | 2,080                     | FALSO      |
| PSAR - Lndet MC - Ent. (B)     | SARSFE - Lndet MC - Ent. (B)   | 0,9880  | 22,00              | 2,074                     | VERDADEIRO |
| PSAR - Lndet full - Ent. (C)   | SARSFE - Lndet full - Ent. (C) | -0,0325 | 23,00              | 2,069                     | VERDADEIRO |
| PSAR - Lndet full - Ent. (C)   | PSAR - Lndet MC - Ent. (C)     | 0,3119  | 23,00              | 2,069                     | VERDADEIRO |
| PSAR - Lndet full - Ent. (C)   | SARSFE - Lndet MC - Ent. (C)   | 0,4135  | 23,00              | 2,069                     | VERDADEIRO |
| SARSFE - Lndet full - Ent. (C) | PSAR - Lndet MC - Ent. (C)     | 0,3445  | 23,00              | 2,069                     | VERDADEIRO |
| SARSFE - Lndet full - Ent. (C) | SARSFE - Lndet MC - Ent. (C)   | 0,4459  | 23,00              | 2,069                     | VERDADEIRO |
| PSAR - Lndet MC - Ent. (C)     | SARSFE - Lndet MC - Ent. (C)   | 0,1042  | 23,00              | 2,069                     | VERDADEIRO |

Na verdade a convergência de  $p$  influenciou o resultado das previsões de forma diferente das outras UE. Conforme a tabela 4.35, os resultados dos experimentos nas outras UE são pareados, ou seja, acontecem par a par conforme a forma funcional; já na UE 247 os resultados estiveram associados ao método de determinação *lndet full*.

Especificamente, na comparação dos resultados nos quais houve convergência de  $p$ , conforme tabela 4.40, os testes de hipótese apresentaram a mesma significância.

Tabela 4.40 – Teste de Hipótese – UE 247 – Experimentos para a Convergência de  $p$ .

| Experimento A                  | Experimento B                  | $T_0^*$ | Graus de Liberdade | $t_{critico} \alpha = 0,05$ | Hipótese   |
|--------------------------------|--------------------------------|---------|--------------------|-----------------------------|------------|
| PSAR - Lndet full - Ent. (A)   | SARSFE - Lndet full - Ent. (A) | 0,1494  | 20,00              | 2,086                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (B)   | SARSFE - Lndet full - Ent. (B) | 0,0345  | 23,00              | 2,069                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (A)   | PSAR - Lndet full - Ent. (B)   | -0,5200 | 23,00              | 2,069                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (A)   | SARSFE - Lndet full - Ent. (B) | -0,4798 | 23,00              | 2,069                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (A) | PSAR - Lndet full - Ent. (B)   | -0,7273 | 18,00              | 2,101                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (A) | SARSFE - Lndet full - Ent. (B) | -0,6793 | 18,00              | 2,101                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (A)   | PSAR - Lndet full - Ent. (C)   | -0,2102 | 21,00              | 2,080                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (A)   | SARSFE - Lndet full - Ent. (C) | -0,2370 | 22,00              | 2,074                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (A)   | PSAR - Lndet MC - Ent. (C)     | 0,0414  | 20,00              | 2,086                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (A)   | SARSFE - Lndet MC - Ent. (C)   | 0,1235  | 20,00              | 2,086                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (A) | PSAR - Lndet full - Ent. (C)   | -0,4453 | 23,00              | 2,069                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (A) | SARSFE - Lndet full - Ent. (C) | -0,4776 | 23,00              | 2,069                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (A) | PSAR - Lndet MC - Ent. (C)     | -0,1372 | 23,00              | 2,069                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (A) | SARSFE - Lndet MC - Ent. (C)   | -0,0331 | 23,00              | 2,069                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (B)   | PSAR - Lndet full - Ent. (C)   | 0,3902  | 19,00              | 2,093                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (B)   | SARSFE - Lndet full - Ent. (C) | 0,3649  | 19,00              | 2,093                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (B)   | PSAR - Lndet MC - Ent. (C)     | 0,6289  | 18,00              | 2,101                       | VERDADEIRO |
| PSAR - Lndet full - Ent. (B)   | SARSFE - Lndet MC - Ent. (C)   | 0,7040  | 18,00              | 2,101                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (B) | PSAR - Lndet full - Ent. (C)   | 0,3461  | 19,00              | 2,093                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (B) | SARSFE - Lndet full - Ent. (C) | 0,3212  | 19,00              | 2,093                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (B) | PSAR - Lndet MC - Ent. (C)     | 0,5819  | 18,00              | 2,101                       | VERDADEIRO |
| SARSFE - Lndet full - Ent. (B) | SARSFE - Lndet MC - Ent. (C)   | 0,6562  | 18,00              | 2,101                       | VERDADEIRO |

Sendo assim, independente do experimento com -0,28% de erro acumulado total e 10,08% de MAPE, conforme tabela 4.35 para efeitos de avaliação assumiu-se que a homogeneidade apresentada com a opção (C) de entrada de dados é mais consistente.

Adicionalmente, nesta opção, como pode ser visto nas figuras 4.36 e 4.37, os erros acumulado total e absoluto mensal, apresentaram comportamento homogêneo no tempo.

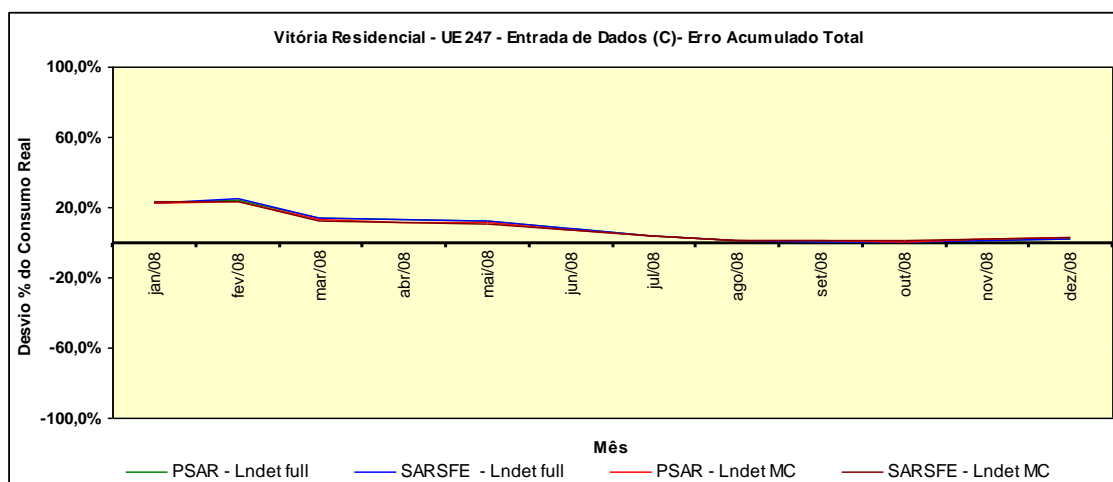


Figura 4.36 – Previsão UE 247 – Erro Acumulado Total – Opção (C).



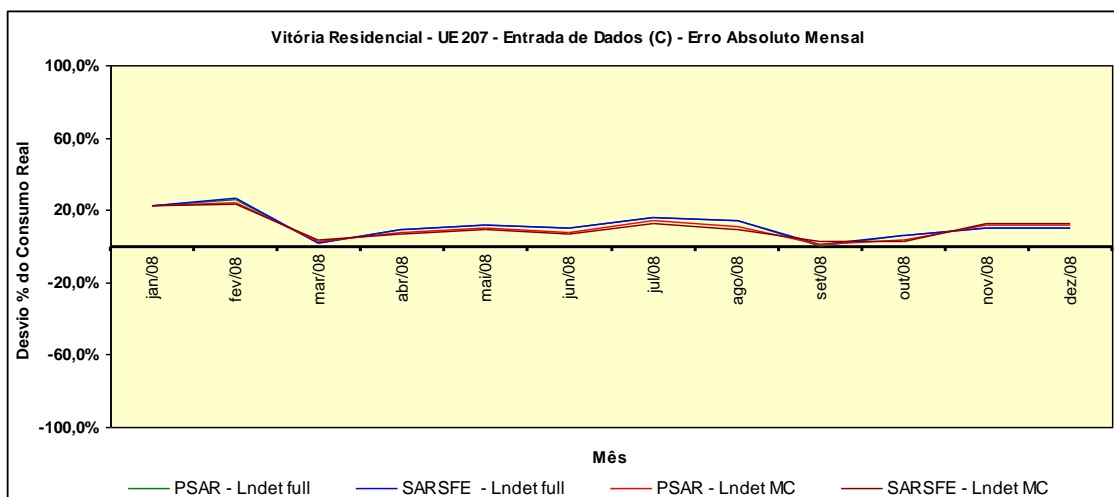


Figura 4.37 – Previsão UE 247 – Erro Absoluto Mensal – Opção (C).

#### d) UE 258

Especificamente na UE 258, todos os sessenta e seis testes de hipótese indicaram igualdade, ou seja, não existem diferenças significativas entre os desvios das doze previsões realizadas para esta unidade. Com isto, a avaliação baseou-se exclusivamente nos índices apresentados na tabela 4.35 e nos gráficos com as previsões.

Inicialmente, observando os gráficos em 4.27 (a) a (c), percebe-se que as previsões a partir das opções de entrada de dados (A) e (B) apresentam o mesmo comportamento no tempo. Ambos os experimentos alcançaram desvio próximo de zero entre o quarto e o quinto evento e deste ponto em diante, ficaram sempre no campo negativo em relação ao valor real.

Na verdade, no meio da série prevista ocorre uma inflexão, a partir do qual o erro acumulado total passa do campo positivo para o negativo. Na figura 4.38, apresenta-se o comportamento para este índice conforme a opção (B) de entrada de dados.

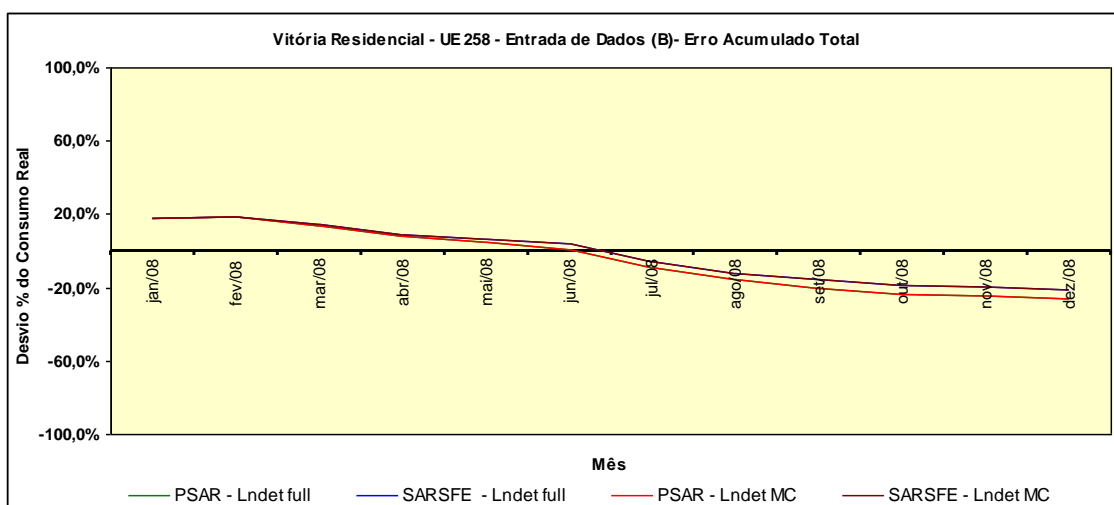


Figura 4.38 – Previsão UE 258 – Erro Acumulado Total – Opção (B).

Conforme tabela 4.35, a opção (C) alcançou os melhores resultados, porém nota-se na figura 4.39, que em nenhum dos experimentos houve desvio igual a zero.

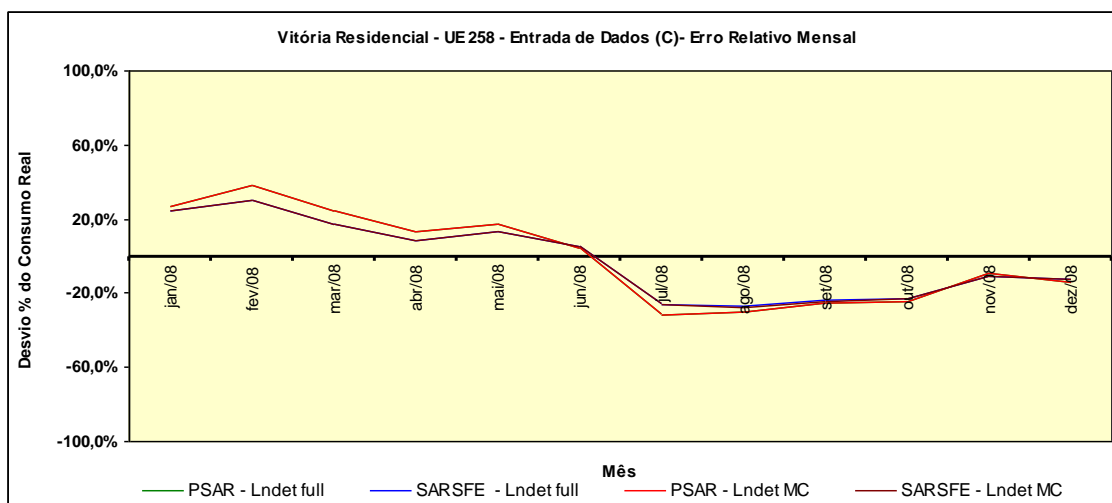


Figura 4.39 – Previsão UE 258 – Erro Relativo Mensal – Opção (C).

A previsão nesta opção apresenta melhor comportamento comparado aos valores reais, como pode ser visto na figura 4.28 (c); com isto atenuam os últimos valores previstos, o que impacta no erro acumulado total, como pode ser visto na figura 4.40.

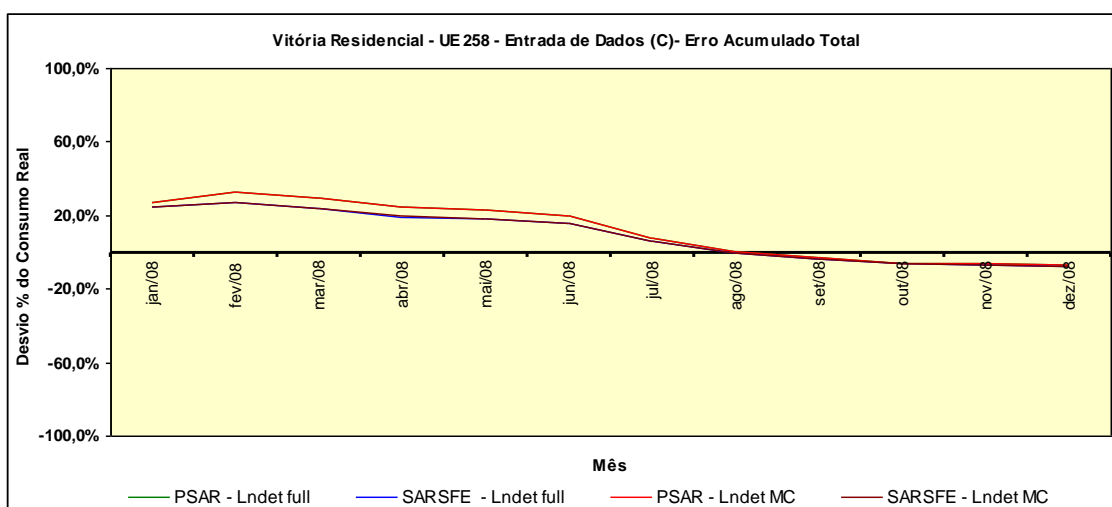


Figura 4.40 – Previsão UE 258 – Erro Acumulado Total – Opção (C).

Ainda na tabela 4.35, nota-se que o erro acumulado total na forma PSAR apresentou ligeira superioridade à SARSFE; enquanto que no MAPE os resultados foram ao contrário, bem como todos os percentuais do erro acumulado total foram negativos, ou seja, qualquer previsão total estaria abaixo do consumo real no período.

#### 4.4.2.2 – CONCLUSÕES COM BASE NA ANÁLISE DAS PREVISÕES

##### a) Estrutura SAR Panel

Em complemento à conclusão do item 4.4.1, percebe-se que, além de mitigar os efeitos dos atrasos espaciais da parcela dos resíduos, a estratégia de aplicação da parcela

espacial influencia na correção dos valores previstos em direção aos valores reais. Na consulta aos gráficos das previsões, apresentado em 4.3.1, observa-se que a aplicação da estrutura SAR *Panel* produz previsões que possuem inflexões ao longo do tempo, diferenciando-as assim das previsões unicamente lineares baseadas no tempo.

Cabe ressaltar que esta percepção ocorre de forma indiscriminada, com mais ou menos intensidade, excetuando os experimentos da UE 207. Porém, mesmo observando estes ajustes, conclui-se que em casos extremos, como as mudanças de patamares na série da variável de interesse, o previsor SAR tem pouca sensibilidade, isto por que, antes de tudo, este estimador é linear, independente das formas funcionais.

Sendo assim, em casos como da UE 258, conforme as figuras 4.27 (a) a (c), nota-se que mesmo existindo inflexões na curva de valores previstos, o modelo SAR *Panel* não foi capaz de melhor posicionar a curva, dentro do horizonte de previsão, ou ainda, de distinguir se este efeito é relacionado com uma sazonalidade.

Com relação à presença de sazonalidade associada aos efeitos espaciais, seja no modelo SAR ou no SEM, cabe ressaltar que não foram encontradas referências ao assunto na bibliografia consultada.

#### **b) Opção de entrada de dados**

Observando a tabela 4.35, conclui-se que a entrada de dados é o elemento de maior influência no processo de previsão. Em sessenta experimentos realizados, dos vinte melhores resultados, quatorze encontram-se na opção (C) de entrada de dados, seja pela visão do erro acumulado total, ou do MAPE. Mesmo na UE 247, onde ocorreu convergência na estimação de  $p$ , os resultados mais homogêneos foram na opção (C).

No tocante à estrutura SAR isto é um item relevante, visto a flexibilidade na entrada de dados em  $X\beta$ ; questão que ganha importância, na medida em que a premissa de analisar e reduzir a influência dos efeitos espaciais, no senso local, torna diferencial para o processo de previsão a possibilidade de capturar o comportamento temporal de  $y$ .

Ressalta-se que o emprego de variáveis explicativas socioeconômicas, apresentam natural dificuldade quanto à sua representação espacial fruto dos processos de coleta de dados (frequência e referência espacial), com isto a flexibilidade da estrutura SAR permitiu combinar arranjos de dados que capacitam o processo de previsão. Particularmente, permitiu montar uma estratégia de entrada de dados, que combina uma estrutura ARX, em  $X\beta$ , com a estrutura de mitigação de efeitos espaciais e que apresentou resultados positivos, visto os resultados obtidos.

Desta forma, a estrutura SAR confere ao modelador, no senso linear, a possibilidade de testar diferentes arranjos com o mesmo cenário de dados, oferecendo assim mais uma possibilidade de estimação, além das formas tradicionais que orbitam as estruturas puramente auto-regressivas e / ou de re-injeção.

***c) Forma Funcional***

Restringindo a análise ao grupo de melhores resultados seria correto afirmar que a forma funcional que apresentou os melhores resultados é a PSAR. Observando a tabela 4.35, onde estão destacados os melhores resultados, em tons de amarelo, em ordem crescente do mais claro para o mais escuro, percebe-se que o erro acumulado total de quatro, das cinco UE, estão na referida forma funcional.

Excluindo deste grupo os resultados da forma funcional SARSFE com opção (A) para a UE 247, pelos motivos apresentados em sua análise individual, percebe-se que nas UE 624, 638 e 207 a forma funcional PSAR é melhor que a SARSRE, e na UE 258 apresentam resultados similares.

Todavia, isto não é um fato, pois ao se observar o conjunto total de experimentos realizados, com exceção da UE 207, todas as UE apresentaram melhores resultados empregando a forma funcional SARSFE e opções de entrada de dados (A) e (B).

Na verdade existem consideráveis variações nos índices de erros quando são comparadas as opções de entrada de dados e não a forma funcional. Por exemplo, considerando a opção (C) como melhor na UE 258, a diferença entre seu erro acumulado total e a dos outros experimentos, orbitou entre 12 e 20 %; entretanto, a diferença de erro acumulado total entre as formas funcionais nesta mesma opção de entrada de dados foi de décimos.

Comparando as formas funcionais com base nos resultados dos testes de hipótese, tem-se o seguinte quadro conforme tabela 4.41.

Tabela 4.41 – Teste de Hipótese – Comparações Entre Formas Funcionais.

| UE    | H <sub>0</sub> Verdadeira | H <sub>0</sub> Falsa |
|-------|---------------------------|----------------------|
| 624   | 28                        | 8                    |
| 638   | 20                        | 16                   |
| 207   | 20                        | 16                   |
| 247   | 20                        | 16                   |
| 258   | 36                        | 0                    |
| Total | 124                       | 56                   |

Ou seja, ao todo existem cento e oitenta testes de hipóteses que relacionam as formas funcionais PSAR e SARSFE, uma contra a outra. Deste total, cento e vinte e

quatro comparações (68,8%) têm a mesma significância e, cinquenta e seis (31,2%) têm diferenças que justifiquem optar por uma forma funcional, como a de melhor performance.

Comparando especificamente as opções (A) e (B) contra a (C), são oitenta; nas quais quarenta e oito (60%) apresentam diferenças que justifiquem optar por uma forma funcional particular. Ou ainda, entre as cinquenta e seis comparações que têm diferenças significantes, quarenta e oito (85,7%) apontam para a opção (C) como a melhor.

Percebe-se com isto que a entrada de dados é mais significativa para os resultados do que a forma funcional empregada na estimação. Isto em parte é justificável sobre a ótica de que a previa análise da dependência espacial e, da conseqüente estacionariedade, aproxima os resultados estimados entre a forma PSAR e a SARSFE.

Como a autocorrelação espacial é análise do desvio da média dos momentos entre as observações do conjunto amostral ao longo do espaço, a identificação de um cluster espacial, com UE com comportamento espacial similar, encapsula os efeitos espaciais, tornando assim mais homogêneo o conjunto destas UE em análise.

Uma vez que o termo intercepto, que é incluído para caracterizar a forma funcional SARSFE, tem base na diferença das médias das observações empregadas, para um conjunto com observações mais homogêneas, o desvio entre os valores reais e os valores transformados no termo intercepto são menores. Sendo assim a influência de seu emprego é mitigado, o que não é verdade para um conjunto amostral mais heterogêneo, que naturalmente apresentaria maiores desvios.

Contudo, a análise previa da dependência espacial é condição necessária para determinar o emprego de estruturas de modelos espaço-temporais, mas não é suficiente para determinar a forma funcional. Sendo assim não é completamente correto afirmar que a melhor forma funcional é a PSAR, principalmente quando o item que apresentou maior influência sobre os estimadores foi a entrada de dados.

Desta forma, acentuando as considerações sobre a flexibilidade da estrutura SAR conclui-se que, mesmo realizando a análise previa da dependência espacial e, da conseqüente estacionariedade, em função das possíveis combinações de entrada de dados o correto é estimar as diversas formas funcionais disponíveis.

#### ***d) Previsão SAR para o consumo de GN***

No geral pôde-se concluir que para a previsão de GN, a estrutura SAR *Panel* apresentou uma ligeira superioridade para prever os valores de consumo total de GN no

período, do que as parcelas mensais, neste caso a distribuição ao longo do tempo. Entre os sessenta experimentos realizados, trinta e três apresentaram variação positiva maior do que 3% entre a diferença dos valores absolutos do erro acumulado total e do MAPE.

Por exemplo, no primeiro experimento da tabela 4.35, o erro acumulado total é de 26,09% contra 29,12% de MAPE, ou seja, pouco acima dos 3% de erro.

Com relação à aplicação da estrutura SAR como ferramenta de estimação de modelos de previsões para o consumo de GN, até este ponto só é possível concluir que:

- A instabilidade estrutural e a sua influência na forma de heterogeneidade são mitigadas na medida em que a aplicação de dados em forma de área, associada à análise da dependência espacial, demandam a investigação e validação da estrutura espacial, para determinar uma matriz  $W_{ij}$  aplicável no tempo;
- Mesmo observando um conjunto diminuto de dados disponíveis foi possível determinar a influência do consumo de GN entre as UE, estimar os parâmetros e constituir uma previsão de consumo de GN de forma local;
- Face às características do consumo nos segmentos estudados, ou seja, infante e não homogêneo sobre o espaço, é necessário estimar os modelos empregando mais de uma forma funcional;
- Permite integrar em um único modelo variáveis explicativas com resoluções espaciais distintas; ou seja, além da análise dos efeitos espaciais, contempla a análise com informações macroeconômicas de vários níveis regionais; e,
- Permite introduzir os efeitos temporais da variável dependente, uma vez que tem flexibilidade de arranjo em  $X\beta$ . Sendo assim pode-se valer de técnicas de estimação temporal, como por exemplo, a análise de componentes principais ou a autocorrelação parcial, para compor modelos mais abrangentes.

Contudo, para avaliar a qualidade da previsão SAR é necessário compará-la a outros modelos de previsão, para avaliar a adequabilidade de seus resultados. Para tal elegeram-se, com base nos testes de hipótese e nos resultados de erro acumulado total e MAPE, os seguintes experimentos para comparação com outros previsores:

- UE 624 e 638 – forma funcional PSAR, com método de determinação *Indet full* e entrada de dados opção (C), pois apresentam diferenças significativas associada aos menores erros acumulado total e MAPE;
- UE 207 – forma funcional PSAR, com método de determinação *Indet MC* e entrada de dados opção (B), pois apresentam respectivamente diferenças

significativas contra a opção (C) de entrada de dados associado aos menores erros acumulado total e MAPE;

- UE 258 – os desvios não apresentam diferenças significativas, sendo assim optou-se pela forma funcional PSAR, com método de determinação *Indet MC* e entrada de dados opção (C), pois apresenta o menor erro acumulado total; e,
- UE 247 – individualmente definiu-se a opção (C) de entrada de dados como a de melhor desempenho. Nesta opção os testes de hipótese não indicaram diferenças significativas entre os desvios dos experimentos, sendo assim, para homogeneizar a comparação com outros previsores, adotou-se a forma funcional PSAR com método de determinação *Indet full*.

#### 4.4.3 – TESTES DE ESTABILIDADE DO MODELO

O teste de estabilidade seguiu o mesmo processo de previsão, empregando os mesmos parâmetros estimados, somente com a substituição, para determinação de  $pWy$  dos valores dos  $y$  originais da massa de teste, pelos  $y$  degradados.

Sendo assim, a degradação gera uma massa de  $y$  desconhecidos, diferente da massa de teste original e desta forma cumpre o papel de fornecer ao modelo um conjunto de dados que, por exemplo, poderia ser fruto da previsão individual de cada UE vizinha.

As figuras 4.41 a 4.45 apresentam os gráficos dos erros relativos mensais a partir dos dados degradados, respectivamente para UE 624, 638, 207, 247 e 258.

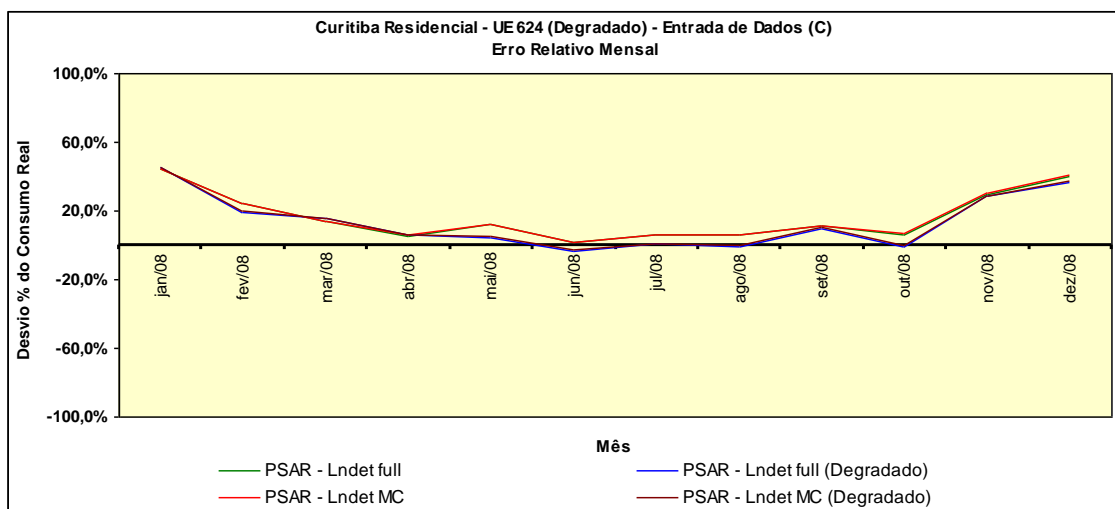


Figura 4.41 – Erro Relativo Mensal SAR UE 624 – Valores Originais x Degradados.

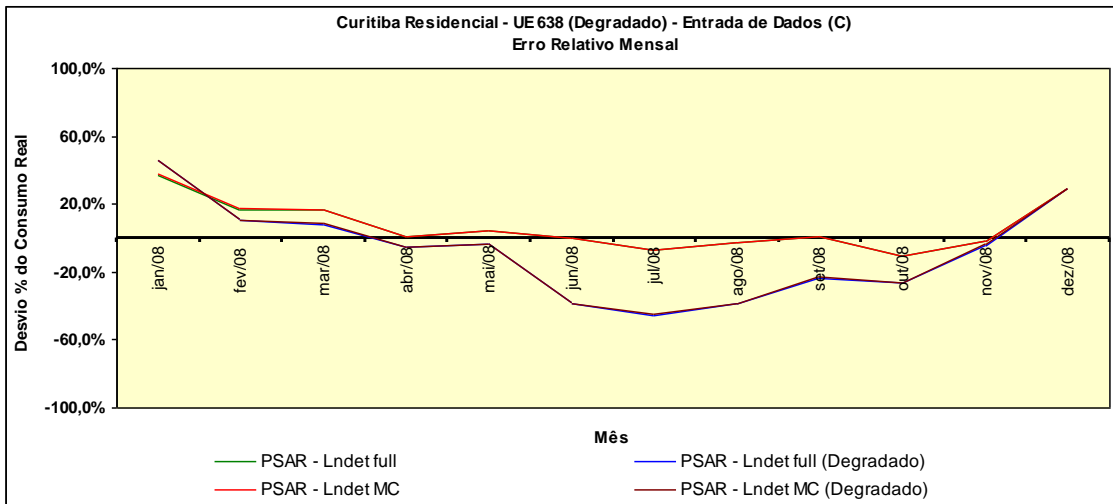


Figura 4.42 – Erro Relativo Mensal SAR UE 638 – Valores Originais x Degradados.

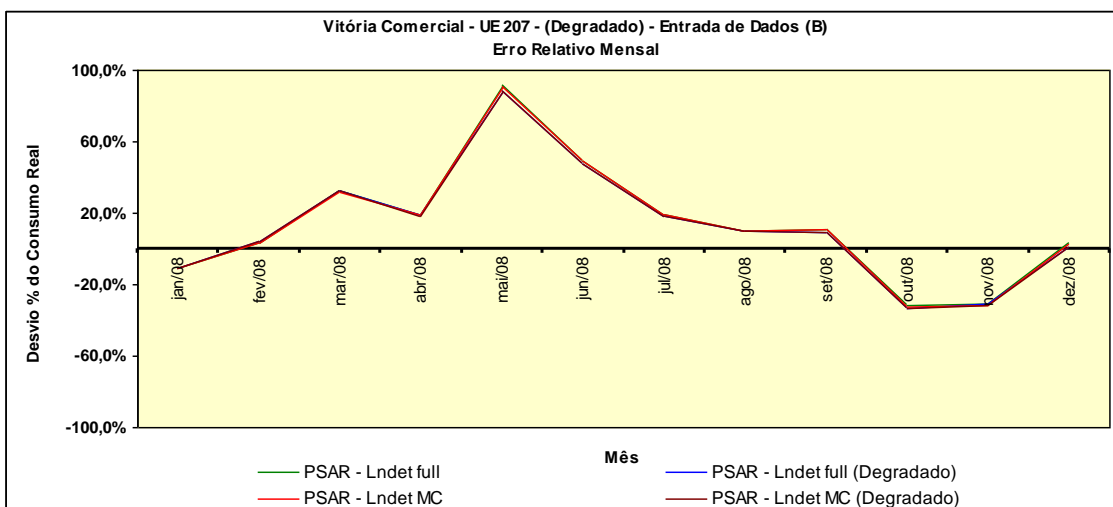


Figura 4.43 – Erro Relativo Mensal SAR UE 207 – Valores Originais x Degradados.

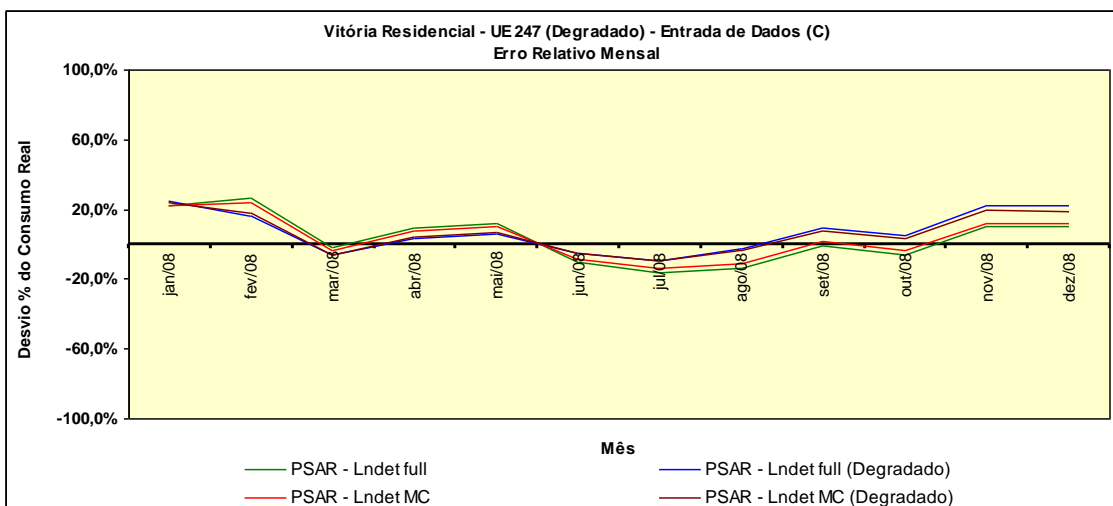


Figura 4.44 – Erro Relativo Mensal SAR UE 247 – Valores Originais x Degradados.



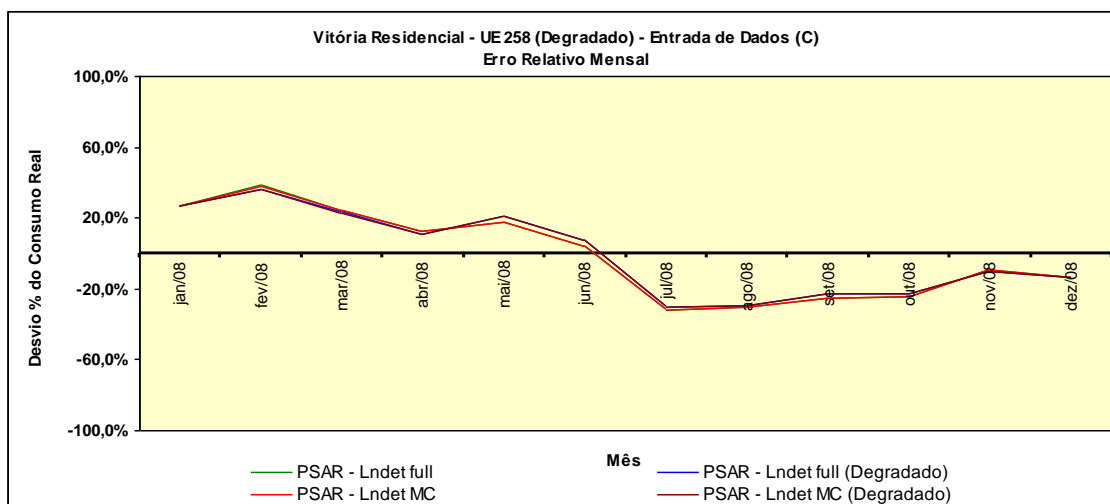


Figura 4.45 – Erro Relativo Mensal SAR UE 258 – Valores Originais x Degradados.

Em complemento aos gráficos da previsão degradada, na tabela 4.42 apresenta-se a comparação entre os resultados do erro acumulado total e do MAPE, entre os experimentos originais e os degradados.

Tabela 4.42 – Comparativo dos Resultados de Degradação.

| Análise Degradação          |                      |                    |                     |         |                    |                     |         |
|-----------------------------|----------------------|--------------------|---------------------|---------|--------------------|---------------------|---------|
| Experimento                 |                      | PSAR - Lndet full  |                     |         | PSAR - Lndet MC    |                     |         |
|                             |                      | Resultado Original | Resultado degradado |         | Resultado Original | Resultado degradado |         |
| 624<br>Entrada de dados (C) | Erro Acumulado Total | 15,07%             | Teste 1             | 13,39%  | 15,33%             | Teste 1             | 13,76%  |
|                             |                      |                    | Teste 2             | 11,47%  |                    | Teste 2             | 11,97%  |
| 624<br>Entrada de dados (C) | MAPE                 | 16,77%             | Teste 1             | 16,40%  | 17,04%             | Teste 1             | 16,66%  |
|                             |                      |                    | Teste 2             | 14,24%  |                    | Teste 2             | 14,44%  |
| 638<br>Entrada de dados (C) | Erro Acumulado Total | 4,19%              | Teste 1             | -1,42%  | 4,25%              | Teste 1             | -1,34%  |
|                             |                      |                    | Teste 2             | -12,71% |                    | Teste 2             | -12,58% |
| 638<br>Entrada de dados (C) | MAPE                 | 10,64%             | Teste 1             | 23,51%  | 10,67%             | Teste 1             | 23,48%  |
|                             |                      |                    | Teste 2             | 23,23%  |                    | Teste 2             | 23,19%  |
| 207<br>Entrada de dados (B) | Erro Acumulado Total | 4,73%              | Teste 1             | 4,96%   | 4,49%              | Teste 1             | 4,69%   |
|                             |                      |                    | Teste 2             | 3,98%   |                    | Teste 2             | 3,82%   |
| 207<br>Entrada de dados (B) | MAPE                 | 26,14%             | Teste 1             | 26,33%  | 25,99%             | Teste 1             | 26,16%  |
|                             |                      |                    | Teste 2             | 25,47%  |                    | Teste 2             | 25,40%  |
| 247<br>Entrada de dados (C) | Erro Acumulado Total | 1,89%              | Teste 1             | 1,56%   | 2,81%              | Teste 1             | 2,61%   |
|                             |                      |                    | Teste 2             | 6,57%   |                    | Teste 2             | 5,71%   |
| 247<br>Entrada de dados (C) | MAPE                 | 11,59%             | Teste 1             | 9,62%   | 10,87%             | Teste 1             | 9,39%   |
|                             |                      |                    | Teste 2             | 11,06%  |                    | Teste 2             | 10,59%  |
| 258<br>Entrada de dados (C) | Erro Acumulado Total | -7,13%             | Teste 1             | -7,42%  | -7,13%             | Teste 1             | -7,40%  |
|                             |                      |                    | Teste 2             | -6,15%  |                    | Teste 2             | -6,18%  |
| 258<br>Entrada de dados (C) | MAPE                 | 21,64%             | Teste 1             | 19,27%  | 21,60%             | Teste 1             | 19,32%  |
|                             |                      |                    | Teste 2             | 21,00%  |                    | Teste 2             | 20,98%  |

Aplicando os testes de hipótese tem-se que trinta e dois testes (80%) de um total de quarenta apresentam o mesmo nível de significância, ou seja, os modelos com dados degradados apresentaram comportamento similar aos originais. Os oito experimentos degradados que não apresentaram similaridade pertencem à UE 638.

#### 4.4.3.1 – CONCLUSÕES COM BASE NOS TESTE DE ESTABILIDADE

##### a) Estrutura SAR

Comparando-se os resultados originais e os degradados, de erro acumulado total e MAPE, verifica-se que os modelos apresentam estabilidade, ou seja, mesmo degradando a influência da vizinhança, o emprego dos parâmetros estimados gerou previsões confiáveis nos novos experimentos.

A determinação da dependência espacial permitiu definir um conjunto de UE, no qual o processo de estimação apresentou capacidade de averiguar e encapsular, a influência espacial dos vizinhos na determinação de  $p$ . Sendo assim, conclui-se que o modelo é robusto, uma vez que independente da estrutura de previsão que possa ser empregada nos vizinhos, existe a garantia de que sua influência está corretamente ponderada em  $pWy$ .

Assim como no caso dos experimentos originais, os previsores degradados apresentaram melhor capacidade de prever os valores totais do período (o consumo total de GN no período) do que as parcelas, neste caso a distribuição ao longo do tempo.

#### 4.5 – COMPARAÇÃO ENTRE ESTRUTURAS DE PREVISÃO

Para efeitos de comparação, os resultados obtidos a partir do modelo SAR *Panel* foram inicialmente comparados com previsores lineares de estrutura ARIMA. Nas figuras 4.46 a 4.50, são apresentados os gráficos dos erros relativos mensais, respectivamente para UE 624, 638, 207, 247 e 258.

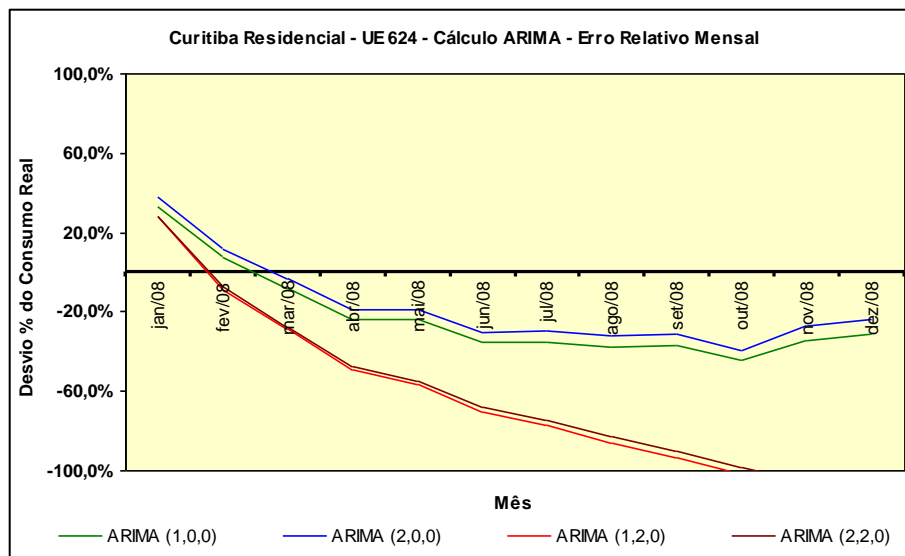


Figura 4.46 – Previsão ARIMA UE 624 – Erro Relativo Mensal.

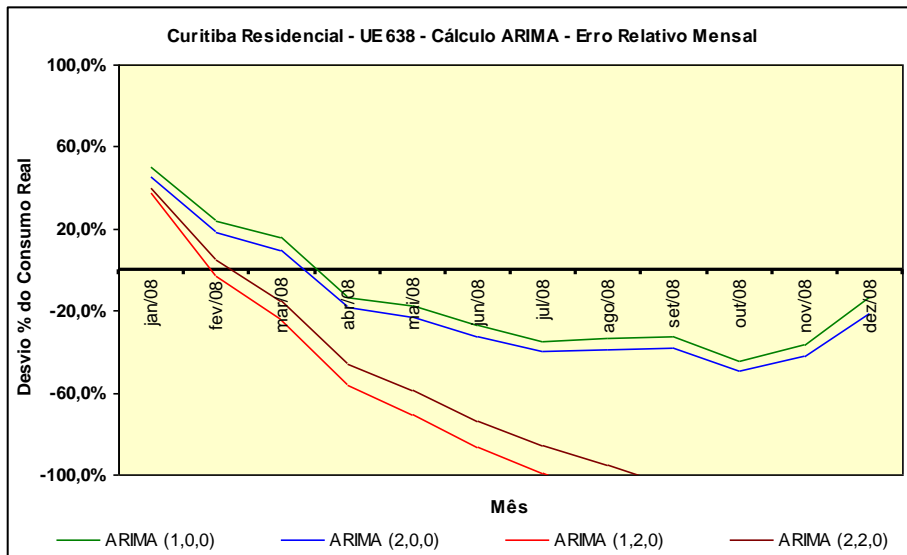


Figura 4.47 – Previsão ARIMA UE 638 – Erro Relativo Mensal.

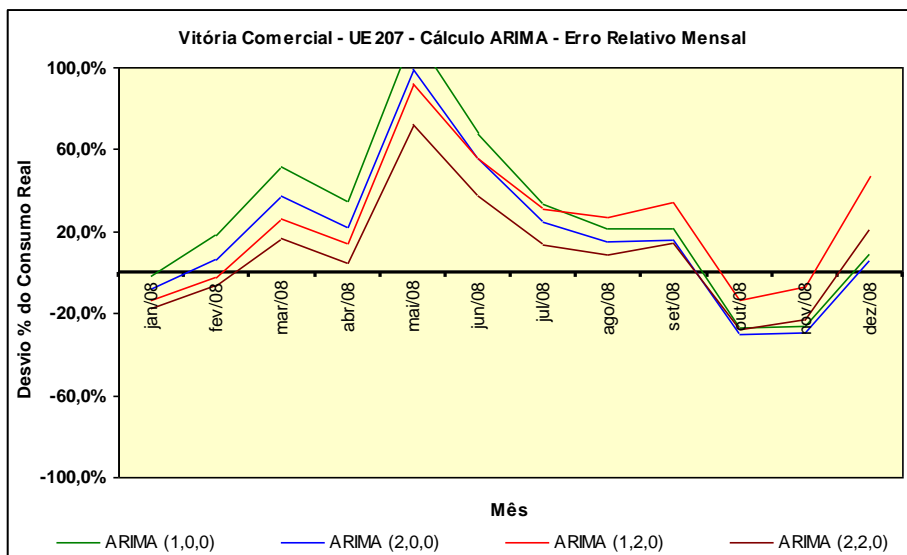


Figura 4.48 – Previsão ARIMA UE 207 – Erro Relativo Mensal.

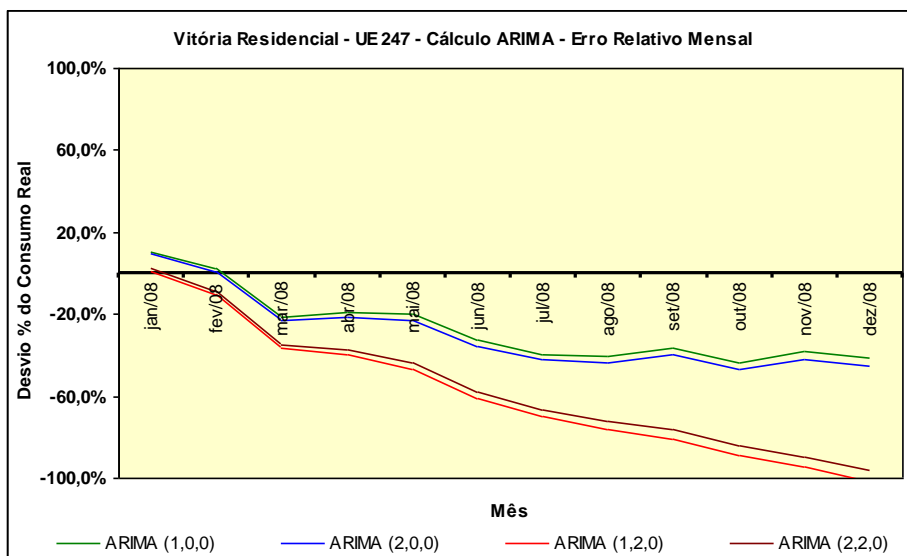


Figura 4.49 – Previsão ARIMA UE 247 – Erro Relativo Mensal.

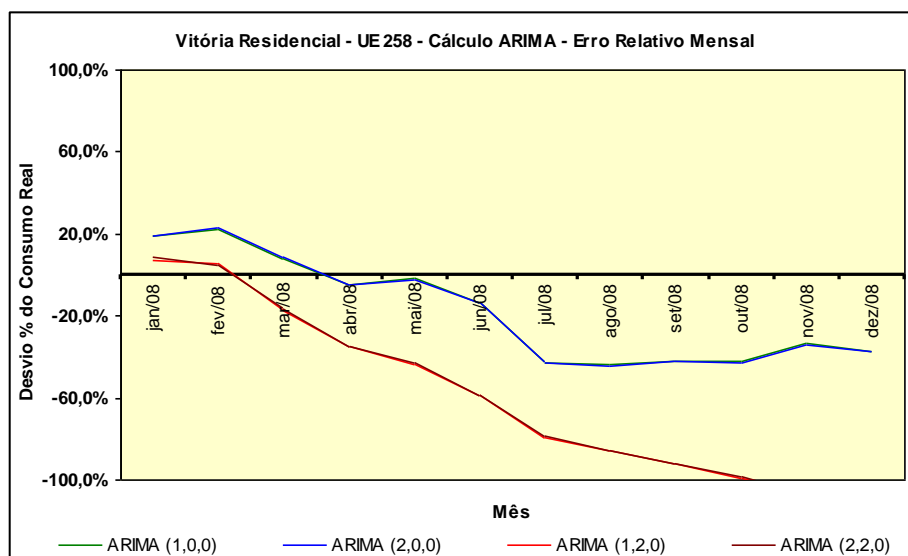


Figura 4.50 – Previsão ARIMA UE 258 – Erro Relativo Mensal.

Em complemento aos gráficos do erro relativo mensal, na tabela 4.43 são apresentados os valores percentuais do erro acumulado total e do MAPE, no qual os melhores resultados estão grifados em tons de amarelo. Todos os gráficos relativos à previsão ARIMA estão disponíveis no Apêndice I.

Tabela 4.43 – ARIMA - Erro Acumulado Total e MAPE por UE.

| ARIMA                        | UE 624               |        | UE 638               |        | UE 207               |        | UE 247               |        | UE 258               |        |
|------------------------------|----------------------|--------|----------------------|--------|----------------------|--------|----------------------|--------|----------------------|--------|
|                              | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   |
| ARIMA (1,0,0) Sazonal lag 12 | -26,02%              | 29,19% | -19,24%              | 28,50% | 15,81%               | 35,82% | -29,10%              | 28,73% | -24,00%              | 25,90% |
| ARIMA (2,0,0) Sazonal lag 12 | -20,39%              | 25,18% | -24,82%              | 31,29% | 8,45%                | 29,07% | -31,90%              | 31,07% | -24,25%              | 26,18% |
| ARIMA (1,2,0) Sazonal lag 12 | -70,09%              | 69,12% | -93,42%              | 90,65% | 17,55%               | 30,36% | -63,06%              | 58,95% | -69,73%              | 62,13% |
| ARIMA (2,2,0) Sazonal lag 12 | -67,50%              | 66,66% | -77,64%              | 76,74% | 2,97%                | 22,02% | -59,60%              | 55,96% | -69,34%              | 61,86% |

#### 4.5.1 – COMPARAÇÃO E CONCLUSÕES SOBRE OS RESULTADOS SAR E ARIMA

No caso do modelo ARIMA pode se constatar, observando a tabela 4.43, que os melhores ajustamentos variaram em função da forma funcional adotada. Três dos cinco experimentos tiveram melhor desempenho em um modelo ARIMA com estimação de constante, atraso de ordem  $k = 1$ , sem transformação (ARIMA 1, 0, 0).

Os outros dois obtiveram os melhores resultados com estimação de constante, atraso de ordem  $k = 2$ , sendo um sem transformação de variáveis (ARIMA 2, 0, 0) e com transformação de variáveis por diferença com atraso 1 e dois passos (ARIMA 2, 2, 0).

Em valores absolutos, as previsões com base no modelo SAR *Panel* foram superiores em quatro casos: 624, 638, 247 e 258. Conforme tabela 4.44, excetuando a UE 207, a previsão empregando a estrutura SAR apresentou melhores resultados, tanto em termos de erro acumulado total quanto em MAPE.

Tabela 4.44 – ARIMA x SAR - Erro Acumulado Total e MAPE por UE.

| Modelo | UE 624               |        | UE 638               |        | UE 207               |        | UE 247               |        | UE 258               |        |
|--------|----------------------|--------|----------------------|--------|----------------------|--------|----------------------|--------|----------------------|--------|
|        | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   |
| ARIMA  | -20,39%              | 25,18% | -19,24%              | 28,50% | 2,97%                | 22,02% | -29,10%              | 28,73% | -24,00%              | 25,90% |
| SAR    | 15,07%               | 16,77% | 4,19%                | 10,64% | 4,49%                | 25,99% | 1,89%                | 11,59% | -7,13%               | 21,60% |

Com base nos testes de hipótese, existem diferenças significativas entre os desvios das previsões no caso das UE 624, 638 e 247, como se observa na tabela 4.45.

Tabela 4.45 – Teste de Hipótese – Comparações entre Previsores ARIMA x SAR *Panel*.

| UE  | Experimento A | Experimento B                | $T_0^*$ | Graus de Liberdade | t crítico $\alpha = 0.05$ | Hipótese   |
|-----|---------------|------------------------------|---------|--------------------|---------------------------|------------|
| 624 | ARIMA (2,0,0) | PSAR - Lndet full - Ent. (C) | 2,192   | 23,00              | 2,069                     | FALSO      |
| 638 | ARIMA (1,0,0) | PSAR - Lndet full - Ent. (C) | 4,023   | 18,00              | 2,101                     | FALSO      |
| 207 | ARIMA (2,2,0) | PSAR - Lndet MC - Ent. (B)   | -0,451  | 21,00              | 2,080                     | VERDADEIRO |
| 247 | ARIMA (1,0,0) | PSAR - Lndet full - Ent. (C) | 3,789   | 14,00              | 2,145                     | FALSO      |
| 258 | ARIMA (1,0,0) | PSAR - Lndet MC - Ent. (C)   | 1,027   | 17,00              | 2,110                     | VERDADEIRO |

Outra ressalva, na comparação entre o predictor ARIMA e o SAR é que no primeiro, as UE 624, 638, 247 e 258, apresentaram erro acumulado total entre 20 e 30 % negativos e, MAPE entre 25 e 30%. Com isto, conclui-se que estes experimentos, além de apresentarem resultados inferiores, podem conduzir a um erro de contratação de GN por parte da CDL, bem como um desvio maior na previsão individual.

No Apêndice J, estão disponíveis todos os gráficos de previsão e erros, empregados nas comparações entre as previsões ARIMA e SAR.

#### 4.5.2 – SIMULAÇÃO COM PREVISOR POR ALISAMENTO EXPONENCIAL

No tocante à simulação com emprego de alisamento exponencial têm-se nas figuras 4.51 a 4.55 os gráficos do erro relativo mensal para as UE de interesse.

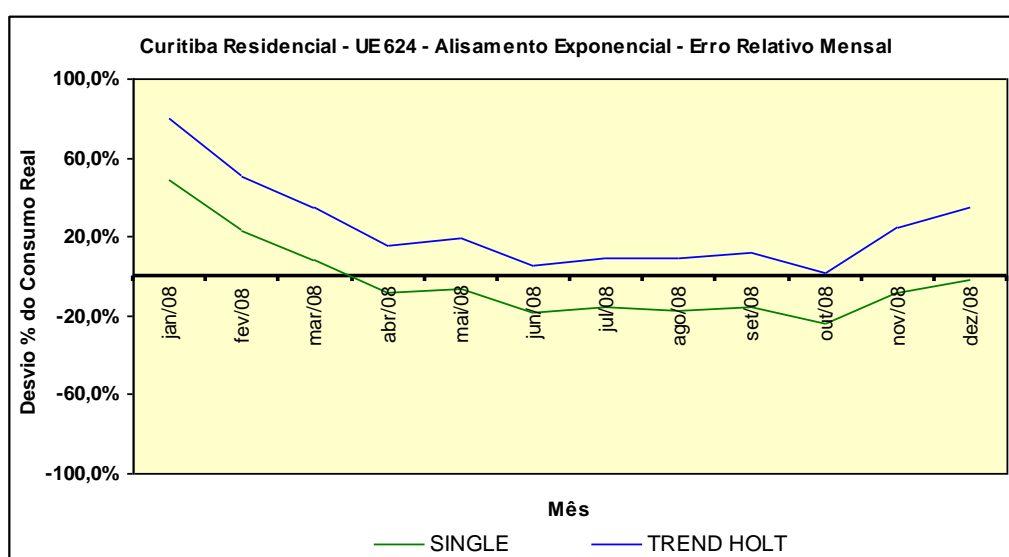


Figura 4.51 – Previsão Alisamento Exponencial - UE 624 – Erro Relativo Mensal.

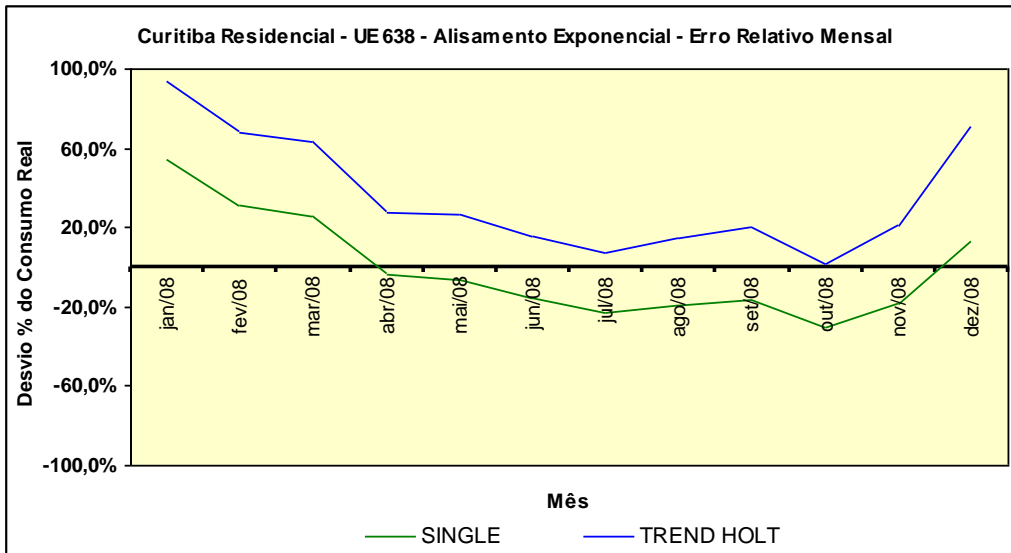


Figura 4.52 – Previsão Alisamento Exponencial - UE 638 – Erro Relativo Mensal.

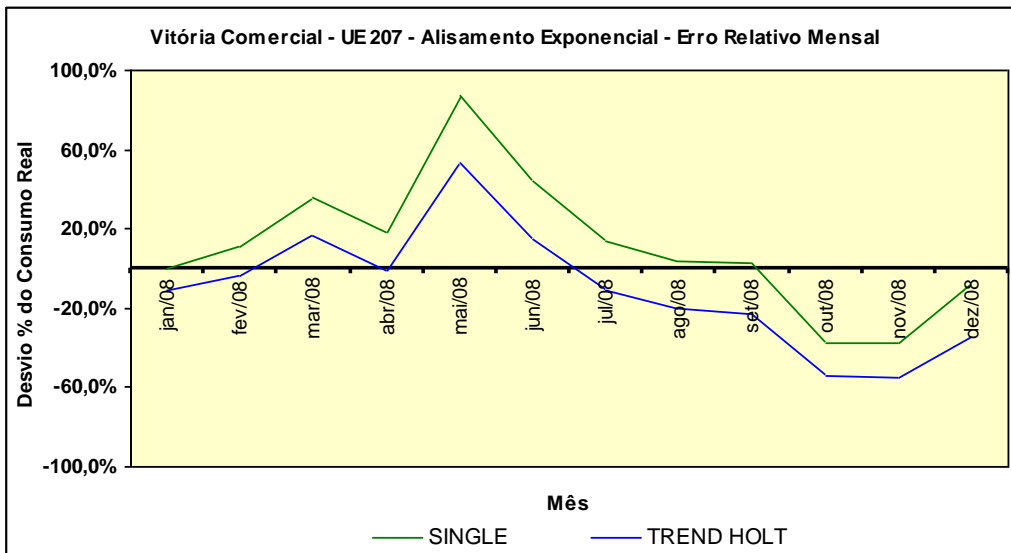


Figura 4.53 – Previsão Alisamento Exponencial - UE 207 – Erro Relativo Mensal.

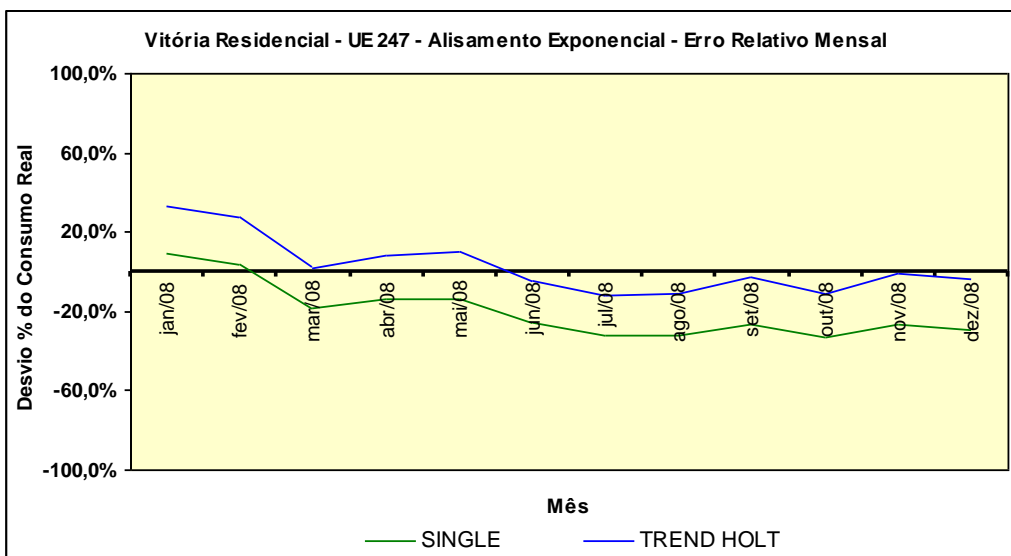


Figura 4.54 – Previsão Alisamento Exponencial - UE 247 – Erro Relativo Mensal.

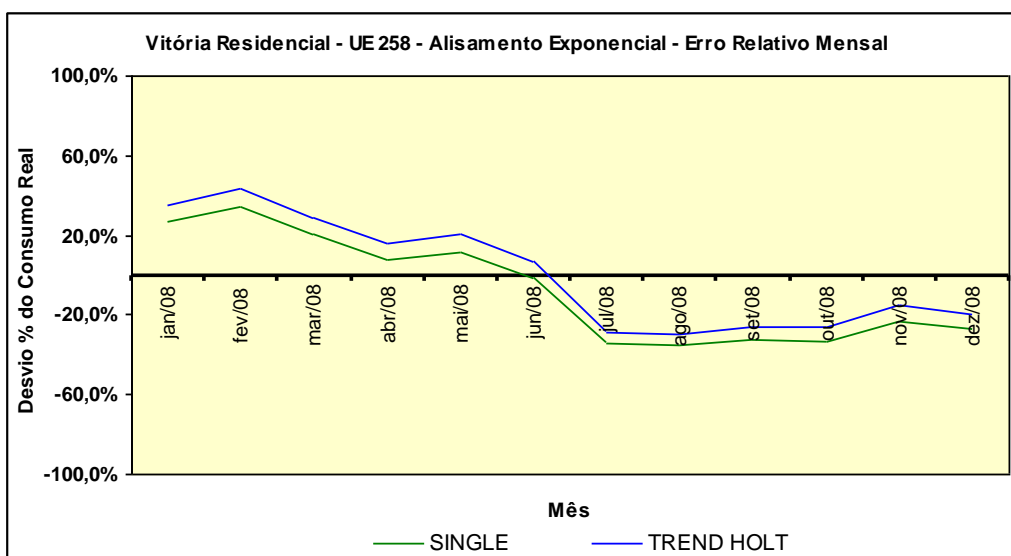


Figura 4.55 – Previsão Alisamento Exponencial - UE 258 – Erro Relativo Mensal.

Em complemento, na tabela 4.46, são apresentados os valores percentuais do erro acumulado total e do MAPE. Todos os gráficos relativos à previsão por alisamento exponencial estão disponíveis no Apêndice K.

Tabela 4.46 – Alisamento Exponencial - Erro Acumulado Total e MAPE por UGR.

| ALISAMENTO EXPONENCIAL | UE 624               |        | UE 638               |        | UE 207               |        | UE 247               |        | UE 258               |        |
|------------------------|----------------------|--------|----------------------|--------|----------------------|--------|----------------------|--------|----------------------|--------|
|                        | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   |
| SINGLE                 | -6,12%               | 16,27% | -6,11%               | 21,41% | 1,65%                | 24,92% | -22,04%              | 22,16% | -14,00%              | 23,95% |
| TREND HOLT             | 21,27%               | 24,76% | 29,99%               | 35,99% | -19,47%              | 24,93% | 0,80%                | 10,55% | -6,48%               | 24,65% |

#### 4.5.2.1 - COMPARAÇÃO ENTRE SAR E ALISAMENTO EXPONENCIAL

Observando a tabela 4.46, percebe-se que os melhores ajustamentos variaram em função da forma funcional empregada no modelo de alisamento exponencial. Fato este similar ao observado com a estrutura ARIMA.

Dos cinco experimentos, em três o melhor resultado se deu com alisamento simples sem componentes de tendência e sazonal, um com alisamento simples com componente de tendência linear e sem componente sazonal e, na UE 258, houve uma divisão. Nesta UE, pela ótica do erro acumulado total a melhor resposta foi com o alisamento simples e, pela ótica do MAPE isto acontece com a forma *Trend Holt*.

Na tabela 4.47, apresenta-se o comparativo entre os resultados obtidos com a previsão empregando alisamento exponencial e modelo espaço-temporal SAR *Panel*.

Tabela 4.47 – Alisamento x SAR - Erro Acumulado Total e MAPE por UGR.

| Modelo                | UE 624               |        | UE 638               |        | UE 207               |        | UE 247               |        | UE 258               |        |
|-----------------------|----------------------|--------|----------------------|--------|----------------------|--------|----------------------|--------|----------------------|--------|
|                       | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   |
| Alisamento SINGLE     | -6,12%               | 16,27% | -6,11%               | 21,41% | 1,65%                | 24,92% | -22,04%              | 22,16% | -14,00%              | 23,95% |
| Alisamento TREND HOLT | 21,27%               | 24,76% | 29,99%               | 35,99% | -19,47%              | 24,93% | 0,80%                | 10,55% | -6,48%               | 24,65% |
| SAR                   | 15,07%               | 16,77% | 4,19%                | 10,64% | 4,49%                | 25,99% | 1,89%                | 11,59% | -7,13%               | 21,60% |

De forma diferente aos resultados ARIMA, nestes cinco experimentos não houve diferença tão acentuada, por exemplo:

- UE 624 – o melhor resultado foi no alisamento exponencial, com diferença absoluta de 9% no erro acumulado total, enquanto no MAPE foi de 0,5%;
- UE 638 – o melhor resultado foi o SAR *Panel* com diferença absoluta de 2% no erro acumulado total, todavia a diferença do MAPE foi alta, cerca de 11%;
- UE 207 – o melhor resultado foi no alisamento exponencial, com diferença absoluta de 3% no erro acumulado total, enquanto no MAPE foi de 1%;
- UE 247 – o melhor resultado foi com alisamento exponencial apresentando diferença absoluta do erro acumulado total e do MAPE de 1%; e,
- UE 258 – considerando o melhor MAPE obtido com alisamento exponencial, 23,95%, o melhor ajustamento se deu com o modelo SAR (-7,13% contra -14,0% e 21,6% contra 23,95%). Entretanto, considerado o melhor erro acumulado total o melhor resultado foi com alisamento exponencial simples com componente de tendência linear, 6,48% contra 7,13%.

Tal situação é corroborada nos testes de hipótese, onde não se identificou diferenças significativas entre as previsões, excetuando a UE 638, como se observa na tabela 4.48.

Tabela 4.48 – Teste de Hipótese – Comparação de Previsores - Alisamento x SAR *Panel*.

| UE  | Experimento A         | Experimento B                | $T_0^*$ | Graus de Liberdade | t crítico $\alpha = 0.05$ | Hipótese   |
|-----|-----------------------|------------------------------|---------|--------------------|---------------------------|------------|
| 624 | Alisamento SINGLE     | PSAR - Lndet full - Ent. (C) | 0,083   | 22,00              | 2,074                     | VERDADEIRO |
| 638 | Alisamento SINGLE     | PSAR - Lndet full - Ent. (C) | 2,973   | 22,00              | 2,074                     | FALSO      |
| 207 | Alisamento SINGLE     | PSAR - Lndet MC - Ent. (B)   | -0,026  | 23,00              | 2,069                     | VERDADEIRO |
| 247 | Alisamento TREND HOLT | PSAR - Lndet full - Ent. (C) | -0,520  | 22,00              | 2,074                     | VERDADEIRO |
| 258 | Alisamento SINGLE     | PSAR - Lndet MC - Ent. (C)   | 0,635   | 22,00              | 2,074                     | VERDADEIRO |
|     | Alisamento TREND HOLT | PSAR - Lndet MC - Ent. (C)   | 0,543   | 23,00              | 2,069                     | VERDADEIRO |

Resumidamente, no tocante ao erro acumulado total os experimentos a partir do predictor por alisamento exponencial apresentaram certa vantagem. Já para o MAPE, com base nos desvios individuais, nota-se pela tabela 4.47, que somente o experimento da UGR 638 apresenta diferença significativa.

Apesar disto, a observação dos gráficos das figuras 4.56 a 4.58, nos quais apresenta-se para as UE 624, 638 e 247, a série de valores previstos e as de teste, para os dois modelos, permitem tecer comentários adicionais. Nota-se que mesmo com o menor erro acumulado total, a previsão SAR *Panel* teve maior capacidade de corrigir as previsões para o centro da meta; ou seja, este modelo apresentou melhor sensibilidade aos dados, visto que a previsão por alisamento tem o perfil de uma reta.



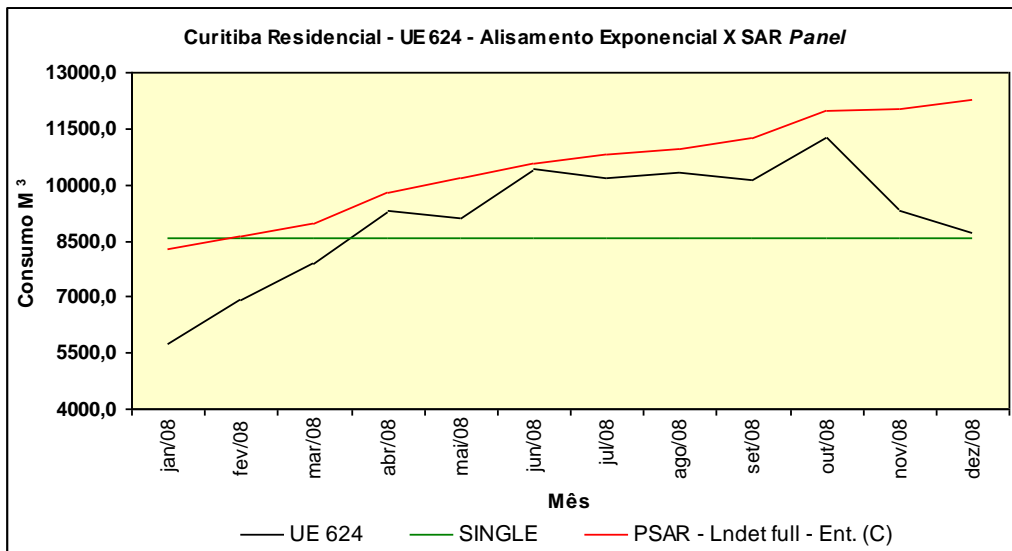


Figura 4.56 – Alisamento Exponencial x SAR - UE 624 – Comparativo da Previsão.

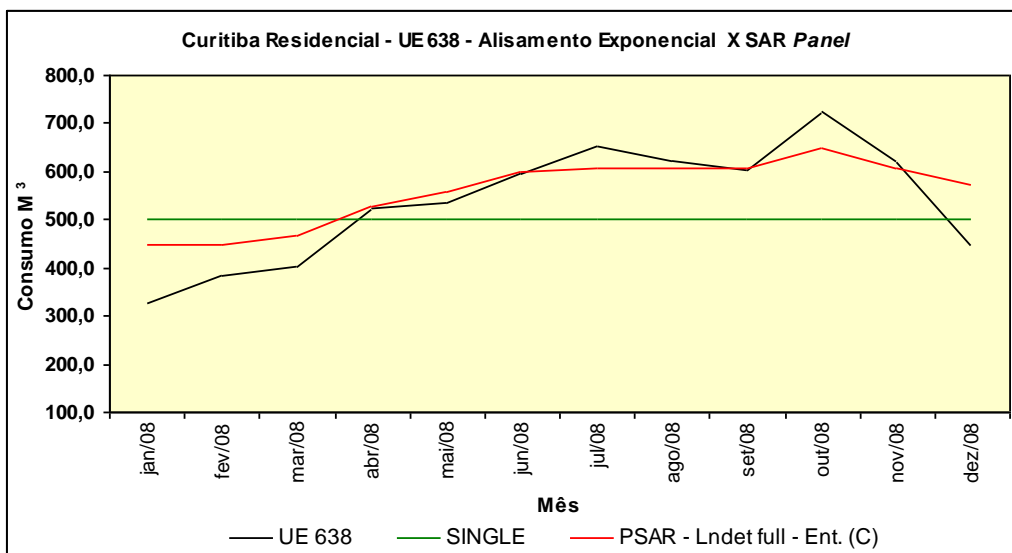


Figura 4.57 – Alisamento Exponencial x SAR - UE 638 – Comparativo da Previsão.

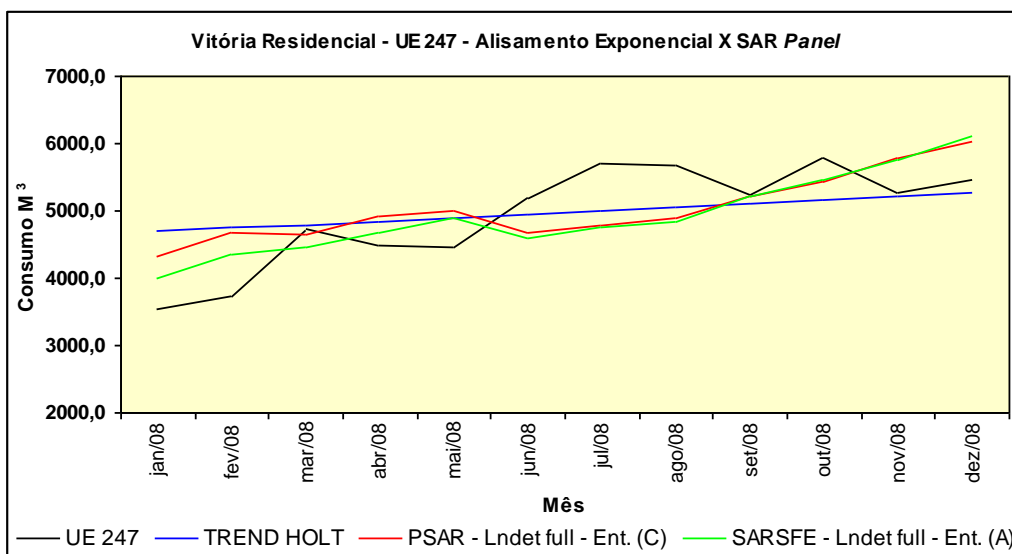


Figura 4.58 – Alisamento Exponencial x SAR - UE 247 – Comparativo da Previsão.

Ressalta-se que, conforme a tabela 4.35, na UE 247 o modelo SAR *Panel* apresentou um resultado com erro acumulado de -0,28% e MAPE de 10,08%, ou seja, melhor que os resultados com alisamento exponencial; bem como o seu erro acumulado manteve-se menor do que o do alisamento até novembro, conforme a figura 4.59.

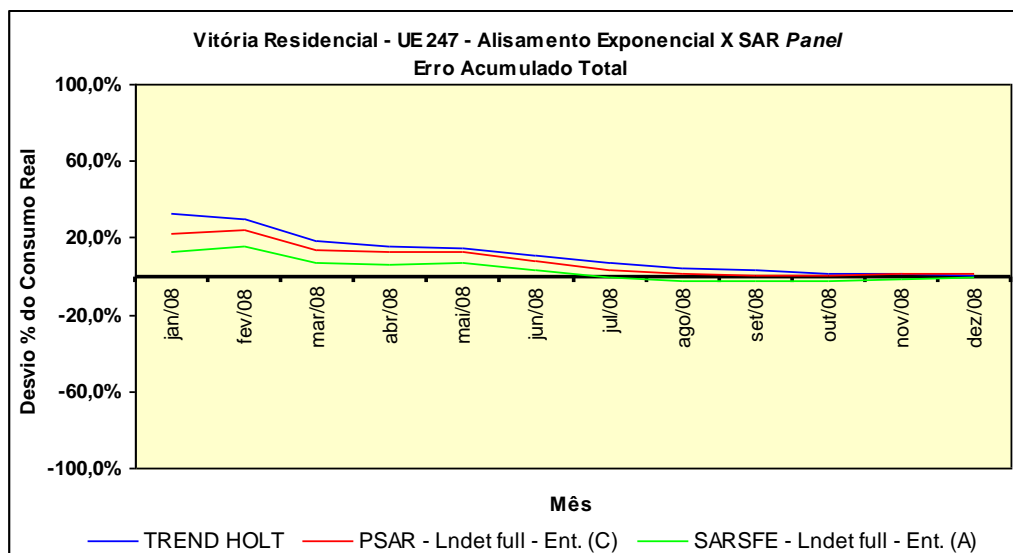


Figura 4.59 – Alisamento Exponencial x SAR - UE 247 – Erro Acumulado Total.

Sobre a estrutura SAR não apresentar boa sensibilidade frente a mudanças abruptas de patamares, como no caso das UE 207, 258 e 624, é natural que a previsão por alisamento exponencial apresente melhor resultado em eventos extremos, pois sua determinação busca posicionar a previsão de forma central a estes eventos.

Isto é visível na figura 4.60, onde o desvio absoluto mensal da previsão SAR *Panel*, na UE 624 apresentou melhor desempenho até outubro de 2008; porém, os dois últimos eventos da série foram suficientes para comprometer o resultado da previsão.

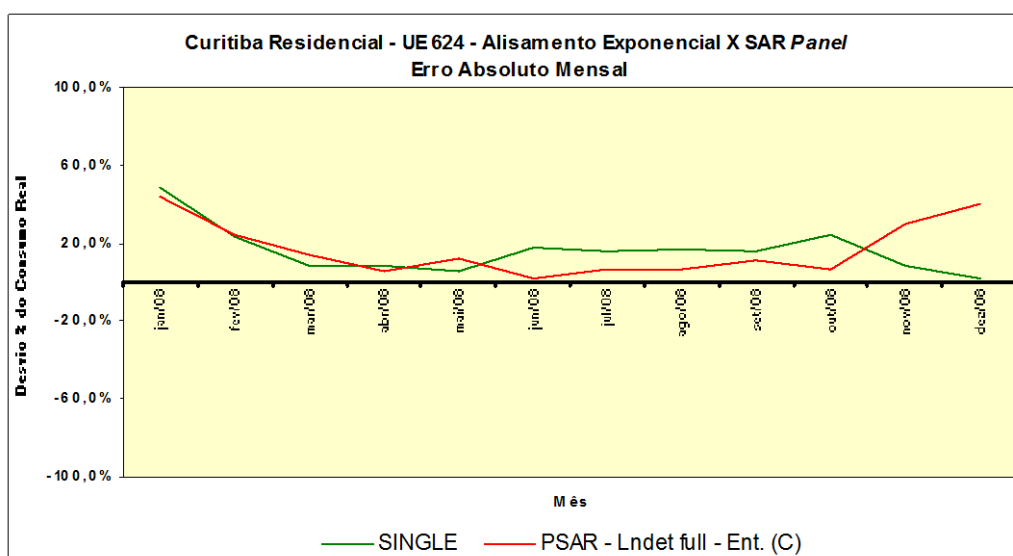


Figura 4.60 – Alisamento Exponencial x SAR - UE 624 – Erro Absoluto Mensal.

Outra desvantagem no previsor por alisamento e que, assim como no ARIMA, as UE 624, 638 e 258 apresentaram erros acumulados totais negativos; ou seja, sua aplicação pode conduzir a uma contratação deficitária de GN por parte da CDL.

Os gráficos das comparações entre as previsões SAR *Panel* e Alisamento Exponencial estão disponíveis no Apêndice J.

#### 4.5.3 – SIMULAÇÃO DO PREVISOR NÃO-LINEAR POR RN

Conforme a metodologia descrita em 3.5.2 a UE 624 foi escolhida para o mapeamento inicial via RN por ter sido a que o previsor SAR *Panel* apresentou os piores resultados, para a qual obtiveram-se os seguintes resultados no treinamento:

- Método TRAINLM – convergiu com os valores padrões de treinamento, apresentando instabilidade (flutuação nas épocas de treino 100 – 1000 épocas);
- Método TRAINBR – convergiu em até 600 épocas apresentando altos valores de desvio (na casa dos décimos). Adicionalmente alteraram-se as seguintes características no experimento, na tentativa de aperfeiçoar o treinamento:
  - Restringiu-se a tolerância de falhas (de 5 p/ 2);
  - Os valores de inicialização do parâmetro  $\mu$  (de 0,005 para 0,001);
  - Os valores para os incrementos dos passos de atualização de  $\mu$  (descida de 0,1 para 0,01 e subida de 10 para 5);
  - A redução da velocidade de treinamento (de 1 para 2); e,
- Método TRAINGDM e TRAINGDX - não convergiram com até 100000 épocas de treinamento, mesmo com aumento da taxa de aprendizagem (0,01 até 0,025).

Para mitigar a ocorrência de mínimos locais de baixa qualidade, repetiram-se os experimentos alterando a função de adaptação GD para GDM, obtendo-se os seguintes resultados:

- Método TRAINLM – mesmo comportamento sem o gradiente por momento;
- Método TRAINBR – convergiu em até 600 épocas apresentando altos valores de desvio (na casa dos décimos). Assim como no caso anterior foram alteradas as seguintes características para aperfeiçoar o treinamento:
  - Restringiu-se a tolerância de falhas (de 5 p/ 2);
  - Os valores de inicialização do parâmetro  $\mu$  (de 0,005 p/ 0,001);
  - Os valores para os incrementos dos passos de atualização de  $\mu$  (descida de 0,1 para 0,01 e subida de 10 para 5);
  - A redução da velocidade de treinamento; e,

- Método TRAINGDM e TRAINGDX – não convergiram com até 200000 épocas de treinamento, mesmo com aumento da taxa de aprendizagem (0,01 até 0,05).

Percebeu-se que os métodos TRAINGDX e TRAINGDM foram mais sensíveis as alterações das configurações no treino, principalmente na taxa de aprendizagem; onde a sua alteração ensejou maiores respostas na redução dos desvios do treinamento.

Nas figuras 4.61 a 4.64 a seguir, são apresentados os gráficos das previsões mensais, para a UE 624, conforme os métodos de treinamento adotados.

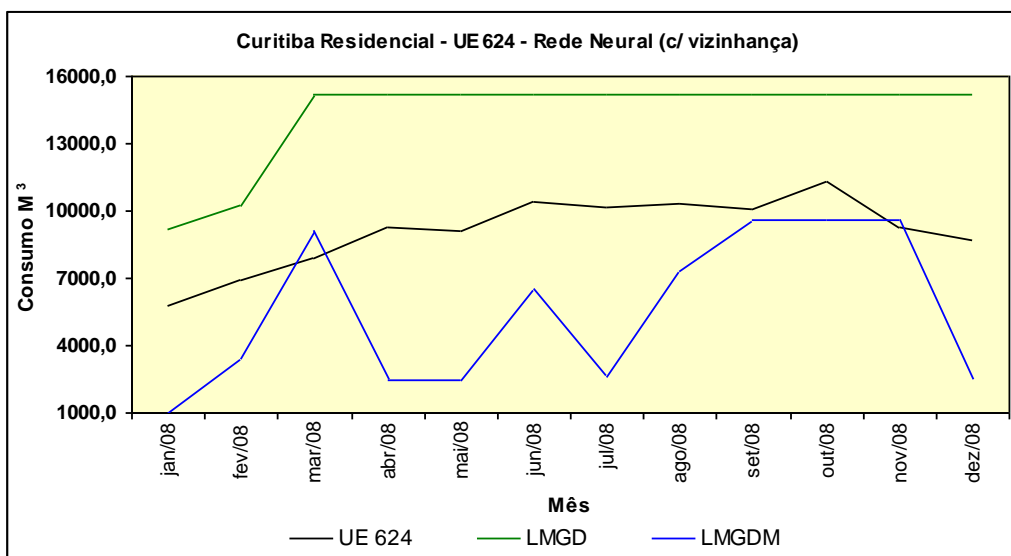


Figura 4.61 – Previsão RN - UE 624 – Método de Treinamento LM GD / GDM.

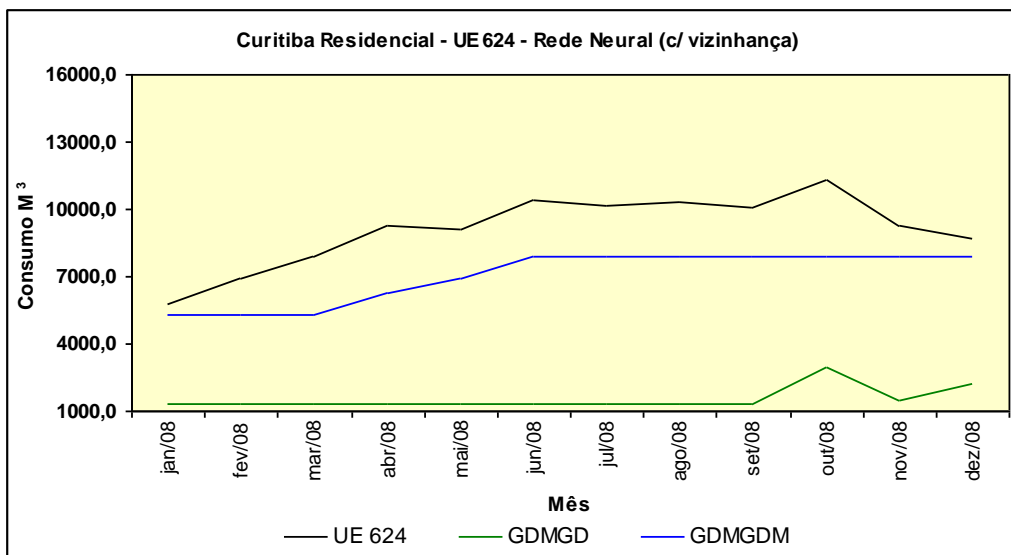


Figura 4.62 – Previsão RN - UE 624 – Método de Treinamento GDM GD / GDM.

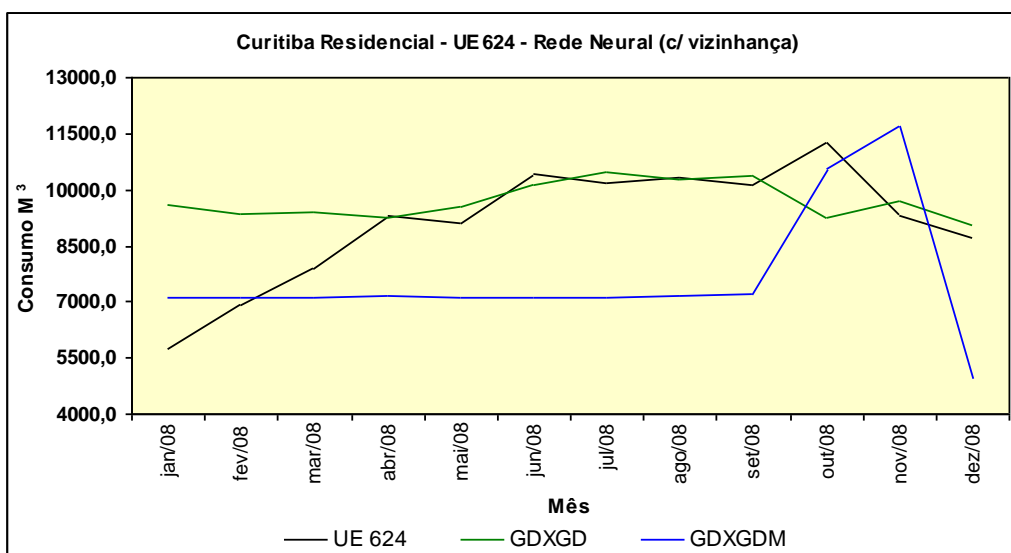


Figura 4.63 – Previsão RN - UE 624 – Método de Treinamento GDX GD / GDM.

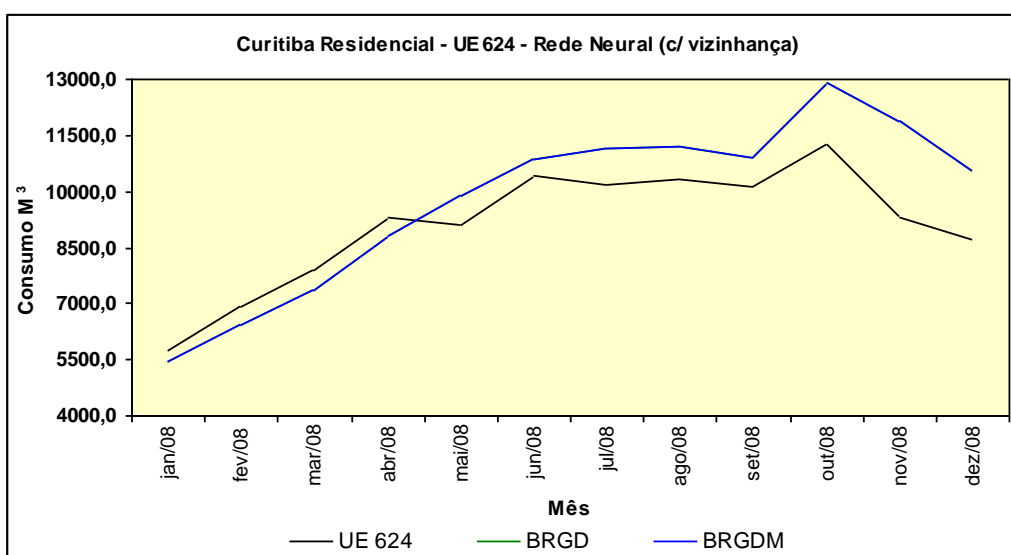


Figura 4.64 – Previsão RN - UE 624 – Método de Treinamento BR GD / GDM.

Comparando o erro acumulado total e o MAPE da previsão RN com a SAR *Panel* na UE 624, conforme a tabela 4.49, nota-se que o método TRAINGDX, com função de adaptação GD e os dois treinamentos BR (grifados em amarelo) apresentaram previsões equivalentes ao melhor experimento com o SAR nesta UE.

Tabela 4.49 – RN x SAR – UE 624 - Erro Acumulado Total e MAPE.

| UE 624<br>RN (c/ vizinhança)<br>x SAR | Método LM            |        | Método GDM           |        | Método GDX           |        | Método BR            |        |
|---------------------------------------|----------------------|--------|----------------------|--------|----------------------|--------|----------------------|--------|
|                                       | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   |
| Fun. Adapt. GD                        | 56,89%               | 58,07% | -83,14%              | 82,90% | 6,56%                | 13,39% | 7,30%                | 10,46% |
| Fun. Adapt. GDM                       | -39,55%              | 44,29% | -22,91%              | 22,37% | -16,42%              | 23,13% | 7,30%                | 10,46% |
| SAR                                   | 15,07%               | 16,77% | 15,07%               | 16,77% | 15,07%               | 16,77% | 15,07%               | 16,77% |

Condição confirmada nos testes de hipóteses, onde na tabela 4.50, observa-se que somente estes experimentos têm a mesma significância do que o predictor SAR.

Tabela 4.50 – Teste de Hipótese UE 624 – Comparações entre Previsores RN x SAR *Panel*.

| Experimento A               | Experimento B                | $T_0^*$ | Graus de Liberdade | t crítico $\alpha = 0,05$ | Hipótese   |
|-----------------------------|------------------------------|---------|--------------------|---------------------------|------------|
| RN (5 FATH x 1 FAL) LMGD    | PSAR - Lndet full - Ent. (C) | 8,2534  | 23,00              | 2,069                     | FALSO      |
| RN (5 FATH x 1 FAL) LMGM    | PSAR - Lndet full - Ent. (C) | 3,0729  | 15,00              | 2,131                     | FALSO      |
| RN (5 FATH x 1 FAL) GDMGD   | PSAR - Lndet full - Ent. (C) | 11,8455 | 21,00              | 2,080                     | FALSO      |
| RN (5 FATH x 1 FAL) GDMGDM  | PSAR - Lndet full - Ent. (C) | 2,0838  | 23,00              | 2,069                     | FALSO      |
| RN (5 FATH x 1 FAL) BRGD    | PSAR - Lndet full - Ent. (C) | -1,1154 | 20,00              | 2,086                     | VERDADEIRO |
| RN (5 FATH x 1 FAL) BRGDM   | PSAR - Lndet full - Ent. (C) | -1,1154 | 20,00              | 2,086                     | VERDADEIRO |
| RN (5 FATH x 1 FAL) GD XGD  | PSAR - Lndet full - Ent. (C) | -0,8336 | 23,00              | 2,069                     | VERDADEIRO |
| RN (5 FATH x 1 FAL) GD XGDM | PSAR - Lndet full - Ent. (C) | 1,7110  | 23,00              | 2,069                     | VERDADEIRO |

Porém até o momento só é possível inferir que, sob alguma condição de treino, a estrutura RN consegue prever com desempenho similar a SAR, mas não se pode avaliar a capacidade de mapear com condições de treinamento iguais (comparar coisas iguais), em situações distintas (UE diferentes), ou qual a influência da vizinhança no processo.

Para tal, replicaram-se os experimentos TRAINGDX GD/GDM, e TRAINBR GD/GDM, para as UE 258 e 207 observando as melhores condições obtidas na UE 624. Nas FIGURAS 4.65 a 4.68 são apresentados os gráficos dos resultados para estas UE.

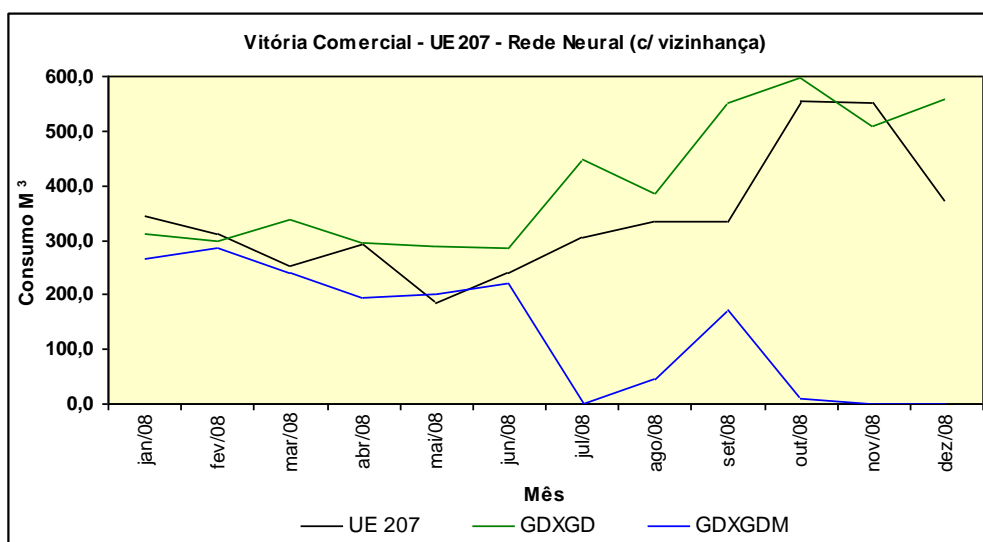


Figura 4.65 – Previsão RN - UE 207 – Método TRAIINGDX GD / GDM.

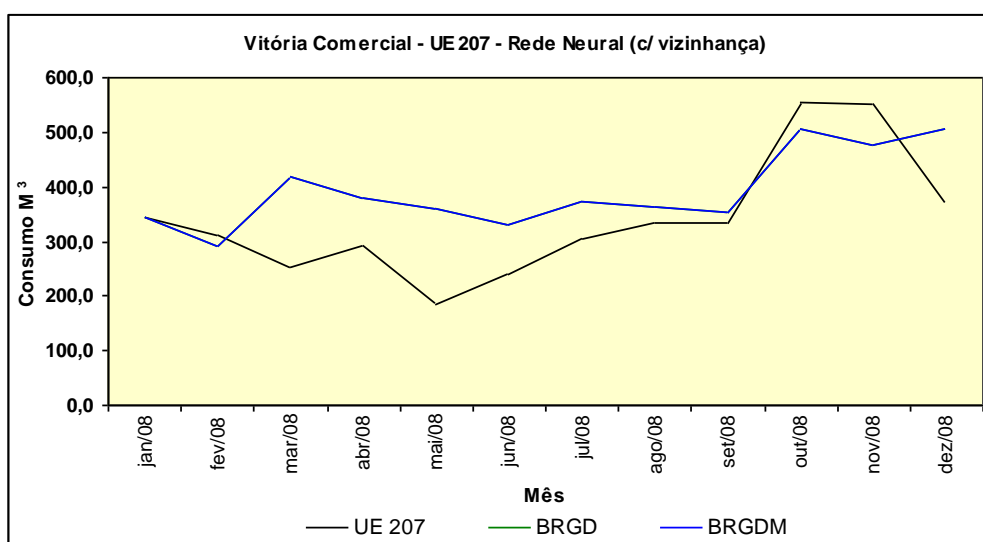


Figura 4.66 – Previsão RN - UE 207 – Método TRAINBR GD / GDM.

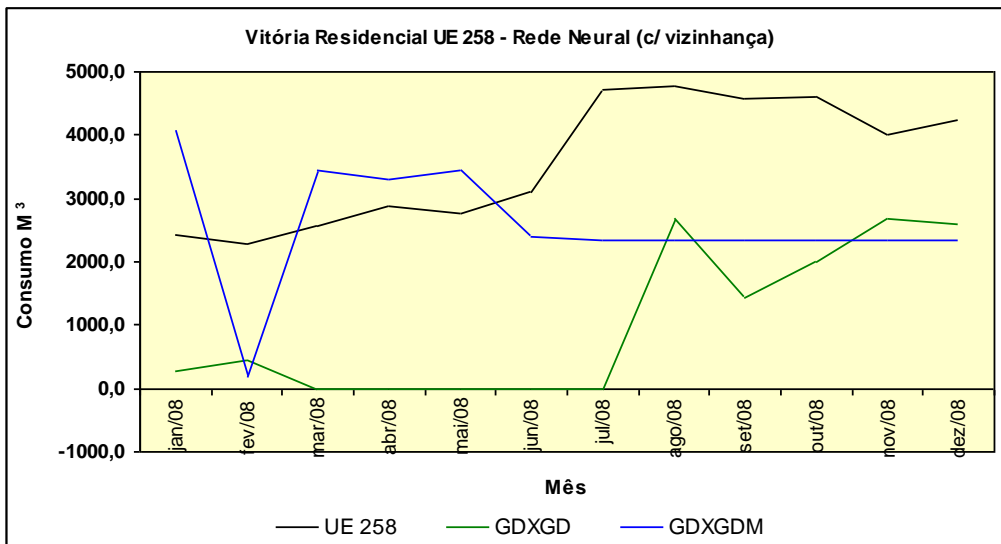


Figura 4.67 – Previsão RN - UE 258 – Método TRAINGDX GD / GDM.

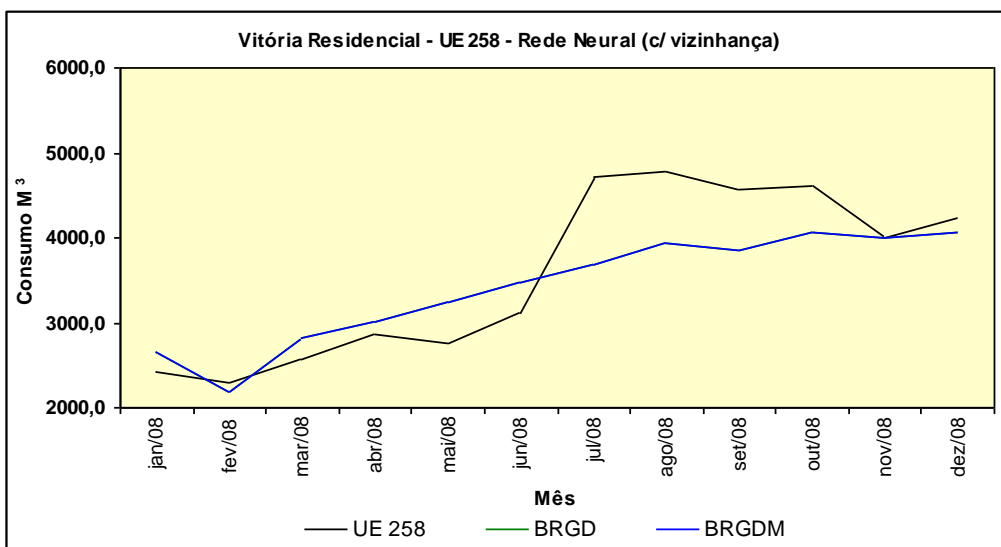


Figura 4.68 – Previsão RN - UE 258 – Método TRAINBR GD / GDM.

Observando as figuras 4.64, 66 e 68, nota-se que em todas as UE ensaiadas com os métodos TRAINBR os resultados apresentaram valores reais positivos. Na verdade neste método de treinamento, a alteração da função de adaptação não alterou significativamente os valores previstos, ou seja, o método TRAINBR proveu maior estabilidade ao experimento.

Condição esta não observada com os métodos TRAINGDX. Na UE 624, figura 4.63, a previsão com este método e função de adaptação GD oscilou ao redor do consumo original; já com função de adaptação GDM apresentou-se como uma reta, com oscilação ao final do período de previsão.

Na UE 207, figura 4.65, a previsão com o método TRAINGDX e função de adaptação GD teve a mesma característica da UE 624; entretanto, a previsão com função de adaptação GDM apresentou valores iguais a zero<sup>43</sup>.

Por último, na UE 258, figura 4.67, a previsão com o método TRAINGDX e função de adaptação GD apresentou valores iguais à zero; enquanto com a aplicação da função de adaptação GDM oscilou no início do período de previsão, estabilizando-se em uma reta na parte final do período, ou seja, ao contrário da UE 624.

Em complemento, na tabela 4.51 são apresentados os valores do erro acumulado total e do MAPE para os experimentos com RN contemplando toda a vizinhança.

Tabela 4.51 – RN - UE 624, 207 e 258 - Erro Acumulado Total e MAPE.

| RII (c/ vizinhança) | UE 624               |        | UE 207               |        | UE 258               |        |
|---------------------|----------------------|--------|----------------------|--------|----------------------|--------|
|                     | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   |
| BRGD                | 7,30%                | 10,46% | 15,30%               | 27,46% | -4,46%               | 10,79% |
| BRGDM               | 7,30%                | 10,46% | 15,30%               | 27,46% | -4,46%               | 10,79% |
| GDXGD               | 6,56%                | 13,39% | 19,50%               | 26,53% | -71,78%              | 75,81% |
| GDXGDM              | -16,42%              | 23,13% | -59,83%              | 51,51% | -28,15%              | 45,29% |

#### 4.5.3.1 – RESULTADOS DOS PREVISORES POR RN SEM A VIZINHANÇA

Em continuidade investigou-se o quanto os experimentos de RN alcançaram estes resultados em função da presença dos dados de consumo das UE vizinhas a de interesse, ou de sua própria capacidade de mapeamento. Para tal replicaram-se os experimentos do tópico anterior com visão puramente temporal e, na tabela 4.52, são apresentados os percentuais do erro acumulado total e do MAPE para estes experimentos.

Tabela 4.52 – RN - UE 624, 207 e 258 - Erro Acumulado Total e MAPE – RN (TEMP).

| RII (s/ vizinhança) | UE 624               |        | UE 207               |        | UE 258               |        |
|---------------------|----------------------|--------|----------------------|--------|----------------------|--------|
|                     | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   |
| BRGD (TEMP)         | -5,46%               | 16,39% | -9,92%               | 22,07% | -17,51%              | 23,75% |
| BRGDM (TEMP)        | -5,46%               | 16,39% | -9,92%               | 22,07% | -17,51%              | 23,75% |
| GDXGD (TEMP)        | 3,60%                | 11,89% | -25,68%              | 28,77% | -44,99%              | 40,65% |
| GDXGDM (TEMP)       | -3,20%               | 27,10% | -36,55%              | 41,28% | -27,67%              | 33,92% |

Nas figuras 4.69 a 4.74, são apresentados respectivamente os gráficos das previsões por UE e método de treinamento.

<sup>43</sup> As previsões iguais a zero são oriundas da camada de saída com função de ativação em linha. Nesta condição o valor simulado pode variar dentro de todo o *range* de dados atribuído na camada de entrada de dados, condição está que permite que a simulação apresente resultados, que desnormalizados sejam negativos, visto a escala original do evento de interesse. Para estes casos, adotou-se na entrada do passo seguinte de simulação da previsão, a informação de valor normalizada em zero conforme a escala original da entrada de dados do evento de interesse.



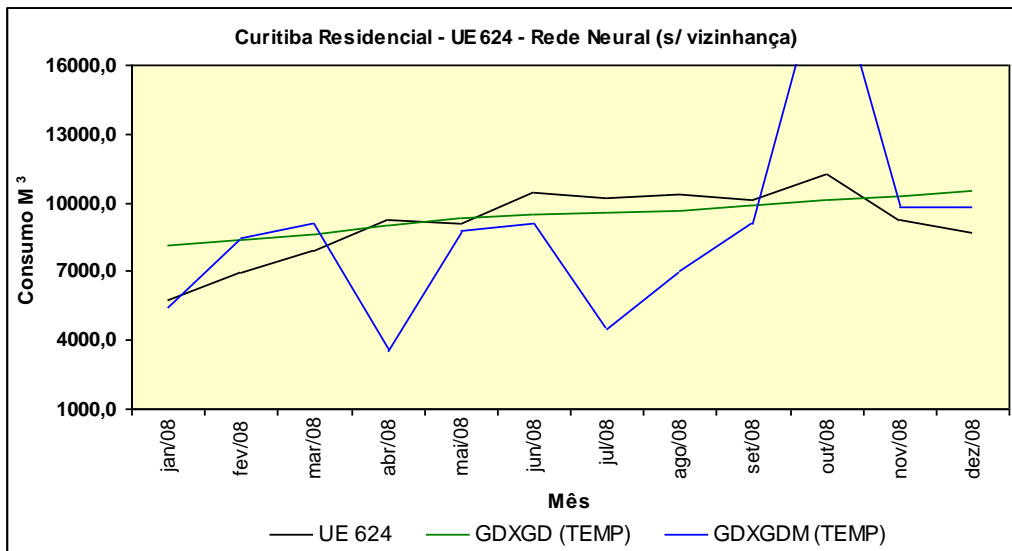


Figura 4.69 – Previsão RN - UE 624 – Método TRAINGDX GD / GDM – (TEMP).

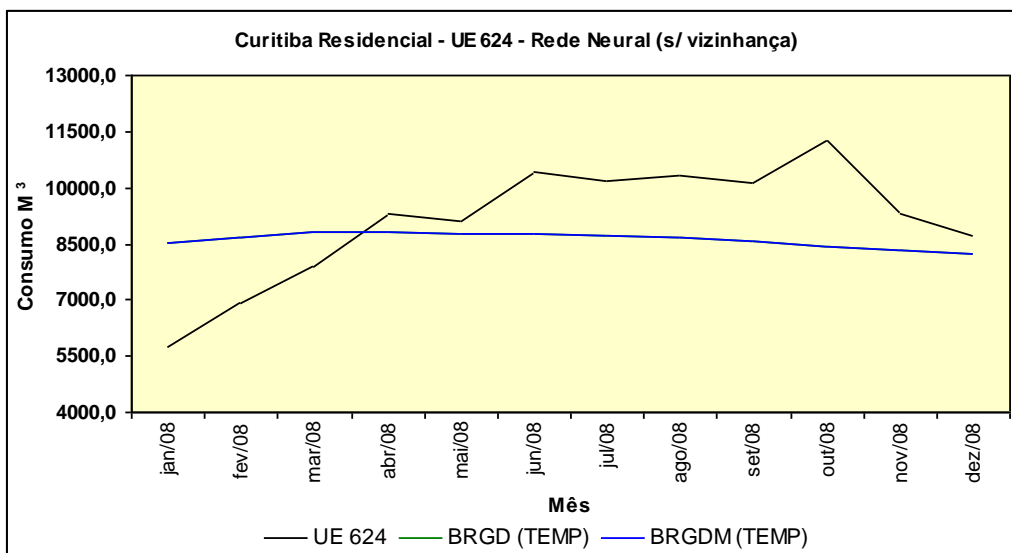


Figura 4.70 – Previsão RN - UE 624 – Método TRAINBR GD / GDM – (TEMP).

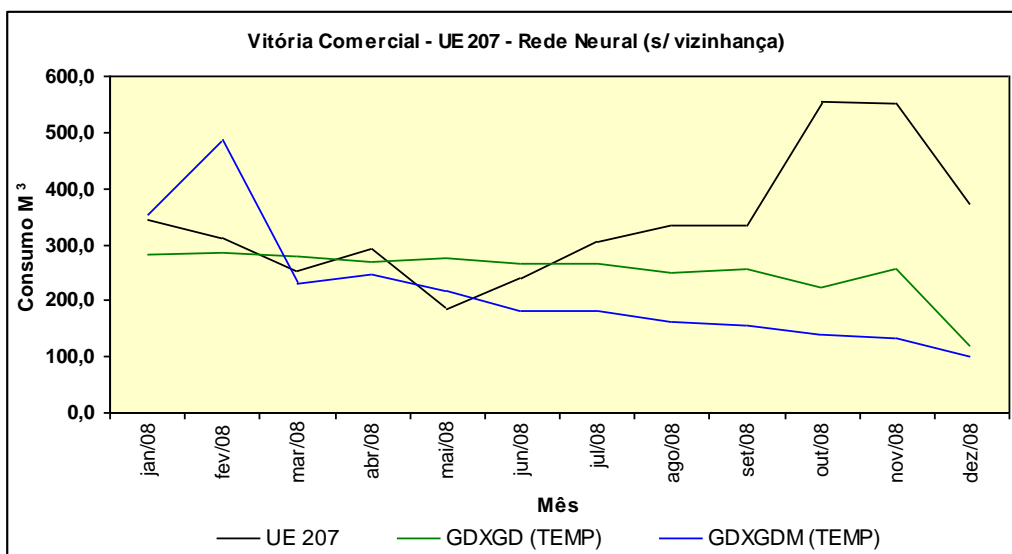


Figura 4.71 – Previsão RN - UE 207 – Método TRAINGDX GD / GDM – (TEMP).

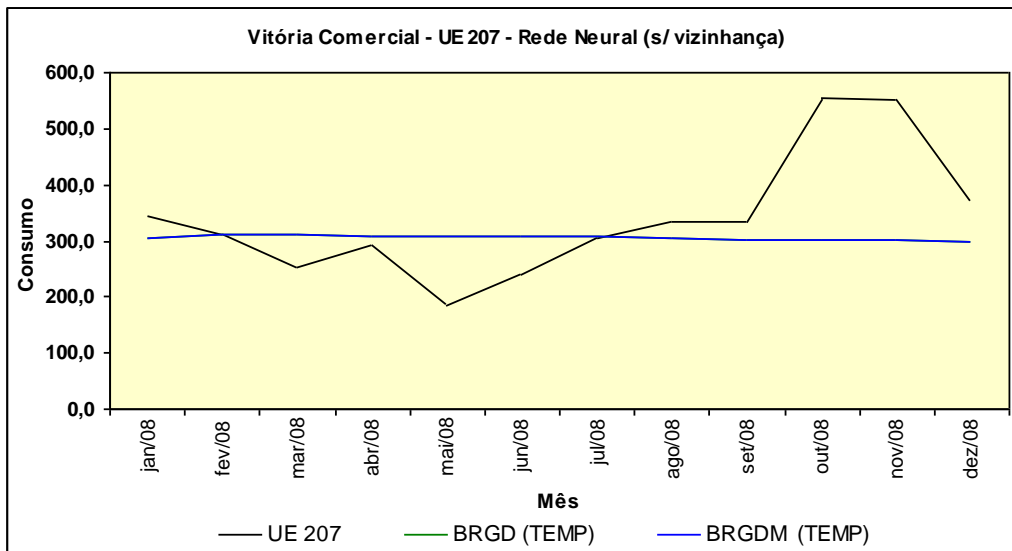


Figura 4.72 – Previsão RN -UE 207 – Método TRAINBR GD / GDM (TEMP).

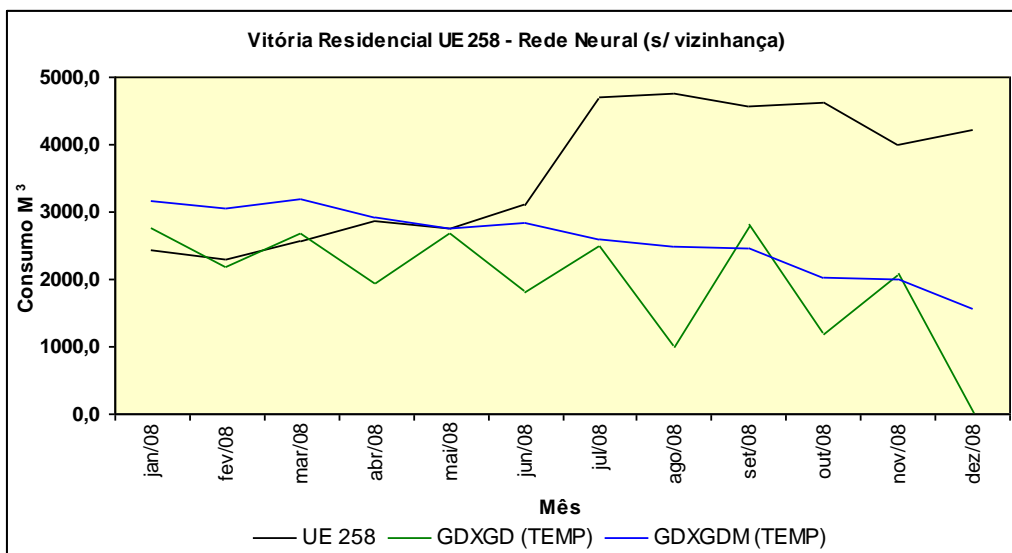


Figura 4.73 – Previsão RN - UE 258 – Método TRAINGDX GD / GDM (TEMP).

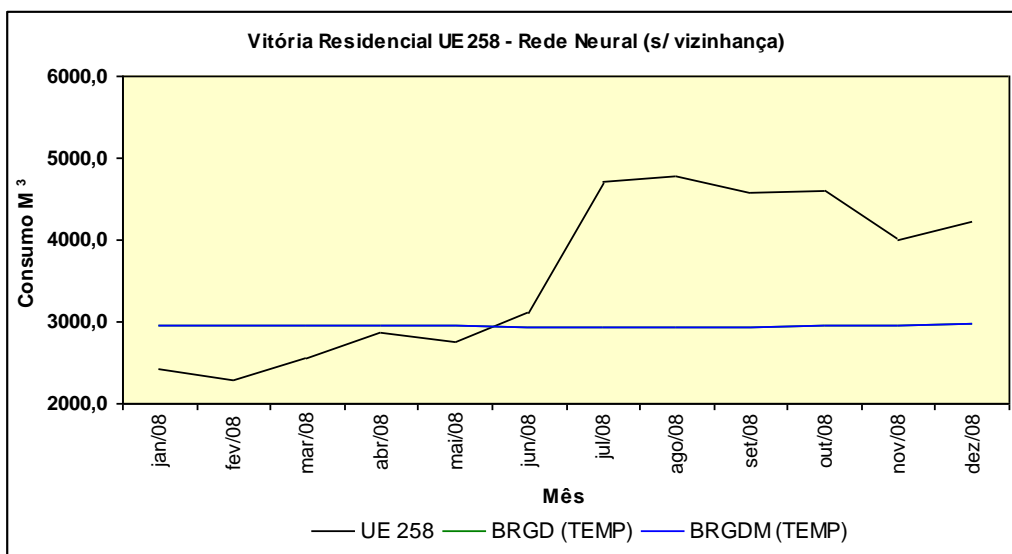


Figura 4.74 – Previsão RN -UE 258 – Método TRAINBR GD / GDM (TEMP).

De forma diferente ao caso dos experimentos com os dados da vizinhança, os perfis das curvas de previsão para as UE ensaiadas com os métodos TRAINBR não acompanharam o perfil do consumo original, aproximando-se de uma reta.

Entretanto, as previsões para todos os experimentos, conforme figuras 4.70, 4.72 e 4.74, apresentaram valores reais positivos, bem como a alteração da função de adaptação não alterou significativamente os resultados finais das previsões. Ou seja, comprova-se que o método TRAINBR provém maior estabilidade ao experimento.

Nos métodos TRAIINGDX manteve-se a característica de instabilidade. Na UE 624, figura 4.69, a previsão com o método TRAIINGDX e função de adaptação GD ficou próximo de uma reta, centrando a previsão entre os extremos, enquanto que com a função de adaptação GDM houve oscilação ao longo de todo o período.

Na UE 207, figura 4.71, as previsões, em ambos os métodos apresentaram previsões com valores reais positivos, entretanto, decaindo com o tempo. Já na UE 258, figura 4.73, a previsão oriunda do método TRAIINGDX e função de adaptação GD apresentou valores iguais a zero, enquanto que na função de adaptação GDM apresentou previsão com valores reais positivos, porém decaindo com o tempo.

No Apêndice L, estão disponíveis os gráficos de previsão e dos erros relativo mensal, acumulado total e absoluto mensal para todos os previsores RN.

#### 4.5.3.2 – ANÁLISE E CONCLUSÕES A CERCA DOS RESULTADOS RN

##### a) Estratégia de Treinamento

Com relação às estratégias de treinamento, a observação dos resultados arranjados par a par na tabela 4.53 permite confirmar a colocação do tópico anterior e concluir que o método de treinamento TRAINBR provém resultados com maior estabilidade. Percebe-se que seus resultados são constantes, independente da função de adaptação, variando somente pela presença dos vizinhos.

Tabela 4.53 – Erro Acumulado Total e MAPE – Comparação de Métodos de Treinamento.

| UE  | Vizinhança | Erro Acumulado |         |         |         | MAPE   |        |        |        |
|-----|------------|----------------|---------|---------|---------|--------|--------|--------|--------|
|     |            | BRGD           | BRGDM   | GDXGD   | GDXGDM  | BRGD   | BRGDM  | GDXGD  | GDXGDM |
| 624 | c/ viz     | 7,30%          | 7,30%   | 6,56%   | -16,42% | 10,46% | 10,46% | 13,39% | 23,13% |
|     | s/ viz     | -5,46%         | -5,46%  | 3,60%   | -3,20%  | 16,39% | 16,39% | 11,89% | 27,10% |
| 207 | c/ viz     | 15,30%         | 15,30%  | 19,50%  | -59,83% | 27,46% | 27,46% | 26,53% | 51,51% |
|     | s/ viz     | -9,92%         | -9,92%  | -25,68% | -36,55% | 22,07% | 22,07% | 28,77% | 41,28% |
| 258 | c/ viz     | -4,46%         | -4,46%  | -71,78% | -28,15% | 10,79% | 10,79% | 75,81% | 45,29% |
|     | s/ viz     | -17,51%        | -17,51% | -44,99% | -27,67% | 23,75% | 23,75% | 40,65% | 33,92% |

Da mesma forma, quando comparados os resultados entre os tipos de funções de adaptação, conforme tabela 4.54, nota-se que dos doze experimentos, nove

apresentaram melhor desempenho em erro acumulado total, e dez em MAPE com o método TRAINBR.

Tabela 4.54 – Erro Acumulado Total e MAPE – Comparação de Funções de Adaptação.

| UE  | Vizinhança | Erro Acumulado |         |         |         | MAPE   |        |        |        |
|-----|------------|----------------|---------|---------|---------|--------|--------|--------|--------|
|     |            | BRGD           | GDXGD   | BRGDM   | GDXGDM  | BRGD   | GDXGD  | BRGDM  | GDXGDM |
| 624 | c/ viz     | 7,30%          | 6,56%   | 7,30%   | -16,42% | 10,46% | 13,39% | 10,46% | 23,13% |
|     | s/viz      | -5,46%         | 3,60%   | -5,46%  | -3,20%  | 16,39% | 11,89% | 16,39% | 27,10% |
| 207 | c/ viz     | 15,30%         | 19,50%  | 15,30%  | -59,83% | 27,46% | 26,53% | 27,46% | 51,51% |
|     | s/viz      | -9,92%         | -25,68% | -9,92%  | -36,55% | 22,07% | 28,77% | 22,07% | 41,28% |
| 258 | c/ viz     | -4,46%         | -71,78% | -4,46%  | -28,15% | 10,79% | 75,81% | 10,79% | 45,29% |
|     | s/viz      | -17,51%        | -44,99% | -17,51% | -27,67% | 23,75% | 40,65% | 23,75% | 33,92% |

Por outro lado o método de treinamento GDX apresentou menor constância, como pode ser visto na comparação entre os erros absolutos mensais das UE 624, 207 e 258 para os casos com vizinhança, conforme as figuras 4.75 a 4.77.

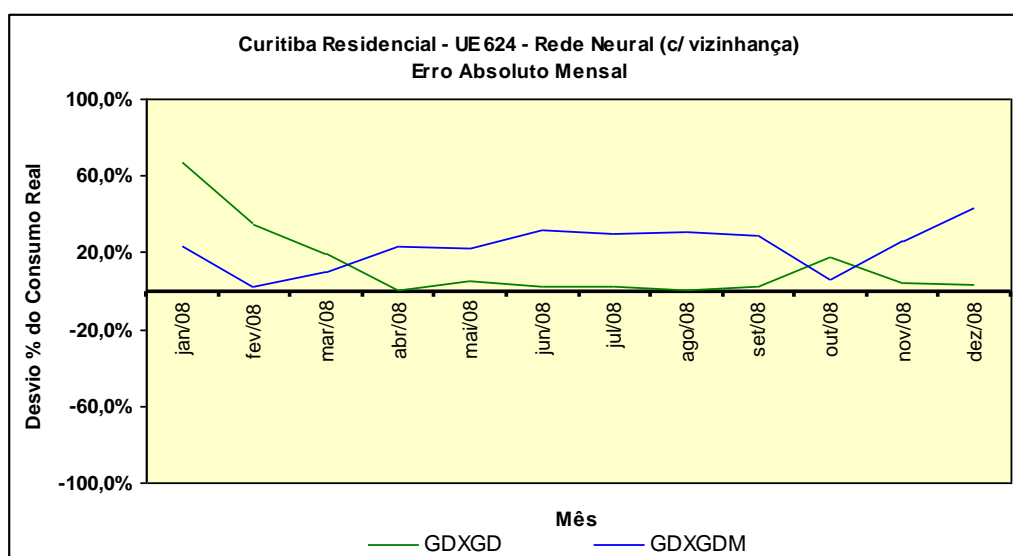


Figura 4.75 – UE 624 – Erro Absoluto Mensal – Experimento RN c/ vizinhança.

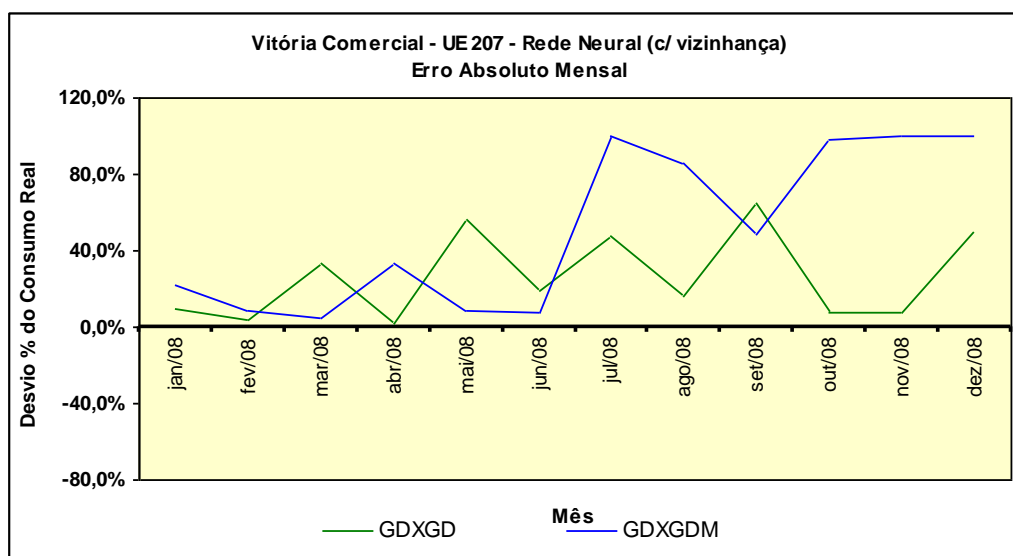


Figura 4.76 – UE 207 – Erro Absoluto Mensal – Experimento RN c/ vizinhança.

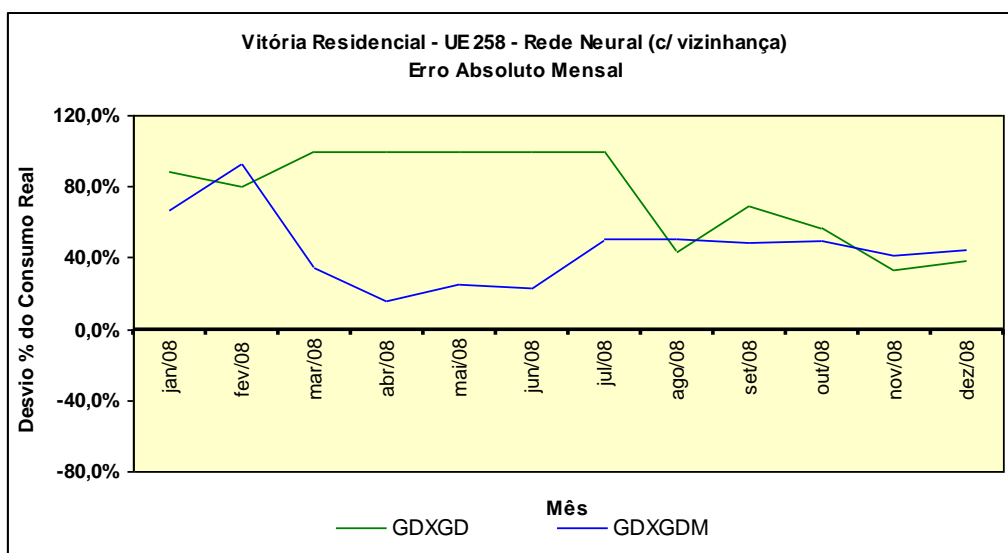


Figura 4.77 – UE 258 – Erro Absoluto Mensal – Experimento RN c/ vizinhança.

Nota-se que na UE 624, independente da função de adaptação, os erros absolutos mensais oscilaram suavemente, no geral entre 0 e 30%; enquanto nos experimentos 258 e 207, oscilaram em até 100%, com comportamentos distintos nos quatro casos. Ou seja, o método TRAINGDX além de prover previsões iguais à zero também apresenta maior oscilação independente da função de adaptação.

### ***b) Capacidade de Mapeamento e Influência da Vizinhança***

Para esta análise os percentuais do erro acumulado total e do MAPE foram rearranjados segundo a ordem das UE e os respectivos resultados com e sem a presença dos dados de consumo da vizinhança, conforme a tabela 4.55.

Tabela 4.55 – Erro Acumulado Total e MAPE por UE e Condição de Vizinhança.

| Método                | Função | UE 624  |        | UE 207  |         | UE 258  |         |
|-----------------------|--------|---------|--------|---------|---------|---------|---------|
|                       |        | c/viz   | sMiz   | c/viz   | sMiz    | c/viz   | sMiz    |
| <b>Erro Acumulado</b> |        |         |        |         |         |         |         |
| BR                    | GD     | 7,30%   | -5,46% | 15,30%  | -9,92%  | -4,46%  | -17,51% |
|                       | GDM    | 7,30%   | -5,46% | 15,30%  | -9,92%  | -4,46%  | -17,51% |
| GDX                   | GD     | 6,56%   | 3,60%  | 19,50%  | -25,68% | -71,78% | -44,99% |
|                       | GDM    | -16,42% | -3,20% | -59,83% | -36,55% | -28,15% | -27,67% |
| <b>MAPE</b>           |        |         |        |         |         |         |         |
| BR                    | GD     | 10,46%  | 16,39% | 27,46%  | 22,07%  | 10,79%  | 23,75%  |
|                       | GDM    | 10,46%  | 16,39% | 27,46%  | 22,07%  | 10,79%  | 23,75%  |
| GDX                   | GD     | 13,39%  | 11,89% | 26,53%  | 28,77%  | 75,81%  | 40,65%  |
|                       | GDM    | 23,13%  | 27,10% | 51,51%  | 41,28%  | 45,29%  | 33,92%  |

No tocante à capacidade de mapear com condições de treinamento iguais (comparar coisas iguais), em situações distintas (UE diferentes), ratificam-se a melhor estabilidade do treinamento baseado no método TRAINBR frente ao TRAINGDX.

Observando a tabela 4.55, nota-se que independente da função de adaptação o método TRAINBR apresentou uma resposta mais estável. As variações, tanto no erro acumulado total quanto no MAPE, sempre foram aos pares; ou seja, o previsor tratou a massa de dados da mesma forma, variando conforme a presença dos dados dos vizinhos.

Este cenário já não foi idêntico para o método TRAINGDX, no qual se apresentou mais sensível à condição de entrada dos dados, principalmente quanto ao *range* da série PC e a característica da camada de saída com função de ativação em linha.

Cabe lembrar que três (dois com e um sem os dados dos vizinhos) dos doze ensaios com o método TRAINGDX, apresentaram previsões mensais de consumo igual a zero; soma-se a isto as citações anteriores sobre a oscilação mensal dos valores previstos.

Quanto à influência da vizinhança, os testes de hipótese não identificaram diferenças significativas entre os experimentos. Conforme a tabela 4.56, dois experimentos (16,7%) apresentaram hipótese de igualdade falsa, ambos onde a estrutura espaço-temporal apresentou melhores resultados.

Tabela 4.56 – Teste de Hipótese – Comparação RN C/ x S / Vizinhança.

| UE  | Experimento A | Experimento B    | $T_{\rho^*}$ | Graus de Liberdade | t crítico $\alpha = 0,05$ | Hipótese   |
|-----|---------------|------------------|--------------|--------------------|---------------------------|------------|
| 624 | RN BRGD       | RN BRGD (TEMP)   | -1,4084      | 23,00              | 2,069                     | VERDADEIRO |
|     | RN BRGDM      | RN BRGDM (TEMP)  | -1,4084      | 23,00              | 2,069                     | VERDADEIRO |
|     | RN GDXGD      | RN GDXGD (TEMP)  | 0,0861       | 18,00              | 2,101                     | VERDADEIRO |
|     | RN GDXGDM     | RN GDXGDM (TEMP) | -0,5606      | 14,00              | 2,145                     | VERDADEIRO |
| 207 | RN BRGD       | RN BRGD (TEMP)   | -0,0883      | 20,00              | 2,086                     | VERDADEIRO |
|     | RN BRGDM      | RN BRGDM (TEMP)  | -0,0883      | 20,00              | 2,086                     | VERDADEIRO |
|     | RN GDXGD      | RN GDXGD (TEMP)  | -0,7740      | 19,00              | 2,093                     | VERDADEIRO |
|     | RN GDXGDM     | RN GDXGDM (TEMP) | 0,6328       | 21,00              | 2,080                     | VERDADEIRO |
| 258 | RN BRGD       | RN BRGD (TEMP)   | -2,4484      | 16,00              | 2,120                     | FALSO      |
|     | RN BRGDM      | RN BRGDM (TEMP)  | -2,4484      | 16,00              | 2,120                     | FALSO      |
|     | RN GDXGD      | RN GDXGD (TEMP)  | 1,7684       | 19,00              | 2,093                     | VERDADEIRO |
|     | RN GDXGDM     | RN GDXGDM (TEMP) | 0,7180       | 21,00              | 2,080                     | VERDADEIRO |

Em outra via, a análise dos valores absolutos de erro acumulado total, mostra que em nove (75%) dos doze ensaios, a estrutura puramente temporal apresentou melhores resultados; já no MAPE não houve diferença, ou seja, 50 % para cada estrutura.

Pela visão hipotética de que o fornecimento de uma maior quantidade de dados tenderia a um melhor treinamento da RN, seria correto esperar que os ensaios puramente temporais apresentassem um desempenho pior. Entretanto, a análise do conjunto total de resultados dos experimentos não apontou para este cenário.

Contudo, excluindo os experimentos com o método TRAINGDX, visto a instabilidade já relatada, revela-se um cenário diferente. Nas figuras 4.78 a 80, onde são apresentadas as curvas do melhor previsor SAR *Panel* versus as previsões RN com o método TRAINBR, é possível notar a influência dos dados das UE vizinhas.

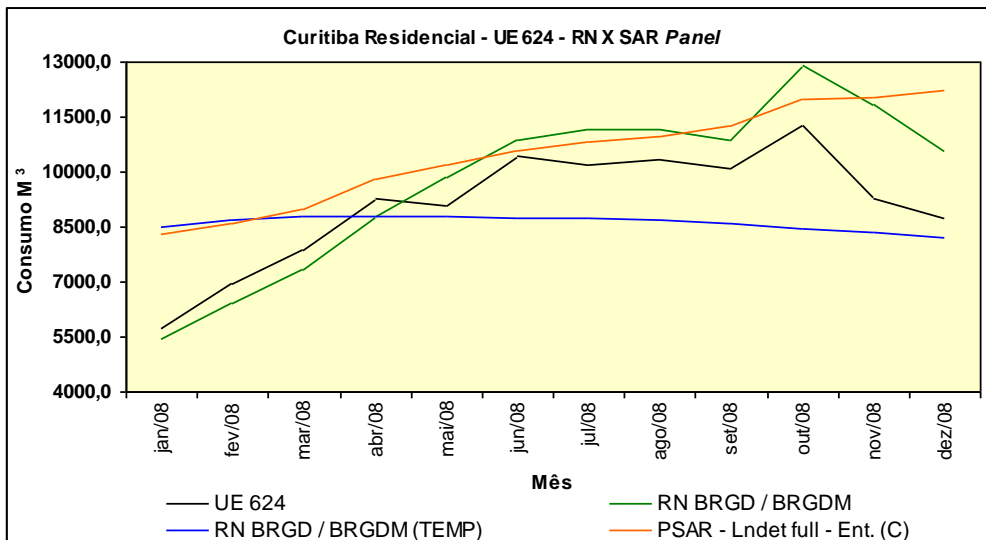


Figura 4.78 – RN x RN (TEMP) x SAR Panel - UE 624 – Comparativo da Previsão.

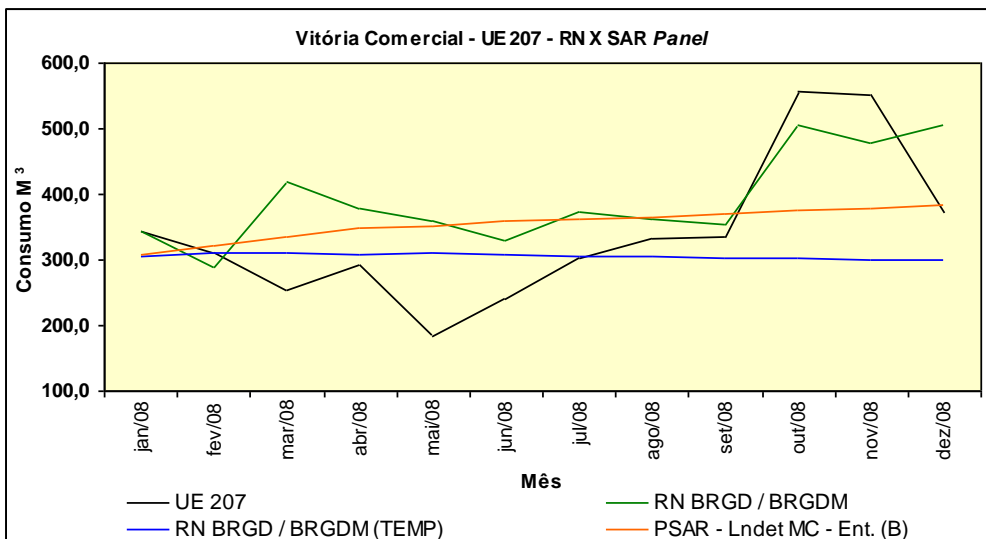


Figura 4.79 – RN x RN (TEMP) x SAR Panel - UE 207 – Comparativo da Previsão.

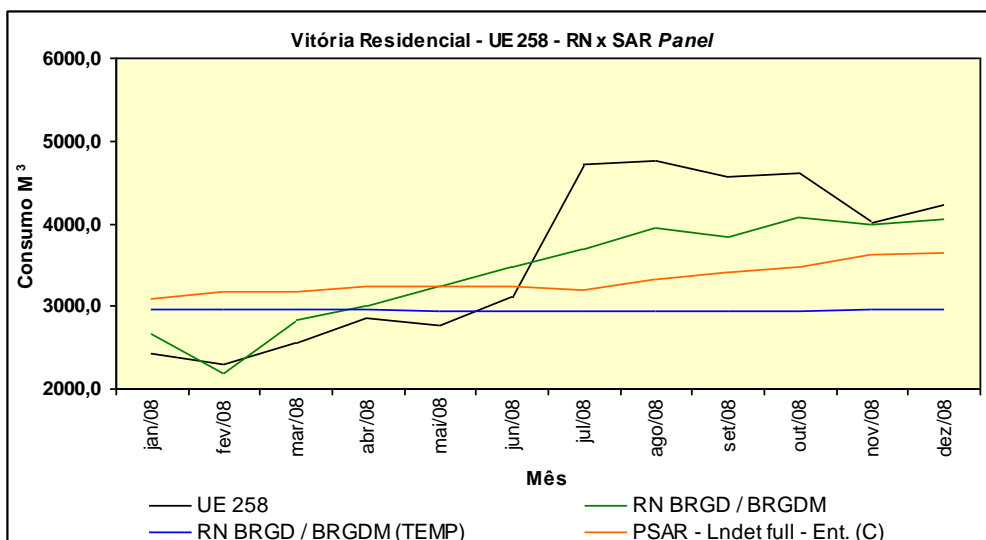


Figura 4.80 – RN x RN (TEMP) x SAR Panel - UE 258 – Comparativo da Previsão.

Em todos os gráficos o comportamento do método TRAINBR, sob o prisma temporal, resultou em projeções com perfil basicamente linear, inclusive com padrão similar aos resultados obtidos nos experimentos ARIMA e Alisamento Exponencial.

Já com a adição do consumo das UE vizinhas, nos três exemplos houve o deslocamento da curva da previsão para a região dos valores reais, a partir do que se conclui que o fornecimento dos dados da vizinhança contribui positivamente com o ajustamento das previsões.

Além disto, as RN com os dados dos vizinhos são menos sensíveis às mudanças de patamar e, sendo assim tem maior capacidade de mapeá-las, em contra ponto ao caso da estrutura SAR que não apresentou boa sensibilidade a esta questão.

### c) Previsão SAR x RN

Na tabela 4.57 a seguir, apresenta-se o comparativo entre os resultados SAR *Panel* e RN com o método TRAINBR, com e sem a presença da vizinhança.

Tabela 4.57 – RN x SAR - Erro Acumulado Total e MAPE por UE.

| Redes Neurais | UE 624               |        | UE 207               |        | UE 258               |        |
|---------------|----------------------|--------|----------------------|--------|----------------------|--------|
|               | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   | Erro Acumulado Total | MAPE   |
| BRGD          | 7,30%                | 10,46% | 15,30%               | 27,46% | -4,46%               | 10,79% |
| BRGDM         | 7,30%                | 10,46% | 15,30%               | 27,46% | -4,46%               | 10,79% |
| BRGD (TEMP)   | -5,46%               | 16,39% | -9,92%               | 22,07% | -17,51%              | 23,75% |
| BRGDM (TEMP)  | -5,46%               | 16,39% | -9,92%               | 22,07% | -17,51%              | 23,75% |
| <b>SAR</b>    | 15,07%               | 16,77% | 4,49%                | 25,99% | -7,13%               | 21,60% |

A análise dos resultados apresentados na tabela permite verificar que:

- UE 624 – a RN apresentou melhor erro acumulado total em ambos os casos, espacial e temporal. Já o MAPE alternou. No caso espacial apresentou MAPE melhor que o SAR e, no caso temporal foi similar ao SAR;
- UE 207 – em ambos os casos, espacial e temporal, a RN apresenta pior desempenho. Já o MAPE variou na faixa de + / - 5%;
- UE 258 – incluindo os dados dos vizinhos a RN apresenta melhores erro acumulado e MAPE, enquanto na visão temporal os dois índices deterioraram.

Quanto à comparação entre os experimentos RN e SAR *Panel*, os testes de hipótese não apontaram diferenças significativas. Conforme a tabela 4.58, dois experimentos (16,7%) apresentaram diferenças significativas; em ambos os casos com a estrutura RN com vizinhos apresentou melhores resultados.



Tabela 4.58 – Teste de Hipótese – Comparações entre Previsores RN C/ x S / Vizinhança.

| UE  | Experimento A              | Experimento B                | $T_0^*$ | Graus de Liberdade | t crítico $\alpha = 0.05$ | Hipótese   |
|-----|----------------------------|------------------------------|---------|--------------------|---------------------------|------------|
| 624 | RN BRGD                    | PSAR - Lndet full - Ent. (C) | -1,1154 | 20,00              | 2,086                     | VERDADEIRO |
|     | RN BRGDM                   | PSAR - Lndet full - Ent. (C) | -1,1154 | 20,00              | 2,086                     | VERDADEIRO |
|     | RN BRGD (TEMP)             | PSAR - Lndet full - Ent. (C) | 0,0853  | 22,00              | 2,074                     | VERDADEIRO |
|     | RN BRGDM (TEMP)            | PSAR - Lndet full - Ent. (C) | 0,0853  | 22,00              | 2,074                     | VERDADEIRO |
| 207 | RN BRGD                    | PSAR - Lndet MC - Ent. (B)   | -0,1649 | 23,00              | 2,069                     | VERDADEIRO |
|     | RN BRGDM                   | PSAR - Lndet MC - Ent. (B)   | -0,1649 | 23,00              | 2,069                     | VERDADEIRO |
|     | PSAR - Lndet MC - Ent. (B) | RN BRGD (TEMP)               | 0,0445  | 21,00              | 2,080                     | VERDADEIRO |
|     | PSAR - Lndet MC - Ent. (B) | RN BRGDM (TEMP)              | 0,0445  | 21,00              | 2,080                     | VERDADEIRO |
| 258 | RN BRGD                    | PSAR - Lndet MC - Ent. (C)   | -2,3485 | 21,00              | 2,080                     | FALSO      |
|     | RN BRGDM                   | PSAR - Lndet MC - Ent. (C)   | -2,3485 | 21,00              | 2,080                     | FALSO      |
|     | RN BRGD (TEMP)             | PSAR - Lndet MC - Ent. (C)   | 0,6805  | 20,00              | 2,086                     | VERDADEIRO |
|     | RN BRGDM (TEMP)            | PSAR - Lndet MC - Ent. (C)   | 0,6805  | 20,00              | 2,086                     | VERDADEIRO |

Resumidamente, o erro acumulado total dos experimentos RN apresentou certa vantagem; já o MAPE, com base nos desvios individuais, somente os experimentos da UE 258 apresentam diferença perceptível. Além disto, a observação dos gráficos das previsões, nas figuras 4.78 a 80, permite que sejam tecidos comentários adicionais.

Observa-se nestas figuras que as previsões considerando o mote espacial, seja pelo predictor SAR ou na própria RN, tiveram uma melhor capacidade de corrigir as previsões para o centro da meta, quando comparado com o caso puramente temporal.

Cabe ressaltar que este ponto vai ao encontro dos comentários realizados quanto ao comportamento da previsão dos experimentos ARIMA e Alisamento.

Comparando os previsores SAR *Panel* e RN, pode-se concluir que ambos têm capacidade de mapear, com condições iguais de treinamento / determinação de parâmetros, em situações distintas (UE diferentes). Porém no geral a estrutura SAR, por ter menos opções a serem testadas (existência de efeitos espaciais fixos e método de determinação de  $p$ ) tem um processo mais simples do que a não-linear por RN.

Como esta última disponibiliza um leque infinito de combinações de estrutura e treino, somente permite inferir que sob alguma combinação, a RN pode apresentar um melhor desempenho do que a estrutura SAR.

De forma específica, os experimentos permitem indicar como opção inicial de busca as RN treinadas com o método TRAINBR, mas isto não é condição satisfatória, pois os resultados podem variar quanto à presença dos dados dos vizinhos.

Sendo assim é possível concluir que, mesmo existindo um método de treino que se mostrou eficiente para mapear o consumo de GN, é necessário testar as condições sobre a presença dos dados da vizinhança, visto a sensibilidade quanto à estrutura de entrada dos dados para treinamento (quantidade e range dos eventos disponíveis).

No Apêndice M, estão disponíveis os gráficos de previsão e de erros para as comparações entre os previsores SAR versus os por RN.

## 5 – CONCLUSÕES E RECOMENDAÇÕES

Quanto ao problema de engenharia cerne desta tese, ou seja, a decisão acerca da determinação da infra-estrutura de distribuição de GN como fator crítico de sucesso para as CDL, conclui-se que o desenvolvimento de um modelo de previsão de demanda com base na perspectiva espacial cumpre seu papel de apoio à decisão para tal problema, na medida em que permitiu:

- Estabelecer porções territoriais de interesse, as UE, em UGR, nas quais foi possível contabilizar o consumo de GN, avaliar sua representatividade e seu crescimento, tanto na dimensão espacial, quanto na temporal;
- Investigar, com base nestas UE, as condições necessárias e suficientes para emprego e desenvolvimento de um modelo de previsão espaço-temporal; e,
- Gerar previsões individuais para estas UE, a partir de modelos baseados em econometria espacial ou em AM, caracterizando assim as diferentes demandas ao longo do tempo e do território.

Por conseqüência, este desenvolvimento fornece subsídios à análise das condicionantes que caracterizam a rede de acesso em diferentes níveis de demanda e porções do espaço.

Com isto, a aplicação da dimensão espacial deixa de ser empregada somente na ilustração das variáveis no tempo, ou seja, passa a considerar a influência dos relacionamentos entre os fenômenos espaciais na análise do ambiente de negócio, onde se encontra o fenômeno de interesse, neste caso o consumo de GN. Desta forma o resultado desta tese contribui com a CDL uma vez que:

- Permite refinar suas provisões de demanda, acarretando em contratação de volumes de GN mais adequados aos segmentos; e,
- Mitiga os riscos no investimento, pois orienta o crescimento da rede para regiões onde existem relações de oferta x consumo mais interessantes.

De forma particular, a análise da dependência espacial e sua conseqüente estacionariedade no tempo, a partir do emprego de técnicas de análise espacial de áreas, permitiu avaliar as condições necessárias e suficientes para atestar o emprego de um previsor de demanda com base em modelos de regressões espaço-temporal.

Toda a fundamentação teórica associada à dimensão espacial permitiu estabelecer as condições de contorno necessárias ao emprego do modelo, principalmente: a formação do BDE e a correta associação dos dados (variáveis explicativas e de

interesse), a análise da UGR pertinente ao estudo e a apuração dos índices de autocorrelação espacial global e local no tempo e a conseqüente definição das UE de interesse à aplicação do modelo.

Sobre o modelo de simulação propriamente dito, ressalta-se que a flexibilidade da entrada de dados na estrutura SAR *Panel*, além de incorporar a influência dos efeitos espaciais, permite compor modelos abrangentes e integrados, pois confere ao modelador a flexibilidade de testar arranjos de variáveis exógenas em resoluções espaciais distintas; ou seja, engloba ao modelo dados macroeconômicos de níveis regionais distintos.

Desta forma, a estrutura SAR confere uma maior capacidade para a estimação dos parâmetros do que as tradicionais formas com estruturas puramente autoregressivas e / ou de reinjeção. Este ponto inclusive constituiu uma vantagem da estrutura SAR, face aos modelos lineares mais usuais, como o ARIMA e o alisamento exponencial.

Comparando estas estruturas observa-se que a previsão SAR apresentou melhor desempenho do que a primeira e desempenho similar à segunda. Não obstante, em ambos os casos, ARIMA e alisamento exponencial, as previsões do modelo SAR *Panel* apresentaram resultados mais atrativos, em função do erro acumulado total da previsão ser positivo, conforme apresentado em 4.5.1 e 4.5.2.1.

Atualmente o consumo de GN nas áreas de concessão das CDL é infante, e com isto os mercados locais são pouco explorados, seja na condição de principal fornecedor ou na de substituto. Neste cenário, a subcontratação é um impeditivo, pois se existe crescimento não previsto, como por exemplo, um novo empreendimento imobiliário, a CDL não tem como atender ao cliente, enquanto que na condição de sobre-capacidade existe alguma possibilidade operacional de executar este atendimento.

Quanto ao mapeamento por RN, pelo grande leque de combinações de estrutura de camadas e treino, somente foi possível concluir que sob alguma combinação, a RN pode mapear o problema de previsão. Apesar dos ensaios realizados indicarem a opção inicial de treinamento das RN pelo método TRAINBR, isto não é condição suficiente, pois os resultados variaram quanto à presença ou ausência dos dados dos vizinhos.

Com relação a esta questão, a da presença dos dados da vizinhança no treinamento das RN, pela hipótese de que fornecer uma maior quantidade de dados possibilitaria um melhor treinamento, seria correto esperar que as RN com os dados das UE vizinhas apresentassem um desempenho melhor.

Entretanto, os testes de hipótese apresentados em 4.5.3.2, tópico b, não identificaram diferenças significativas entre estes experimentos. Sendo assim, mesmo existindo um método de treino, que se mostrou eficiente para mapear o consumo de GN, é necessário testar as duas condições de presença dos dados da vizinhança.

Contudo, a observação dos gráficos que sobrepõem às previsões SAR *Panel* e por RN, com e sem os dados de vizinhança, também no item 4.5.3.2, tópico b, demonstram que houve um direcionamento da curva da previsão das RN com dados da vizinhança na direção da curva dos valores de testes; a partir disso se conclui que fornecer os dados da vizinhança contribui com o ajustamento das previsões.

Quanto à comparação dos previsores SAR *Panel* e RN, os testes de hipóteses, apresentados no item 4.5.3.2, tópico c, apontaram que ambos os modelos têm capacidade de mapear, com condições iguais de treinamento / estimação de parâmetros, em situações distintas (UE diferentes).

Entretanto, como a carga de trabalho até as etapas de: análise das condições necessárias e suficientes para emprego do modelo espaço-temporal, definição das variáveis para compor o modelo e definição das UE de interesse; são iguais para as duas formas, o modelo SAR *Panel* torna-se mais atrativo nas etapas de estimação e previsão, pois apresenta menos opções a serem testadas (forma funcional e método de determinação de  $\ln|I - pw|$ ) do que a RN, e com isto tem um processo mais simples.

Desta forma, indica-se, frente aos resultados alcançados, a estrutura SAR *Panel* como a opção a ser adotada no caso da previsão espaço-temporal de consumo de GN.

Certamente, seja o previsor linear ou não, o seu desenvolvimento não extingue a necessidade de incorporar dados específicos à previsão de demanda. Como os modelos se valem de dados históricos, não é possível capturar inflexões específicas de consumo (mudanças de patamar) de forma prévia.

Mesmo que as RN tenham apresentado maior capacidade de mapear tal situação, é possível que o motivador da inflexão não ocorra de forma cíclica, como por exemplo, o impacto de um shopping ou de um condomínio em uma dada UE.

Sendo assim o trabalho de prospecção e análise de possibilidades de novos negócios sempre contribuirá com o refinamento da demanda; ainda assim seus resultados poderão ser melhor alocados espacialmente à demanda.

Por último, com relação ao conjunto de dados coletados para a tese, independente dos descartes provenientes das análises estatísticas, dos ajustes de frequência e dos

procedimentos de *clean-up* de dados; todas as variáveis aplicadas ao modelo, excetuando as informações de preço e consumo de EE e GN, são de fontes públicas.

Apesar das reduções de quantidade de variáveis disponíveis, associada a um consumo infante e não homogêneo na AE, para os segmentos estudados, foi possível avaliar a dependência e a heterogeneidade espacial como condições para desenvolver um modelo previsão empregando tanto o modelo espaço-temporal SAR *Panel*, quanto uma abordagem por RN.

Com relação ao exposto conclui-se que os objetivos propostos à tese foram integralmente cumpridos, tanto o principal quanto os secundários.

## 5.1 – PRINCIPAIS CONTRIBUIÇÕES

As seguintes contribuições foram elencadas como principais:

- a) **Previsão de Demanda de Energéticos** – como ressaltado no item 2.6, a dimensão espacial tem sido empregada prioritariamente para representar variáveis e resultados na sua respectiva posição no espaço. Fruto disto, não se observa a aplicação desta dimensão do dado direcionada à previsão de demanda de energia e com ela, o emprego de condicionantes e modelos de regressões previstos na econometria espacial. Neste sentido entende-se que esta tese, ao estabelecer um processo de previsão *bottom up* a partir da previsão local de demanda, permitiu atribuir espacialmente à previsão de demanda, que hora é feita em caráter *top down* e sem a correta distribuição espacial. Para tal este processo contemplou nos itens 4.2.2 e 4.2.3 a avaliação das condicionantes espaciais (dependência e heterogeneidade) e, no item 2.6.3 os possíveis métodos espaço-temporais aplicáveis, bem como observou as condições básicas à previsão de demanda nos segmentos de energia, entre elas a natureza sócio-econômica dos dados e a indisponibilidade dos mesmos em níveis espaciais e temporais adequados.
- b) **No âmbito da Econometria Espacial** – quanto aos métodos de econometria espacial identificaram-se algumas contribuições particulares relacionadas a:
  - **Extensão da análise dos efeitos espaciais para dados em painel** – as contribuições usuais restringem-se a avaliação das condicionantes espaciais em um senso *cross section*, tanto na avaliação do atraso espacial (*spatial lag*), quanto nos erros espacialmente correlacionados (*spatial error*); com isto, a observação dos efeitos mitigados dos resíduos se dá por

testes a posteriori da estimação. Esta tese propôs, em 3.2.2 e 3.2.3, um método para avaliar a dependência e a heterogeneidade espacial à luz de dados em painel, estabelecendo como condição necessária para aplicar o modelo SAR *Panel*, a existência de atraso espacial entre a variável de interesse nas diversas UE vizinhas e, em qual grau de predominância, global ou local; além disso, estabeleceu a condição suficiente para determinar o emprego em caráter temporal, ou seja, se existe dependência espacial ao longo da série da variável dependente na UE de interesse. Método este que foi corroborado quando da realização dos testes de estrutura, proposto em 3.4.1, pelos quais constatou-se, conforme 4.4.3, que a estrutura SAR mitigou os efeitos dos atrasos espaciais dos resíduos;

- **Estimação do modelo** – as contribuições usuais focam a pesquisa orientada ao desempenho dos métodos de otimização e estimação dos parâmetros dos modelos espaço-temporais. Como ressaltado em 3.3.3, não se encontrou citação de pesquisa, que avalie o impacto da flexibilidade do termo  $X\beta$  à estimação do modelo. Esta tese demonstrou que o impacto da forma como se dá a entrada de dados em  $X\beta$  tem grande influência na estimação dos parâmetros. A maximização do escalar  $p$  apresentou comportamento similar por tipo de experimento, ou seja, não diferiu visto o método de determinação  $\ln|I - pW|$ , bem como os valores apresentaram certa estabilidade não permitindo determinar uma forma funcional como a de melhor desempenho. Porém, com base no apresentado na tabela 4.35, conclui-se que a entrada de dados foi o elemento de maior influência no processo de previsão, pois entre os sessenta experimentos realizados, quatorze dos vinte melhores resultados, seja pelo erro acumulado total ou pelo MAPE, associam-se à forma de entrada de dados.
- **Flexibilidade de arranjos de dados** – Fruto das opções de entrada de dados, onde  $X\beta$  ganhou contornos de estrutura ARX, esta tese permitiu, conforme descrito em 4.4.2.2, tópico b, caracterizar a flexibilidade que a estrutura SAR confere ao modelador para estabelecer, com o mesmo cenário de variáveis explicativas, diferentes arranjos no senso linear. Isto amplia o leque de modelos de estimação, além das formas puramente auto-regressivas e / ou de re-injeção.

- **Cálculo da previsão** – conforme ressaltado em 3.3.4, não se encontrou citação, ou exemplo aplicado de como empregar os parâmetros estimados para gerar uma previsão real. Neste sentido, conforme formulado em (3.21), (3.22) e (3.23), esta tese permitiu apresentar de forma concreta a aplicação dos parâmetros estimados, em uma previsão real observando: o equilíbrio dos termos do modelo, o processo de previsão de  $y$  em  $n$  para  $t + 1$  e, por último as considerações sobre o efeito temporal oriundo da previsão da variável dependente nas UE vizinhas sobre a matriz  $W_{ij}$  para UE de interesse no mesmo  $t$ .
- c) **No âmbito das RN** – foi materializada, no item 4.5.3, a capacidade das RN servirem ao propósito da previsão de demanda quando observa-se de forma conjunta as condicionantes temporais e espaciais, neste caso o comportamento de consumo das UE vizinhas a uma dada área qualquer de interesse. Neste sentido esta tese proveu uma análise das estratégias de treinamento para uma RN aplicada com estas condicionantes de uso, indicando entre outros: a estrutura da rede, a forma de entrada de dados e a indicação das funções de treino e adaptação de melhor desempenho de treinamento.

## 5.2 – TRABALHOS FUTUROS

Propõem-se os seguintes possíveis tópicos para continuidade desta tese:

- **Sazonalidade, tendência e ciclos senoidais das séries da variável de interesse** – como observado em 4.4.2.2, tópico a, a literatura espacial não aborda como tratar estas questões no âmbito do desenvolvimento do modelo, sendo assim cabe avaliar o impacto que o seu tratamento prévio traz à determinação da dependência espacial e previsão com o modelo SAR *Panel*. Mesmo considerando o horizonte de previsão de 12 meses cabe tal avaliação, pois estas particularidades podem se apresentar em ciclos menores que um ano e inferindo diretamente sobre a frequência de previsão mensal. Por extensão, avaliar também tal consideração no treinamento das RN com os dados da vizinhança;
- **Análise da estacionariedade da dependência espacial** – aprimorar a forma da análise da estacionariedade da dependência espacial, a partir de desenvolvimento de código computacional para realizar os testes globais de Moran I e locais de Moran e Estatística  $G_i$  e  $G_i^*$  (Anselin, 1995), observando o

fornecimento de dados em painel, conforme a metodologia descrita em 3.2.2 e 3.2.3. Bem como, tomando como base a normalização *Z Score* do índice, determinar critérios estatísticos aplicados à etapa de quantificação da flutuação temporal do índice de dependência espacial local ao longo do espaço;

- Fornecimento da matriz  $W_{ij}$  ao treinamento das RN – conforme observado no item 4.5.3.2, tópico b, o fornecimento dos dados da vizinhança resulta em melhorias, quando da comparação das previsões por RN com dados da vizinhança contra as RN puramente temporal. Com isto, cabe ainda avaliar qual a melhora na capacidade de mapeamento que o previsor neural, com os dados da vizinhança, tem a partir do fornecimento das relações de contiguidade ao processo de treinamento da RN. Inclusive na via da avaliação do ganho na sensibilidade da detecção das mudanças de patamar da série histórica, frente ao previsor com estrutura SAR *Panel*; e,
- Previsão das UE em regimes de transição – a estacionariedade da dependência espacial foi definida como condição necessária à aplicação do modelo SAR *Panel* em caráter local, sendo assim o mesmo não tem motivador para uso em UE que não apresentam efeitos associados ao atraso espacial. Entretanto, o experimento com RN contemplou o fornecimento dos dados dos vizinhos de forma global, ou seja, para todas as UE constantes da área de estudo. Sendo assim, cabe avaliar se o emprego da via não-linear para uma UE que não apresenta atraso espacial permite aprimorar a previsão a partir da captura dos erros correlacionados espacialmente; condição esta que não pode ser avaliada com o emprego do modelo SEM *Panel*, em virtude das incapacidades relacionadas no item 2.6.3.



## 6 – REFERÊNCIAS BIBLIOGRÁFICAS

- ALCAMO, J., 2001, **Scenarios as Tools for International Environmental Assessments**. In: Environmental Issue Report n. 24, European Environment Agency, Denmark. ISBN 92-9167-402-8.
- ALLEN, J. F., 1983, “Maintaining knowledge about temporal intervals”, **Communications of the ACM**, v. 26, n. 11(Nov), pp. 832-843.
- AL-FATTAH, S. M., STARTZMAN, R. A., 2000, “Forecasting World Natural Gas Supply”, **Journal of Petroleum Technology**, v. 52, n. 5 (Mai), pp. 62-72.
- ANDRADE, A. L., SELEME, A., RODRIGUES, L. H., *et al.*, 2006, **Pensamento Sistêmico: Caderno de Campo – O Desafio da Mudança Sustentada nas Organizações e na Sociedade**. Porto Alegre, Bookman. ISBN: 85-36307-00-5
- ANSELIN, L., 1988, **Spatial econometrics: methods and models**. Boston, Kluwer Academic. ISBN: 90-247-3735-4
- ANSELIN, L. **Spatial Econometrics**. [S.3.]: Virtual Paper, 1999. Disponível em: <[http://www.csiss.org/learning\\_resources/content/papers/baltchap.pdf](http://www.csiss.org/learning_resources/content/papers/baltchap.pdf)> Acesso em: 03 fev. 2012, 15:45:00.
- ANSELIN, L., 1999, “Interactive techniques and exploratory spatial data analysis”. In: Longley, P. A., Maguire, D. J., Goodchild, M. F., *et al.* (eds), **Geographical Information Systems: Principles, Techniques, Management and Applications**, v. 1, 2nd ed., chapter 17, pp. 253-266, New York, USA, John Wiley & Sons. ISBN: 0-471-32182-6.
- ANSELIN L., HUDAK S., 1992, “Spatial econometrics in practice: A review of software options”, **Regional Science and Urban Economics**, v. 22, n. 3 (Set), pp. 509-536.
- ANSELIN, L., LOZANO-GRACIA, N., 2008, “Errors in variables and spatial effects in hedonic house price models of ambient air quality”, **Journal Empirical Economics**, v. 34, n.1 (Fev), pp. 5-34.

- ANSELIN, L., LE GALLO, J., JAYET, H., 2008, “Spatial panel econometrics”. In: **The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice** (3rd edition), v. 46, part II, **Advanced Studies in Theoretical and Applied Econometrics**, Heidelberg, Germany, Springer-Verlag, pp. 625-660. ISBN: 978-3-540-75889-1.
- ARNOLD, M. **GMM Estimation of the Autoregressive Parameter in a Spatial Autoregressive Error Model using Regression Residuals**. [S.5]: Virtual Paper, 2007. Disponível em: <[https://eldorado.tu-dortmund.de/bitstream/2003/24712/1/TR\\_25-arnold.pdf](https://eldorado.tu-dortmund.de/bitstream/2003/24712/1/TR_25-arnold.pdf)> Acesso em: 09 fev. 2012, 14:54:00.
- ARONOFF, S., 2004, **Geographic Information Systems: A management Perspective**. 2nd ed. Canada, WDL Publications. ISBN: 0-921804-91-1
- ASCHE, F., OSMUNDSEN, P., TVETERÅS, R., 2001, “Market integration for natural gas in Europe”, **International Journal of Global Energy Issues**, v. 16, n. 4, pp. 300-312.
- AZEVEDO, S. J. S., 2006, “A contribuição da análise sócio-demográfica para a definição de demandas sociais – O exemplo das telecomunicações em campinas”. In: **Demografia dos Negócios: Campo de Estudo, Perspectivas e Aplicações**, v. 3, **Demographics**, ABEP, pp. 189-211. ISBN 85-85543-14-0.
- BAILEY, T. C., GATRELL, A. C., 1995, **Interactive Spatial Data Analysis**. 1st ed. England, Longman Scientific & Technical. ISBN: 0-582-24493-1.
- BAKIRTZIS, A. G., PETRIDIS, V. KIARTZIS, S. J., *et al.*, 1996, “A neural network short-term load forecasting model for the Greek power system”, **IEEE Transactions on Power Systems**, v. 11, n. 2 (Mai), pp. 858–863.
- BALTAGI, B. H., 2001, **Econometric Analysis of Panel Data**. 2nd ed. England, John Wiley & Sons. ISBN 0-471-49937-4.
- BALTAGI, B. H., EGGER, P., PFAFERMAYR, M., 2009, **A generalized spatial panel data model with random effects**. Report: n. 113 Center for Policy Research of Maxwell School of Syracuse University, New York, USA.

- BALTAGI, B. H., BRESSON, G., PIROTTE, A., 2010, **Forecasting with Spatial Panel Data**. Report: IZA DP n. 4242 IZA – Institute of Study of Labor, Bonn, Germany.
- BALTAGI, B. H., PIROTTE, A., 2010, **Panel Data Inference Under Spatial Dependence**”, Report: n. 123 Center for Policy Research of Maxwell School of Syracuse University, New York, USA.
- BANDEIRA, M. I. V. Q. B., 2005, **Metodologia de Priorização da Expansão da Rede de Distribuição de Gás Natural com Base na Sintaxe Espacial: uma Aplicação no Setor Residencial em Fortaleza**. Msc. Dissertação, Universidade Federal do Ceará, Fortaleza, Ceará.
- BURROUGH, P.A., MCDONNELL, R. A. 1998, **Principles of Geographical Information Systems (Spatial Information Systems)**. 2nd ed. England: Oxford University Press. ISBN: 0-198-23365-5.
- CAIO, L. S., BERMAN, C., 1998, “Análise das metodologias de previsão de mercado de energia elétrica face ao novo perfil de planejamento no ambiente pós-privatização”. In: **Anais do III Congresso de Planejamento Energético**. Disponível em: < <http://www.iee.usp.br/biblioteca/producao/1998/trabalho/anal1.pdf>. Acesso em: 01 fev. 2012, 11:42:00.
- CÂMARA, G., MONTEIRO, A. M., FUCKS, S. D, *et al.*, 2004, “Análise espacial e geoprocessamento”. In: Fucks, S. D., Carvalho, M. S., Câmara, G., Monteiro, A. V. M. (Eds). **Análise Espacial de Dados Geográficos**, 1 ed., capítulo 1, pp. 235-280, Brasília, DF, Embrapa Informações Tecnológicas. ISBN: 85-7383-260-6.
- CÂMARA, G., CARVALHO, M. S., CRUZ, O. G., *et al.*, 2004, “Análise espacial de áreas”. In: Fucks, S. D., Carvalho, M. S., Câmara, G., Monteiro, A. V. M. (Eds). **Análise Espacial de Dados Geográficos**, 1 ed., capítulo 5, pp. 235-280, Brasília, DF, Embrapa Informações Tecnológicas. ISBN: 85-7383-260-6.

- CASTRO, J. H., 2005, “A utilização das geotecnologias na atividade de planejamento estratégico de uma empresa”. In: **Anais do XXII Congresso Brasileiro de Cartografia**, Macaé, Rio de Janeiro, 26-30 Set.
- CASTRO, J. H., 2007, **O Emprego da Informação Espacial em Planejamento Estratégico: Uma Abordagem na Previsão de Demanda de Energia**. Msc. Dissertação, Universidade do Estado do Rio de Janeiro - UERJ, Rio de Janeiro.
- CHATFIELD, C., 2000, **Time-Series Forecasting**. 1st ed. Florida, USA, Chapman & Hall / CRC Press. ISBN: 1-58488-063-5.
- CHIAVENATO, I., SAPIRO, A., 2009, **Planejamento Estratégico: Fundamentos e Aplicações**. 2 ed. Rio de Janeiro, Elsevier - Campus.
- COUCLELIS, H., 1999, “Space, Time, Geography”. In: Longley, P. A., Maguire, D. J., Goodchild, M. F., *et al.* (eds), **Geographical Information Systems: Principles, Techniques, Management and Applications**, v. 1, 2nd ed., chapter 2, pp. 29-38, New York, USA, John Wiley & Sons. ISBN: 0-471-32182-6.
- COUGO, P. S., 1997, **Modelagem Conceitual e Projeto de Banco de Dados**. 1 ed. Rio de Janeiro, Campus. ISBN: 85-35201-58-0
- CRASSIE, N. A. C., 1993, **Statistics for Spatial Data – Revised edition**. New York, USA, John Wiley & Sons. ISBN: 0-471-00255-0.
- DEAL, B., SCHUNK, D., 2004, “Spatial dynamic modeling and urban land use transformation: a simulation approach to assessing the costs of urban sprawl”, **Ecological Economics**, v. 51, n. 1-2 (Nov), pp. 79-95.
- DENG, J. S., WANG, K., HONG, Y., *et al.*, 2009, “Spatio-temporal dynamics and evolution of land use change and landscape pattern in response to rapid urbanization”, **Landscape and Urban Planning**, v. 92, n. 3-4 (Set), pp. 187-198.
- DOS SANTOS, A. C. K., 1994, **Introdução a Modelagem Computacional na Educação**. Rio Grande, Editora FURG.

- DUBIN, R., 2004, “Spatial lags and spatial errors revisited: some Monte Carlo evidence”. In: **Spatial and Spatiotemporal Econometrics**, v. 18, **Advances in Econometrics**, London, UK, Elsevier, pp. 75–98. ISBN: 0-7623-1148-7.
- EGENHOFER, M. J., GOLLEDGE, R. G., 1998, **Spatial and Temporal Reasoning in Geographic Information Systems**. 1st ed. New York, Oxford Press. ISBN 0-19-510342-4.
- EGGING, R., GABRIEL, S. A., HOLZ, F., *et al.*, 2008, “A complementarity model for the European natural gas market”, **Energy Policy**, v. 36, n. 7 (Jul), pp. 2385-2414.
- ELHORST, J. P., 2001, **Panel Data Models Extended to Spatial Error Autocorrelation or a Spatially Lagged Dependent Variable**. In: Report 01C05, Research Institute SOM (Systems, Organisations and Management), University of Groningen, Netherlands.
- ELHORST, J. P., 2003, “Specification and estimation of spatial panel data models”, **International Regional Science Review**, v. 26, n. 3 (Jul), pp. 244-268.
- ELHORST, J. P., 2010, “Spatial panel data models”. In: Fischer M. M., Getis, A. (Eds.), **Handbook of Applied Spatial Analysis**, 1st ed., chapter C.2, pp. 377-408, Heidelberg, Germany, Springer-Verlag. ISBN: 978-3-642-03646-0
- EPE – EMPRESA DE PESQUISA ENERGÉTICA, 2008. **Balanco Energético Nacional 2008**. Empresa de Pesquisa Energética, Rio de Janeiro, Brasil.
- EPE – EMPRESA DE PESQUISA ENERGÉTICA, 2009. **Balanco Energético Nacional 2009: Resultados Preliminares**. Empresa de Pesquisa Energética, Rio de Janeiro, Brasil.
- ERWIG, M., GUTING, R.H., SCHNEIDER, M., *et al.*, 1999, “Spatio-temporal data types: an approach to modeling and querying moving objects in databases”, **GeoInformatica**, v. 3, n. 3(Jan), pp. 269-296.
- FISCHER, M. M., 2006, **Spatial Analysis and GeoComputation**, Heidelberg, Germany, Springer Berlin. ISBN: 3-540-35729-7.

- FLINGLETON, B., 2008, “A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors”, **Spatial Economic Analysis**, v. 3, n. 1 (Fev), pp. 27-44.
- FLINGLETON, B., Le GALLO, J., 2008, “Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties”, **Papers in Regional Science**, v. 87, n. 3 (Aug), pp. 319-339.
- FORLIZZI, L., GUTING, R.H., NARDELLI, E., *et al.*, 2000, “A data model and data structures for moving objects databases”. In: **Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data**, pp. 319 – 330, Dallas, Texas, USA, Mai. ISBN: 1-58113-217-4.
- FU, C. W., NGUYEN, T. T., 2003, “Models for long-term energy forecasting”. In: **Proceedings of the Power Engineering Society General Meeting – IEEE v.4**, pp. 1- 12, Canada, Jul. ISBN: 0-7803-7989-6.
- GALTON, A., 2004, “Fields and objects in space, time, and space-time”, **Spatial Cognition and Computation**, v. 4, n. 1, pp. 39-68.
- GATALSKY, P., ANDRIENKO, N., ANDRIENKO, G., 2004, “Interactive analysis of event data using space-time cube”. In: **Proceedings of the Eighth International Conference on Information Visualization**, pp. 145-152, Sankt Augustin, Germany, Jul. ISBN: 0-7695-2177-0.
- GETIS, A., 2009, “Spatial weights matrices”, **Geographical Analysis**, v. 41, n. 4 (Out), pp.404–410.
- GIVISIEZ, G. H. N., RIOS-NETO, E. L. G., SAWYER, D. O., 2006, “Projeção de demanda demográfica por domicílios: aplicação da metodologia das taxas de chefia baseada em modelos de Idade- Período-Coorte”. In: **Demografia dos Negócios: Campo de Estudo, Perspectivas e Aplicações**, v. 3, **Demographicas**, ABEP, pp. 213-255. ISBN 85-85543-14-0.
- GODET, M., 2000, “The art of scenarios and strategic planning: tools and pitfalls”, **Technological Forecasting and Social Change**, v. 65, n.1 (Set), pp. 3–22.

- GRIFFITHS, T., FERNANDES, A., PATON, N., *et al.*, 2002, “TRIPOD: A spatio-historical object database system”. In: **Mining Spatio-Temporal Information Systems**, v. 699, **The Springer International Series in Engineering and Computer Science**, Kluwer Academic Publisher, pp. 127-146. ISBN: 978-1-4020-7170-1.
- GRUMBACH, R. J. S., 2000, **Prospectiva: a Chave para o Planejamento Estratégico**. 2 ed. Rio de Janeiro, Catau.
- GRUMBACH, R. J. S., **Método Grumbach de Gestão Estratégica**. [S.2.]: Virtual book, 2010. Disponível em: < [http://www.mp.ba.gov.br/pga/instrumentos/aplicacao\\_metodo\\_grumbach.pdf](http://www.mp.ba.gov.br/pga/instrumentos/aplicacao_metodo_grumbach.pdf).> Acesso em: 01 fev. 2012, 13:25:00.
- HAKKERT, R., 2006, “Demografia dos negócios: campo de estudo, tendências e possibilidades”. In: **Demografia dos Negócios: Campo de Estudo, Perspectivas e Aplicações**, v. 3, **Demographics**, ABEP, pp. 19-73. ISBN 85-85543-14-0.
- HÖFFLER, F., KÜBLER, M., 2007, “Demand for storage of natural gas in northwestern Europe: Trends 2005–30”, **Energy Policy**, v. 35, n. 10 (Out), pp. 5206-5219.
- K., EGENHOFER, M. J., 1997, “Qualitative representation of change”. In: **Proceedings of the spatial information theory: a theoretical basis for GIS, International Conference COSIT 97**, v. 1329, **Lecture Notes in Computer Science**, Springer, pp. 15-33. ISBN 3-540-63623-4.
- HORNSBY, K. EGENHOFER, M. J., 2000, “Identity-based change: a foundation for spatio-temporal knowledge representation”, **International Journal of Geographical Information Science**, v. 14, n.3 (Abr), pp. 207-224.
- HORNSBY, K.; EGENHOFER, M. J, 2002, “Modeling moving objects over multiple granularities”, **Annals of Mathematics and Artificial Intelligence**, v. 36, n. 1-2, pp. 177-194.

- HUNTINGTON, H. G., 2007, “Industrial natural gas consumption in the United States: an empirical model for evaluating future trends”, **Energy Economics**, v. 29, n.4 (Jul), p.743–759.
- IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2008, **Contas Nacionais Trimestrais**. In: Série Relatórios Metodológicos, v. 28, 2 ed. IBGE, Rio de Janeiro. ISBN 978-85-240-4044-3.
- JANNUZI, G. M., SWISHER, J. N. P., 1997, **Planejamento Integrado de Recursos Energéticos: meio-ambiente, conservação e energia e fontes renováveis**. 1 ed. Campinas, Autores Associados. ISBN 85-85701-50-1.
- JANSEN, L. J. M., CARRAI, G., MORANDINI, L., *et al.*, 2006, “Analysis of the spatio-temporal and semantic aspects of land-cover/use change dynamics 1991–2001 in Albania at national and district levels”, **Environmental Monitoring and Assessment**, v. 119, n. 1-3, pp.107-136.
- JOHANNESSON, G., STEWART, J., BARR, C., *et al.*, 2006, **Spatial statistical procedures to validate input data in energy models**, Reports: NREL/TP-620-39497, National Renewable Energy Laboratory and UCRL-TR-218702, Lawrence Livermore National Laboratory, Colorado, USA.
- KAMERSCHEN, D. R., PORTER, D. V., 2004, “The demand for residential, industrial and total electricity, 1973–1998”, **Energy Economics**, v. 26, n.1 (Jan), pp. 87–100.
- KAPOOR, M., KELEJIAN, H. H., PRUCHA, I. R., 2007, “Panel data models with spatially correlated error components”, **Journal of Econometrics**, v. 140, n. 1 (Set), pp. 97-130.
- KHOTANZAD, A., AFKHAMI-ROHANI, R., MARATUKULAM, D., 1998, “ANNSTLF - artificial neural network short-term load forecaster generation three”, **IEEE Transactions on Power Systems**, v. 13, n. 4 (Nov), pp. 1413-1422.



- KHOTANZAD, A., ELRAGAL, H., LU, T., 2000, “Combination of artificial neural-network forecasters for prediction of natural gas consumption”, **IEEE Transactions on Neural Networks**, v. 11, n. 2 (Mar), pp. 464-473.
- KIRKWOOD, C. W., 1997, **Strategic Decision Making: Multiobjective Decision Analysis with Spreadsheets**. 1st ed. Belmont, Duxbury Press. ISBN: 0-534-51692-0
- KIRKWOOD, C. W. **System Dynamics Methods: A Quick Introduction**. [S.1.]: Virtual Book, 1998. Disponível em: < <http://www.public.asu.edu/~kirkwood/sysdyn/SDIntro/SDIntro.htm>.> Acesso em: 01 fev. 2012, 10:59:00.
- KRICHENE, N., 2005, **A Simultaneous Equations Model for World Crude Oil and Natural Gas Markets**. Report: WP/05/32 International Monetary Fund - IMF, Washington, D.C. Disponível em: < <http://www.imf.org/external/pubs/ft/wp/2005/wp0532.pdf> > Acesso em: 07 fev. 2012, 17:50:00.
- KRUGMAN, P. R.; FUJITA, M.; VENABLES, A. J., 2002, **Economia Espacial – Urbanização, prosperidade econômica e desenvolvimento humano no mundo**. São Paulo, Editora Futura. ISBN: 85-7413-123-7.
- LAURET, P., FOCK, E., RANDRIANARIVONY, R. N., *et al.*, 2008, “Bayesian neural network approach to short time load forecasting”, **Energy Conversion and Management**, v. 49, n. 5 (Mai), pp.1156–1166.
- LEE, L., 2007, “GMM and 2SLS estimation of mixed regressive, spatial autoregressive models”, **Journal of Econometrics**, v. 37, n. 2 (Abr), pp. 489-514.
- LESAGE, J. P. **The Theory and Practice of Spatial Econometrics**. [S.4.]: Virtual Book, 1999. Disponível em: < [www.spatial-econometrics.com/html/sbook.pdf](http://www.spatial-econometrics.com/html/sbook.pdf).> Acesso em: 03 fev. 2012, 16:55:00.
- LIU, H., LIU, D., ZHENG, G., *et al.*, 2004, “Research on Natural Gas load forecasting based on support vector regression”. In: **Proceedings of the 5th World Congress on Intelligent Control and Automation – WCICA 2004** v.4, pp. 3591A – 3595, China, Jun. ISBN: 0-7803-8273-0.

- LIU, H., LIU, D., LIANG, Y., *et al.*, 2004, “Research on Natural Gas load forecasting based on least squares support vector regression”. In: **Proceedings of the International Conference on Machine Learning and Cybernetics** v.5, pp. 3124 – 3128, China, Jun. ISBN: 0-7803-8403-2.
- LIU, J., ZHAN, J., DENG, X., 2005, “Spatio-temporal patterns and driving forces of urban land expansion in China during the economic reform era”, **Ambio: A Journal of the Human Environment**, v. 34, n. 6 (Ago), pp. 450-455.
- LLOYD, C. D., 2006, **Local Models for Spatial Analysis**. 1st ed. Abingdon, UK, CRC Press. ISBN: 0-415-31681-2.
- LYNESS, F. K., 1984, “Gas demand forecasting”, **Journal of the Royal Statistical Society. Series D (The Statistician)** v. 33, n. 1 (Mar) pp. 9-21.
- MADDALA, G. S., TROST, R. P., LI, H., *et al.*, 1997, “Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage estimators”, **Journal of Business & Economic Statistics**, v. 15, n. 1 (Jan), pp.90-100.
- MAKRIDAKIS S., WHEELWRIGHT, S. C., HYNDMAN R. J., 1998, **Forecasting, Methods and Applications**. 3rd ed. USA, John Wiley & Sons. ISBN 0-471-53233-9.
- MALLIARIS, M. E., MALLIARIS S. G., 2005, “Forecasting Energy Product Prices”. In: **Proceedings of the International Conference on Neural Network** v. 5, pp. 3284-3289, Canada, Ago. ISBN: 0-7803-9048-2.
- MASSARA, V. M., 2007, **A Dinâmica Urbana na Otimização da Infra-Estrutura para o Gás Natural**. Dsc. Tese, Universidade de São Paulo - USP, São Paulo.
- MELICIANI, V., PERACCHI, F., 2006, “Convergence in pre-capita GDP across European regions: a reappraisal”, **Journal Empirical Economics**, v. 31, n.3 (Jun), pp. 549-568.

- MENNIS, J., LIU, J. W., 2005, “Mining association rules in spatio-temporal data: an analysis of urban socioeconomic and land cover change”, **Transactions in GIS**, v. 9, n.1. pp. 5–17.
- MISHRA, S., PATRA, S. K., 2008, “Short term load forecasting using a novel recurrent neural network”. In: **Proceedings of the TENCON 2008 - IEEE Region 10 Conference Publication**, pp. 1-6, Hyderabad, India, 19-21 Nov. 2008.
- MISHRA, S., PATRA, S. K., 2008, “Short term load forecasting using neural network trained with genetic algorithm & particle swarm optimization”. In: **Proceedings of the First International Conference on Emerging Trends in Engineering and Technology – ICETET 2008**, pp. 606 – 611, Maharashtra, India, 16 -18 Jul. 2008. E-ISBN: 978-0-7695-3267-7
- MITCHELL, T. M., 1997, **Machine Learning**. 1st ed. USA, McGraw Hill. ISBN: 0-070-42807-7.
- MONTGOMERY, D. C., RUNGER, G. C., 2009, **Estatística Aplicada e Probabilidade para Engenheiros**. 4 ed. Rio de Janeiro, Brasil, Livros Técnicos e Científicos. ISBN: 978-8-521-61664-1.
- MORETTIN P. A., TOLOI, C. M. C., 2004, **Análise de Séries Temporais**. 2 Ed. Revista e Ampliada, São Paulo, Editora Edgard Blücher. ISBN: 85-212-0389-6.
- NETO, S. L. R., 2000, **Um Modelo Conceitual de Sistemas de Apoio a Decisão Espacial para Gestão de Desastres por inundações**. Dsc. Tese, Escola Politécnica da Universidade de São Paulo, São Paulo.
- NETO, S. L. R., RODRIGUES, M., 2001, “Um modelo conceitual para integração de modelos científicos e informações geográficas”. In: **Anais do III Workshop Brasileiro de Geoinformática - GeoInfo 2001**, v. 1, pp. 71-78, Rio de Janeiro, 4-5 Out.
- PACE, R. K., BARRY, R., 1997, “Quick computation of spatial autoregressive estimators”, **Geographical Analysis**, v. 29, n. 3 (Jul), pp. 232-247.

- PACE, R. K., BARRY, R., 1998, "Simulating mixed regressive spatially autoregressive estimators", **Computational Statistics**, v. 13, pp. 397-418.
- PACE, R. K., BARRY, R., 1999, "Monte Carlo estimates of the log determinant of large sparse matrices", **Linear Algebra and its Applications**, v. 289, n. 1-3 (Mar), pp. 41-54.
- PAI, P., HONG, W., 2005, "Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms", **Electric Power Systems Research**, v. 74, n. 3 (Jun), pp. 417-425.
- PANAGIOTIDISA, T., RUTLEDGE, E., 2007, "Oil and gas markets in the UK: evidence from a cointegrating approach", **Energy Economics**, v. 29, n. 2 (Mar), pp. 329-347.
- PAPALEXOPOULOS, A. D., HESTERBERG, T. C., 1990, "A regression-based approach to short-term system load forecasting", **IEEE Transactions on Power Systems**, v. 5, n. 4 (Nov), pp. 1535-1547.
- PARTSINEVELOU, P., STEFANIDIS, A., AGOURIS, P., 2001, "Automated spatiotemporal scaling for video generalization". In: **Proceedings of the IEEE International Conference on Image Processing (ICIP)** v. 1, pp. 177-180, Greece, Oct. ISBN: 0-7803-6725-1.
- PEDROSA, B. M., CÂMARA, G., 2007, "Modelagem dinâmica e sistemas de informações geográficas". In: Meirelles, M. S. P., Câmara, G., Almeida, C. M. (Eds.), **Geomática – Modelos e Aplicações Ambientais**, 1 ed., capítulo 5, pp. 235-280, Brasília, DF, Embrapa Informações Tecnológicas. ISBN: 978-85-7383-386-7.
- PERSAUD, A. J., KUMAR, U., 2001, "An eclectic approach in energy forecasting: a case of natural resources Canada's (NRCan's) oil and gas outlook", **Energy Policy**, v. 29, n.4 (Mar), p.303-313.
- PEUQUET, D. J., 2001, "Making space for time: issues in space-time data representation", **GeoInformatica**, v. 5, n.1 (Mar), pp.11-32.

- PEUQUET, D. J., DUAN, N., 1995, “An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data”, **International Journal of Geographical Information Systems**, v. 9, n. 1, pp. 7-24.
- PINDYCK, R. S., RUBINFELD, D. L., 2002, **Microeconomia**. 5. ed. São Paulo. Prentice Hall. ISBN: 85.87918-11-7.
- QUAN, Z., CAIXIN, S., GUOQING, C., *et al*, 2002, “GIS based distribution system spatial load forecasting and the optimal planning of substation location and capacity”. In: **Proceedings of the International Conference on Power System Technology, PowerCon 2002**, v. 2, pp. 885- 889, China, 13-17 Out. ISBN: 0-7803-7459-2
- RICHARDSON, H. W., 1975, **Economia Regional – Teoria da Localização, Estrutura Urbana e Crescimento Regional**. Rio de Janeiro, Zahar editores.
- SÁNCHEZ-ÚBEDA, F., BERZOSA, A., 2007, “Modeling and forecasting industrial end-use natural gas consumption”, **Energy Economics**, v. 29, n. 4 (Jul), pp.710–742.
- SANTOS, E. M., ZAMALLOA, G. C., VILLANUEVA, L. D. *et al.*, 2002, **Gás Natural: Estratégias para uma Energia Nova no Brasil**. 1 ed. São Paulo, Annablume. ISBN: 85-7419-285-6.
- SCHMIDT, C. A. J., LIMA, M. A. M., 2004, “A demanda por energia elétrica no Brasil”, **Revista Brasileira de Economia**, v. 58, n.1 (Jan), pp. 67-98.
- SEETHA, H., SARAVANAN. R., 2007, “Short term electric load prediction using Fuzzy BP”, **Journal of Computing and Information Technology**, v.15, n. 3, pp. 267–282.
- SHORT, W., 2007, **Region in Energy Market Models**. Report: NREL/TP-640-40506, National Renewable Energy Laboratory, Colorado, USA.

- SIEMEK, J., NAGY, S., RYCHLICKI, S., 2003, “Estimation of natural-gas consumption in Poland based on the logistic-curve interpretation”, **Applied Energy**, v. 75, n. 1-2 (Mai-Jun), pp. 1-7.
- SILVA, J. X., 2001, **Geoprocessamento para análise ambiental**. Rio de Janeiro: Sermograf – Artes gráficas e editora LTDA. ISBN 85-902162-1-7.
- SKOVSMOSE, O., 1988, “Mathematics as part of technology - elements of philosophy of an applied oriented mathematical education”, **Educational Studies in Mathematics**, v.19, n.1 (Mai). pp. 23-41.
- SNODGRASS, R. T., 2000, **Developing Time-oriented Database Applications in SQL**. California, Morgan Kaufmann Publishers. ISBN: 1-55860-436-7.
- TOLMASQUIM, M. T., GUERREIRO, A., GORINI, R., 2007, “Matriz energética brasileira: uma prospectiva”. **Revista Novos Estudos - CEBRAP**, n.79 (Nov), pp. 47-69.
- VAONA, A., 2008, “Regional evidence on financial development, finance term structure and growth”, **Journal Empirical Economics**, v. 34, n.1 (Fev), pp. 185-201.
- VERBEEK, M., 2004, **A Guide to Modern Econometrics**. 2nd ed. England, John Wiley & Sons. ISBN 0-470-85773-0.
- VIANNA, J. C. T., 2005, “O pensamento sistêmico e a modelagem”, **Revista Didática Sistêmica**, v. 1(Out), PP. 90-105.
- WERBOS, P. J., 1988, “Generalization of backpropagation with application to a recurrent gas market model”, **Neural Networks**, v. 1, n.4, pp. 339-356.
- WILLIS, H. L., **Spatial Electric Load Forecaasting**. 2nd ed. New York USA, Marcel Dekker. ISBN 0-8247-0840-7.
- WOOLDRIDGE, J. M., 2002, **Econometric Analysis of Cross Section and Panel Data**. 1st ed. Massachusetts, USA, The MIT Press. ISBN: 0-262-23219-7.

- WORBOYS, M. F., 1994, "A unified model for spatial and temporal information", **The Computer Journal**, v. 37, n.1, pp. 26-33.
- WORBOYS, M. F., DUCKHAM, M., 2004, **GIS: A Computing Perspective**. 2nd ed. New York, CRC Press. ISBN: 0-415-28375-2.
- WORBOYS, M. F., HORNSBY, K., 2004, "From objects to events: GEM, the geospatial event model". In: **Proceedings of the GIScience 2004**, pp.327-344, USA, Out.
- YANG, Y., ZHENG, G., LIU, D., 2001, "BP-GA mixed algorithms for short-term load forecasting". In: **Proceedings of the ICII 2001 International Conferences on Info-tech and Info-net v.4**, pp. 334-339, Beijing. ISBN: 0-7803-7010-4.
- YOO, H., PIMMEL, R. L., 1999, "Short term load forecasting using a self-supervised adaptive neural network", **IEEE Transactions on Power Systems**, v.14, n. 2 (Mai), pp. 779-784.

## APÊNDICE (A) - CONTEXTUALIZAÇÃO DA INDÚSTRIA DO GN

Em linhas gerais, a Indústria do GN é constituída pelas atividades de exploração, produção, processamento, transporte e distribuição.

A exploração e o desenvolvimento da produção englobam o estudo, reconhecimento e delimitação das jazidas potenciais; entretanto as técnicas geológicas, geofísicas e geoquímicas fornecem prognósticos sobre a presença de óleo e gás. O único meio de ter certeza da existência comercial destes insumos é perfurando.

O processamento do gás é realizado nas UPGN, onde o mesmo é desidratado e fracionado gerando basicamente o metano e o etano, que formam o GN, o propano e o butano que formam o GLP e o Líquido de Gás Natural.

A forma mais usual de transporte do GN é no estado gasoso através de gasodutos, que no Brasil são classificados em transporte, transferência e distribuição. Inicialmente o GN é transportado até os *city-gates* das CDL, onde localizam-se os equipamentos para transferir o gás dos gasodutos de transporte para os de distribuição, que são empregados pelas CDL para levar o GN até o consumo final. Os gasodutos de transferência destinam-se a levar o GN entre unidades produtivas da mesma empresa.

O GN pode ainda ser transportado no estado líquido, na forma de Gás Natural Liquefeito (GNL), ou, comprimido, na forma de Gás Natural Comprimido (GNC). Na figura A.1 encontra-se um sinóptico da cadeia do GN onde são apresentadas as cinco etapas, bem como os entroncamentos produtivos.

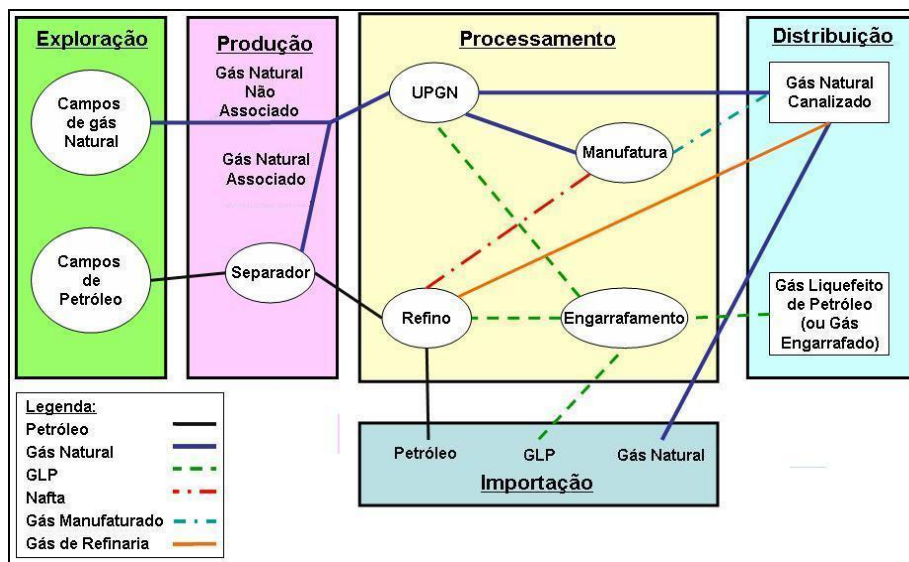


Figura A.1 – Cadeia produtiva do Gás Natural - Fonte: Adaptada de SANTOS *et al.* (2002)

A distribuição, última etapa da cadeia, é responsável por levar o GN até o consumidor final. As CDL's utilizam os mesmos modais de transporte, porém em menor escala, visto os volumes distribuídos e a distância percorrida. Ressalta-se que por



determinação constitucional a distribuição de GN cabe às unidades federativas brasileiras, e estas constituem as concessões de distribuição às respectivas CDL.

O modelo de comercialização do GN é por contrato, tanto nas relações entre os produtores com os distribuidores, quanto destes para seus consumidores finais. Devido à sua própria natureza tangibilizam acordos de longa duração. No Brasil, os instrumentos contratuais existentes na cadeia produtiva do GN são basicamente de dois tipos: contratos de comercialização (compra/venda) e os de transporte.

O GN, seja de proveniência nacional ou estrangeira, pode ser vendido diretamente a CDL ou através de um atravessador. Entretanto, a venda ao consumidor final é exclusividade da distribuidora; em qualquer uma dessas situações, deve-se constituir um contrato de comercialização formalizando a operação.

No tocante aos segmentos de consumo podem ser identificados como principais usuários os segmentos: Industrial, Veicular, Comercial, Residencial e Termoelétrico.

#### **A.1 – O GN NO SEGMENTO INDUSTRIAL**

O segmento industrial representa a maior fatia de consumo de GN no mercado nacional, em parte pelo fato que a implementação da distribuição de GN nas diversas CDL foi iniciada por este segmento.

Tal característica tem um aspecto técnico muito importante: cada ramal da rede atende a particularidades de pressão e vazão específicos, desta forma o projeto da rede de distribuição tem que considerar a limitação máxima de carga possível por ramal.

Sendo assim, é natural que se busque atender os maiores consumidores potenciais de GN em volume, neste caso, o segmento industrial. Na prática o que se observa é que as CDL chegam a estabelecer ramais específicos para atender a indústria.

Outro aspecto importante é a estratégia de financiamento das redes, que fez com que as CDL procurassem naturalmente priorizar os consumidores com maiores demandas de GN e com fluxo de fornecimento estável; condições estas desejáveis ao desenvolvimento de soluções de engenharia de dutos.

Além destes aspectos, podem ser citadas como exemplo de benefícios da adoção do GN, pelo segmento industrial, as seguintes vantagens:

- Possuir combustão completa não produzindo cinzas;
- Apresentar melhor regulagem da chama em fornos de queima;
- Melhor distribuição de calor e temperatura mais constante, reduzindo às variações nos processos industriais;

- Não exigir tratamento dos gases de combustão e possui rápida dispersão;
- Reduzir áreas de tancagem e armazenamento; e,
- Reduzir o tempo de parada de equipamentos para manutenção.

De certo o crescimento do consumo do GN na indústria depende basicamente da capacidade deste combustível deslocar seus principais energéticos concorrentes, em especial o óleo combustível.

Como destaques citam-se em maior escala a indústria química; ferro gusa e aço e, em menor grau, papel e celulose. Outros setores de elevado destaque no emprego do GN em seus processos produtivos são os de cerâmica, vidro, e alimentos e bebidas.

## **A.2 – O GN NO SEGMENTO VEICULAR**

Regulamentada pela portaria ANP Nº 32, DE 6.3.2001, a revenda varejista de GNV tinha o objetivo inicial de substituir o diesel como combustível para veículos pesados em centros urbanos. Isto ocorreu entre 1982 e 1991 e teve como principais dificuldades: a falta de rede de postos de abastecimento, a baixa diferença percebida entre o valor do GNV e o diesel e a falta de tecnologia de conversão.

Em um segundo momento, entre 1992 e 1994, houve a liberação do uso do GNV para as frotas de empresas e taxis. O alto consumo de combustível destes consumidores viabilizava economicamente seu uso; entretanto, o não crescimento da rede de postos de abastecimento restringiu o acesso criando a falta do produto.

Entre 1997 e 2005 ocorre a elevação do preço da gasolina e a liberação do emprego do GNV para veículos particulares. Associados ao baixo interesse pela aquisição de novos automóveis movidos a álcool permeiam a popularização do consumo do GNV. É o período onde cresce o processo de conversão das frotas.

A partir de 2006 o segmento de transporte experimenta o ressurgimento da indústria do álcool, a entrada dos biocombustíveis e a adoção por parte das montadoras da estratégia de veículos dual e total flex. Acentua-se também a questão ambiental, onde apesar do GNV contar com incentivo dos programas de controle de poluição, que tem o diesel como alvo principal, teve sua adoção comprometida uma vez que o diferencial entre o preço deste e o diesel é menor.

No sentido de dar garantias às montadoras e impulsionar a multiplicação da frota, a Petrobras em 2004 atrelou o preço do GNV ao diesel, na proporção fixa correspondente a 55% do preço do diesel para os próximos 10 anos.

### **A.3 – O GN NO SEGMENTO RESIDENCIAL**

O segmento residencial tipicamente de varejo apresenta como característica intrínseca o consumo em menores volumes, porém em grande escala de consumidores.

Em geral, este segmento não é o foco estratégico da maior parte das CDL, uma vez que apresenta reduzida taxa de consumo por habitante que justifique os investimentos de expansão da rede; bem como, as CDL's, como já salientado, preferem priorizar o mercado industrial como âncora para capilarizar a rede.

Existem outros dois pontos que contribuem com a redução do crescimento deste segmento: as adequações de infra-estrutura e a falta de equipamentos domésticos à GN.

As adequações na infra-estrutura residencial para receber o GN são forte limitador da adoção do energético por demandar do consumidor investimento financeiro.

Salvo exceções, em regiões que contam com CDL's há várias décadas, ou as edificações mais novas, que seguem novos padrões de projeto determinados em códigos municipais de construção civil, na grande maioria do país é necessário por parte do consumidor investimento para adequar uma instalação.

O outro ponto é o pequeno espectro de possibilidades de uso pela ausência de equipamentos a GN, como por exemplo, condicionadores de ar, aquecedores e geladeiras, que fomentem o uso e maximizem o investimento por parte do consumidor.

Na verdade apesar de existirem iniciativas no sentido de desenvolver tecnologia associada ao setor o fato é que não existe uma indústria consolidada de equipamentos a GN para uso doméstico. Tal realidade restringe o seu emprego à cocção de alimentos e ao aquecimento de água para o banho, indicando claramente o GLP e a EE como combustíveis concorrentes, e mais recentemente o aquecimento solar fruto do desenvolvimento da indústria nacional de placas e *boilers* de emprego residencial.

Apesar desta preferência inicial pelo segmento industrial, pelas razões já apresentadas, é possível verificar na análise das CDL empregadas nesta tese, que o setor experimenta crescimento na base de clientes e no consumo nos últimos anos.

### **A.4 – O GN NO SEGMENTO COMERCIAL**

O segmento comercial do GN apresenta as mesmas particularidades inerentes ao segmento residencial. Em especial pode-se identificar uma maior demanda unitária em função da característica do tipo e volume consumido.

Com relação aos combustíveis concorrentes, além do GLP e da EE, percebe-se a presença do óleo diesel empregado em alguns sistemas *backup*, como os de segurança

contra incêndio e geração auxiliar de EE, principalmente em grandes estabelecimentos comerciais como shoppings, serviços de entretenimento e rede hoteleira.

Nestes casos para o resfriamento de ar; uso de *boilers* e aquecimento de água; e, para a co-geração com produção simultânea de frio. As principais dificuldades para a penetração do GN neste segmento são os baixos investimentos em expansão na rede de distribuição, o desenvolvimento de uma indústria produtora de equipamentos à GN e a baixa promoção ao uso de combustíveis limpos.

#### **A.5 – O GN NO SEGMENTO TERMO-ELÉTRICO**

Até 1990, foi neste segmento que se verificou a mais importante inovação tecnológica propiciando o emprego do GN: a evolução das turbinas aeroderivadas (aviação) para as atuais turbinas a GN à geração de eletricidade. Entretanto, as indefinições decorrentes do processo de reforma do setor elétrico constituíram barreiras à difusão do GN como insumo à produção termelétrica.

Entre 1993 e 2003, período conhecido como a 1ª Reforma Elétrica Brasileira, houve alterações em aspectos institucionais, regras de comercialização, de despacho e na política de preços. Neste período, foram dados vários passos à reestruturação da estrutura industrial e do desenho institucional do setor elétrico brasileiro.

Apesar disto, não havia ofertas de preço no sistema brasileiro, principalmente pelo cálculo da oportunidade das hidroelétricas considerar que todas as usinas pertenciam a um único dono, assegurando assim a predominância da geração hidrelétrica. Tal situação gerou um desequilíbrio, pois as usinas depreciadas operavam com boas margens individuais, mas, criaram um *markup* de preço baixo, reduzindo o interesse por novos empreendimentos.

Isto influenciou a não execução do Plano Prioritário Termoelétrico, PPT, que teve seu sinal de preço comprometido pela indefinição das regras de comercialização da energia térmica a ser gerada e a incerteza quanto ao preço do GN.

A partir de 2004 estabelece-se um novo modelo que regulamenta as regras de comercialização, o processo de outorga e de autorizações de concessões; e que objetiva estabelecer a modicidade tarifária e garantir a segurança do abastecimento.

A idéia geral do novo modelo era criar um ambiente de contratação entre todos os geradores e todos os distribuidores, mediante um contrato com a Câmara de Comercialização de Energia Elétrica. Esta Câmara seria responsável por comercializar

as sobras de energia fruto da diferença entre a energia contratada pela distribuidora e o consumo real de seu mercado, bem como pelas compensações dos déficits.

Em longo prazo este modelo resgata a diretriz do planejamento da oferta de energia, independente da fonte, institucionalizando-se este planejamento através da criação da Empresa de Pesquisa Energética.

No curto prazo o maior desafio é lidar com as conseqüências de um possível aumento de demanda elétrica sobre a demanda e a infra-estrutura de transporte de GN, onde, três pontos críticos vêm sendo mitigados de forma a equacionar a questão da oferta de GN e a capacidade de transporte:

- Os regimes contratuais que presidem a comercialização do GN (lastro físico);
- Regras operacionais de despacho térmico que permitam equalizar a oferta; e,
- O aumento da capacidade de transporte de GN.

O crescimento da participação termoelétrica na matriz energética, independente do combustível, é um papel previsto dentro das atribuições da EPE. De forma cíclica são apresentados o Plano Decenal da Expansão de Energia e o Nacional da Expansão.

Entretanto, tais planos são de nível estratégico, não constituindo um conjunto de ações de fato. Na prática os empreendimentos elencados no PDE podem não se realizar, assim como é possível que outros, oriundos de iniciativa privada, venham a ser incluídos e ofertados em leilões.

No âmbito do PDE, com abrangência de médio e longo prazo, as novas iniciativas e os atrasos são atualizados anualmente no corpo do plano, compondo um novo cenário decenal de empreendimentos, demanda e oferta no macro cenário.

Por outro lado, no curto prazo, a demanda real de GN é conseqüência do plano de despacho do ONS, que é executado considerando: o lastro hídrico dos reservatórios, o balanço do fluxo de energia entre os subsistemas do SIN e a manutenção da menor custo marginal para a energia gerada.

Neste conjunto de particularidades do segmento termo-elétrico cabe a CDL, em alguns casos, somente criar a infra-estrutura necessária ao atendimento do empreendimento e a prestação do serviço da operação dentro de sua área de concessão.

Sendo assim não tem ação de planejamento de rede considerando o crescimento da demanda de mercado de sua área de atuação, visto que o planejamento de novos empreendimentos de geração e do despacho das usinas existentes não é de sua competência; desta forma este segmento em particular não será abordado na tese.

## APÊNDICE (B) - PROCEDIMENTO DE ESPACIALIZAÇÃO DE CLIENTES

Para realizar a espacialização no ArcGIS é preciso criar um *Address Locator*, que contém uma espécie de modelo para referenciar os dados, neste caso composto de um conjunto de dados que contém informações básicas acerca dos atributos dos endereços, índices e consultas necessárias para realizar a espacialização dos clientes.

Para efetuar a criação do *Address Locator* é necessário definir um estilo. O ArcGIS possui vários estilos, conforme apresentado na tabela B1 a seguir.

Tabela B.1 – Estilos de Adress Locator.

| Estilo               | Geometria           | Representação dos dados de referência                                       | Parâmetros para busca do endereço                                 | Exemplo            | Aplicações  |
|----------------------|---------------------|---|---|--------------------|---|
| US Streets           | linhas              | Intervalo de endereços para os dois lados do segmento de rua                | Todos os elementos de endereço em um único campo                  | 320 Laranjal R.    | Localizar uma casa em lado específico da rua                      |
| US Alphanumeric      | linhas              | Endereço que contenha informação da zona de grade                           | Todos os elementos de endereço em um único campo                  | N2W1700 County Rd. | Usado em algumas regiões dos EUA                                  |
| US Hyphenated        | linhas              | Informações de cruzamentos entre ruas                                       | Todos os elementos de endereço em um único campo                  | 105-30 Union St.   | Usado em algumas regiões dos EUA                                  |
| US One Range         | linhas              | Um intervalo para cada segmento de rua                                      | Todos os elementos de endereço em um único campo                  | 2 Nunes Machado R. | Localizar uma casa em uma rua onde o lado da rua não é necessário |
| US One Address       | Pontos ou Polígonos | Cada feição representa um endereço  | Todos os elementos de endereço em um único campo                  | 71 Cherry Ln.      | Localizar parcelas, prédios ou pontos de endereço                 |
| Single Field         | Pontos ou Polígonos | Cada feição representa um lugar em particular                               | Variável única, especificada pelo usuário                         | Cabrillo College   | Procurar por nomes de lugares identificados por campo único       |
| US Cities with State | Pontos ou Polígonos | Cidades em relação a um Estado  | Nome da Cidade e o nome do Estado ou sua abreviação               | Rice, WA           | Localizar uma cidade específica em um país                        |
| World Cities         | Pontos ou Polígonos | Cidades em relação ao País  | Nome da cidade e do País  | Lima, Peru         | Localizar uma cidade específica no mundo                          |
| ZIP de 5 dígitos     | Pontos ou Polígonos | Região com Zip Code ou centróide  | Zip Code de 5 dígitos   | 22066              | Localizar uma região específica que use o sistema Zip Code        |
| ZIP +4 Range         | Pontos ou Polígonos | Cada feição representa um ZIP code e uma extensão de intervalo de 4 dígitos | ZIP codes de 5 dígitos e extensões de 4 dígitos em campo separado | 63703, 0078        | Localizar uma região específica que use o código ZIP+4            |

Cada um dos estilos possui características específicas, a escolha depende do tipo de dado que se quer localizar, bem como do formato do endereço.

Nesta tese, a tabela de atributos apresenta um campo denominado *NOINICIO* (número inicial do trecho de logradouro), um campo denominado *NOFIM* (número final do trecho de logradouro) e *NM VIA* (nome do logradouro), conforme pode ser observado no exemplo apresentado na figura B.1.

| OBJECTID_1* | Shape*   | OBJECTID | CHAVE           | CODVIA | NOINICIO | NOFIM | COD_LOGGTM | STATUS  | NM VIA                            |
|-------------|----------|----------|-----------------|--------|----------|-------|------------|---------|-----------------------------------|
| 8217        | Polyline | 8430     | N262B0008000135 | N262B  | 00080    | 00135 | 1609       | OFICIAL | R. ALFAZEMA                       |
| 3383        | Polyline | 3311     | E226E0000000055 | E226E  | 00000    | 00055 | 461        | OFICIAL | R. ALFERES ALFREDO ANTÔNIO MÜLLER |
| 3384        | Polyline | 3312     | E226E0005500430 | E226E  | 00055    | 00430 | 461        | OFICIAL | R. ALFERES ALFREDO ANTÔNIO MÜLLER |
| 13626       | Polyline | 13558    | S010 0000000120 | S010   | 00000    | 00120 | 3773       | OFICIAL | R. ALFERES ANGELO SAMPAIO         |
| 13627       | Polyline | 13559    | S010 0012000260 | S010   | 00120    | 00260 | 3773       | OFICIAL | R. ALFERES ANGELO SAMPAIO         |
| 13628       | Polyline | 13560    | S010 0026000420 | S010   | 00260    | 00420 | 3773       | OFICIAL | R. ALFERES ANGELO SAMPAIO         |
| 13629       | Polyline | 13561    | S010 0042000585 | S010   | 00420    | 00585 | 3773       | OFICIAL | R. ALFERES ANGELO SAMPAIO         |
| 13630       | Polyline | 13562    | S010 0058500745 | S010   | 00585    | 00745 | 3773       | OFICIAL | R. ALFERES ANGELO SAMPAIO         |
| 13631       | Polyline | 13563    | S010 0074500920 | S010   | 00745    | 00920 | 3773       | OFICIAL | R. ALFERES ANGELO SAMPAIO         |
| 13632       | Polyline | 13564    | S010 0092001080 | S010   | 00920    | 01080 | 3773       | OFICIAL | R. ALFERES ANGELO SAMPAIO         |
| 13633       | Polyline | 13565    | S010 0108001230 | S010   | 01080    | 01230 | 3773       | OFICIAL | R. ALFERES ANGELO SAMPAIO         |
| 13634       | Polyline | 13566    | S010 0123001405 | S010   | 01230    | 01405 | 3773       | OFICIAL | R. ALFERES ANGELO SAMPAIO         |
| 13635       | Polyline | 13567    | S010 0140501645 | S010   | 01405    | 01645 | 3773       | OFICIAL | R. ALFERES ANGELO SAMPAIO         |
| 13636       | Polyline | 13568    | S010 0164501830 | S010   | 01645    | 01830 | 3773       | OFICIAL | R. ALFERES ANGELO SAMPAIO         |
| 13637       | Polyline | 13569    | S010 0183001965 | S010   | 01830    | 01965 | 3773       | OFICIAL | R. ALFERES ANGELO SAMPAIO         |
| 13638       | Polyline | 13570    | S010 0196502090 | S010   | 01965    | 02090 | 3773       | OFICIAL | R. ALFERES ANGELO SAMPAIO         |
| 13639       | Polyline | 13571    | S010 0209002230 | S010   | 02090    | 02230 | 3773       | OFICIAL | R. ALFERES ANGELO SAMPAIO         |

Figura B.1 – Atributos da camada de informação de logradouros de Curitiba

Com base na tabela B1 verifica-se que o estilo adequado aos dados de Curitiba é o *US One Range*, que não faz distinção entre lado direito ou esquerdo da via. No caso de Vitória, a tabela tem outras características e adotou-se o estilo *US Streets*, que considera o lado da via. Para entender isto na prática, tome-se como exemplo a figura B.2, onde estão representados alguns logradouros da malha viária. O trecho em azul é parte de um logradouro com indicação do número inicial ( $NI=NOINICIO$ ) e final ( $NF=NOFIM$ ).

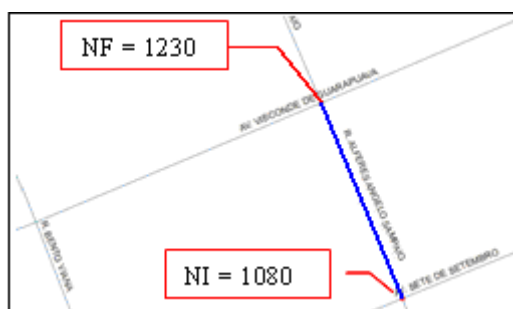


Figura B.2 – Estrutura da camada logradouro.

Para efetuar o procedimento de geocodificação é necessária, além do *Address Locator*, construído com o estilo adequado, uma tabela com os endereços e serem especializados. Esta tabela deve conter um campo com a estrutura definida pelo estilo do localizador. A figura B.3 apresenta parte da tabela de endereços dos clientes de Curitiba. Observa-se que o campo *NOME* contém o número e o nome do logradouro.

| CODIGO | NIP | SEGNUM | NOME                               |
|--------|-----|--------|------------------------------------|
| 26     | 28  | 26     | 0 PÇ. OSWALDO CRUZ                 |
| 37     | 41  | 37     | 0 PÇ. OUVADOR PARDINHO             |
| 206    | 293 | 206    | 10 AV. PRESIDENTE AFFONSO CAMARGO  |
| 129    | 193 | 129    | 10 R. JOSE RIETMEYER               |
| 70     | 80  | 70     | 1100 R. ALFERES ANGELO SAMPAIO     |
| 232    | 321 | 232    | 1100 R. BENTO VIANA                |
| 217    | 305 | 217    | 1148 AL. DOUTOR CARLOS DE CARVALHO |
| 145    | 215 | 145    | 1154 AL. DOUTOR CARLOS DE CARVALHO |
| 219    | 308 | 219    | 1168 R. VICENTE MACHADO            |

Figura B.3 – Parte da tabela de endereços de clientes de Curitiba

Assim, o que o *Address Locator* faz é buscar na tabela da camada espacial o nome do logradouro e em seguida analisar em qual trecho o ponto correspondente ao cliente está localizado, como por exemplo, os endereços em destaque na figuras B4.

| CÓDIGO | NIP | NOME                               |
|--------|-----|------------------------------------|
| 26     | 28  | 0 PÇ. OSWALDO CRUZ                 |
| 37     | 41  | 0 PÇ. OUVIDOR PARDINHO             |
| 206    | 293 | 10 AV. PRESIDENTE AFFONSO CAMARGO  |
| 129    | 193 | 10 R. JOSE RIETMEYER               |
| 55     | 63  | 100 R. DOMINGOS STRAPASSON         |
| 244    | 333 | 1093 AL. JULIA DA COSTA            |
| 70     | 80  | 1100 R. ALFERES ANGELO SAMPAIO     |
| 232    | 321 | 1100 R. BENTO VIANA                |
| 217    | 305 | 1148 AL. DOUTOR CARLOS DE CARVALHO |
| 145    | 215 | 1154 AL. DOUTOR CARLOS DE CARVALHO |
| 219    | 308 | 1168 R. VICENTE MACHADO            |
| 156    | 227 | 1177 R. ALFERES ANGELO SAMPAIO     |

| OBJECTID 1* | Shape *  | CHAVE           | CODVIA | NOINICIO | NOFIM | NMVA                      |
|-------------|----------|-----------------|--------|----------|-------|---------------------------|
| 8217        | Polyline | N262B0008000135 | N262B  | 00080    | 00135 | R. ALFAZEMA               |
| 3383        | Polyline | E226E0000000055 | E226E  | 00000    | 00055 | R. ALFERES ANGELO SAMPAIO |
| 3384        | Polyline | E226E0005500430 | E226E  | 00055    | 00430 | R. ALFERES ANGELO SAMPAIO |
| 13626       | Polyline | S010 0000000120 | S010   | 00000    | 00120 | R. ALFERES ANGELO SAMPAIO |
| 13627       | Polyline | S010 0012000260 | S010   | 00120    | 00260 | R. ALFERES ANGELO SAMPAIO |
| 13628       | Polyline | S010 0026000420 | S010   | 00260    | 00420 | R. ALFERES ANGELO SAMPAIO |
| 13629       | Polyline | S010 0042000585 | S010   | 00420    | 00585 | R. ALFERES ANGELO SAMPAIO |
| 13630       | Polyline | S010 0058500745 | S010   | 00585    | 00745 | R. ALFERES ANGELO SAMPAIO |
| 13631       | Polyline | S010 0074500920 | S010   | 00745    | 00920 | R. ALFERES ANGELO SAMPAIO |
| 13632       | Polyline | S010 0092001080 | S010   | 00920    | 01080 | R. ALFERES ANGELO SAMPAIO |
| 13633       | Polyline | S010 0108001230 | S010   | 01080    | 01230 | R. ALFERES ANGELO SAMPAIO |
| 13634       | Polyline | S010 0123001405 | S010   | 01230    | 01405 | R. ALFERES ANGELO SAMPAIO |
| 13635       | Polyline | S010 0140501645 | S010   | 01405    | 01645 | R. ALFERES ANGELO SAMPAIO |
| 13636       | Polyline | S010 0164501830 | S010   | 01645    | 01830 | R. ALFERES ANGELO SAMPAIO |
| 13637       | Polyline | S010 0183001965 | S010   | 01830    | 01965 | R. ALFERES ANGELO SAMPAIO |
| 13638       | Polyline | S010 0196502090 | S010   | 01965    | 02090 | R. ALFERES ANGELO SAMPAIO |

Figura B.4 – Endereço a ser espacializado (a) x trecho de logradouro que o endereço está contido (b).

A partir desta identificação é feita uma interpolação para localizar o ponto ao longo do trecho, conforme pode ser visto na figura B.5 a seguir.

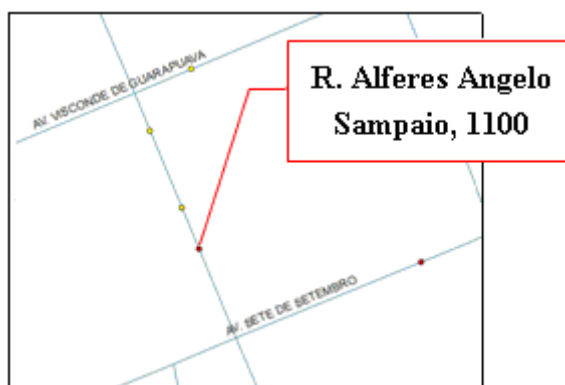


Figura B.5 – Resultado do processo de geocodificação.

Destaca-se que alguns critérios auxiliares foram adotados ao longo do processo:

- Pontos cujo endereço não apresentava numeração do cliente, somente o nome do logradouro – foram localizados no primeiro trecho do logradouro;
- Pontos cujo número do cliente era maior do que a numeração do logradouro – assumiu-se colocá-lo no último trecho do logradouro;
- Endereços referentes às rodovias, por exemplo, BR101 km 2,6 – estimou-se a distância ao longo da rodovia para localizar o cliente.



## APÊNDICE (C) - INTERPOLAÇÃO DE VARIÁVEIS - PREÇO DE ÓLEO COMBUSTÍVEL

As pesquisas de frequência semanal de preço para os óleos combustíveis do tipo A1, A2 e B1, foram separadas por mês originário da pesquisa e, calculada a média de preço mensal com base no número de eventos (pesquisas semanais) mensais.

Nas figuras C.1 e C.2 a seguir é possível notar que na região sudeste somente as pesquisas referentes aos óleos A1 e A2 forneceram uma curva satisfatória para emprego como variável explicativa do preço do referido combustível. Já na região sul esta condição somente foi satisfeita para o óleo do tipo A1.

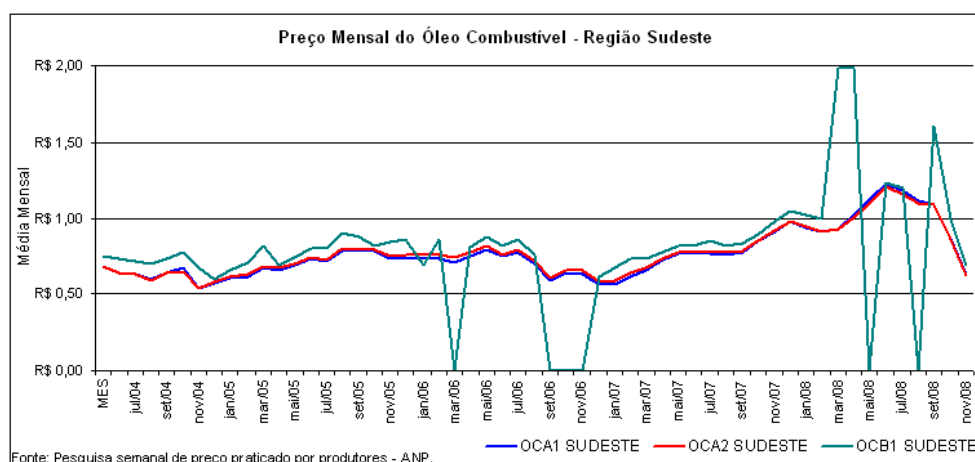


Figura C.1 – Evolução do Preço Mensal do Óleo Combustível – Região Sudeste – Fonte: ANP.

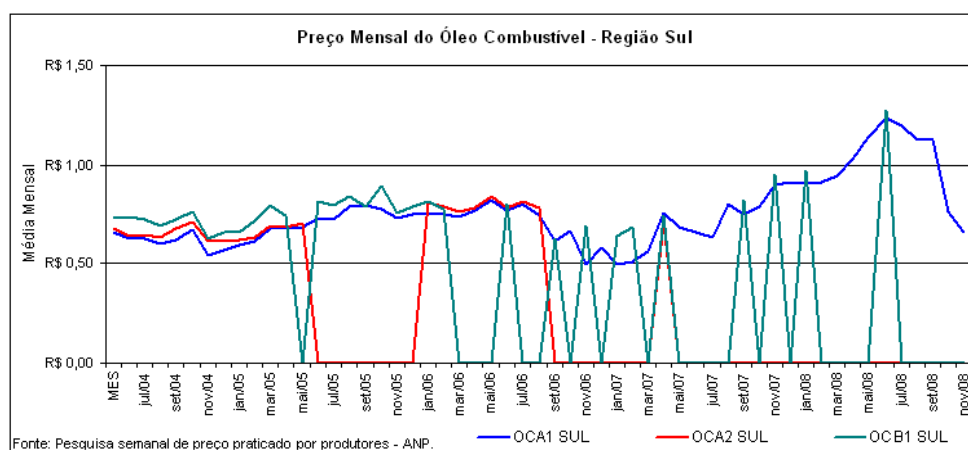


Figura C.2 – Evolução do Preço Mensal do Óleo Combustível – Região Sul – Fonte: ANP.

Tal inconsistência se deu a falta de dados pesquisados para os meses que se apresentaram com média zero; em função disto na continuação desta tese adotou-se o mesmo critério empregado para as outras séries, ou seja, foram desconsideradas as variáveis que apresentaram valores zerados na série histórica, neste caso os óleos B1 nas duas regiões e o A2 no sul.

## APÊNDICE (D) - GRÁFICOS DE QUANTIDADE DE PONTOS DE CONSUMO E MAPAS DE DISPOSIÇÃO DE CLIENTES

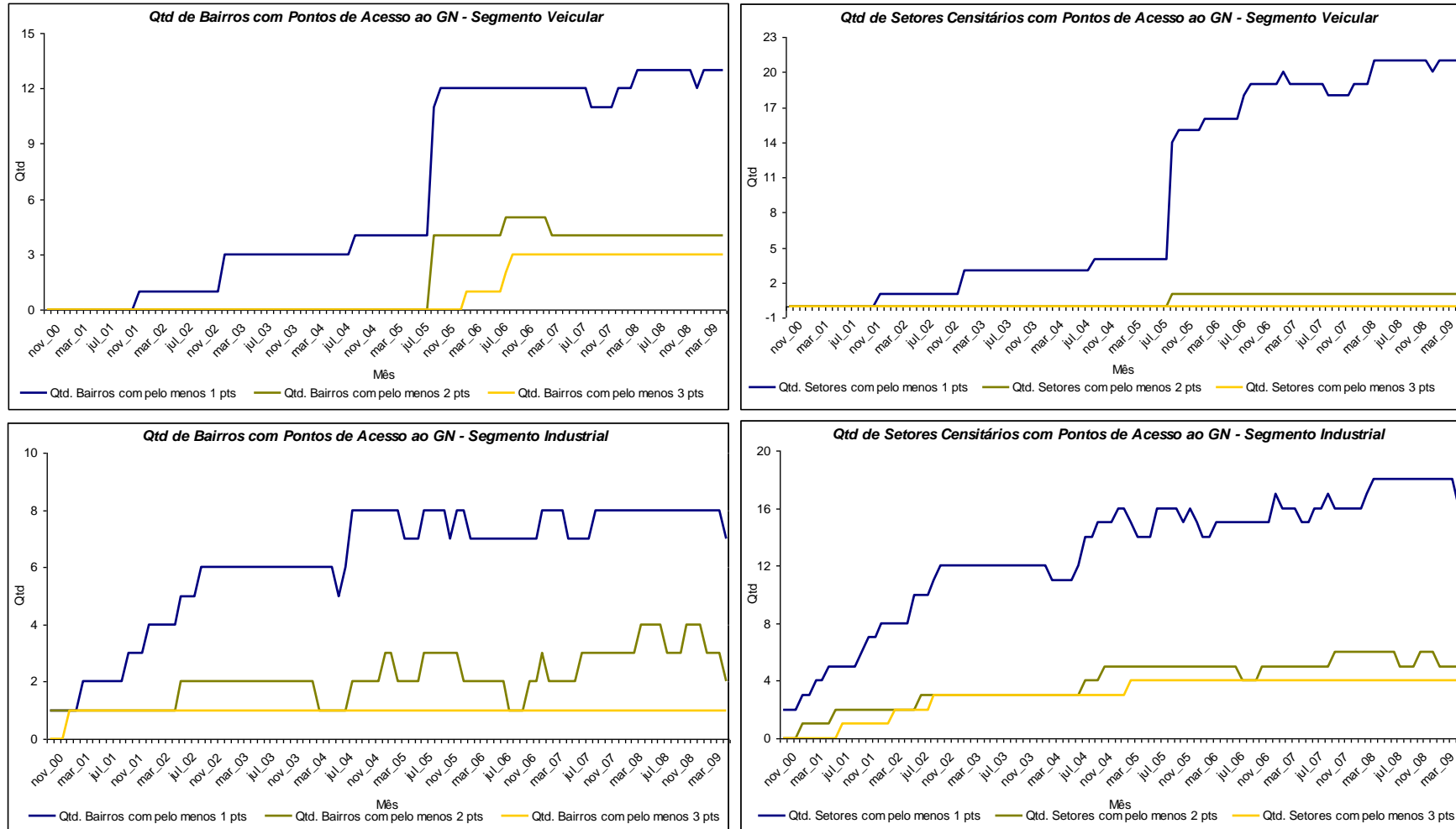


Figura D.1 – Gráficos de Quantidade de Pontos de Consumo por Bairros e Setores – Segmentos Industrial e Veicular - Curitiba.

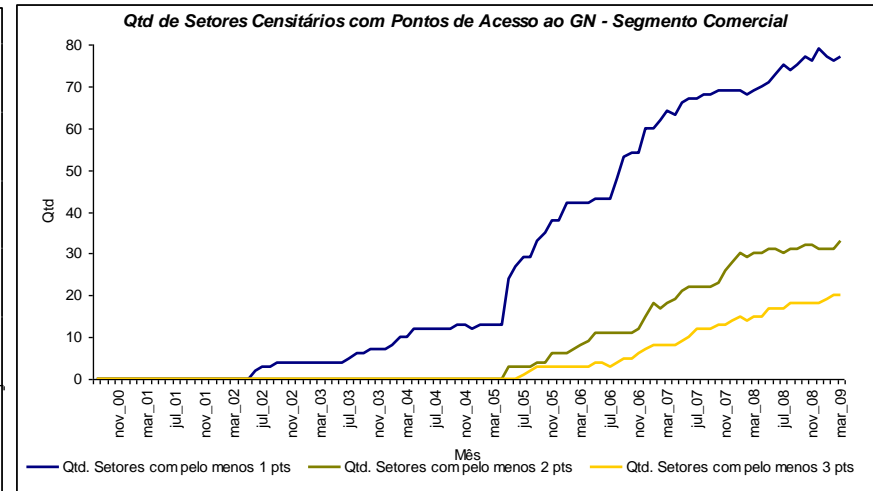
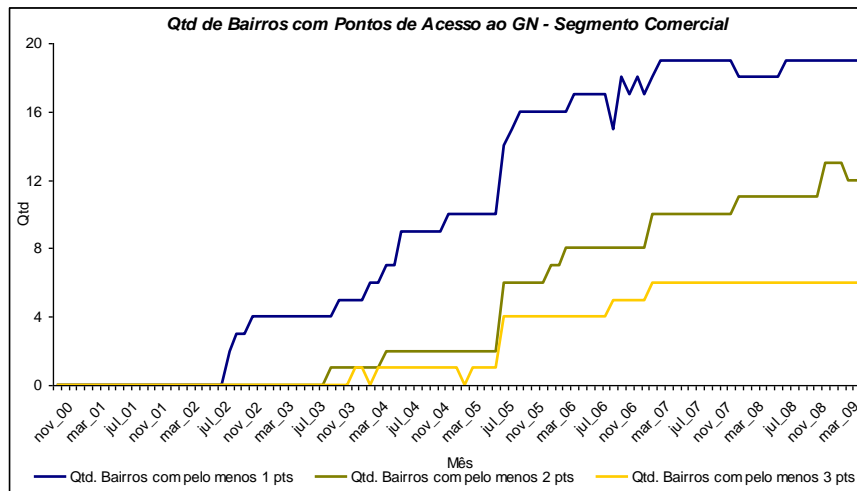
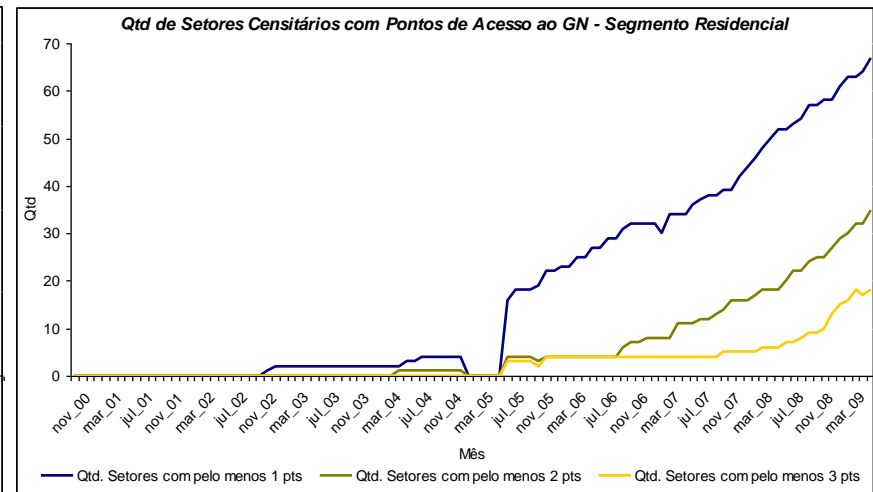
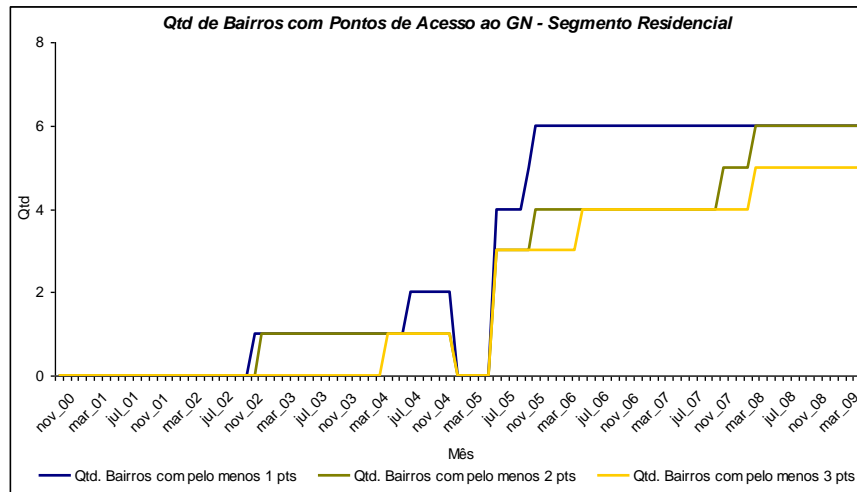


Figura D.2 – Gráficos de Quantidade de Pontos de Consumo por Bairros e Setores – Segmentos Residencial e Comercial - Curitiba.

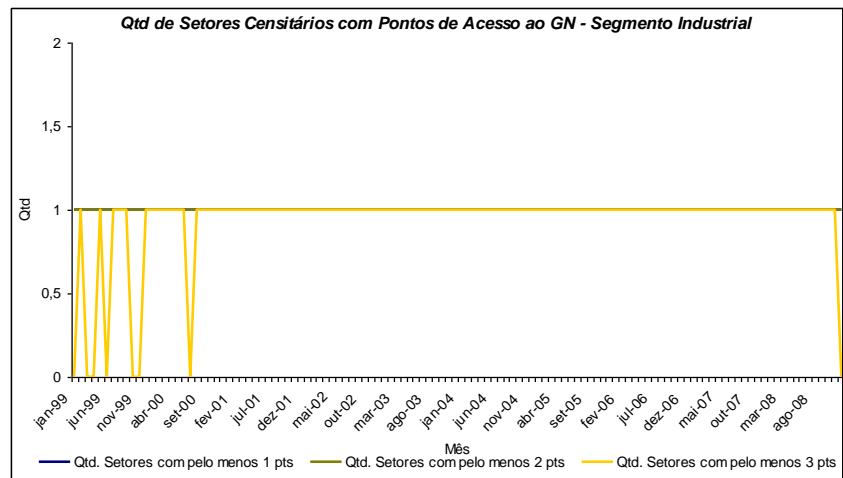
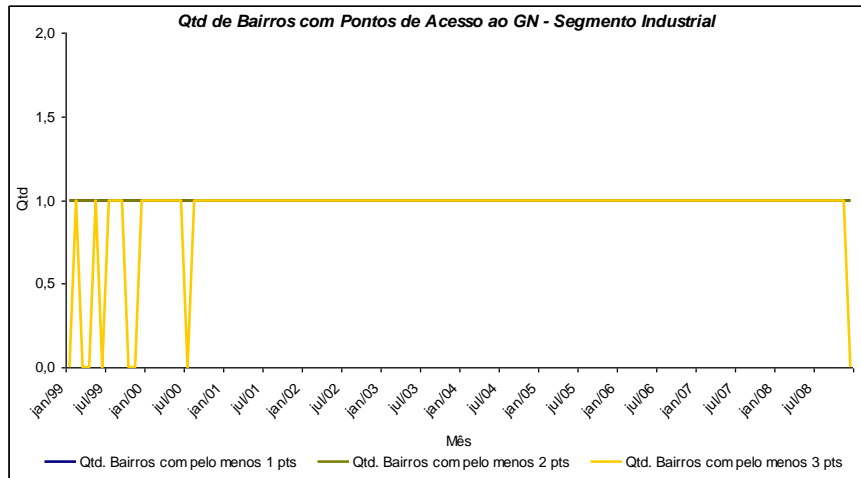
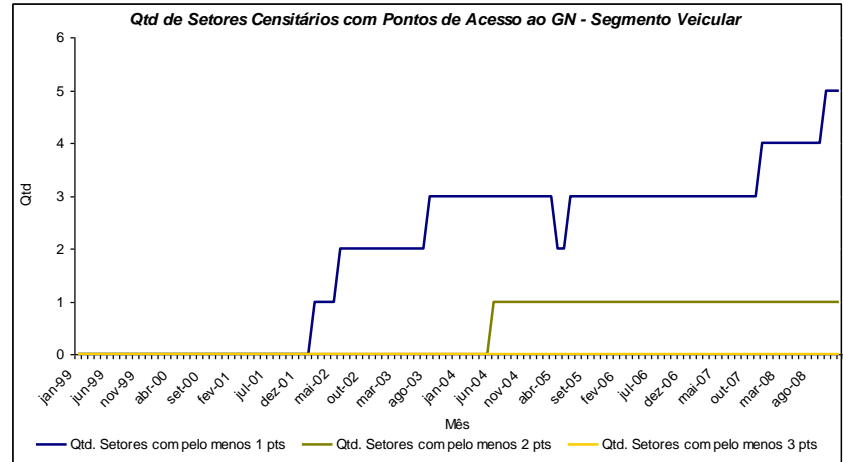
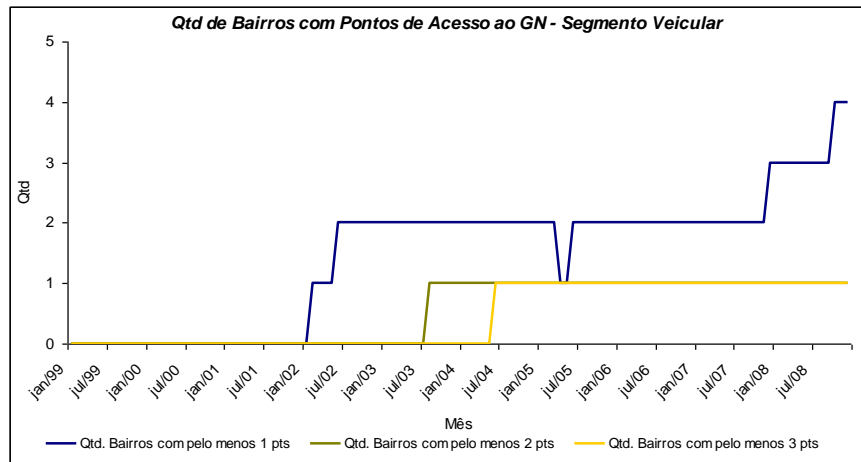


Figura D.3 – Gráficos de Quantidade de Pontos de Consumo por Bairros e Setores – Segmentos Industrial e Veicular -Vitória.

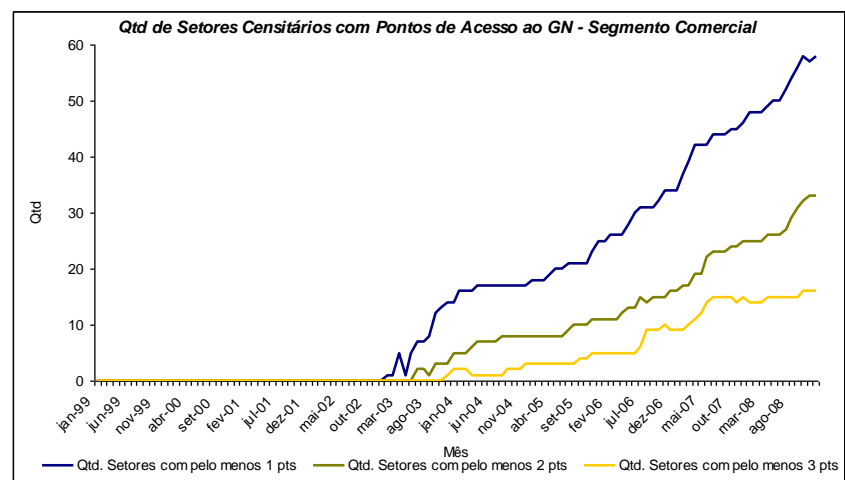
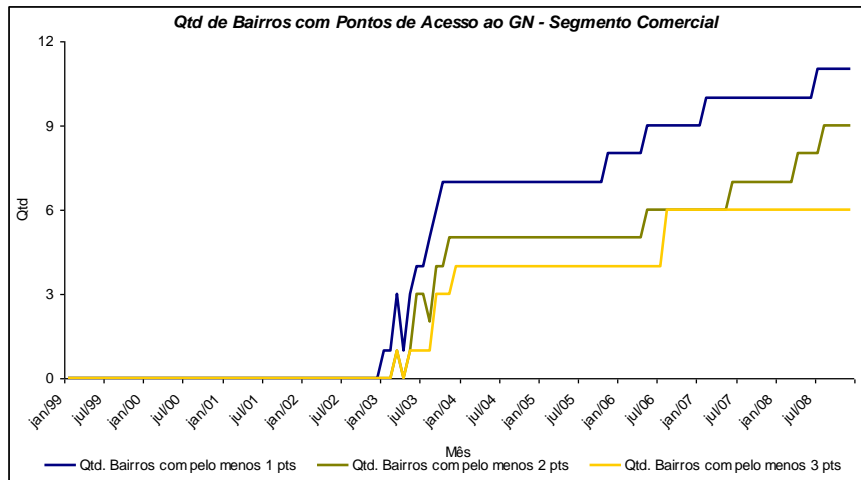
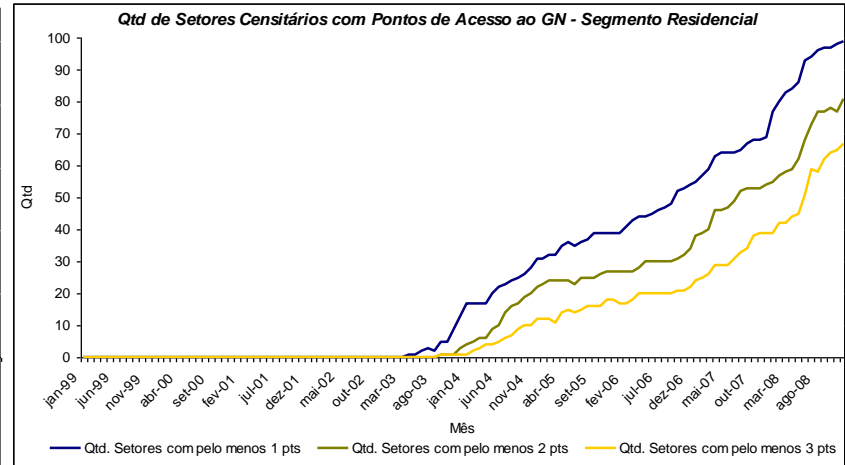
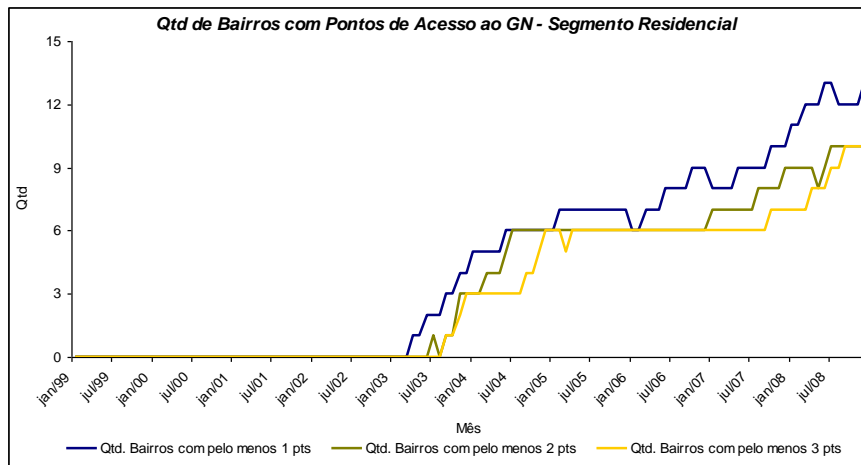


Figura D.4 – Gráficos de Quantidade de Pontos de Consumo por Bairros e Setores – Segmentos Residencial e Comercial -Vitória.

## APÊNDICE (E) - GRÁFICOS DE ÍNDICE MORAN GLOBAL

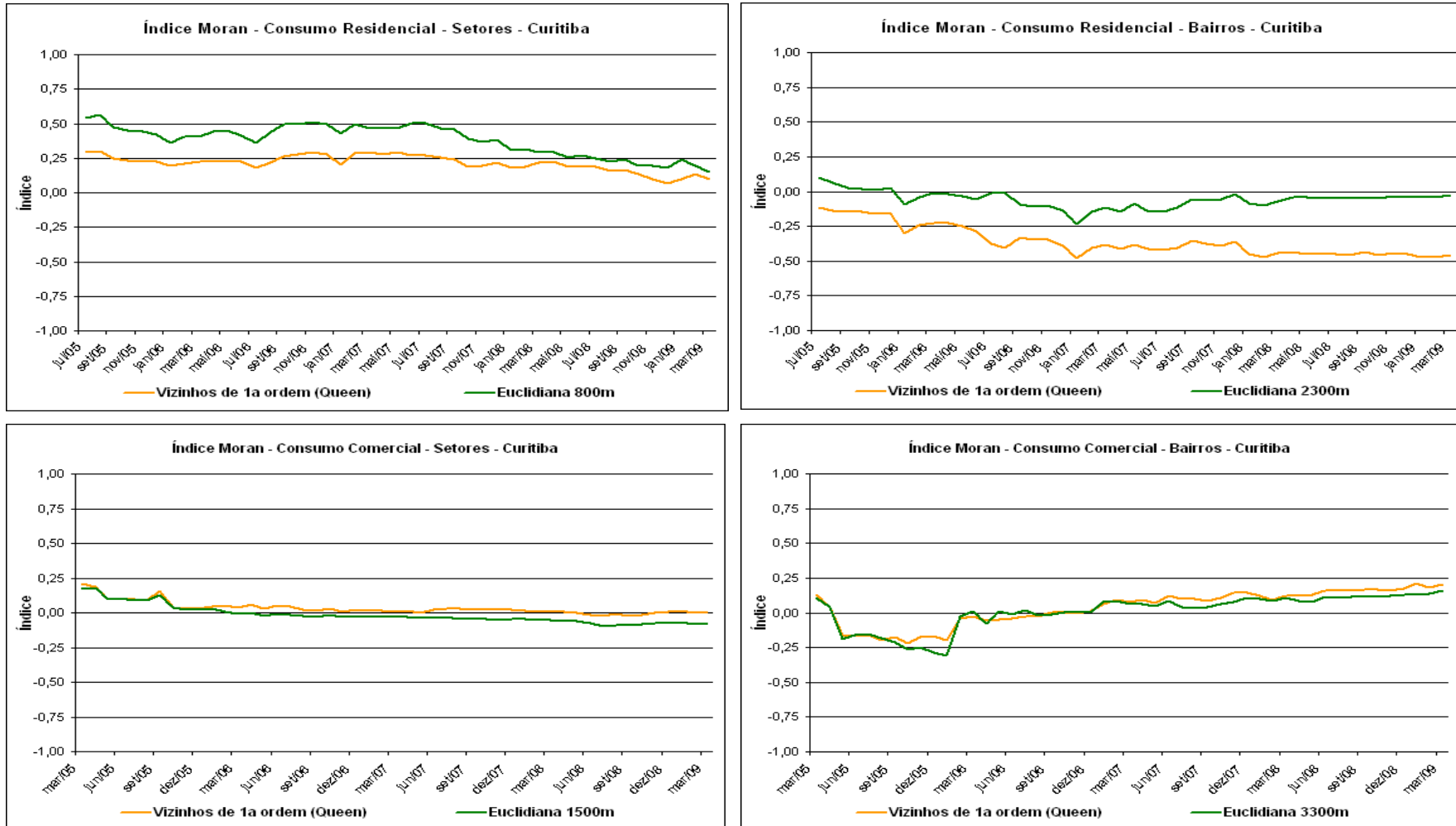


Figura E.1 – Gráficos do Índice de Moran Global por Bairros e Setores – Segmentos Residencial e Comercial - Curitiba.

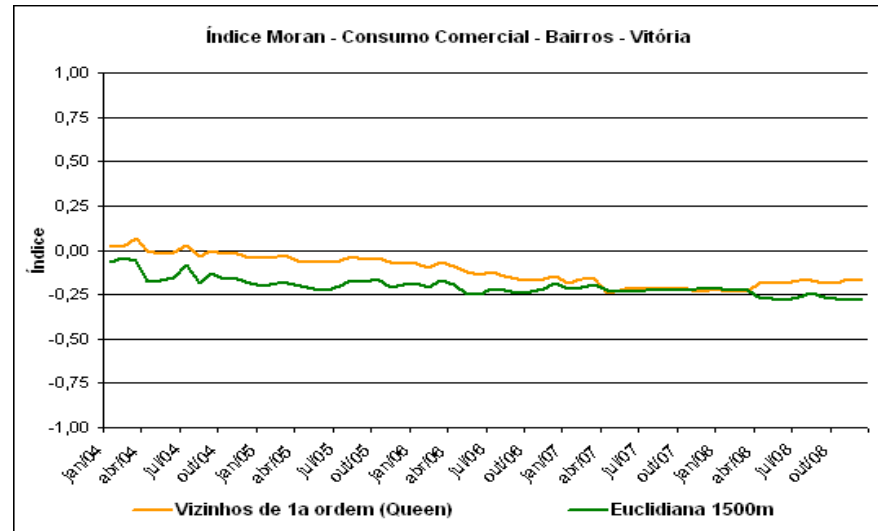
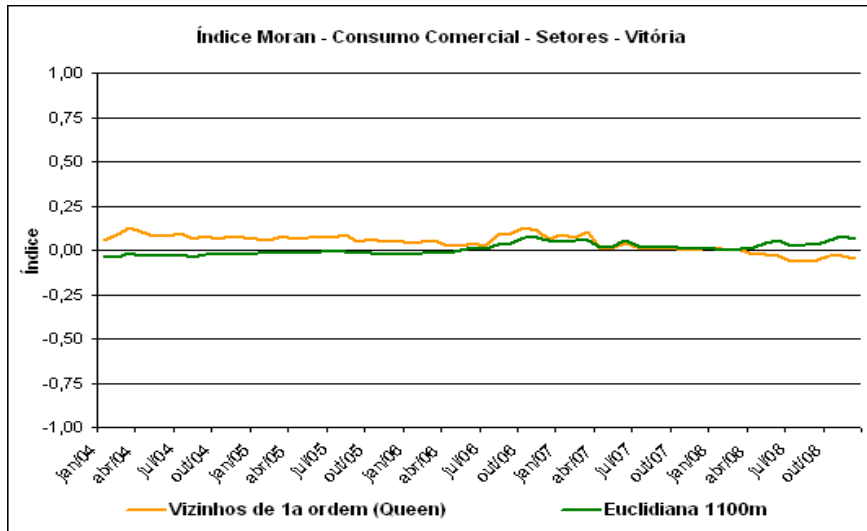
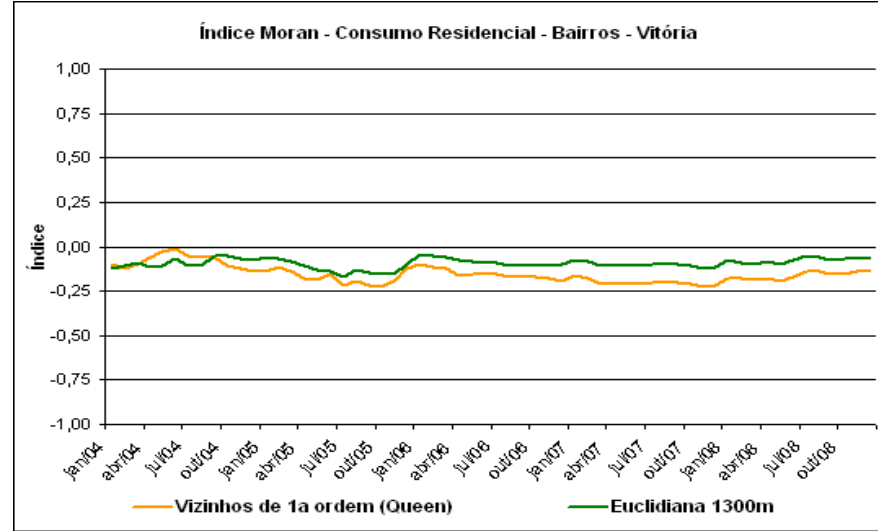
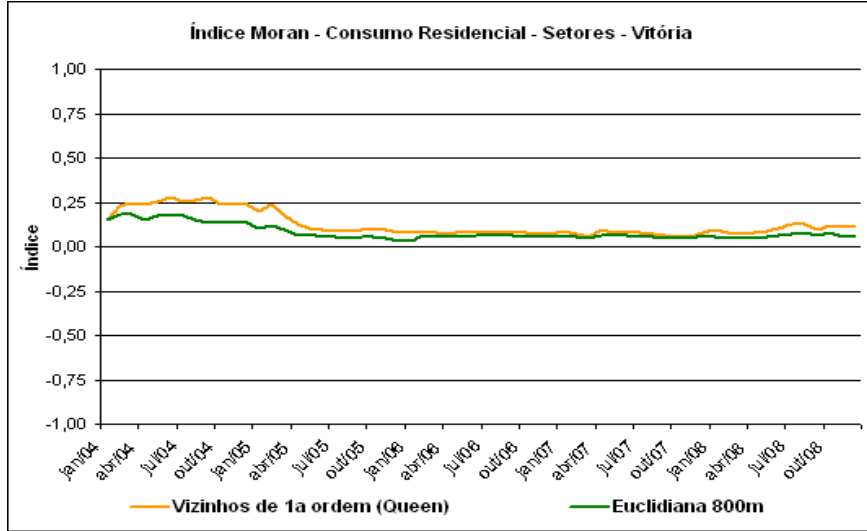


Figura E.2 – Gráficos do Índice de Moran Global por Bairros e Setores – Segmentos Residencial e Comercial - Vitória.

## APÊNDICE (F) - RESUMO DA DETERMINAÇÃO DAS PCA

### F.1 – PCA PARA O SEGMENTO RESIDENCIAL DE CURITIBA – OPÇÃO 1

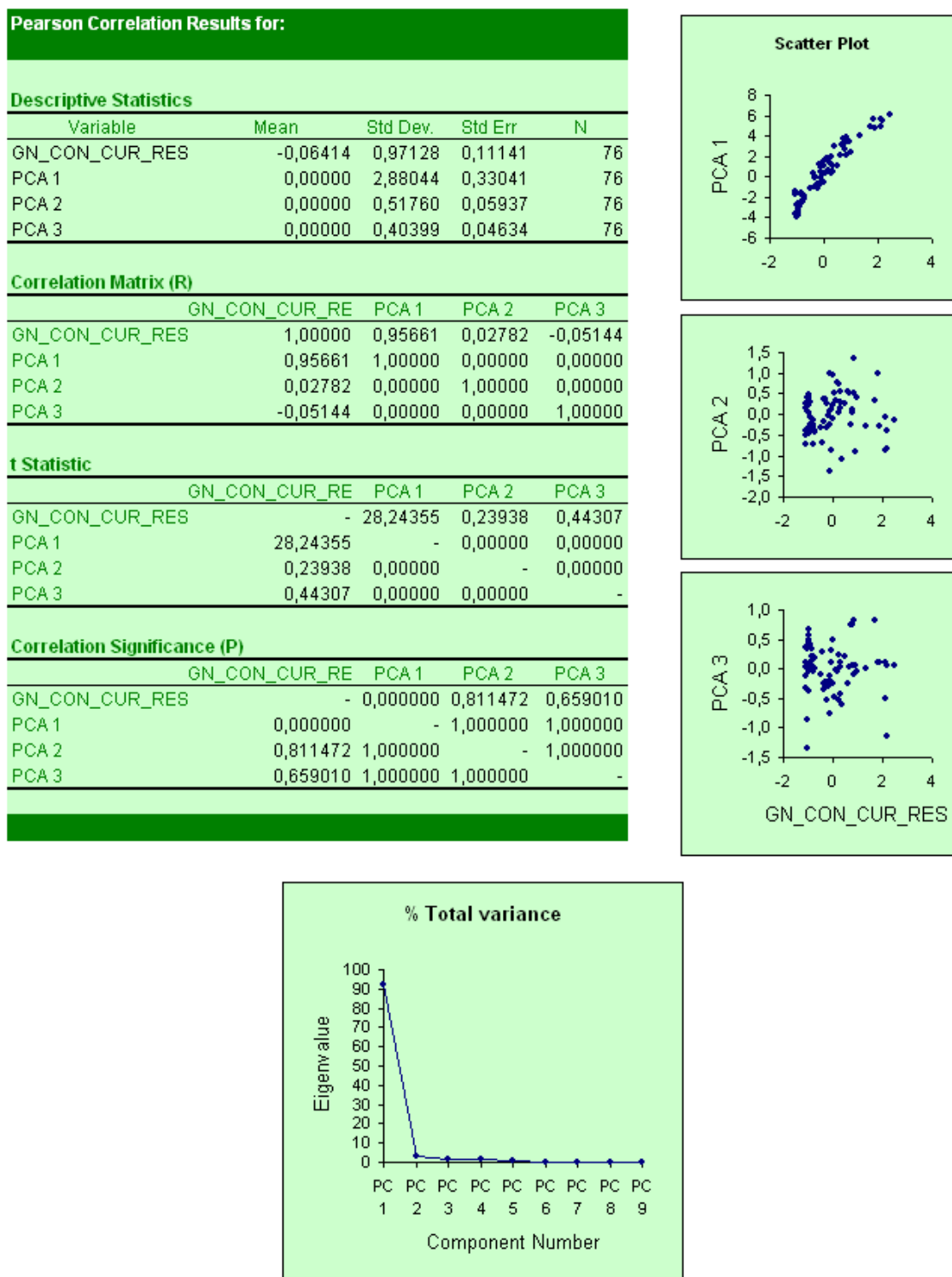


Figura F.1 – Análise de Correlação - Variável Dependente x PCA Variável Explicativas  
Segmento Residencial – Curitiba (opção 1) – Ferramenta STATISTIXL EXCEL



## F.2 – PCA PARA O SEGMENTO RESIDENCIAL DE VITÓRIA – OPÇÃO 1



Figura F.2 – Análise de Correlação - Variável Dependente x PCA Variável Explicativas  
Segmento Residencial – Vitória (opção 1) – Ferramenta STATISTIXL EXCEL

Figura F.8 – Análise de Correlação - Variável Dependente x PCA Variável Explicativas  
Segmento Residencial – Vitória (opção 1) – Ferramenta STATISTICAL 7.

### F.3 – PCA PARA O SEGMENTO RESIDENCIAL DE VITÓRIA – OPÇÃO 2

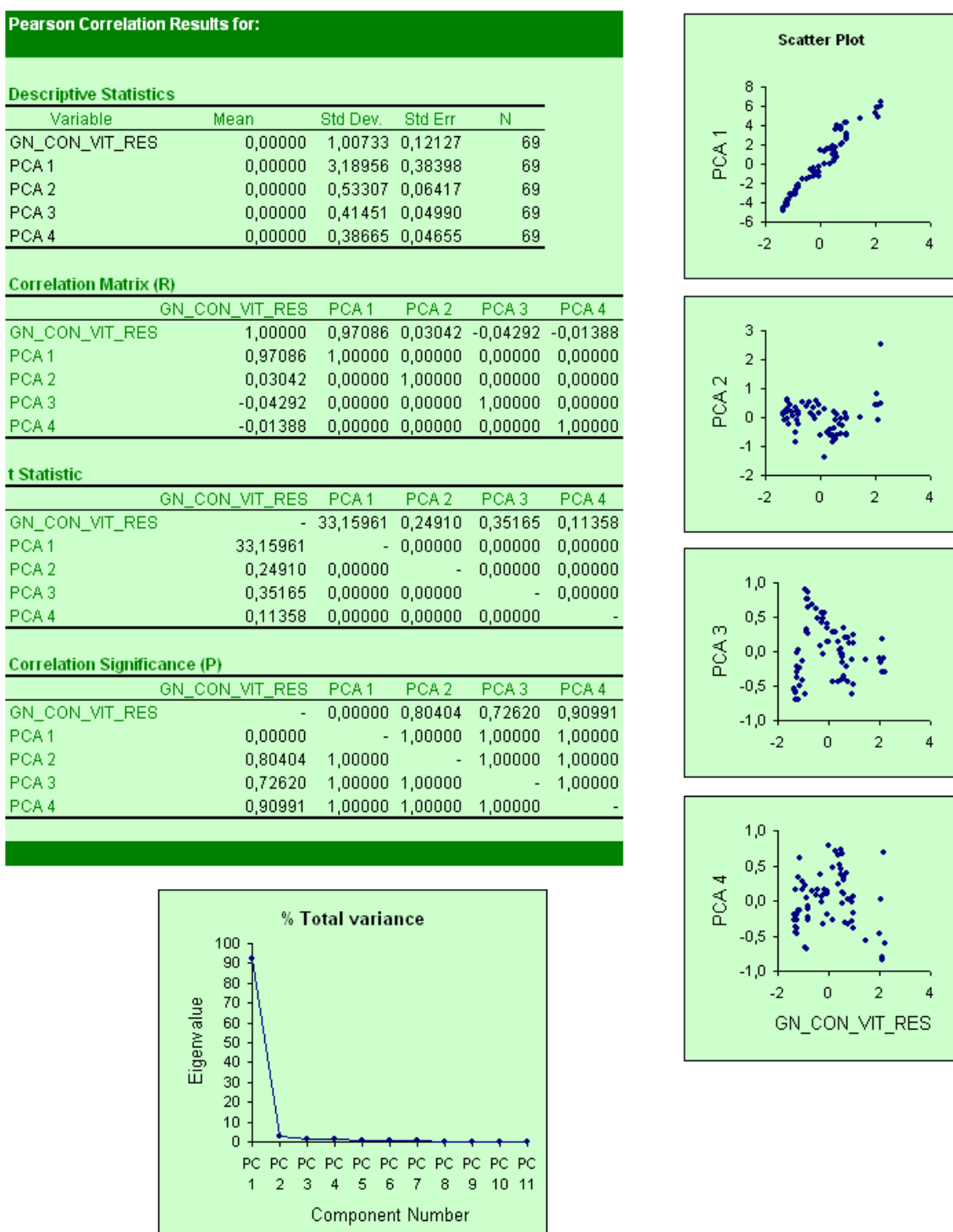


Figura F.3 – Análise de Correlação - Variável Dependente x PCA Variável Explicativas Segmento Residencial – Vitória (opção 2) – Ferramenta STATISTIXL EXCEL

Figura F.11 – Análise de Correlação - Variável Dependente x PCA Variável Explicativas Segmento Residencial – Vitória (opção 2) – Ferramenta STATISTICAL 7.

### F.4 – PCA PARA O SEGMENTO RESIDENCIAL DE VITÓRIA – OPÇÃO 1 (MANUTENÇÃO DE TODAS AS VARIÁVEIS EXPLICATIVAS)

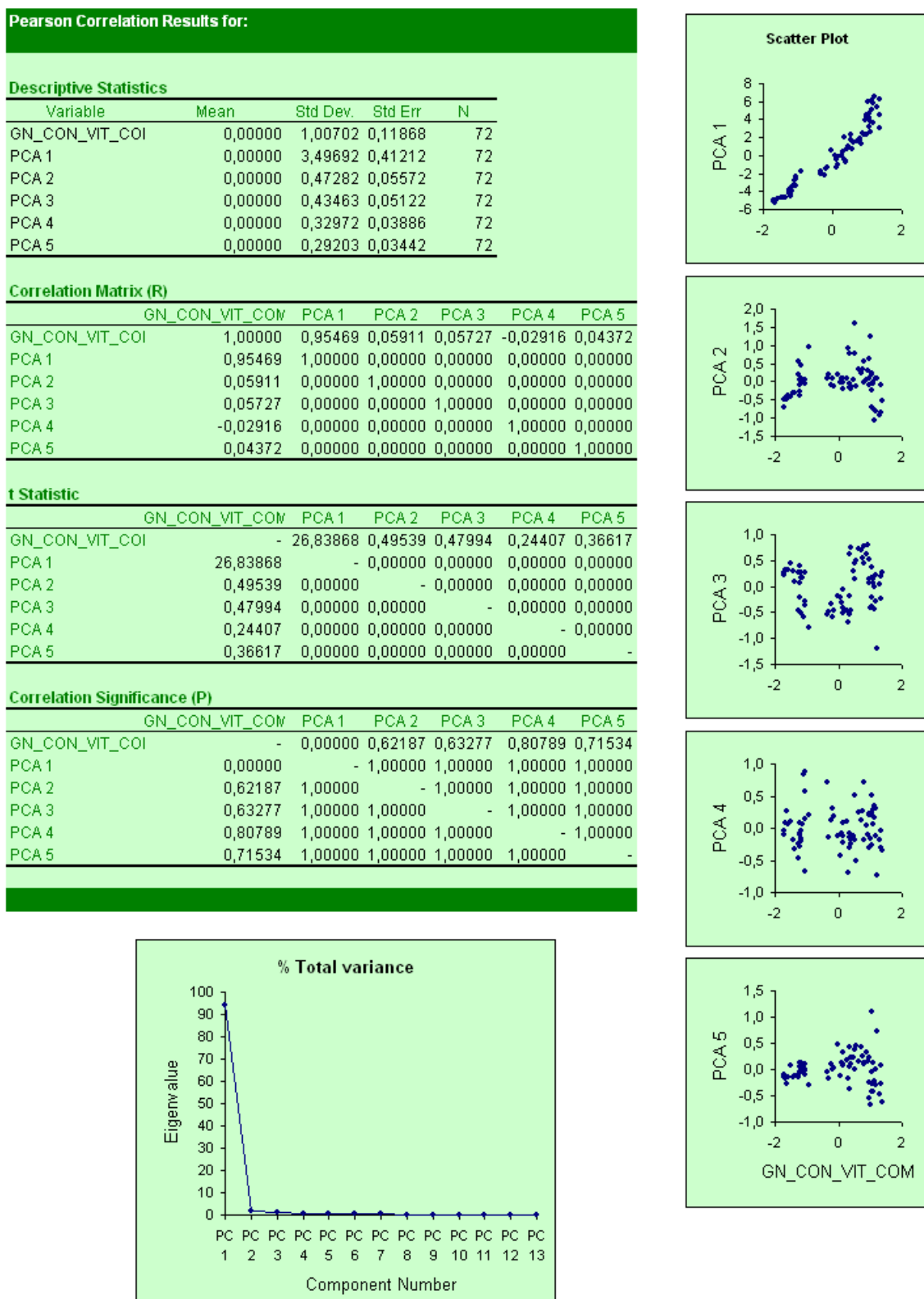


Figura F.4 – Análise de Correlação - Variável Dependente x PCA Variável Explicativas  
Segmento Comercial – Vitória (opção 1) – Ferramenta STATISTIXL EXCEL

## F.5 – PCA PARA O SEGMENTO RESIDENCIAL DE VITÓRIA – OPÇÃO 2

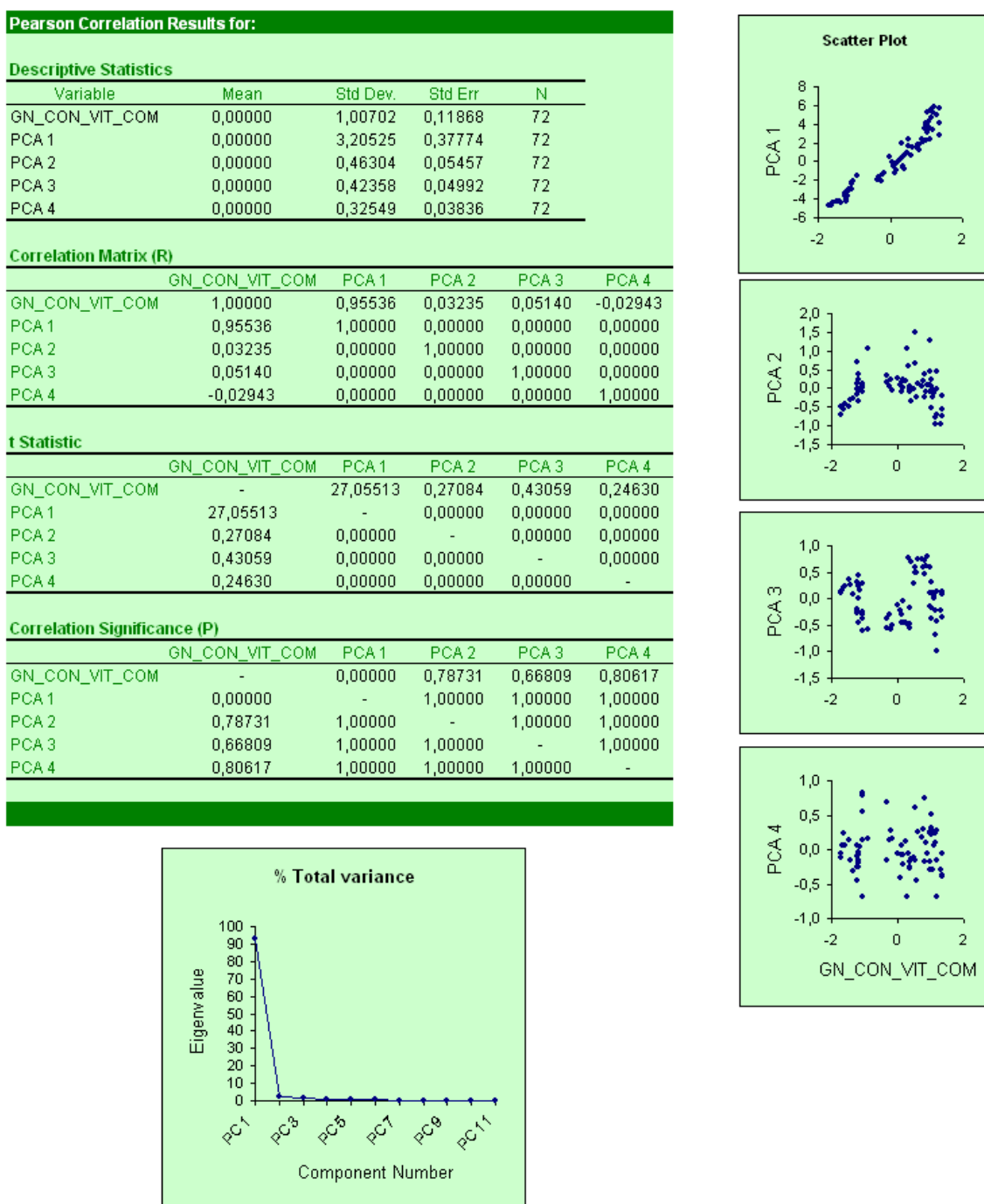
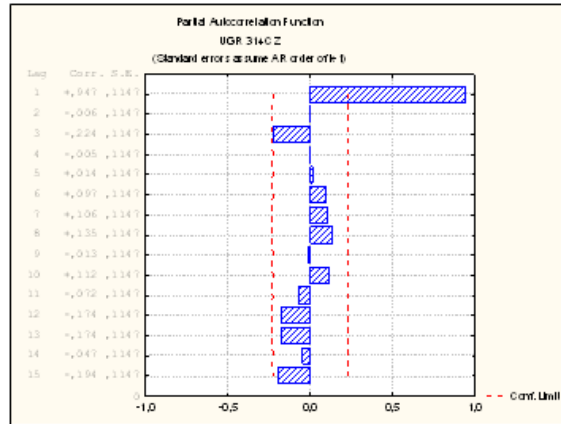


Figura F.5 – Análise de Correlação - Variável Dependente x PCA Variável Explicativas  
Segmento Comercial – Vitória (opção 2) – Ferramenta STATISTIXL EXCEL

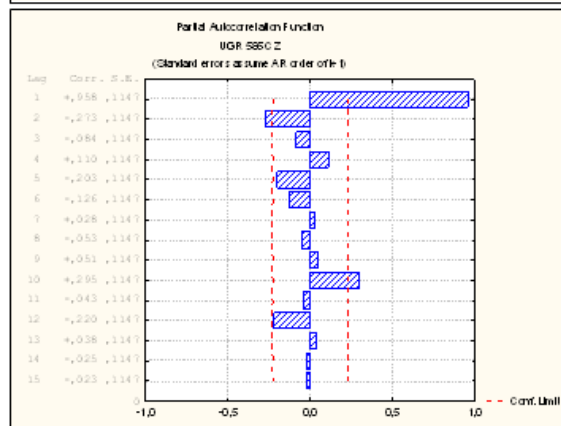
# APÊNDICE (G) - AUTOCORRELAÇÃO PARCIAL VARIÁVEL DE INTERESSE

## G.1 - SEGMENTO RESIDENCIAL DE CURITIBA - UE 624

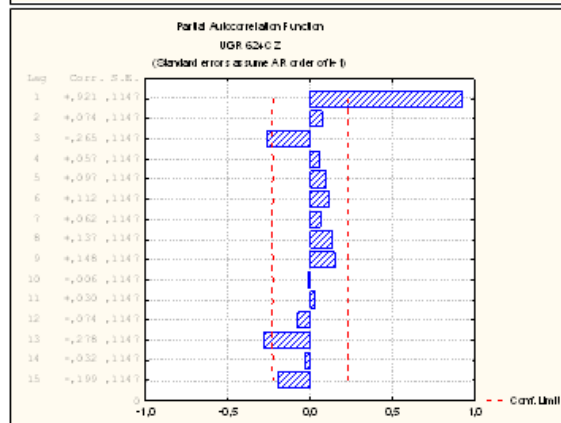
| UE 314 | Lag | Partial-Auto r. | Std.Err. |
|--------|-----|-----------------|----------|
|        | 1   | 0,94731         | 0,11471  |
|        | 2   | -0,00588        | 0,11471  |
|        | 3   | -0,22410        | 0,11471  |
|        | 4   | -0,00502        | 0,11471  |
|        | 5   | 0,01357         | 0,11471  |
|        | 6   | 0,09693         | 0,11471  |
|        | 7   | 0,10632         | 0,11471  |
|        | 8   | 0,13530         | 0,11471  |
|        | 9   | -0,01325        | 0,11471  |
|        | 10  | 0,11173         | 0,11471  |
|        | 11  | -0,07152        | 0,11471  |
|        | 12  | -0,17447        | 0,11471  |
|        | 13  | -0,17378        | 0,11471  |
|        | 14  | -0,04689        | 0,11471  |
|        | 15  | -0,19356        | 0,11471  |



| UE 585 | Lag | Partial-Auto r. | Std.Err. |
|--------|-----|-----------------|----------|
|        | 1   | 0,95779         | 0,11471  |
|        | 2   | -0,27272        | 0,11471  |
|        | 3   | -0,08375        | 0,11471  |
|        | 4   | 0,11037         | 0,11471  |
|        | 5   | -0,20327        | 0,11471  |
|        | 6   | -0,12570        | 0,11471  |
|        | 7   | 0,02784         | 0,11471  |
|        | 8   | -0,05336        | 0,11471  |
|        | 9   | 0,05086         | 0,11471  |
|        | 10  | 0,29479         | 0,11471  |
|        | 11  | -0,04300        | 0,11471  |
|        | 12  | -0,21991        | 0,11471  |
|        | 13  | 0,03766         | 0,11471  |
|        | 14  | -0,02523        | 0,11471  |
|        | 15  | -0,02269        | 0,11471  |



| UE624 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,92147         | 0,11471  |
|       | 2   | 0,07422         | 0,11471  |
|       | 3   | -0,26453        | 0,11471  |
|       | 4   | 0,05736         | 0,11471  |
|       | 5   | 0,09682         | 0,11471  |
|       | 6   | 0,11187         | 0,11471  |
|       | 7   | 0,06190         | 0,11471  |
|       | 8   | 0,13663         | 0,11471  |
|       | 9   | 0,14806         | 0,11471  |
|       | 10  | -0,00641        | 0,11471  |
|       | 11  | 0,03019         | 0,11471  |
|       | 12  | -0,07366        | 0,11471  |
|       | 13  | -0,27823        | 0,11471  |
|       | 14  | -0,03233        | 0,11471  |
|       | 15  | -0,19888        | 0,11471  |



| UE626 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,91045         | 0,11471  |
|       | 2   | 0,04397         | 0,11471  |
|       | 3   | -0,23948        | 0,11471  |
|       | 4   | 0,01303         | 0,11471  |
|       | 5   | 0,08600         | 0,11471  |
|       | 6   | 0,01880         | 0,11471  |
|       | 7   | 0,04430         | 0,11471  |
|       | 8   | 0,06678         | 0,11471  |
|       | 9   | -0,01721        | 0,11471  |
|       | 10  | -0,02307        | 0,11471  |
|       | 11  | -0,02963        | 0,11471  |
|       | 12  | 0,01036         | 0,11471  |
|       | 13  | -0,17314        | 0,11471  |
|       | 14  | 0,00970         | 0,11471  |
|       | 15  | -0,10388        | 0,11471  |

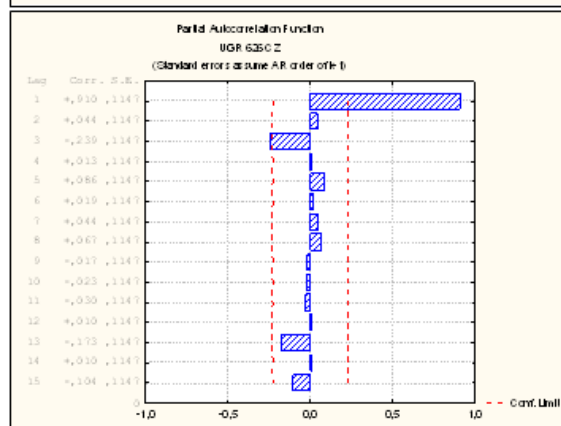
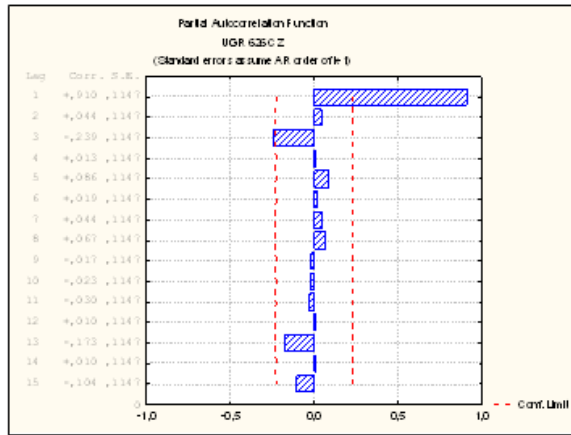


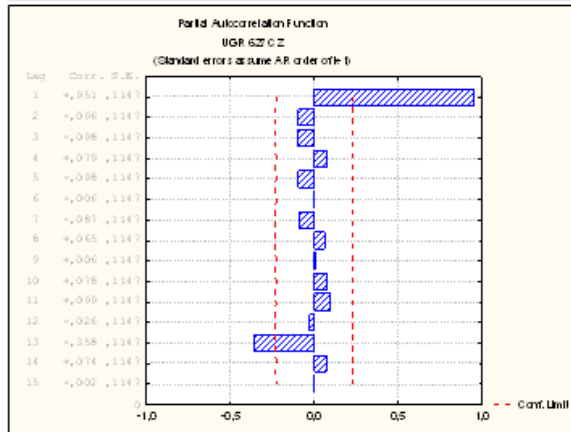
Figura G.1 – Autocorrelação Parcial UE 624 – Segmento Residencial de Curitiba.

## G.2 - SEGMENTO RESIDENCIAL DE CURITIBA - UE 638

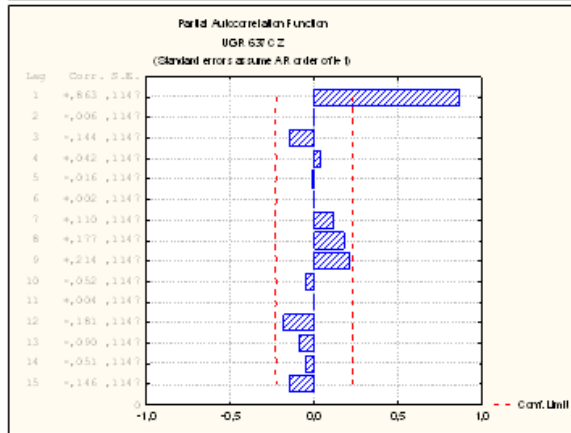
| UE626 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,91045         | 0,11471  |
|       | 2   | 0,04397         | 0,11471  |
|       | 3   | -0,23948        | 0,11471  |
|       | 4   | 0,01303         | 0,11471  |
|       | 5   | 0,08600         | 0,11471  |
|       | 6   | 0,01880         | 0,11471  |
|       | 7   | 0,04430         | 0,11471  |
|       | 8   | 0,06678         | 0,11471  |
|       | 9   | -0,01721        | 0,11471  |
|       | 10  | -0,02307        | 0,11471  |
|       | 11  | -0,02963        | 0,11471  |
|       | 12  | 0,01036         | 0,11471  |
|       | 13  | -0,17314        | 0,11471  |
|       | 14  | 0,00970         | 0,11471  |
|       | 15  | -0,10388        | 0,11471  |



| UE627 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,95144         | 0,11471  |
|       | 2   | -0,09630        | 0,11471  |
|       | 3   | -0,09795        | 0,11471  |
|       | 4   | 0,07879         | 0,11471  |
|       | 5   | -0,09778        | 0,11471  |
|       | 6   | -0,00570        | 0,11471  |
|       | 7   | -0,08671        | 0,11471  |
|       | 8   | 0,06461         | 0,11471  |
|       | 9   | 0,00589         | 0,11471  |
|       | 10  | 0,07840         | 0,11471  |
|       | 11  | 0,09940         | 0,11471  |
|       | 12  | -0,02628        | 0,11471  |
|       | 13  | -0,35818        | 0,11471  |
|       | 14  | 0,07378         | 0,11471  |
|       | 15  | -0,00187        | 0,11471  |



| UE637 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,86303         | 0,11471  |
|       | 2   | -0,00574        | 0,11471  |
|       | 3   | -0,14430        | 0,11471  |
|       | 4   | 0,04218         | 0,11471  |
|       | 5   | -0,01581        | 0,11471  |
|       | 6   | 0,00207         | 0,11471  |
|       | 7   | 0,11018         | 0,11471  |
|       | 8   | 0,17679         | 0,11471  |
|       | 9   | 0,21386         | 0,11471  |
|       | 10  | -0,05190        | 0,11471  |
|       | 11  | 0,00368         | 0,11471  |
|       | 12  | -0,18090        | 0,11471  |
|       | 13  | -0,09036        | 0,11471  |
|       | 14  | -0,05125        | 0,11471  |
|       | 15  | -0,14565        | 0,11471  |



| UE638 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,94434         | 0,11471  |
|       | 2   | -0,18888        | 0,11471  |
|       | 3   | -0,03253        | 0,11471  |
|       | 4   | 0,03692         | 0,11471  |
|       | 5   | 0,00699         | 0,11471  |
|       | 6   | 0,03529         | 0,11471  |
|       | 7   | -0,00964        | 0,11471  |
|       | 8   | 0,24611         | 0,11471  |
|       | 9   | -0,01386        | 0,11471  |
|       | 10  | 0,14623         | 0,11471  |
|       | 11  | 0,04094         | 0,11471  |
|       | 12  | -0,15353        | 0,11471  |
|       | 13  | -0,26114        | 0,11471  |
|       | 14  | -0,09075        | 0,11471  |
|       | 15  | -0,11444        | 0,11471  |

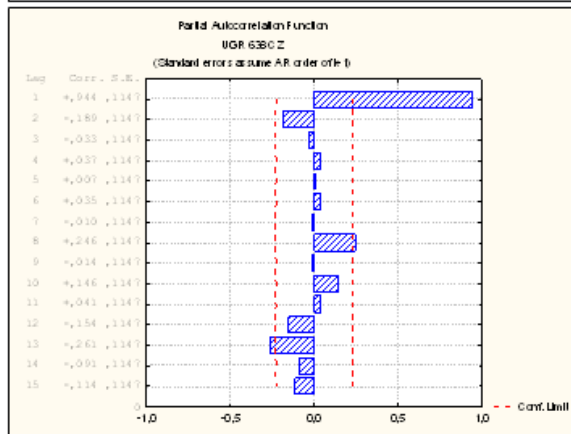
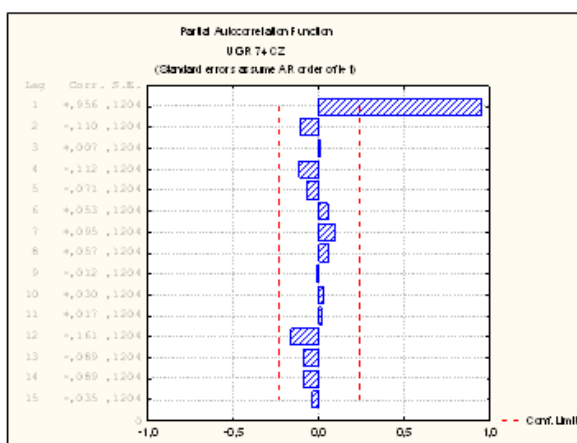


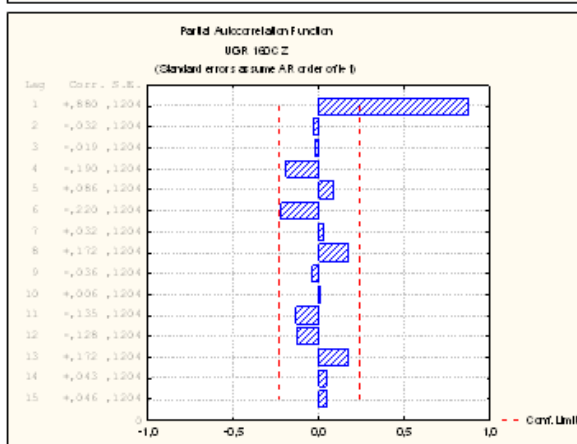
Figura G.2 – Autocorrelação Parcial UE 638 – Segmento Residencial de Curitiba.

### G.3 - SEGMENTO RESIDENCIAL DE VITÓRIA - UE 247

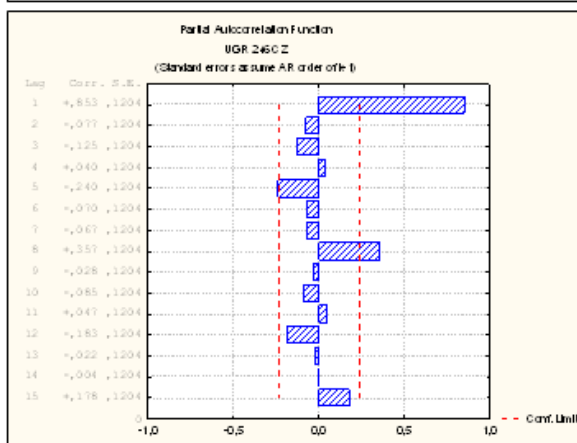
| UE74 | Lag | Partial-Auto r. | Std.Err. |
|------|-----|-----------------|----------|
|      | 1   | 0,95644         | 0,12039  |
|      | 2   | -0,11003        | 0,12039  |
|      | 3   | 0,00697         | 0,12039  |
|      | 4   | -0,11224        | 0,12039  |
|      | 5   | -0,07096        | 0,12039  |
|      | 6   | 0,05274         | 0,12039  |
|      | 7   | 0,09472         | 0,12039  |
|      | 8   | 0,05731         | 0,12039  |
|      | 9   | -0,01185        | 0,12039  |
|      | 10  | 0,03020         | 0,12039  |
|      | 11  | 0,01684         | 0,12039  |
|      | 12  | -0,16076        | 0,12039  |
|      | 13  | -0,08908        | 0,12039  |
|      | 14  | -0,08862        | 0,12039  |
|      | 15  | -0,03527        | 0,12039  |



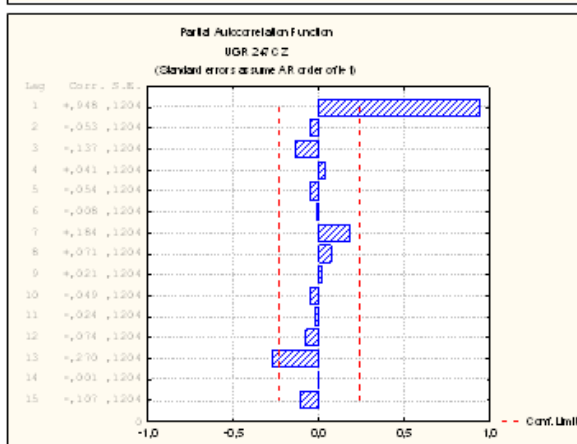
| UE160 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,87996         | 0,12039  |
|       | 2   | -0,03232        | 0,12039  |
|       | 3   | -0,01893        | 0,12039  |
|       | 4   | -0,19033        | 0,12039  |
|       | 5   | 0,08559         | 0,12039  |
|       | 6   | -0,21998        | 0,12039  |
|       | 7   | 0,03159         | 0,12039  |
|       | 8   | 0,17226         | 0,12039  |
|       | 9   | -0,03555        | 0,12039  |
|       | 10  | 0,00590         | 0,12039  |
|       | 11  | -0,13475        | 0,12039  |
|       | 12  | -0,12802        | 0,12039  |
|       | 13  | 0,17165         | 0,12039  |
|       | 14  | 0,04281         | 0,12039  |
|       | 15  | 0,04564         | 0,12039  |



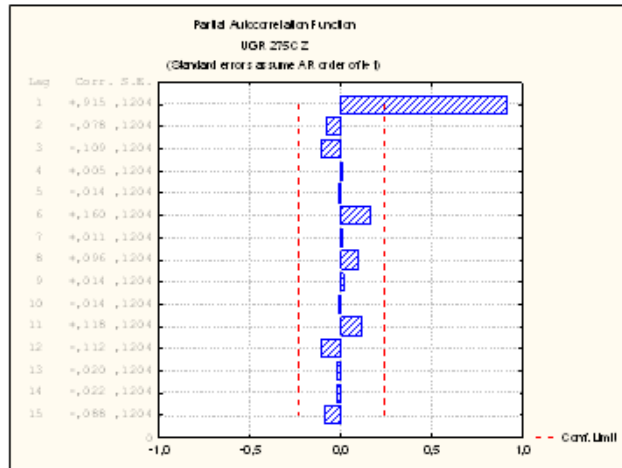
| UE246 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,85275         | 0,12039  |
|       | 2   | -0,07666        | 0,12039  |
|       | 3   | -0,12464        | 0,12039  |
|       | 4   | 0,03969         | 0,12039  |
|       | 5   | -0,24047        | 0,12039  |
|       | 6   | -0,06992        | 0,12039  |
|       | 7   | -0,06744        | 0,12039  |
|       | 8   | 0,35711         | 0,12039  |
|       | 9   | -0,02834        | 0,12039  |
|       | 10  | -0,08528        | 0,12039  |
|       | 11  | 0,04727         | 0,12039  |
|       | 12  | -0,18274        | 0,12039  |
|       | 13  | -0,02230        | 0,12039  |
|       | 14  | -0,00408        | 0,12039  |
|       | 15  | 0,17764         | 0,12039  |



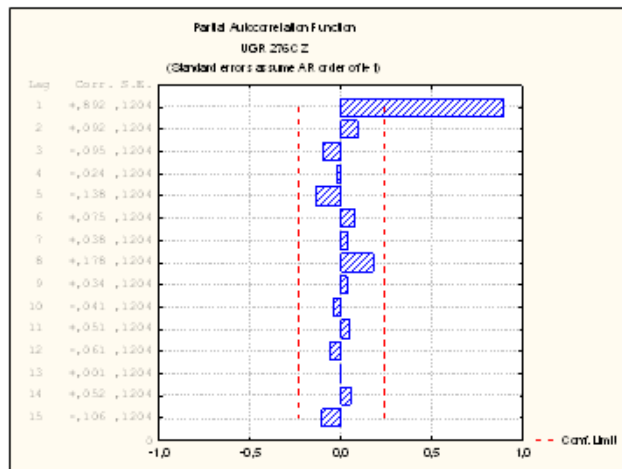
| UE247 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,94776         | 0,12039  |
|       | 2   | -0,05261        | 0,12039  |
|       | 3   | -0,13669        | 0,12039  |
|       | 4   | 0,04099         | 0,12039  |
|       | 5   | -0,05400        | 0,12039  |
|       | 6   | -0,00828        | 0,12039  |
|       | 7   | 0,18353         | 0,12039  |
|       | 8   | 0,07141         | 0,12039  |
|       | 9   | 0,02060         | 0,12039  |
|       | 10  | -0,04867        | 0,12039  |
|       | 11  | -0,02384        | 0,12039  |
|       | 12  | -0,07419        | 0,12039  |
|       | 13  | -0,27024        | 0,12039  |
|       | 14  | -0,00115        | 0,12039  |
|       | 15  | -0,10692        | 0,12039  |



| UE275 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,91501         | 0,12039  |
|       | 2   | -0,07824        | 0,12039  |
|       | 3   | -0,10914        | 0,12039  |
|       | 4   | 0,00479         | 0,12039  |
|       | 5   | -0,01380        | 0,12039  |
|       | 6   | 0,16021         | 0,12039  |
|       | 7   | 0,01096         | 0,12039  |
|       | 8   | 0,09587         | 0,12039  |
|       | 9   | 0,01445         | 0,12039  |
|       | 10  | -0,01386        | 0,12039  |
|       | 11  | 0,11826         | 0,12039  |
|       | 12  | -0,11199        | 0,12039  |
|       | 13  | -0,01980        | 0,12039  |
|       | 14  | -0,02203        | 0,12039  |
|       | 15  | -0,08815        | 0,12039  |



| UE276 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,89205         | 0,12039  |
|       | 2   | 0,09181         | 0,12039  |
|       | 3   | -0,09485        | 0,12039  |
|       | 4   | -0,02351        | 0,12039  |
|       | 5   | -0,13767        | 0,12039  |
|       | 6   | 0,07506         | 0,12039  |
|       | 7   | 0,03782         | 0,12039  |
|       | 8   | 0,17766         | 0,12039  |
|       | 9   | 0,03418         | 0,12039  |
|       | 10  | -0,04149        | 0,12039  |
|       | 11  | 0,05088         | 0,12039  |
|       | 12  | -0,06126        | 0,12039  |
|       | 13  | 0,00088         | 0,12039  |
|       | 14  | 0,05226         | 0,12039  |
|       | 15  | -0,10560        | 0,12039  |



| UE282 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,95674         | 0,12039  |
|       | 2   | -0,01495        | 0,12039  |
|       | 3   | -0,18495        | 0,12039  |
|       | 4   | 0,04180         | 0,12039  |
|       | 5   | -0,02749        | 0,12039  |
|       | 6   | -0,09875        | 0,12039  |
|       | 7   | -0,03017        | 0,12039  |
|       | 8   | 0,15313         | 0,12039  |
|       | 9   | -0,06900        | 0,12039  |
|       | 10  | -0,08140        | 0,12039  |
|       | 11  | 0,06955         | 0,12039  |
|       | 12  | -0,06950        | 0,12039  |
|       | 13  | -0,13884        | 0,12039  |
|       | 14  | -0,02105        | 0,12039  |
|       | 15  | -0,10043        | 0,12039  |

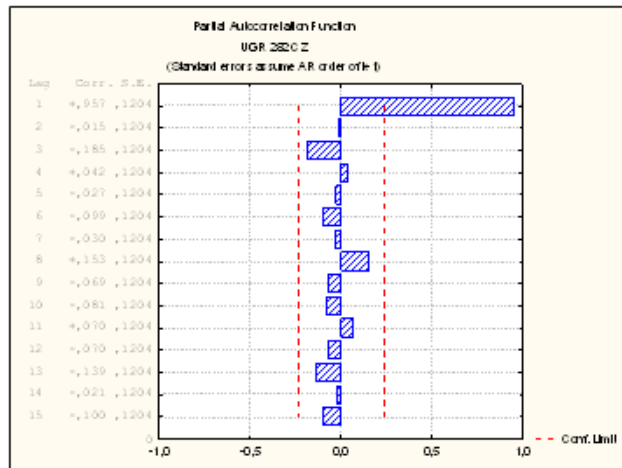
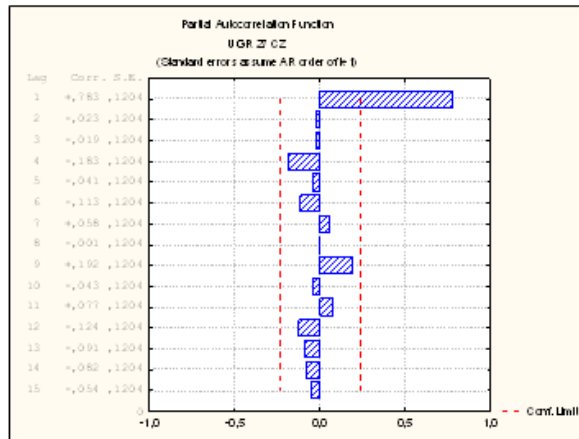


Figura G.3 – Autocorrelação Parcial UE 247 – Segmento Residencial de Vitória.

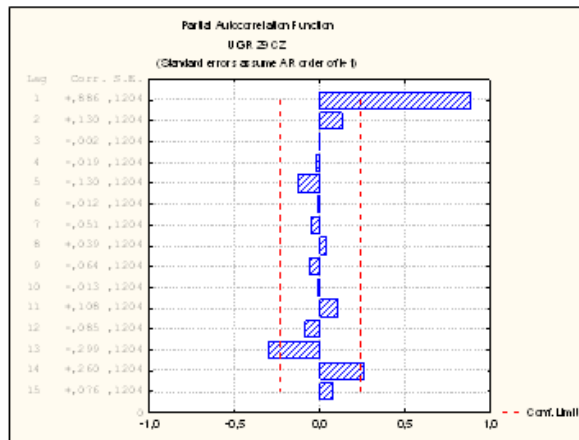


## G.4 - SEGMENTO RESIDENCIAL DE VITÓRIA - UE 258

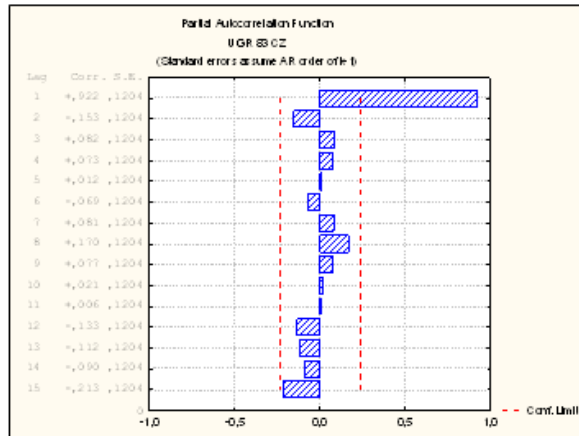
| UE27 | Lag | Partial-Auto r. | Std.Err. |
|------|-----|-----------------|----------|
|      | 1   | 0,78315         | 0,12039  |
|      | 2   | -0,02342        | 0,12039  |
|      | 3   | -0,01860        | 0,12039  |
|      | 4   | -0,18306        | 0,12039  |
|      | 5   | -0,04147        | 0,12039  |
|      | 6   | -0,11260        | 0,12039  |
|      | 7   | 0,05761         | 0,12039  |
|      | 8   | -0,00066        | 0,12039  |
|      | 9   | 0,19229         | 0,12039  |
|      | 10  | -0,04270        | 0,12039  |
|      | 11  | 0,07722         | 0,12039  |
|      | 12  | -0,12384        | 0,12039  |
|      | 13  | -0,09143        | 0,12039  |
|      | 14  | -0,08211        | 0,12039  |
|      | 15  | -0,05407        | 0,12039  |



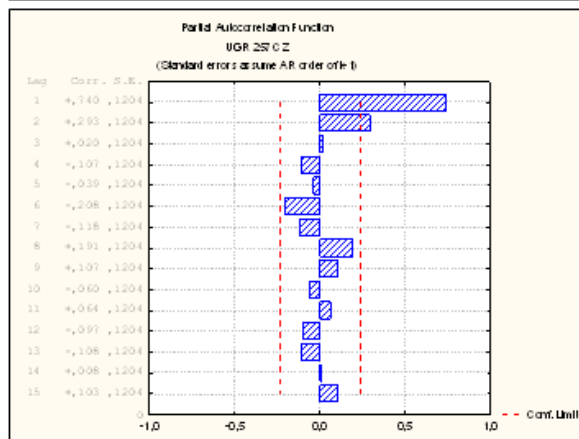
| UE29 | Lag | Partial-Auto r. | Std.Err. |
|------|-----|-----------------|----------|
|      | 1   | 0,88615         | 0,12039  |
|      | 2   | 0,13008         | 0,12039  |
|      | 3   | -0,00196        | 0,12039  |
|      | 4   | -0,01867        | 0,12039  |
|      | 5   | -0,13007        | 0,12039  |
|      | 6   | -0,01208        | 0,12039  |
|      | 7   | -0,05090        | 0,12039  |
|      | 8   | 0,03946         | 0,12039  |
|      | 9   | -0,06375        | 0,12039  |
|      | 10  | -0,01330        | 0,12039  |
|      | 11  | 0,10803         | 0,12039  |
|      | 12  | -0,08462        | 0,12039  |
|      | 13  | -0,29868        | 0,12039  |
|      | 14  | 0,25983         | 0,12039  |
|      | 15  | 0,07558         | 0,12039  |



| UE83 | Lag | Partial-Auto r. | Std.Err. |
|------|-----|-----------------|----------|
|      | 1   | 0,92196         | 0,12039  |
|      | 2   | -0,15288        | 0,12039  |
|      | 3   | 0,08179         | 0,12039  |
|      | 4   | 0,07282         | 0,12039  |
|      | 5   | 0,01195         | 0,12039  |
|      | 6   | -0,06937        | 0,12039  |
|      | 7   | 0,08074         | 0,12039  |
|      | 8   | 0,16971         | 0,12039  |
|      | 9   | 0,07712         | 0,12039  |
|      | 10  | 0,02116         | 0,12039  |
|      | 11  | 0,00564         | 0,12039  |
|      | 12  | -0,13337        | 0,12039  |
|      | 13  | -0,11247        | 0,12039  |
|      | 14  | -0,08980        | 0,12039  |
|      | 15  | -0,21350        | 0,12039  |



| UE257 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,73981         | 0,12039  |
|       | 2   | 0,29292         | 0,12039  |
|       | 3   | 0,02031         | 0,12039  |
|       | 4   | -0,10676        | 0,12039  |
|       | 5   | -0,03872        | 0,12039  |
|       | 6   | -0,20834        | 0,12039  |
|       | 7   | -0,11818        | 0,12039  |
|       | 8   | 0,19115         | 0,12039  |
|       | 9   | 0,10740         | 0,12039  |
|       | 10  | -0,06046        | 0,12039  |
|       | 11  | 0,06372         | 0,12039  |
|       | 12  | -0,09666        | 0,12039  |
|       | 13  | -0,10758        | 0,12039  |
|       | 14  | 0,00805         | 0,12039  |
|       | 15  | 0,10321         | 0,12039  |



| UE258 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,87675         | 0,12039  |
|       | 2   | -0,01990        | 0,12039  |
|       | 3   | -0,17475        | 0,12039  |
|       | 4   | 0,01042         | 0,12039  |
|       | 5   | -0,11890        | 0,12039  |
|       | 6   | 0,08991         | 0,12039  |
|       | 7   | 0,20397         | 0,12039  |
|       | 8   | 0,04990         | 0,12039  |
|       | 9   | 0,02216         | 0,12039  |
|       | 10  | -0,01808        | 0,12039  |
|       | 11  | -0,05081        | 0,12039  |
|       | 12  | 0,03806         | 0,12039  |
|       | 13  | -0,10509        | 0,12039  |
|       | 14  | -0,00609        | 0,12039  |
|       | 15  | -0,07928        | 0,12039  |

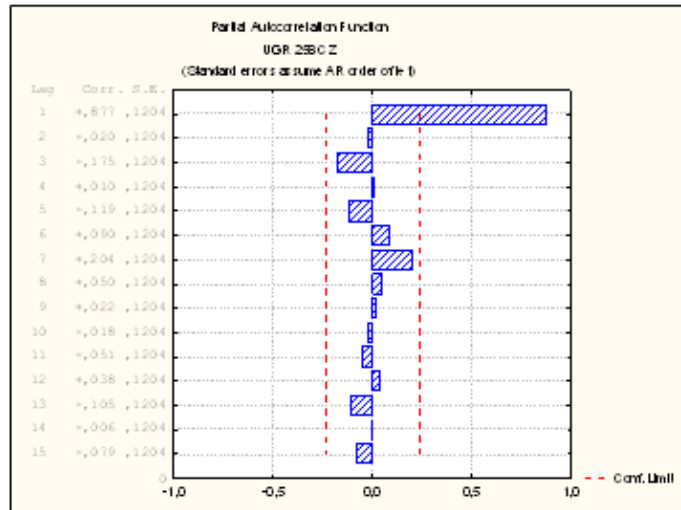
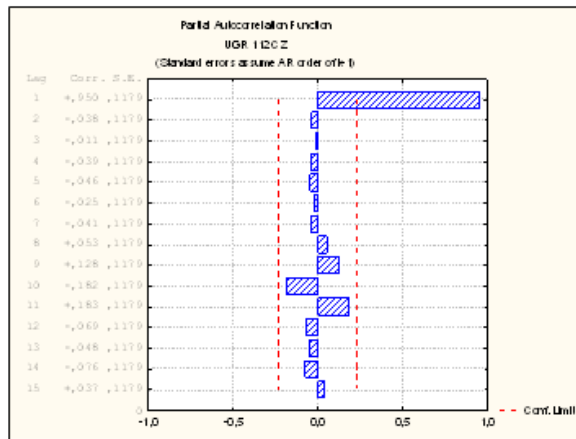


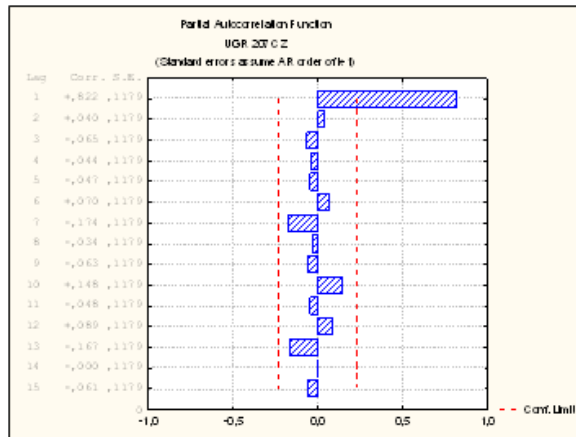
Figura G.4 – Autocorrelação Parcial UE 258 – Segmento Residencial de Vitória.

## G.5 - SEGMENTO COMERCIAL DE VITÓRIA - UE 207

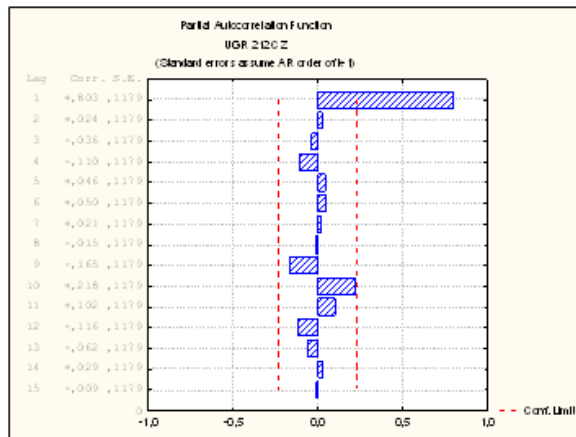
| UE112 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,95014         | 0,11785  |
|       | 2   | -0,03765        | 0,11785  |
|       | 3   | -0,01077        | 0,11785  |
|       | 4   | -0,03858        | 0,11785  |
|       | 5   | -0,04578        | 0,11785  |
|       | 6   | -0,02503        | 0,11785  |
|       | 7   | -0,04103        | 0,11785  |
|       | 8   | 0,05278         | 0,11785  |
|       | 9   | 0,12800         | 0,11785  |
|       | 10  | -0,18187        | 0,11785  |
|       | 11  | 0,18275         | 0,11785  |
|       | 12  | -0,06852        | 0,11785  |
|       | 13  | -0,04815        | 0,11785  |
|       | 14  | -0,07575        | 0,11785  |
|       | 15  | 0,03737         | 0,11785  |



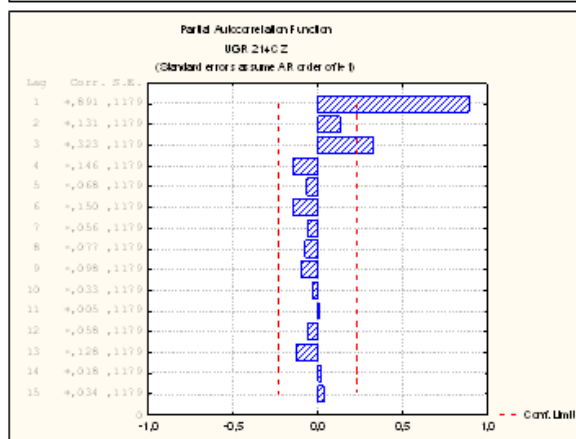
| UE207 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,82196         | 0,11785  |
|       | 2   | 0,03996         | 0,11785  |
|       | 3   | -0,06452        | 0,11785  |
|       | 4   | -0,04393        | 0,11785  |
|       | 5   | -0,04662        | 0,11785  |
|       | 6   | 0,06975         | 0,11785  |
|       | 7   | -0,17353        | 0,11785  |
|       | 8   | -0,03369        | 0,11785  |
|       | 9   | -0,06262        | 0,11785  |
|       | 10  | 0,14774         | 0,11785  |
|       | 11  | -0,04845        | 0,11785  |
|       | 12  | 0,08876         | 0,11785  |
|       | 13  | -0,16653        | 0,11785  |
|       | 14  | -0,00021        | 0,11785  |
|       | 15  | -0,06051        | 0,11785  |



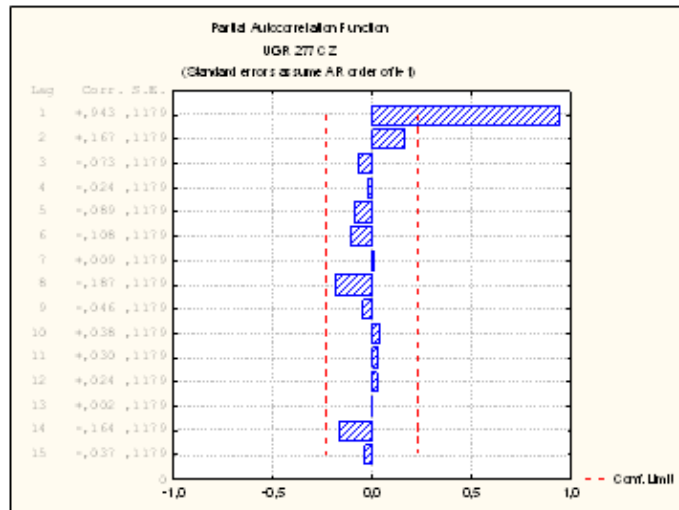
| UE212 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,80288         | 0,11785  |
|       | 2   | 0,02439         | 0,11785  |
|       | 3   | -0,03572        | 0,11785  |
|       | 4   | -0,11037        | 0,11785  |
|       | 5   | 0,04571         | 0,11785  |
|       | 6   | 0,05032         | 0,11785  |
|       | 7   | 0,02129         | 0,11785  |
|       | 8   | -0,01505        | 0,11785  |
|       | 9   | -0,16507        | 0,11785  |
|       | 10  | 0,21847         | 0,11785  |
|       | 11  | 0,10225         | 0,11785  |
|       | 12  | -0,11593        | 0,11785  |
|       | 13  | -0,06197        | 0,11785  |
|       | 14  | 0,02877         | 0,11785  |
|       | 15  | -0,00942        | 0,11785  |



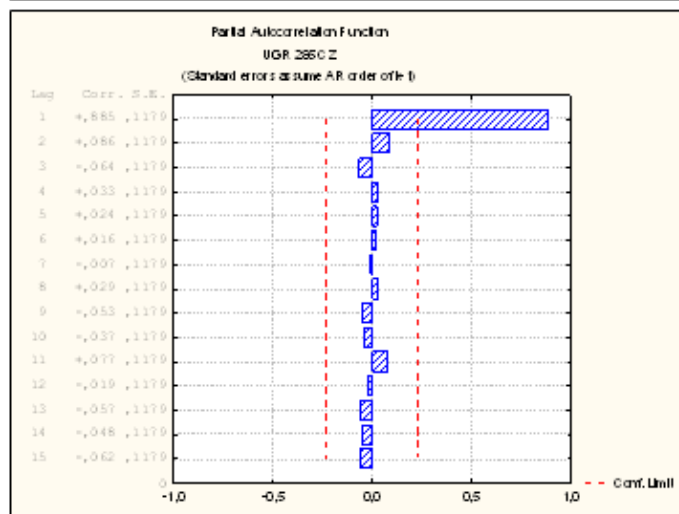
| UE214 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,89081         | 0,11785  |
|       | 2   | 0,13149         | 0,11785  |
|       | 3   | 0,32250         | 0,11785  |
|       | 4   | -0,14637        | 0,11785  |
|       | 5   | -0,06849        | 0,11785  |
|       | 6   | -0,14999        | 0,11785  |
|       | 7   | -0,05630        | 0,11785  |
|       | 8   | -0,07729        | 0,11785  |
|       | 9   | -0,09820        | 0,11785  |
|       | 10  | -0,03267        | 0,11785  |
|       | 11  | 0,00516         | 0,11785  |
|       | 12  | -0,05776        | 0,11785  |
|       | 13  | -0,12797        | 0,11785  |
|       | 14  | 0,01828         | 0,11785  |
|       | 15  | 0,03425         | 0,11785  |



| UE277 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,94348         | 0,11785  |
|       | 2   | 0,16671         | 0,11785  |
|       | 3   | -0,07276        | 0,11785  |
|       | 4   | -0,02420        | 0,11785  |
|       | 5   | -0,08947        | 0,11785  |
|       | 6   | -0,10763        | 0,11785  |
|       | 7   | 0,00871         | 0,11785  |
|       | 8   | -0,18738        | 0,11785  |
|       | 9   | -0,04583        | 0,11785  |
|       | 10  | 0,03848         | 0,11785  |
|       | 11  | 0,03037         | 0,11785  |
|       | 12  | 0,02416         | 0,11785  |
|       | 13  | 0,00162         | 0,11785  |
|       | 14  | -0,16397        | 0,11785  |
|       | 15  | -0,03696        | 0,11785  |



| UE285 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,88502         | 0,11785  |
|       | 2   | 0,08553         | 0,11785  |
|       | 3   | -0,06408        | 0,11785  |
|       | 4   | 0,03261         | 0,11785  |
|       | 5   | 0,02446         | 0,11785  |
|       | 6   | 0,01644         | 0,11785  |
|       | 7   | -0,00657        | 0,11785  |
|       | 8   | 0,02870         | 0,11785  |
|       | 9   | -0,05319        | 0,11785  |
|       | 10  | -0,03668        | 0,11785  |
|       | 11  | 0,07718         | 0,11785  |
|       | 12  | -0,01929        | 0,11785  |
|       | 13  | -0,05667        | 0,11785  |
|       | 14  | -0,04838        | 0,11785  |
|       | 15  | -0,06229        | 0,11785  |



| UE334 | Lag | Partial-Auto r. | Std.Err. |
|-------|-----|-----------------|----------|
|       | 1   | 0,80195         | 0,11785  |
|       | 2   | -0,21032        | 0,11785  |
|       | 3   | 0,04573         | 0,11785  |
|       | 4   | 0,31977         | 0,11785  |
|       | 5   | -0,11641        | 0,11785  |
|       | 6   | 0,02823         | 0,11785  |
|       | 7   | 0,17256         | 0,11785  |
|       | 8   | -0,10617        | 0,11785  |
|       | 9   | 0,05056         | 0,11785  |
|       | 10  | 0,07412         | 0,11785  |
|       | 11  | -0,00667        | 0,11785  |
|       | 12  | 0,01512         | 0,11785  |
|       | 13  | 0,01197         | 0,11785  |
|       | 14  | 0,01041         | 0,11785  |
|       | 15  | -0,00404        | 0,11785  |

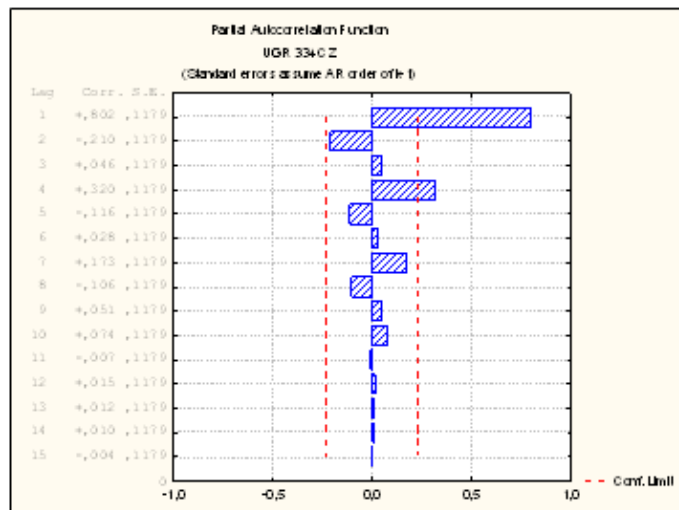


Figura G.5 – Autocorrelação Parcial UE 207 – Segmento Comercial de Vitória.

# APÊNDICE (H) - PREVISÃO SAR *PANEL* – GRÁFICOS DE ERRO

## H.1 - SEGMENTO RESIDENCIAL DE CURITIBA - UE 624

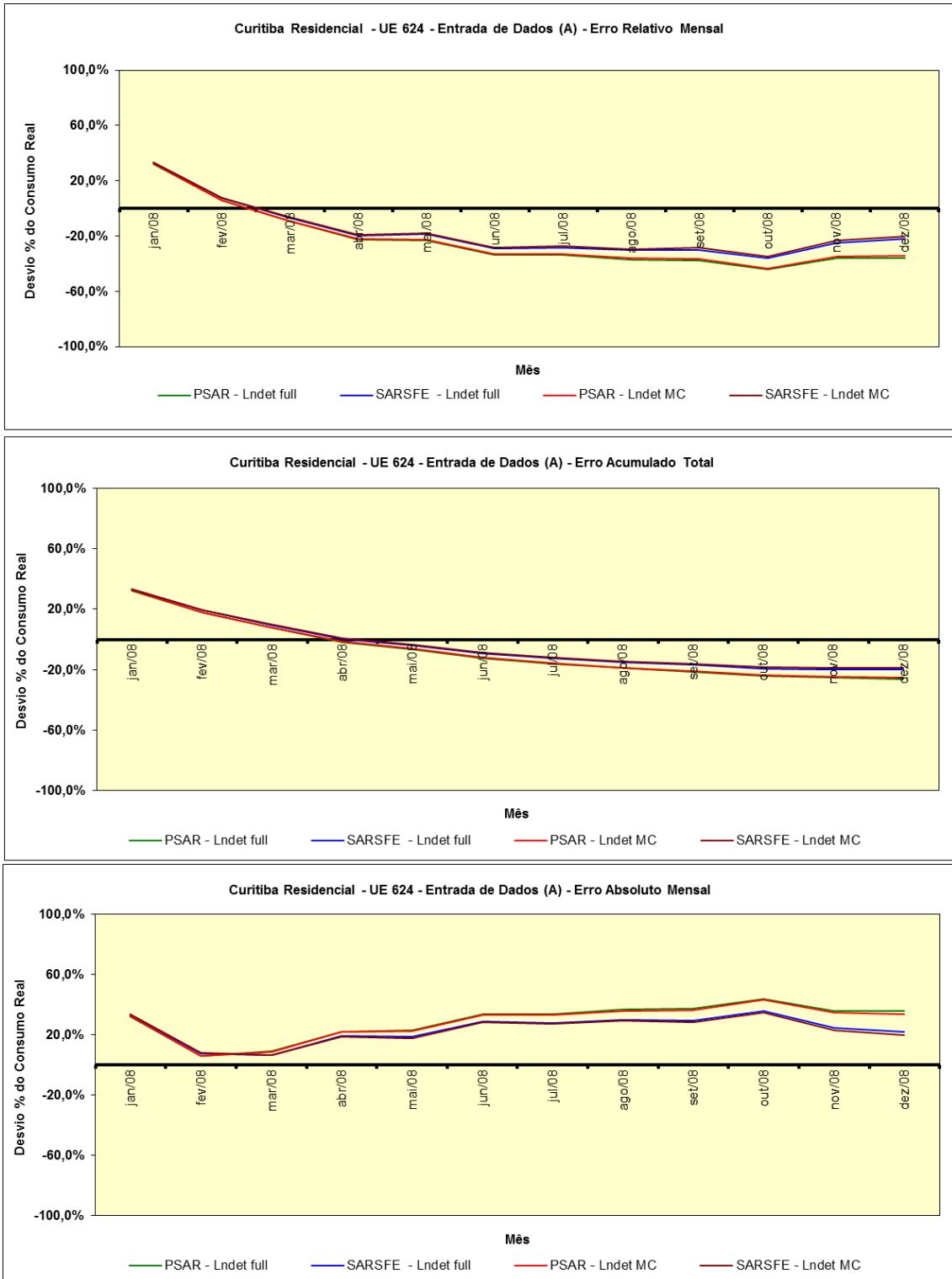


Figura H.1 – Previsão UE 624 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (A).

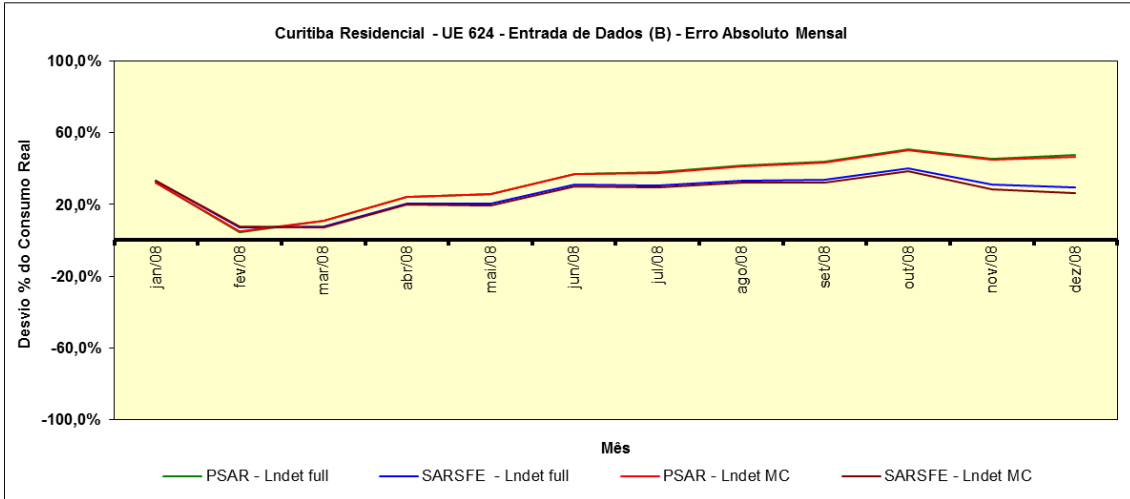
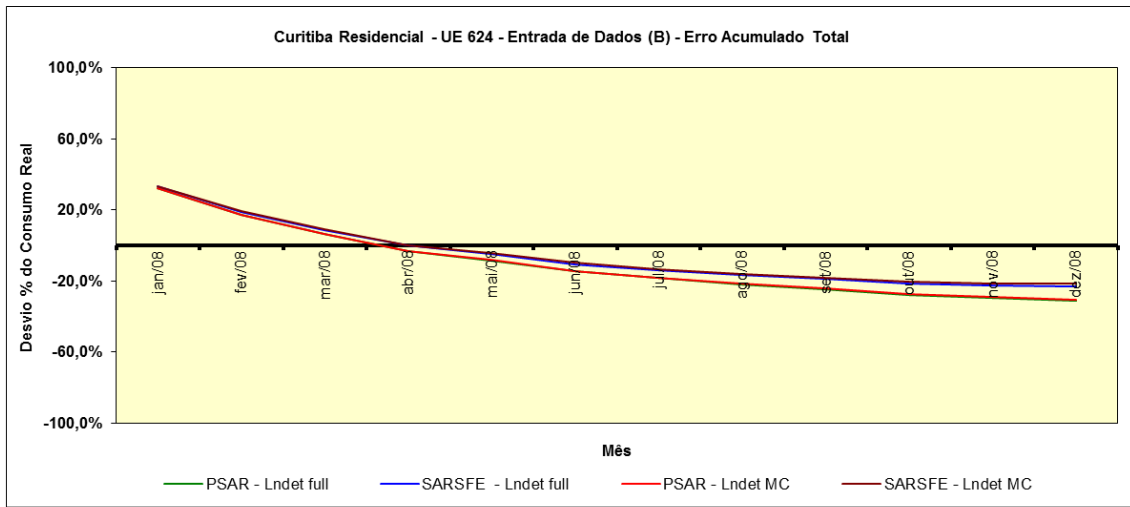
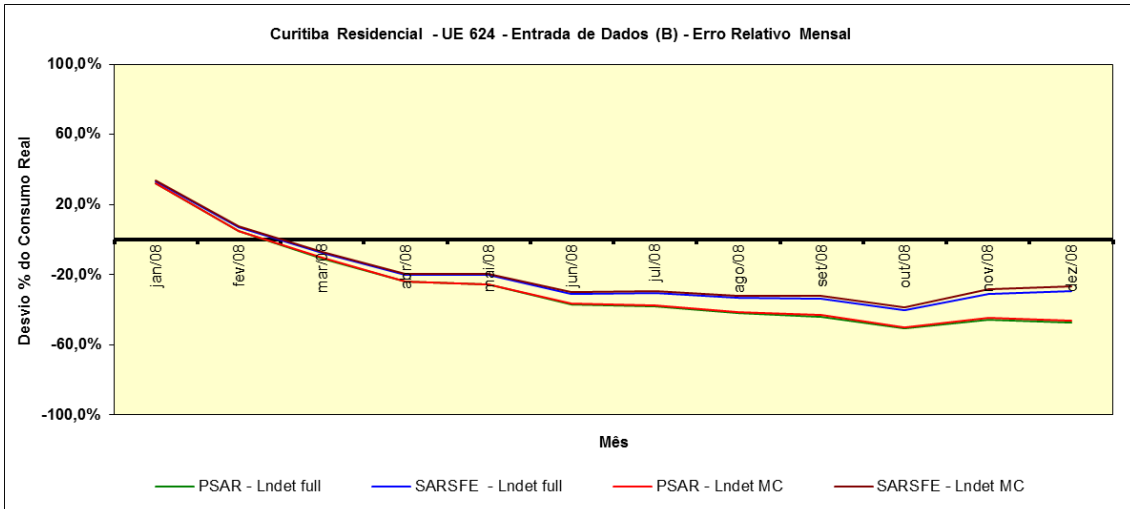


Figura H.2 – Previsão UE 624 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (B).

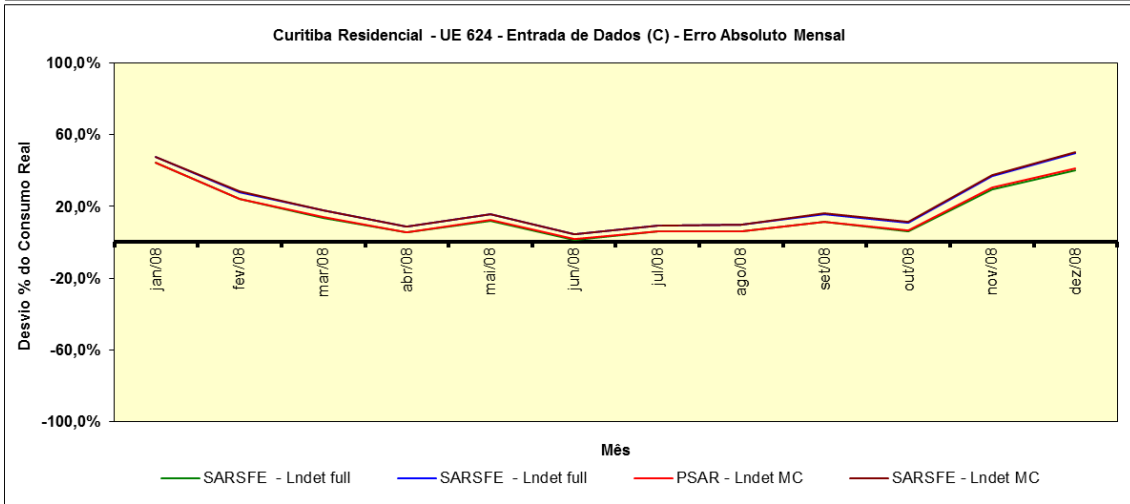
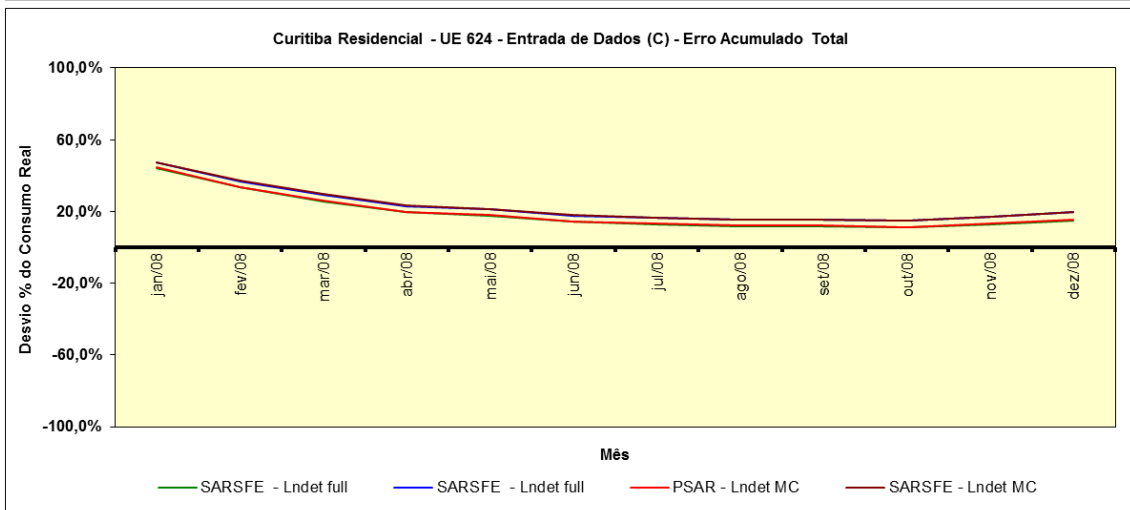
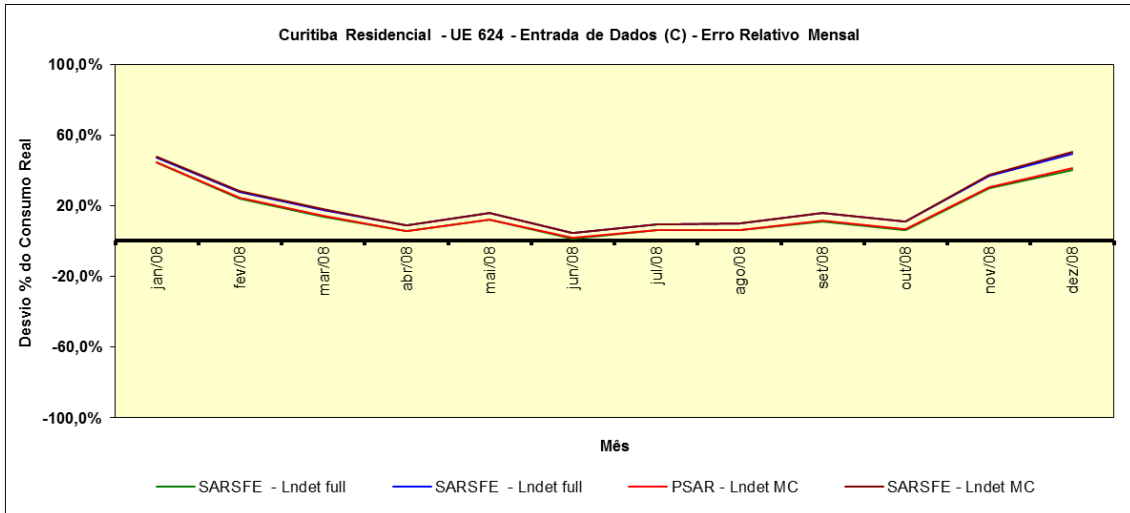


Figura H.3 – Previsão UE 624 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (C).

## H.2 - SEGMENTO RESIDENCIAL DE CURITIBA - UE 638

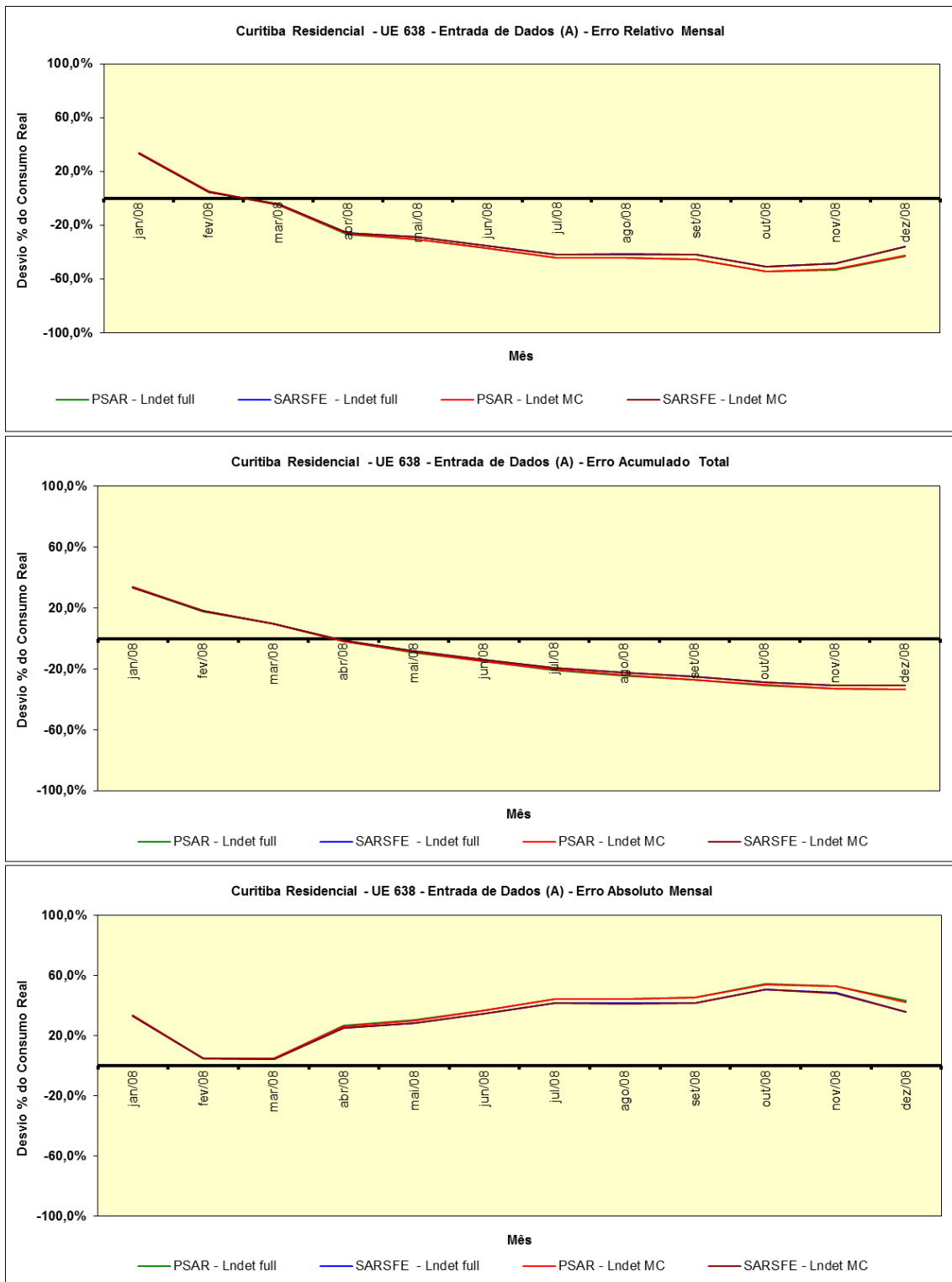


Figura H.4 – Previsão UE 638 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (A).



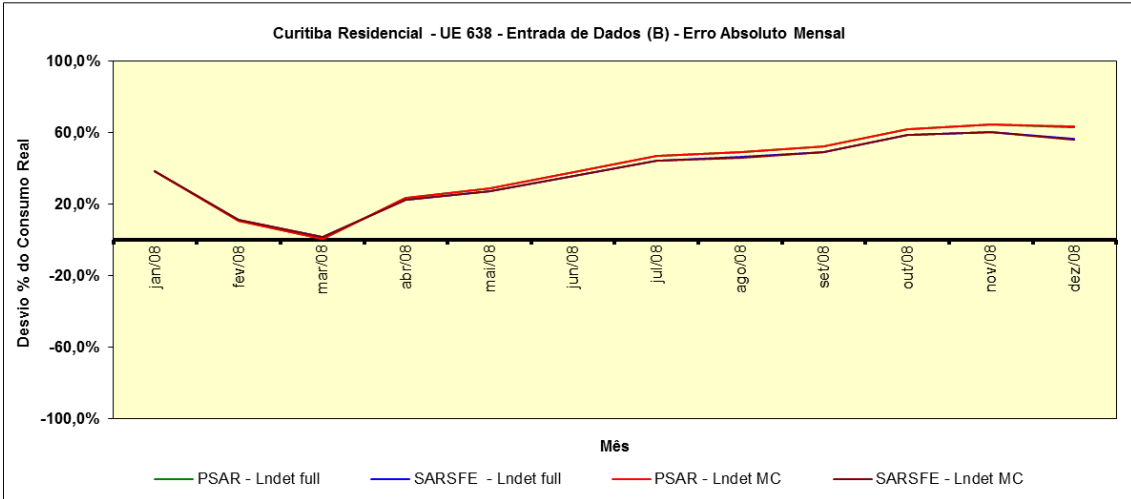
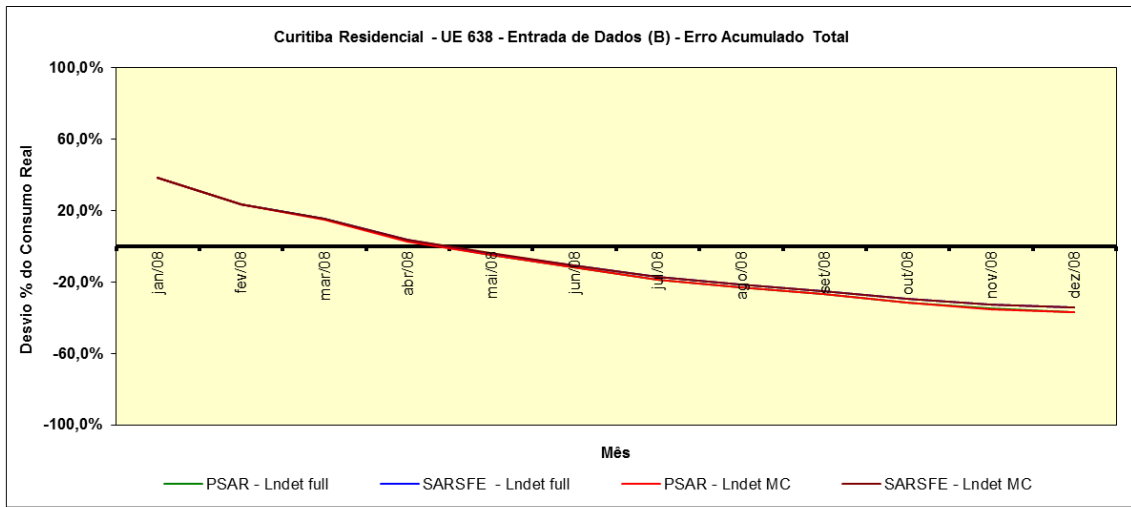
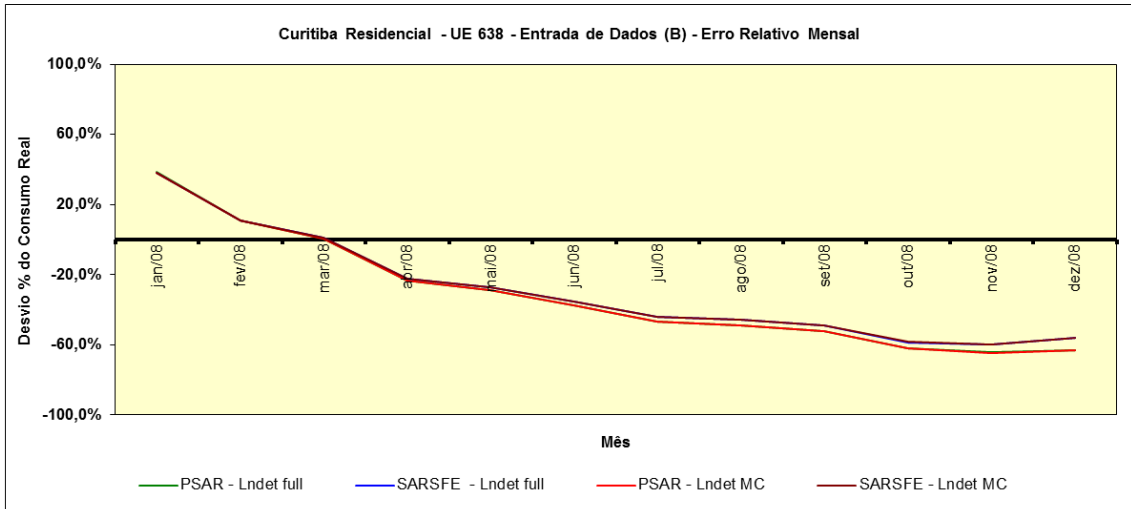


Figura H.5 – Previsão UE 638 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (B).

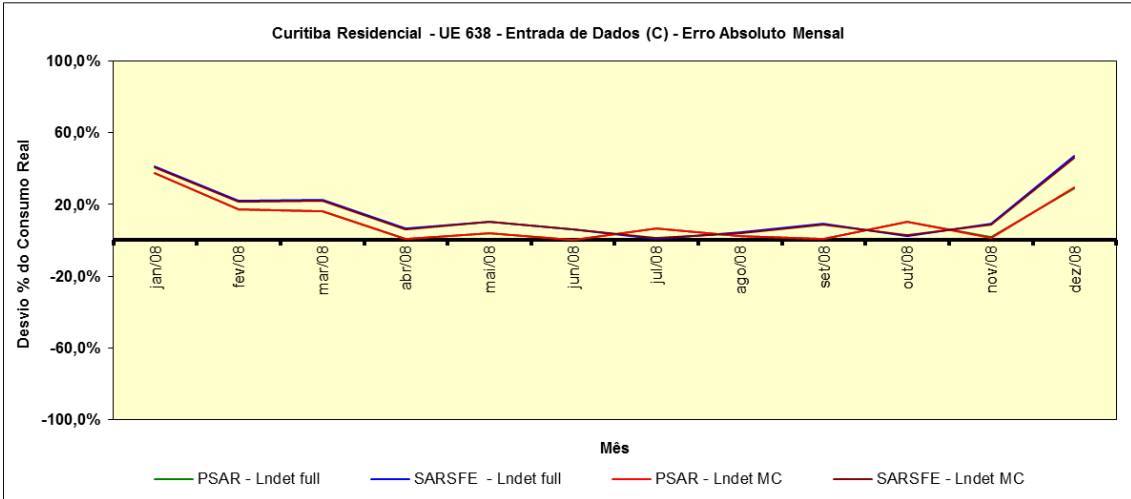
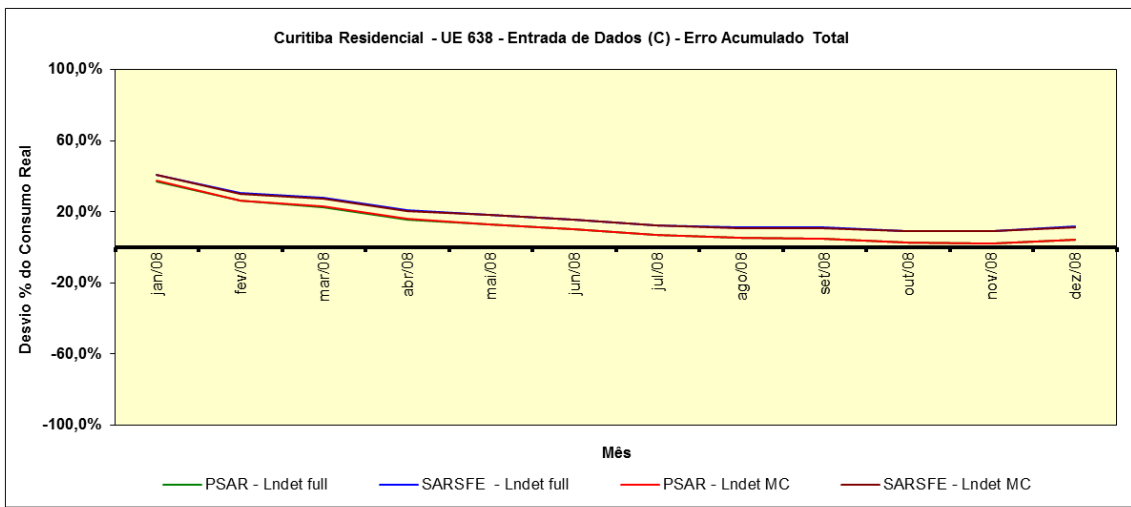
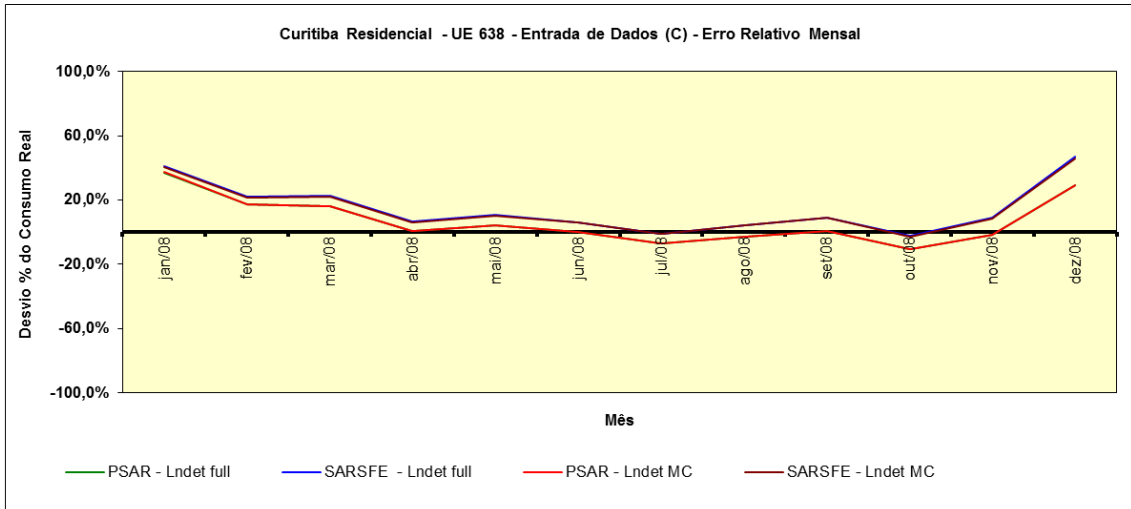


Figura H.6 – Previsão UE 638 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (C).

### H.3 - SEGMENTO COMERCIAL DE VITÓRIA - UE 207

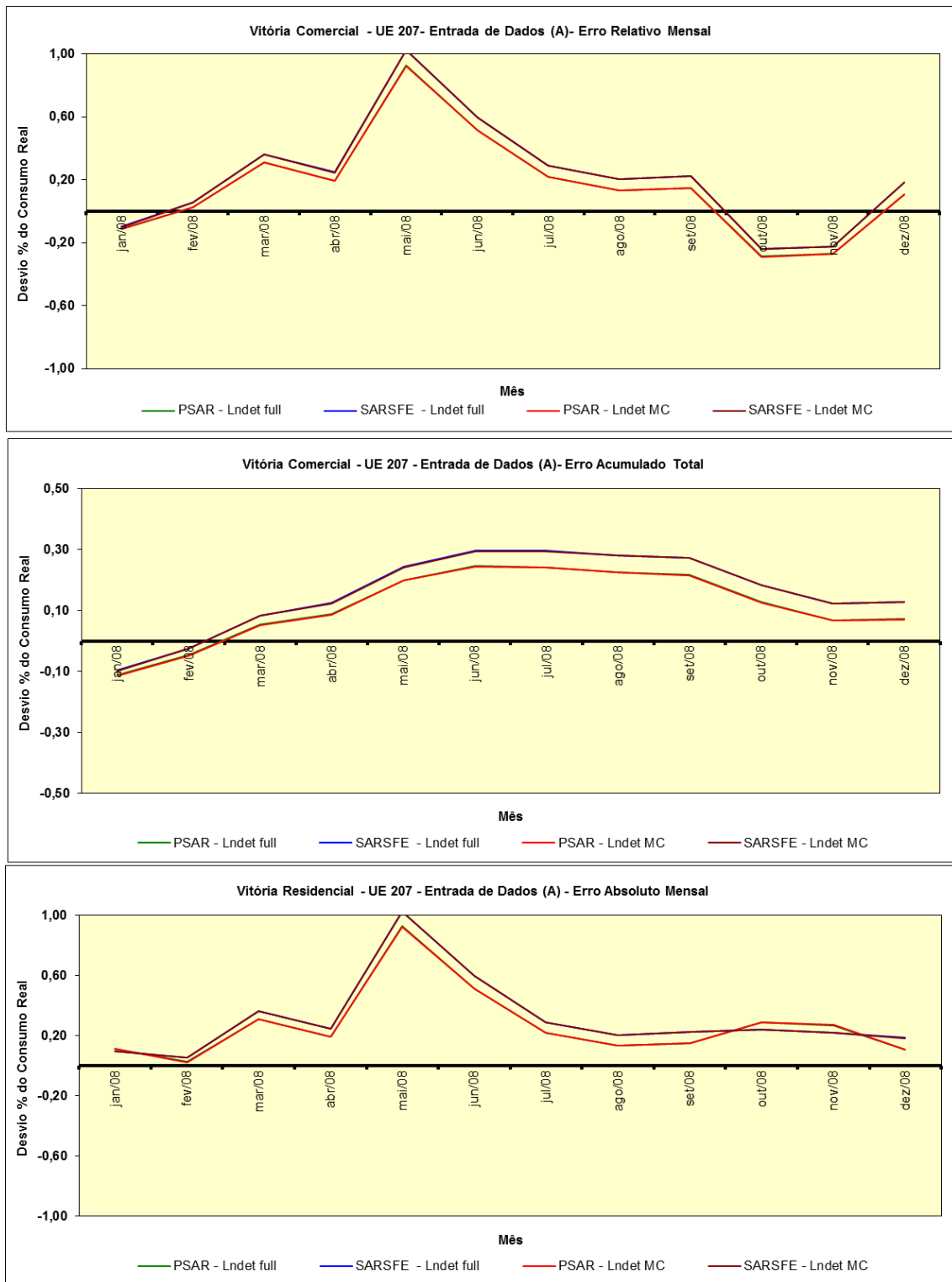


Figura H.7 – Previsão UE 207 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (A).

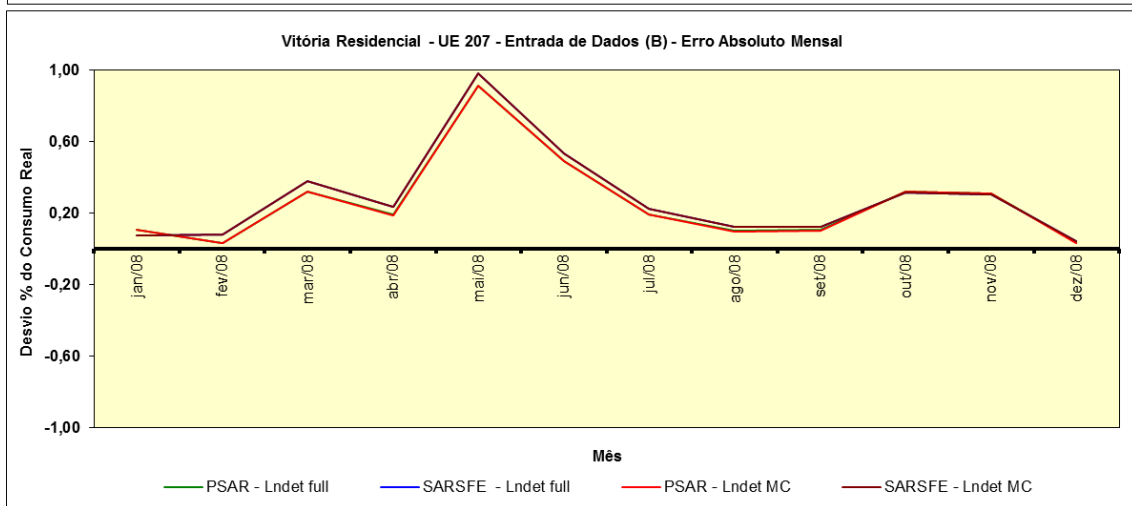
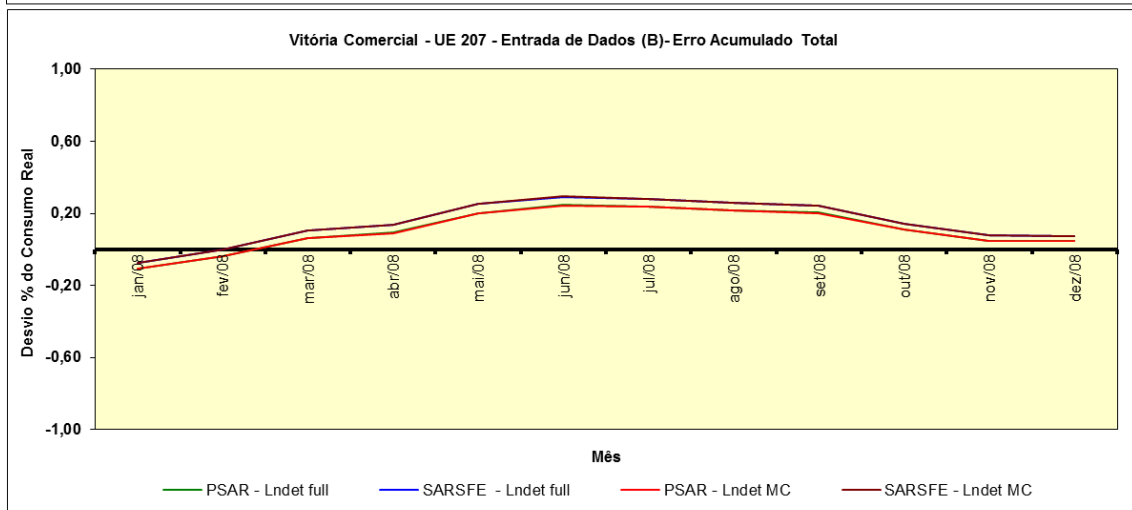
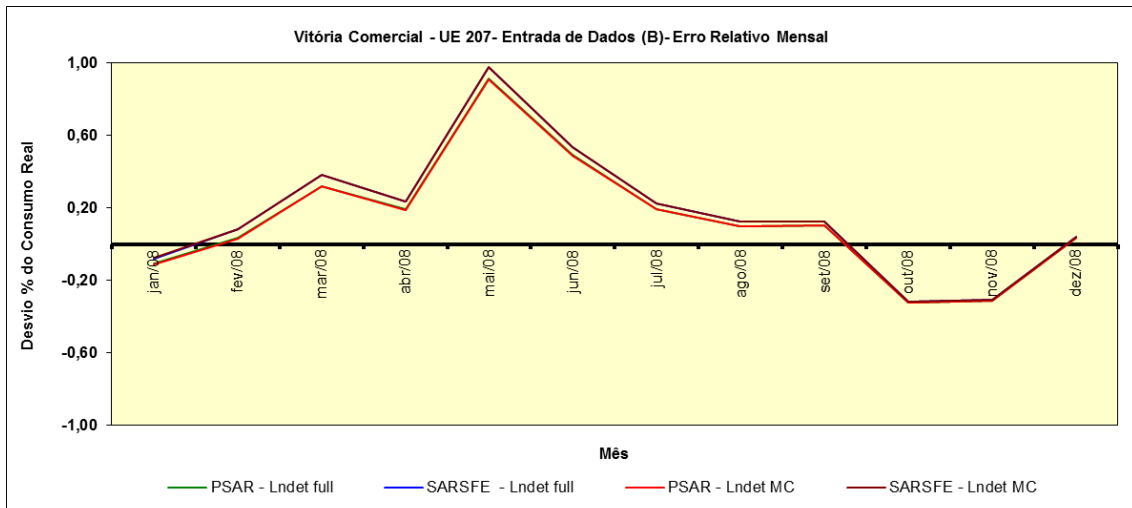


Figura H.8 – Previsão UE 207 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (B).

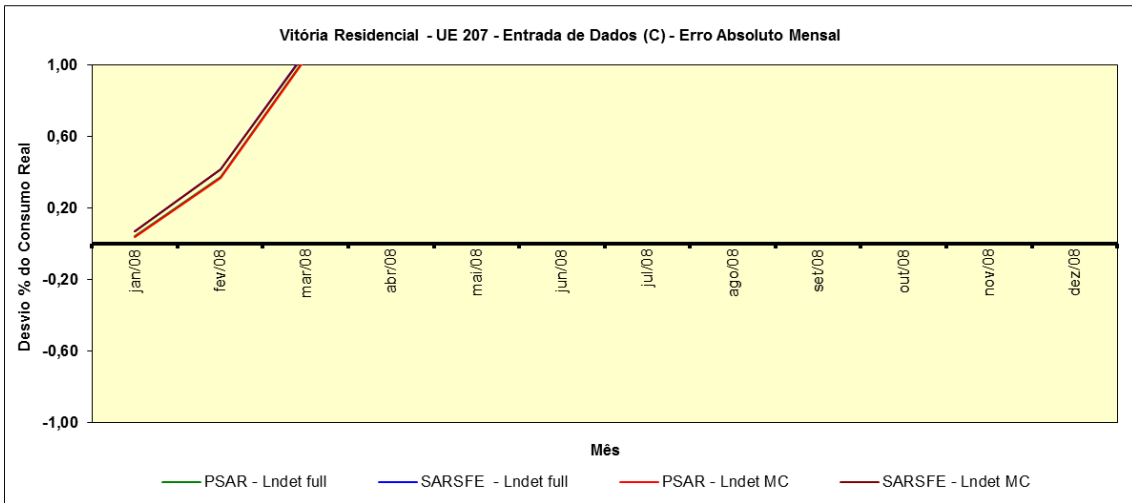
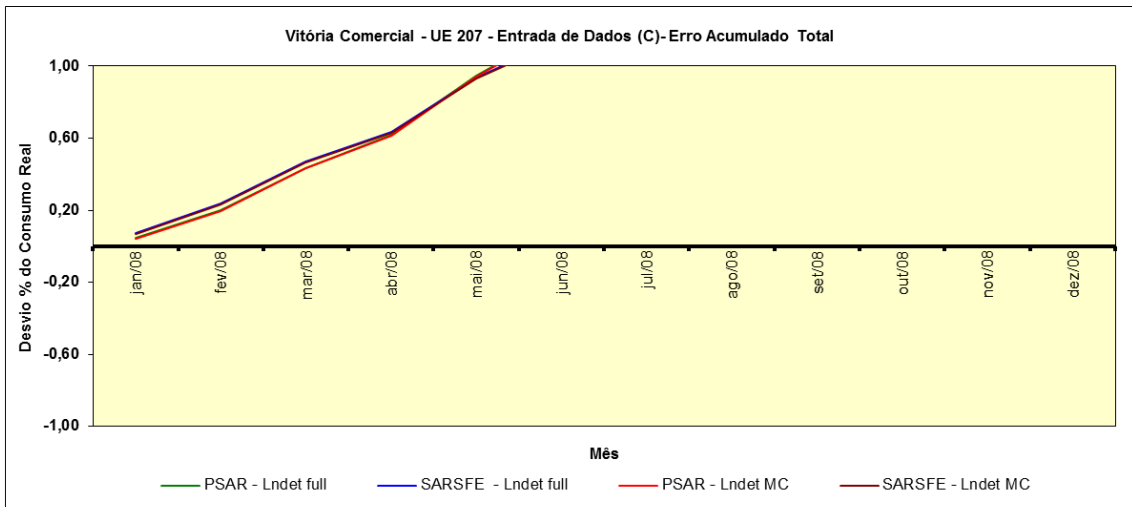
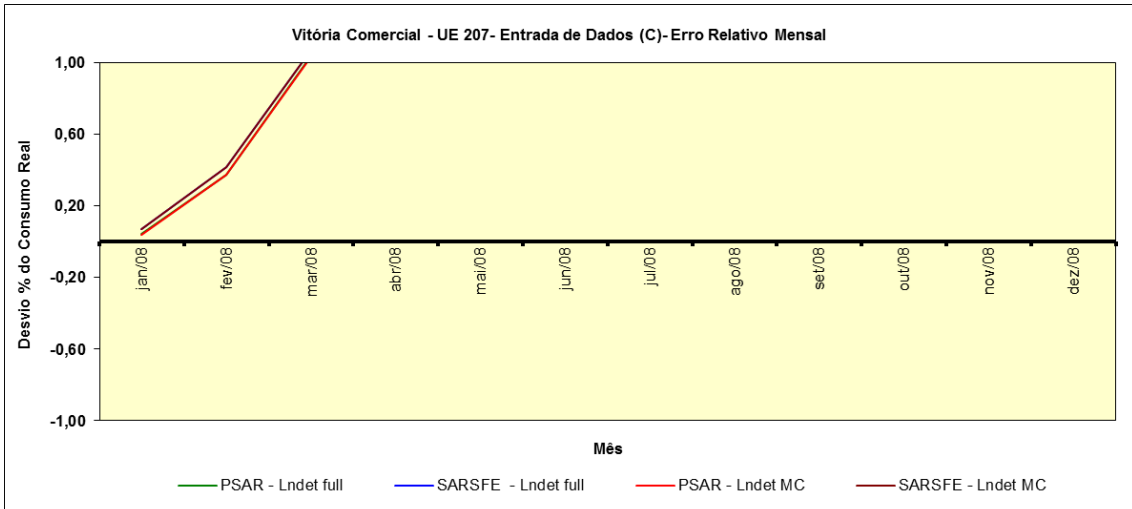


Figura H.9 – Previsão UE 207 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (C).

#### H.4 - SEGMENTO RESIDENCIAL DE VITÓRIA - UE 247

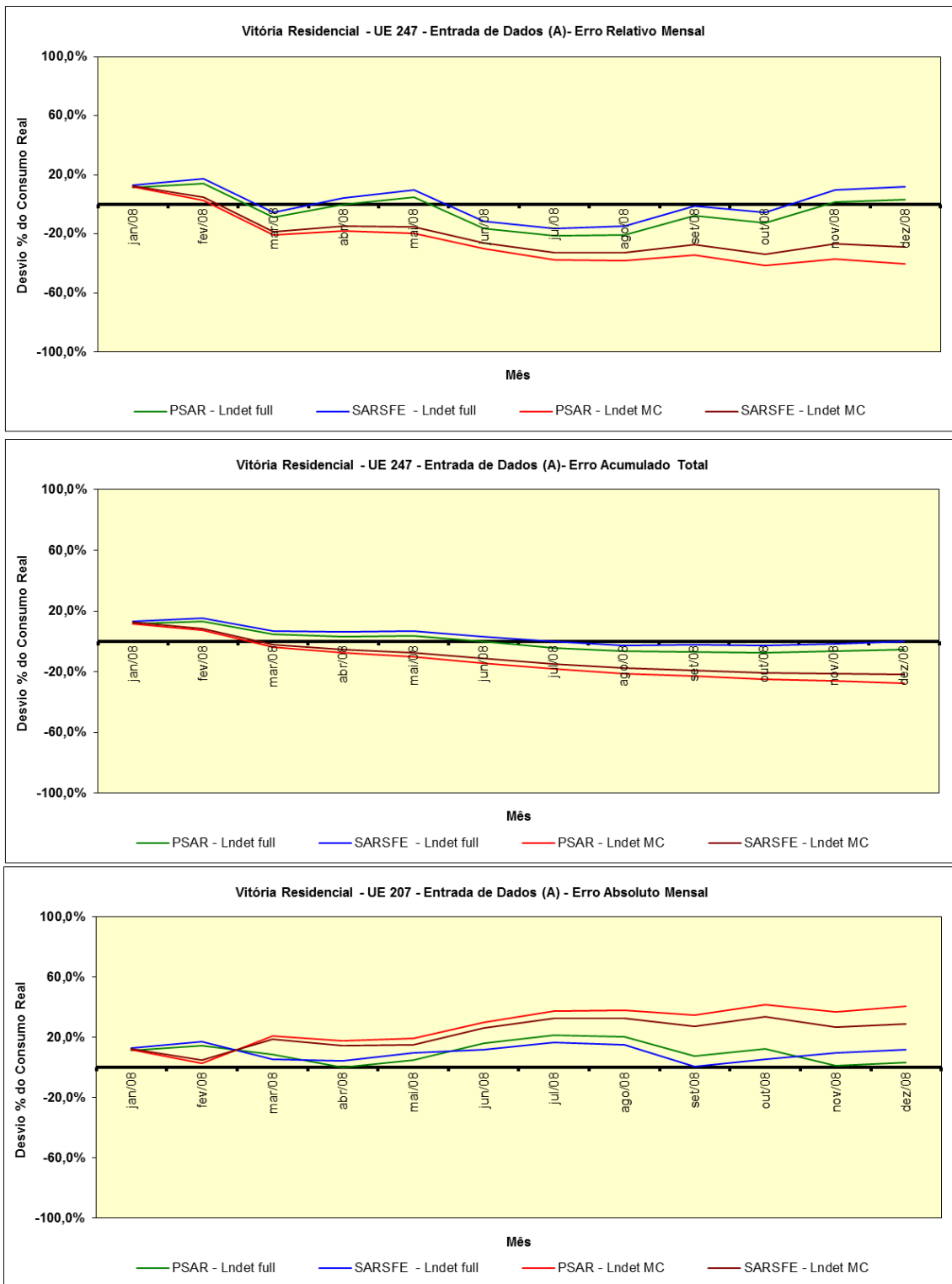


Figura H.10 – Previsão UE 247 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (A).

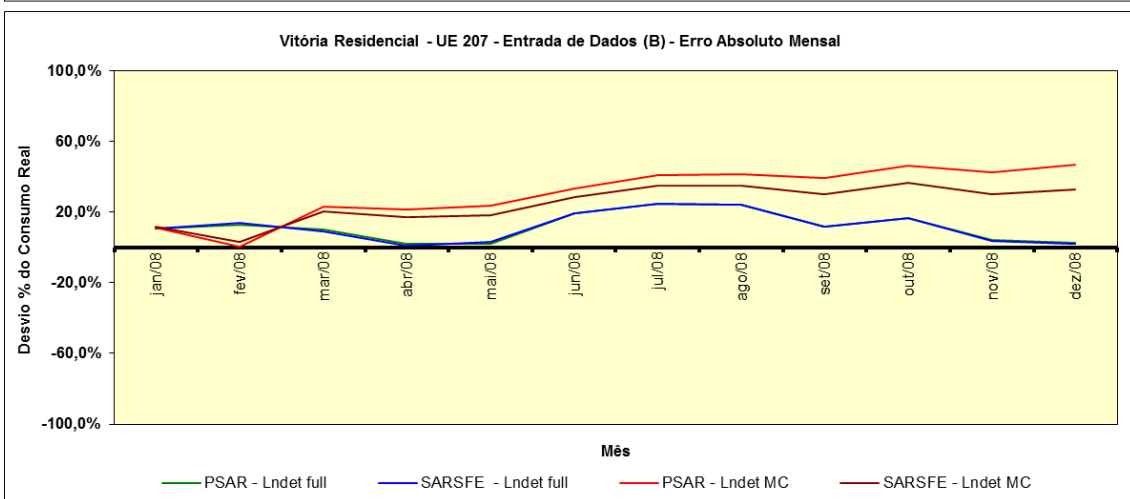
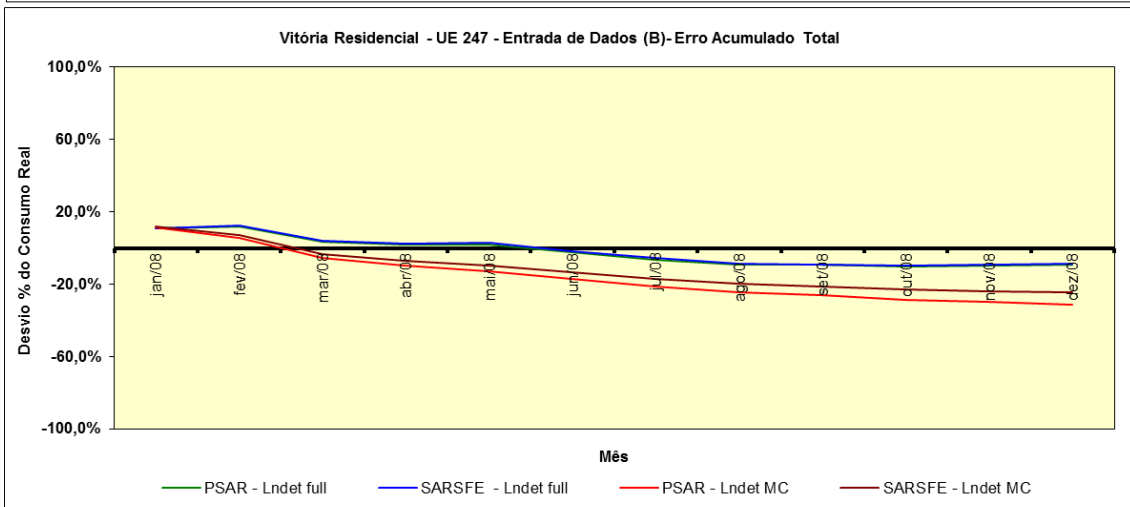
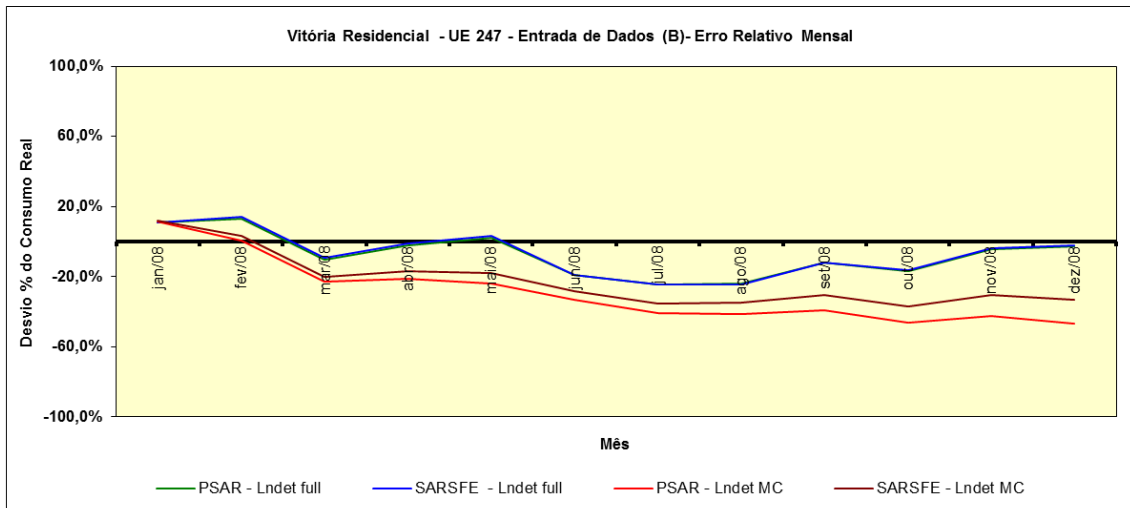


Figura H.11 – Previsão UE 247 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (B).

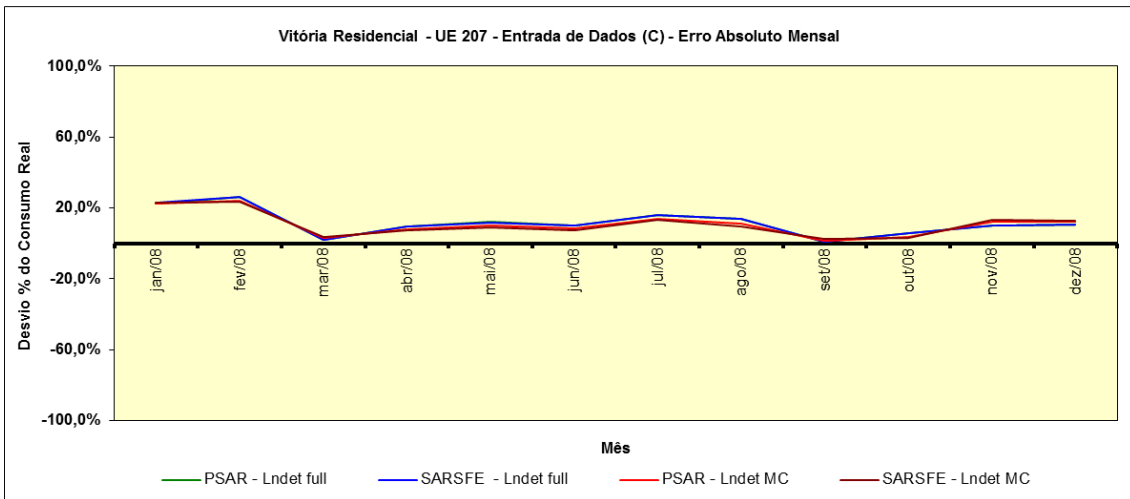
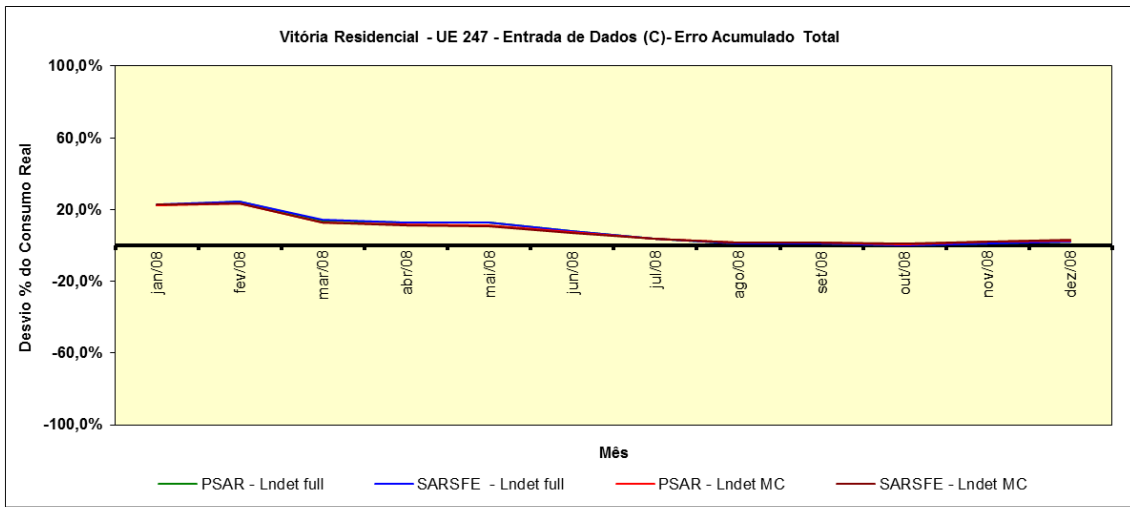
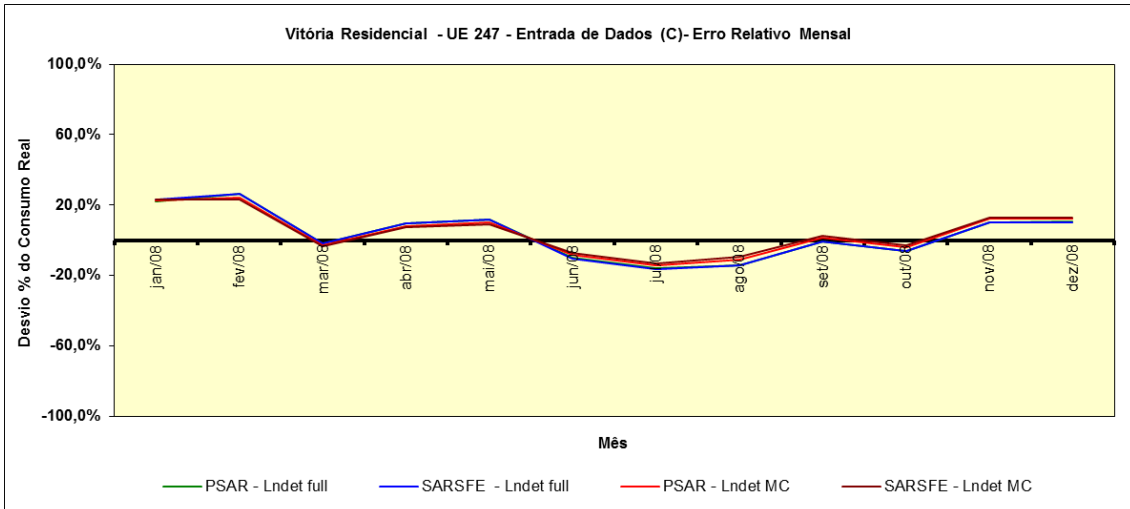


Figura H.12 – Previsão UE 247 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (C).



## H.5 - SEGMENTO RESIDENCIAL DE VITÓRIA - UE 258

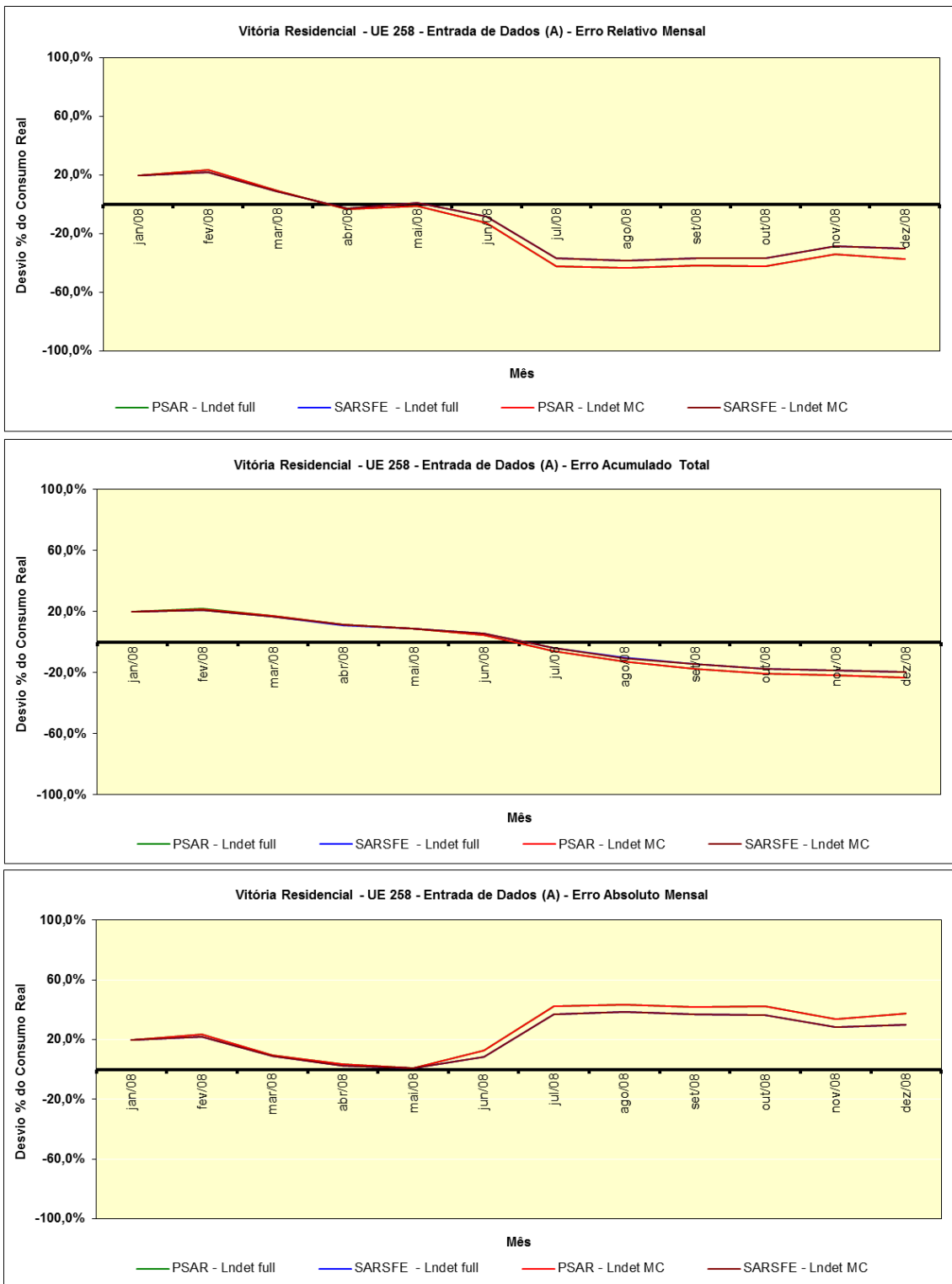


Figura H.13 – Previsão UE 258 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (A).

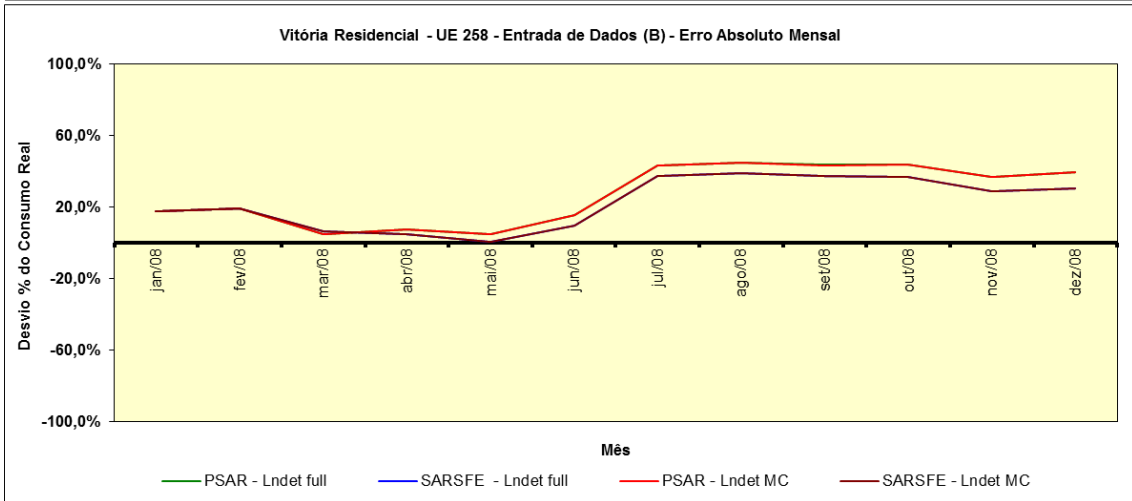
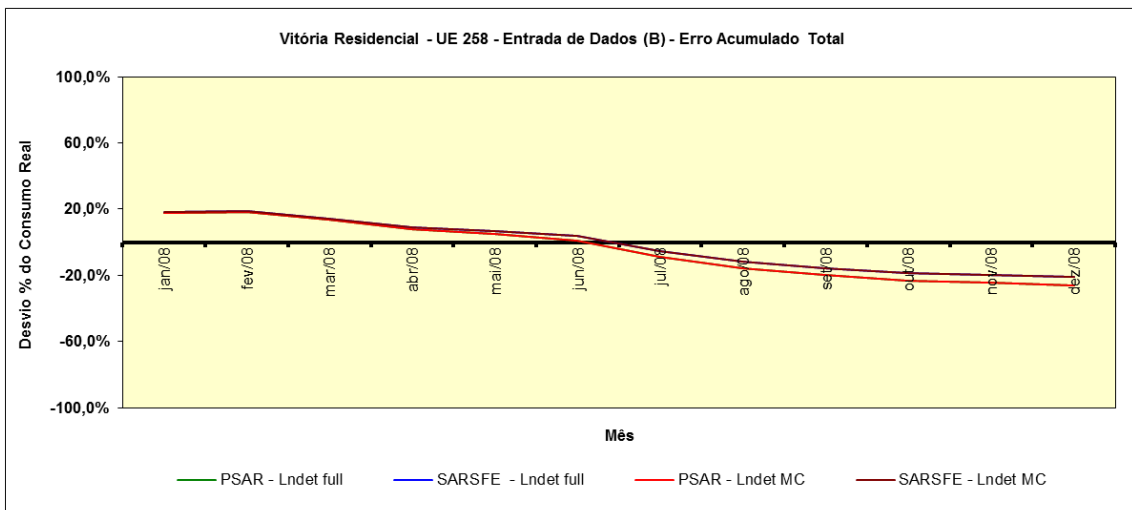
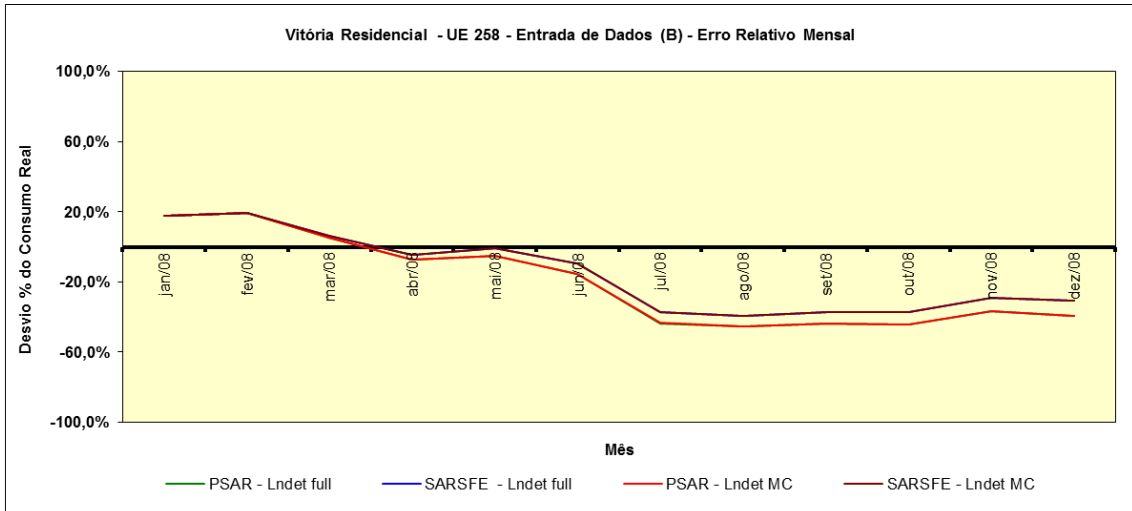


Figura H.14 – Previsão UE 258 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (B).

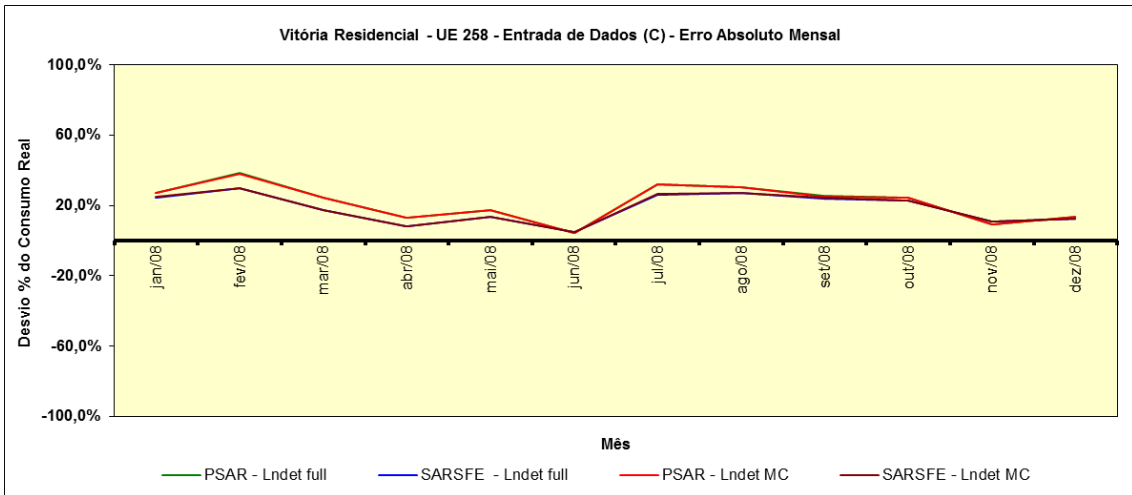
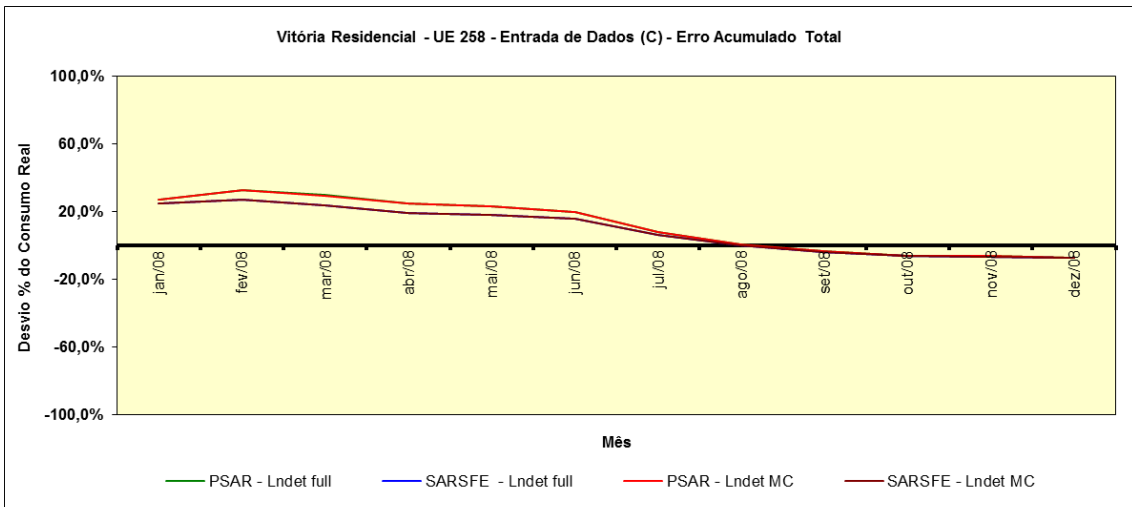
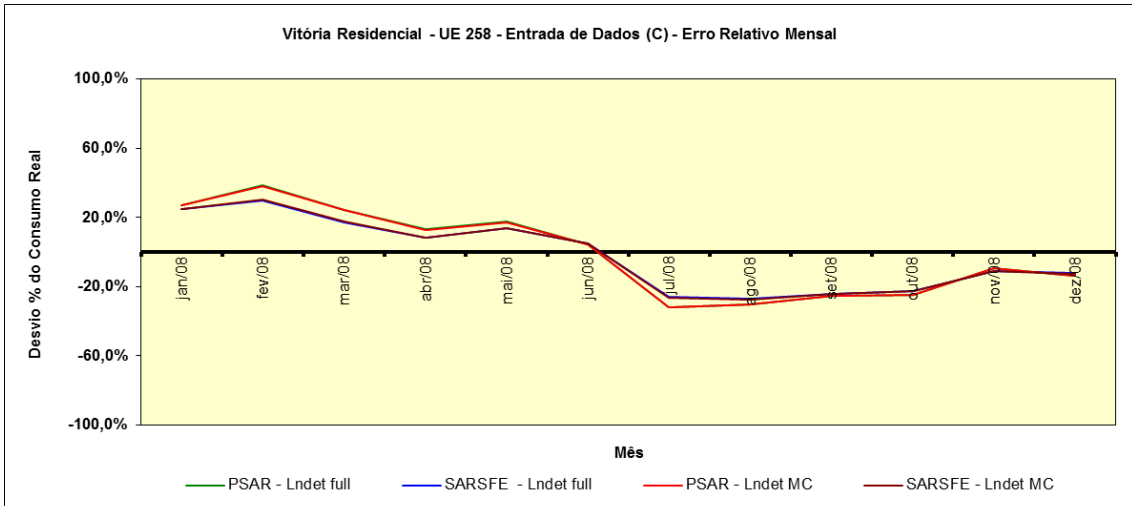


Figura H.15 – Previsão UE 258 – Gráficos de Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal) – Opção de Entrada de Dados (C).

## APÊNDICE (I) - PREVISÃO ARIMA POR UE

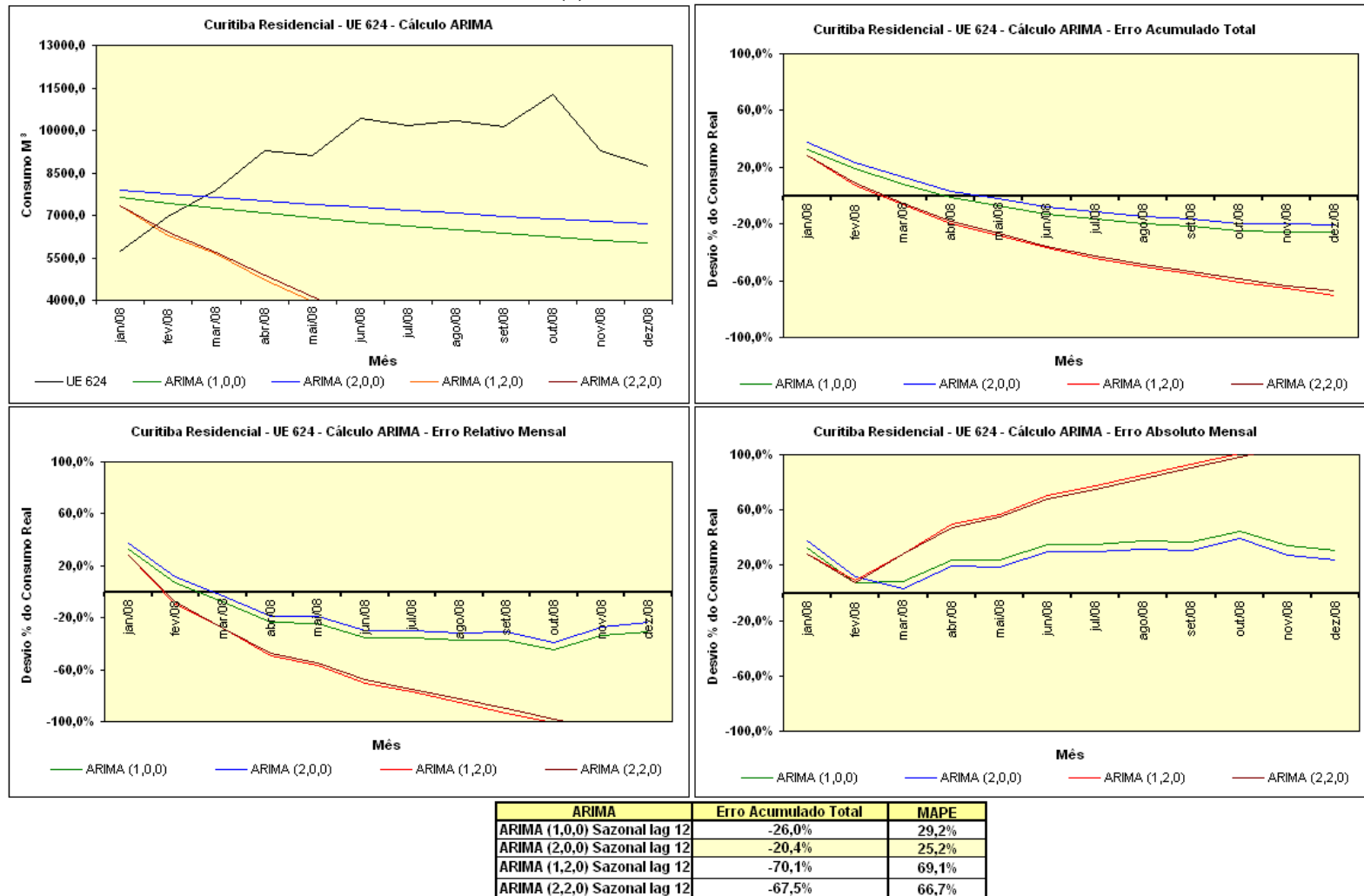
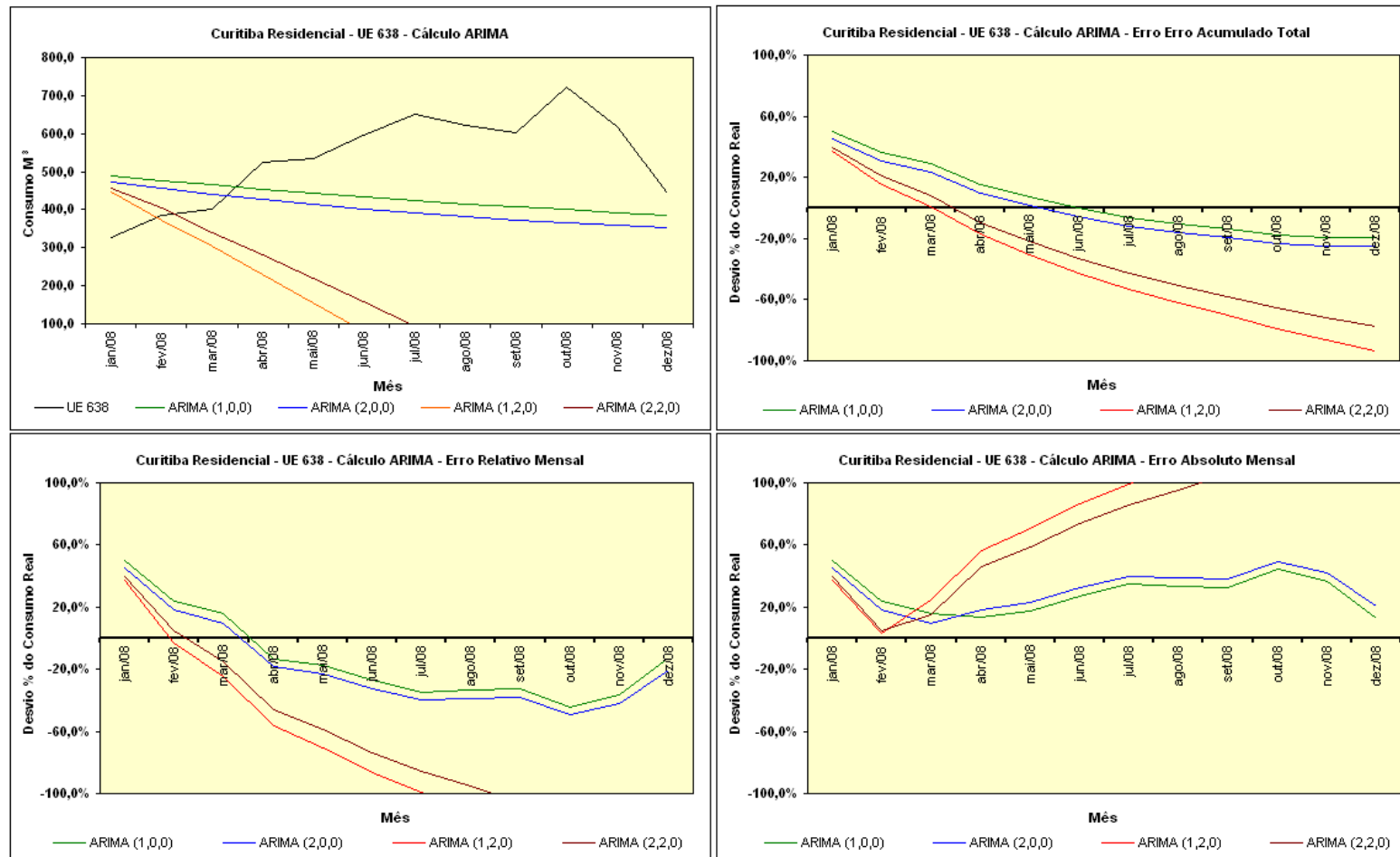
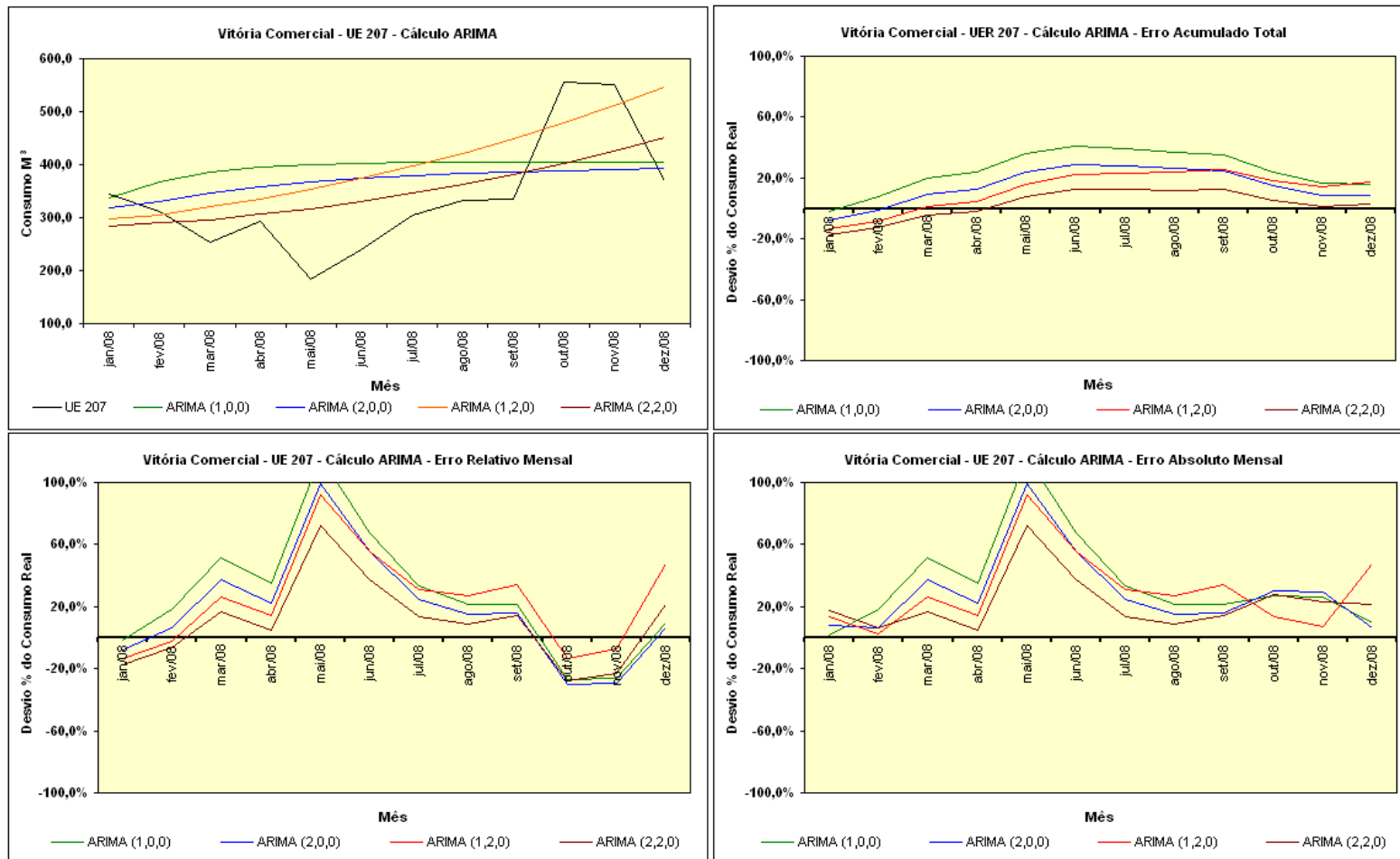


Figura J.1 – Previsão ARIMA UE 624– Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



| ARIMA                        | Erro Acumulado Total | MAPE  |
|------------------------------|----------------------|-------|
| ARIMA (1,0,0) Sazonal lag 12 | -19,2%               | 28,5% |
| ARIMA (2,0,0) Sazonal lag 12 | -24,8%               | 31,3% |
| ARIMA (1,2,0) Sazonal lag 12 | -93,4%               | 90,6% |
| ARIMA (2,2,0) Sazonal lag 12 | -77,6%               | 76,7% |

Figura J.2 – Previsão ARIMA UE 638– Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



| ARIMA                        | Erro Acumulado Total | MAPE  |
|------------------------------|----------------------|-------|
| ARIMA (1,0,0) Sazonal lag 12 | 15,8%                | 35,8% |
| ARIMA (2,0,0) Sazonal lag 12 | 8,4%                 | 29,1% |
| ARIMA (1,2,0) Sazonal lag 12 | 17,6%                | 30,4% |
| ARIMA (2,2,0) Sazonal lag 12 | 3,0%                 | 22,0% |

Figura J.3 – Previsão ARIMA UE 207– Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).

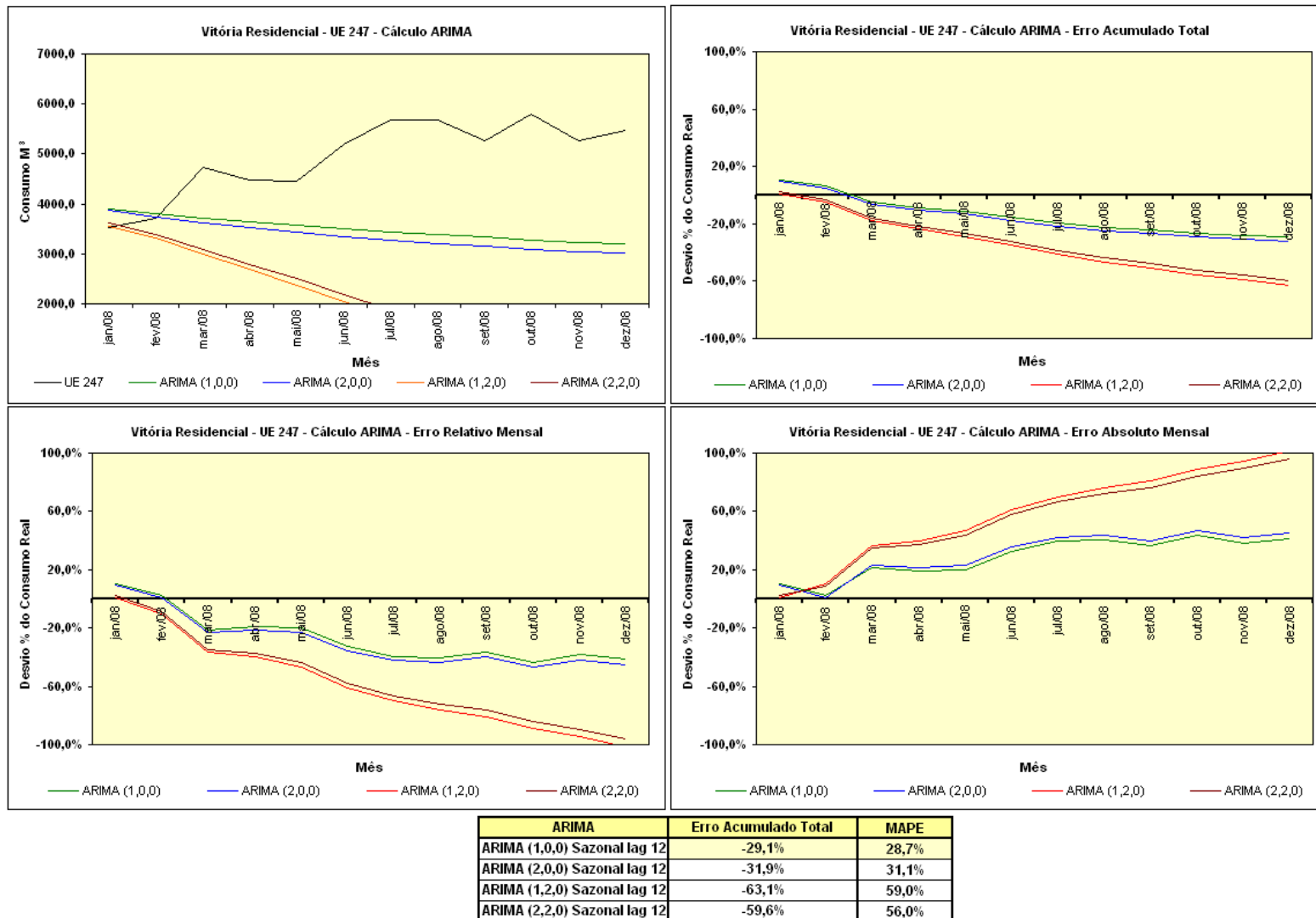
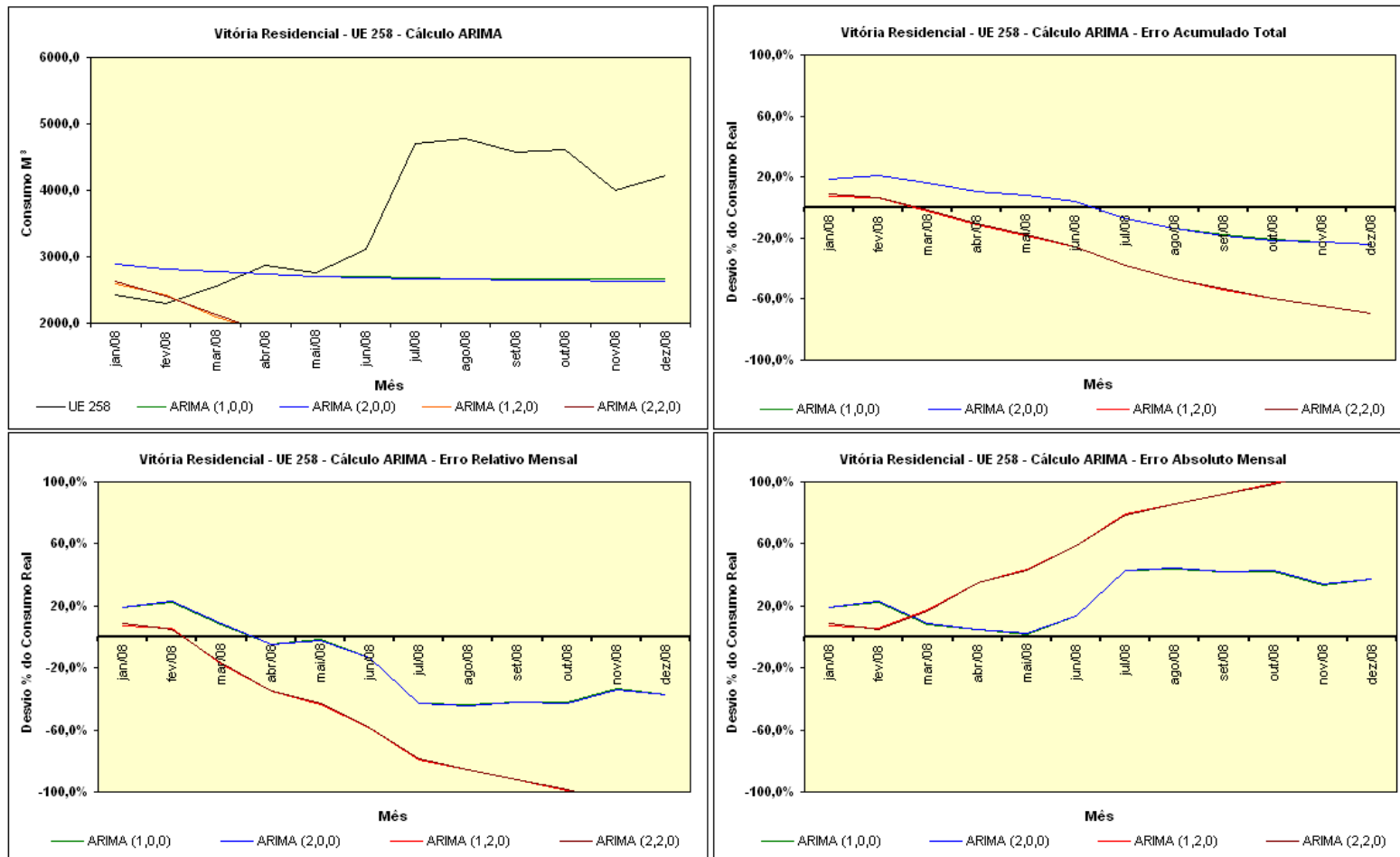


Figura J.4 – Previsão ARIMA UE 247– Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).

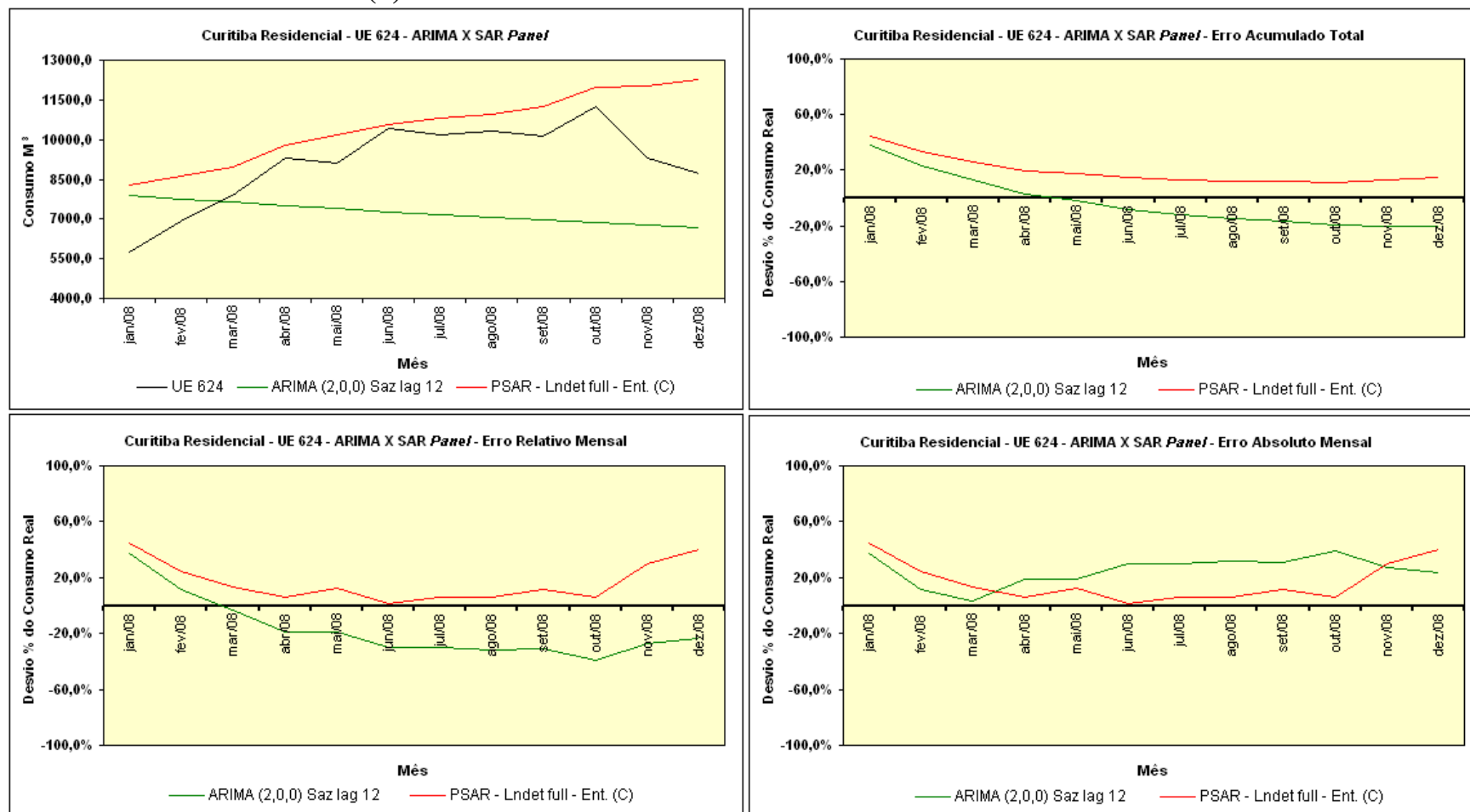


| ARIMA                        | Erro Acumulado Total | MAPE  |
|------------------------------|----------------------|-------|
| ARIMA (1,0,0) Sazonal lag 12 | -24,0%               | 25,9% |
| ARIMA (2,0,0) Sazonal lag 12 | -24,2%               | 26,2% |
| ARIMA (1,2,0) Sazonal lag 12 | -69,7%               | 62,1% |
| ARIMA (2,2,0) Sazonal lag 12 | -69,3%               | 61,9% |

Figura J.5 – Previsão ARIMA UE 258– Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).

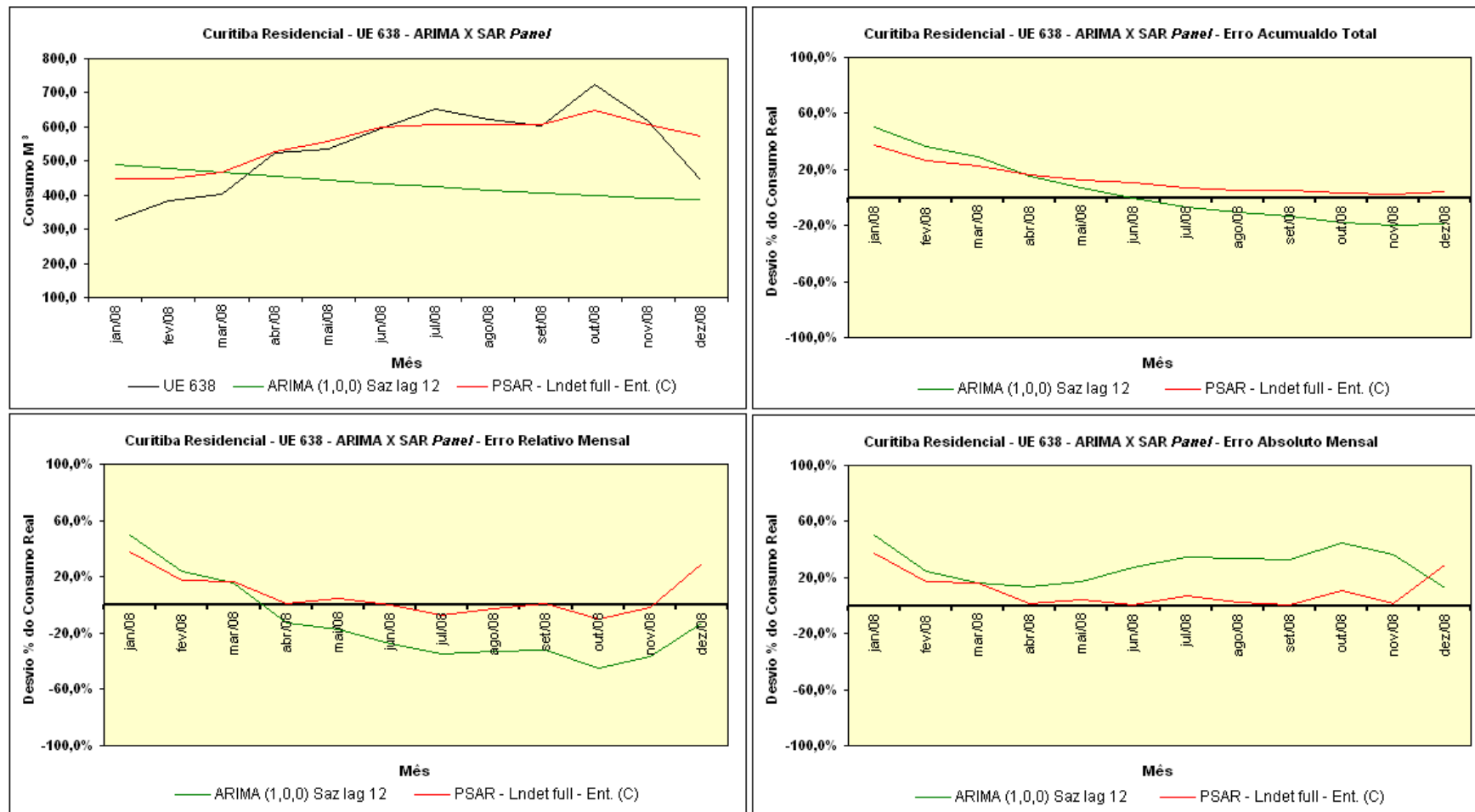


## APÊNDICE (J) - PREVISÕES SAR *Panel* x ARIMA / ALISAMENTO EXPONENCIAL



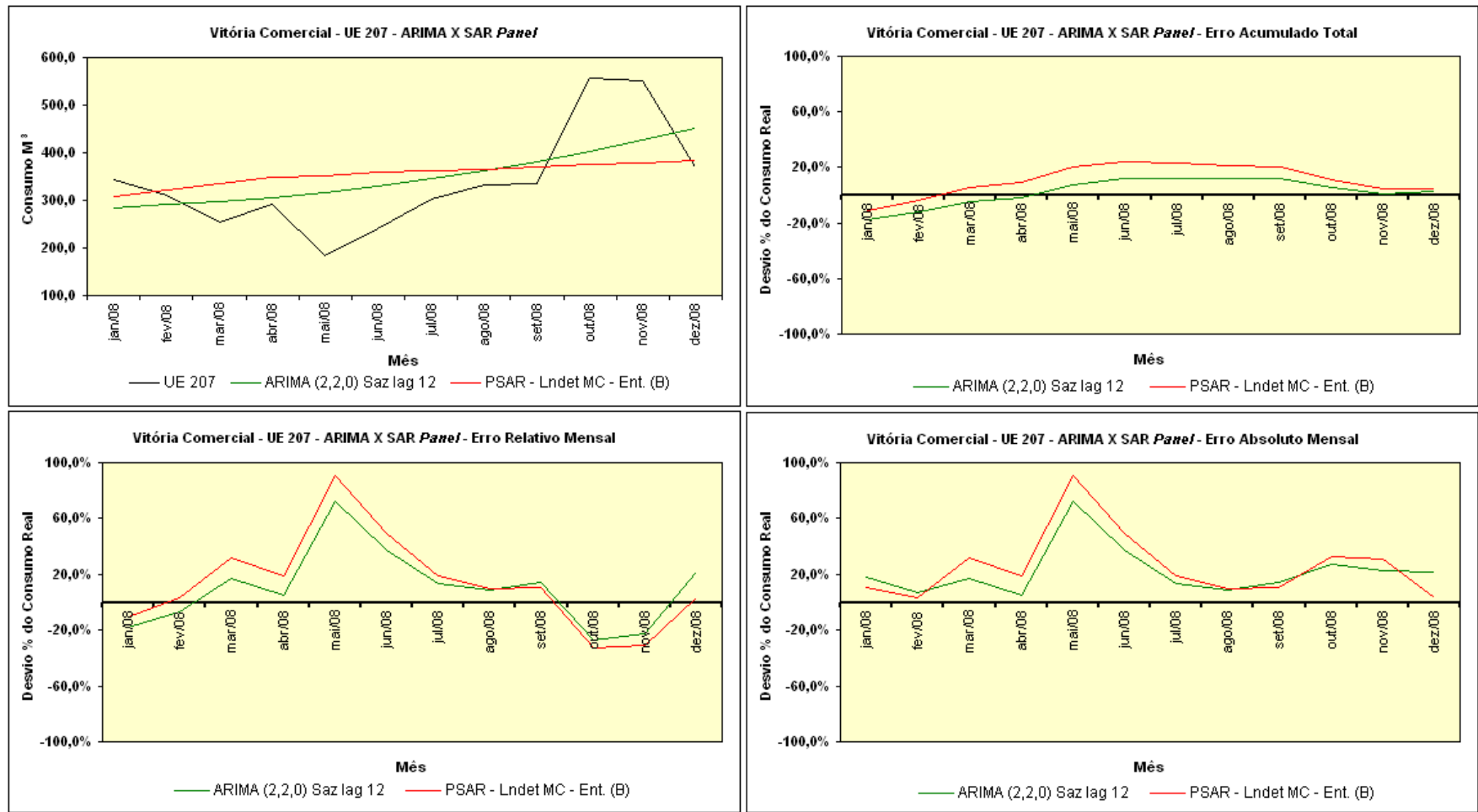
| Modelo                       | Erro Acumulado Total | MAPE   |
|------------------------------|----------------------|--------|
| ARIMA (2,0,0) Saz lag 12     | -20,39%              | 25,18% |
| PSAR - Lndet full - Ent. (C) | 15,07%               | 16,77% |

Figura K.1 – Previsão SAR *Panel* x ARIMA UE 624 – Gráficos Comparativos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



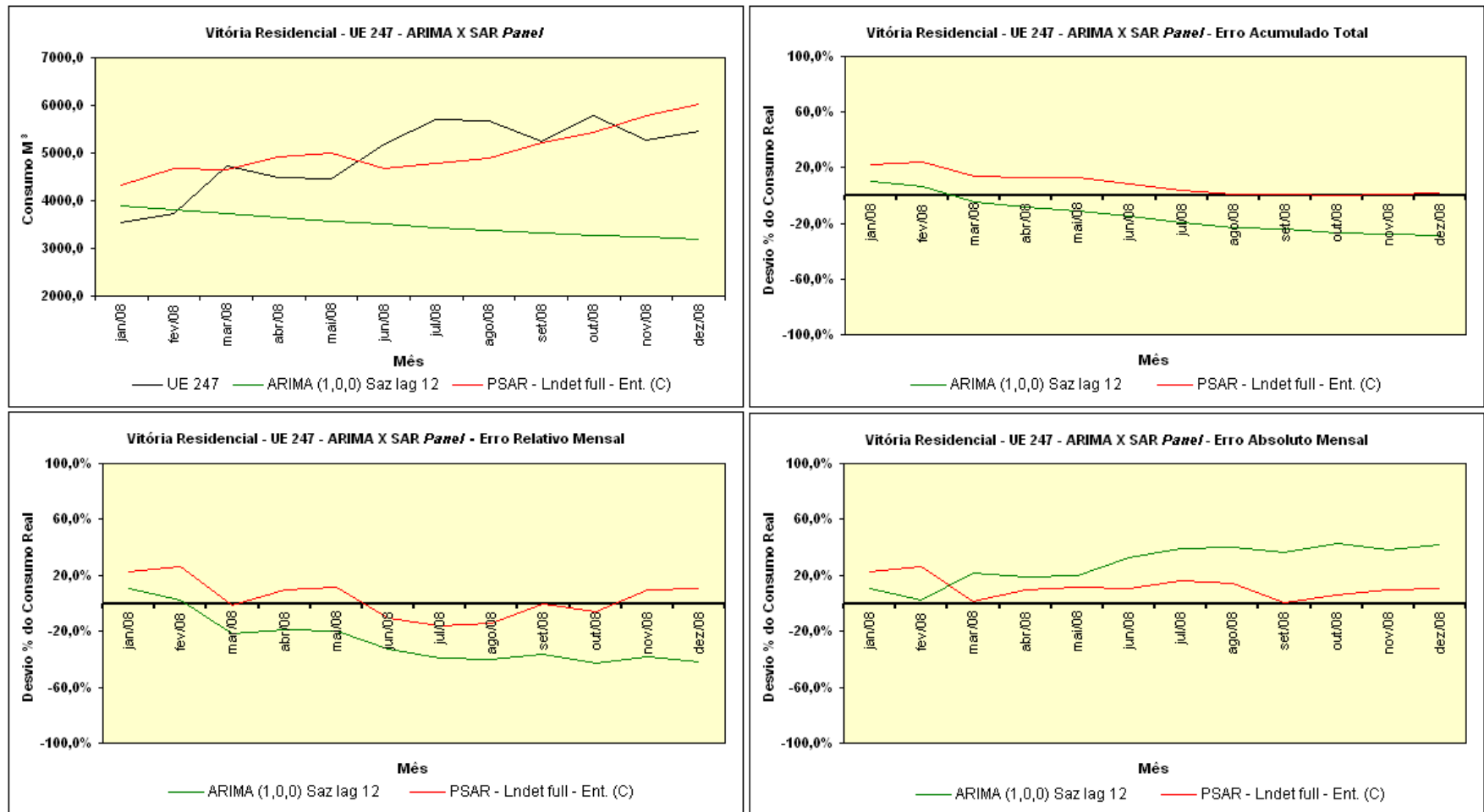
| Modelo                       | Erro Acumulado Total | MAPE   |
|------------------------------|----------------------|--------|
| ARIMA (1,0,0) Saz lag 12     | -19,24%              | 28,50% |
| PSAR - Lndet full - Ent. (C) | 4,19%                | 10,64% |

Figura K.2 – Previsão SAR Panel x ARIMA UE 638 – Gráficos Comparativos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



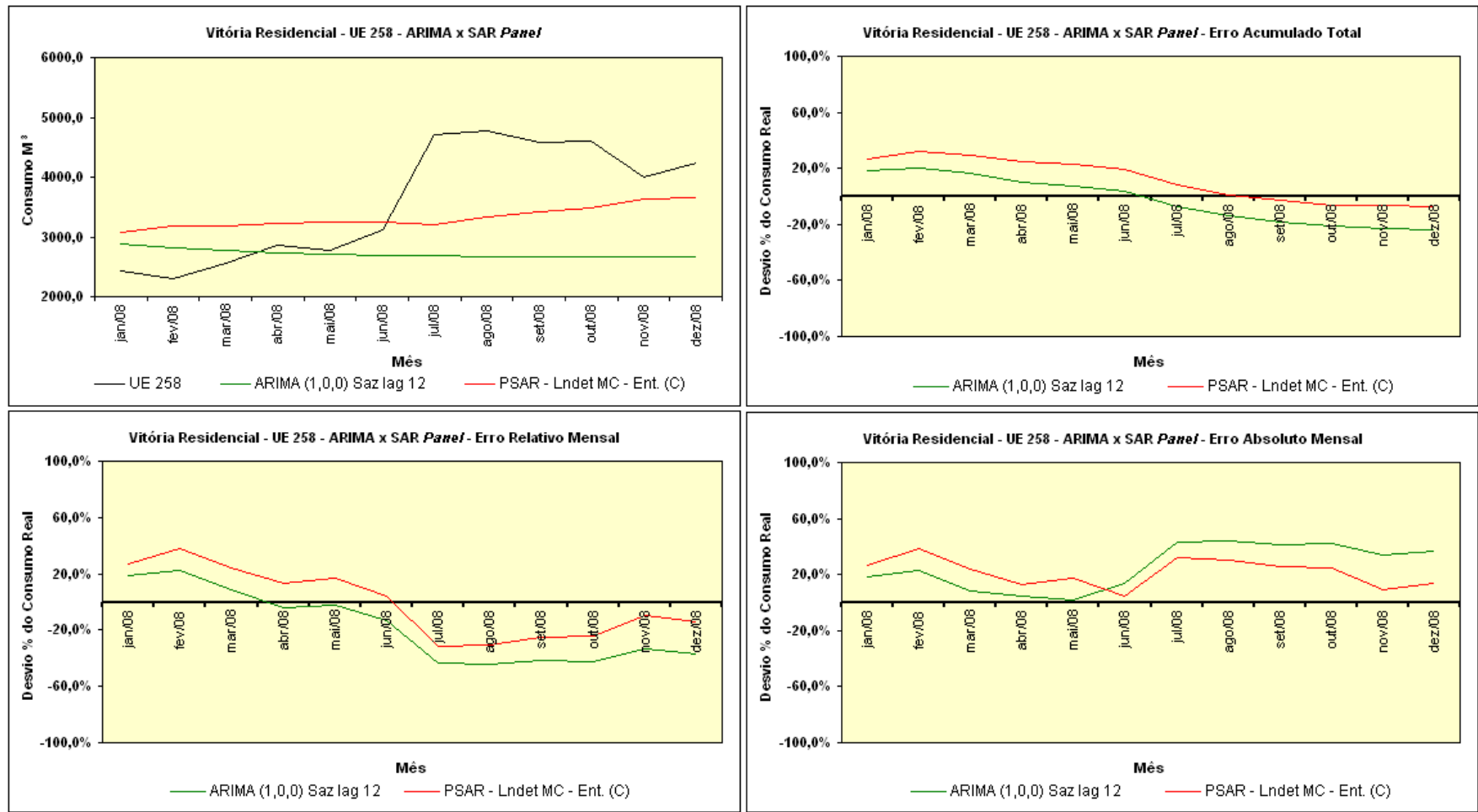
| Modelo                     | Erro Acumulado Total | MAPE   |
|----------------------------|----------------------|--------|
| ARIMA (2,2,0) Saz lag 12   | 2,97%                | 22,02% |
| PSAR - Lndet MC - Ent. (B) | 4,49%                | 25,99% |

Figura K.3 – Previsão SAR Panel x ARIMA UE 207 – Gráficos Comparativos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



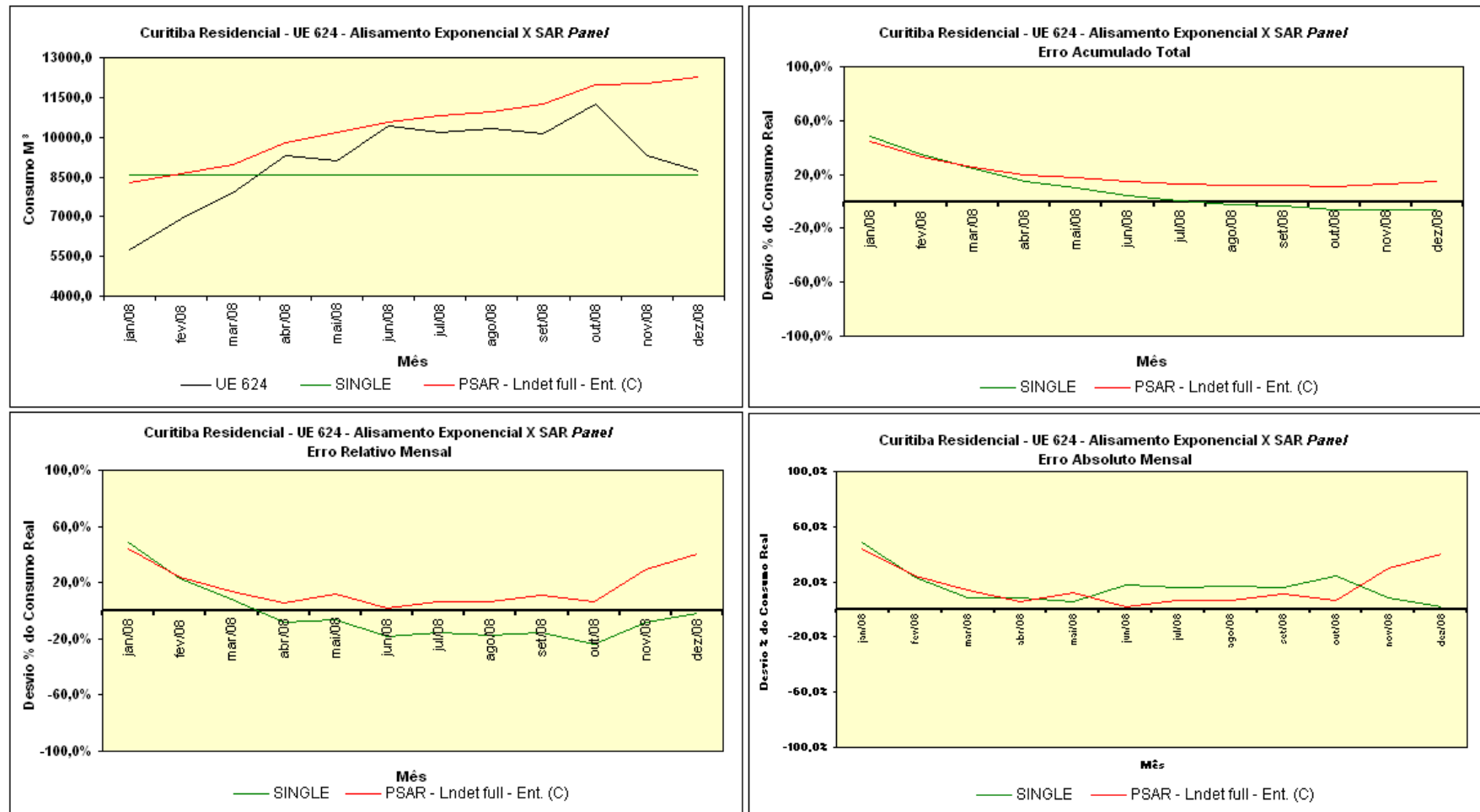
| Modelo                       | Erro Acumulado Total | MAPE   |
|------------------------------|----------------------|--------|
| ARIMA (1,0,0) Saz lag 12     | -29,10%              | 28,73% |
| PSAR - Lndet full - Ent. (C) | 1,89%                | 11,59% |

Figura K.4 – Previsão SAR Panel x ARIMA UE 247 – Gráficos Comparativos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



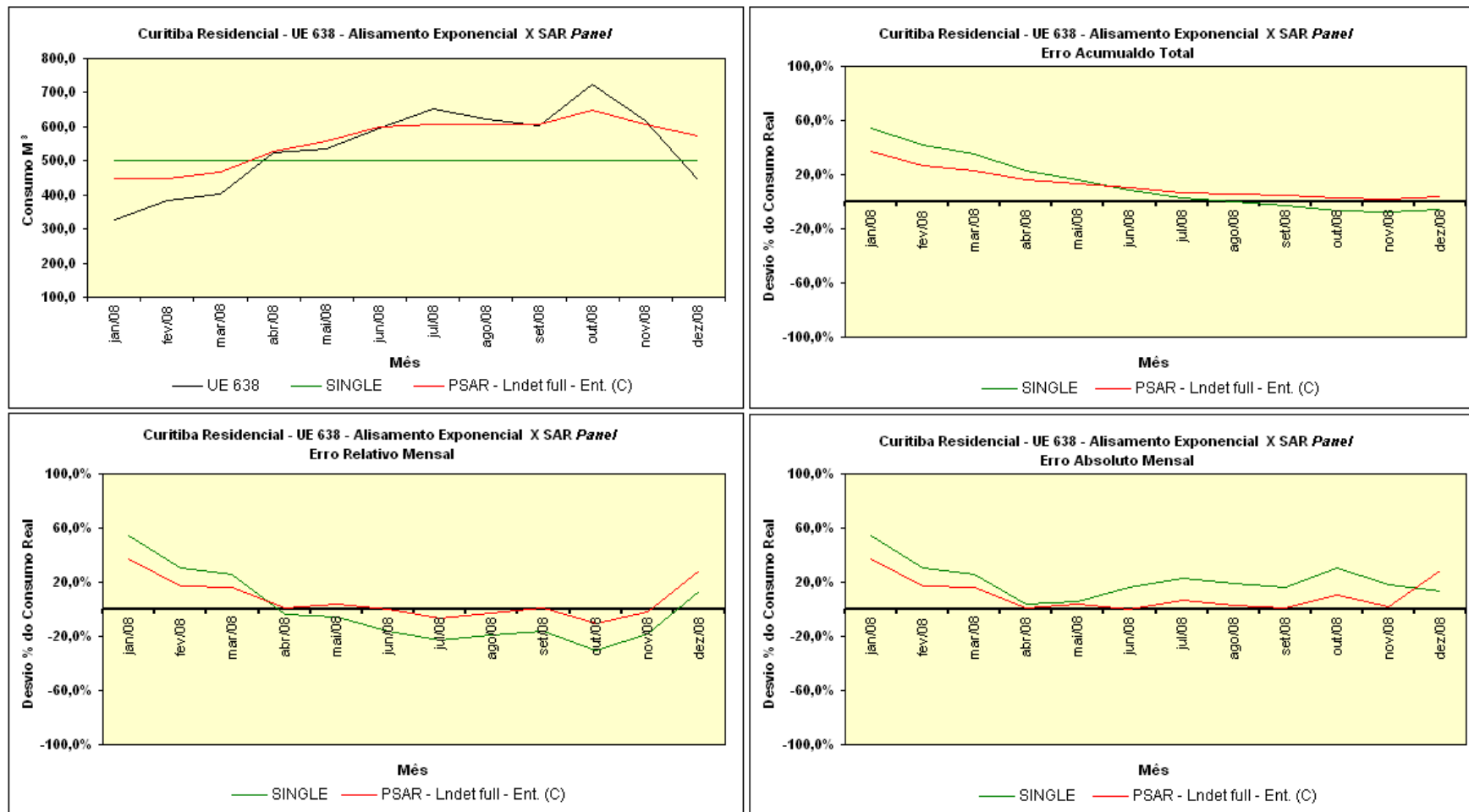
| Modelo                     | Erro Acumulado Total | MAPE   |
|----------------------------|----------------------|--------|
| ARIMA (1,0,0) Saz lag 12   | -24,00%              | 25,90% |
| PSAR - Lndet MC - Ent. (C) | -7,13%               | 21,60% |

Figura K.5 – Previsão SAR Panel x ARIMA UE 258 – Gráficos Comparativos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



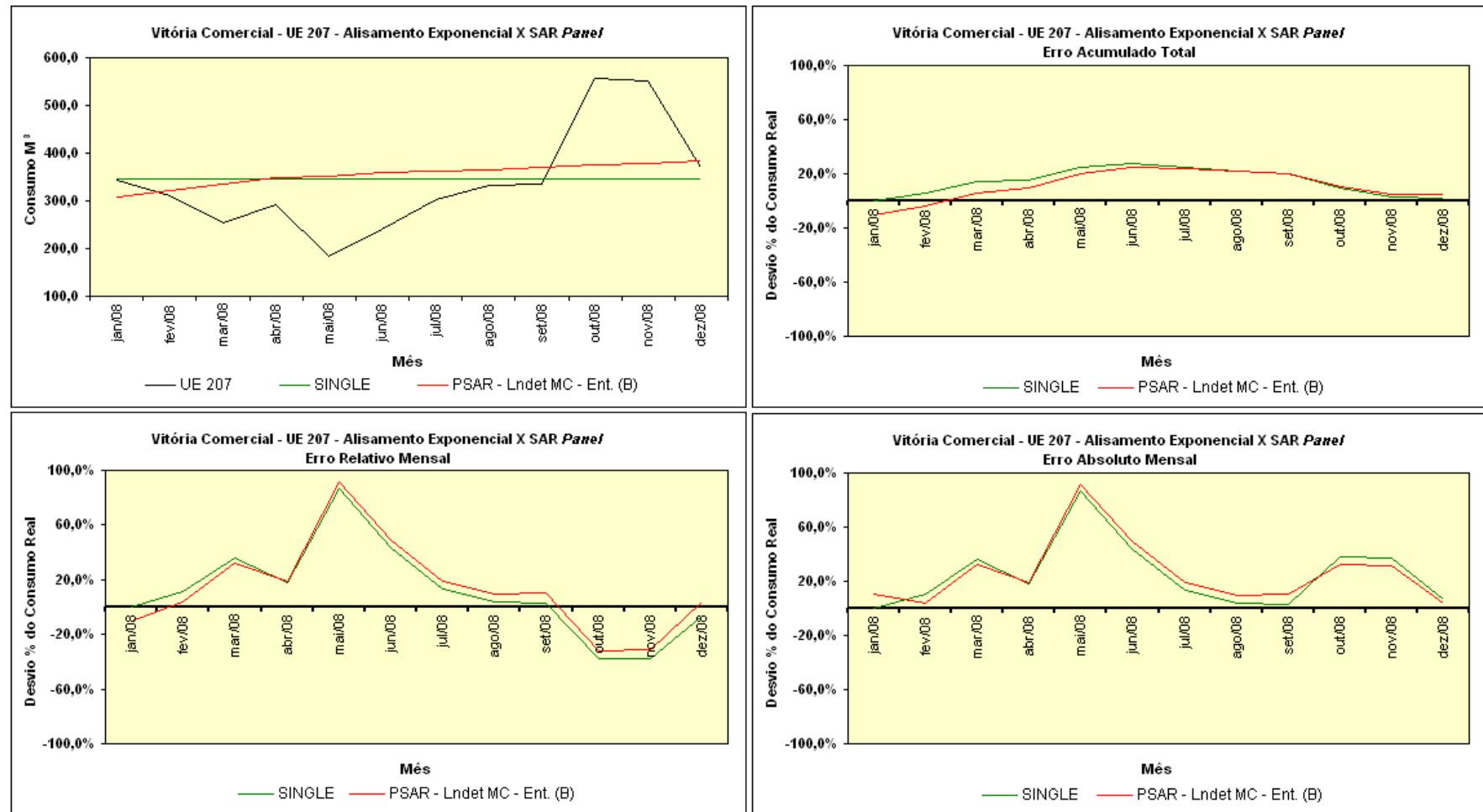
| Modelo                       | Erro Acumulado Total | MAPE   |
|------------------------------|----------------------|--------|
| Alisamento SINGLE            | -6,12%               | 16,27% |
| PSAR - Lndet full - Ent. (C) | 15,07%               | 16,77% |

Figura K.6 – Previsão SAR Panel x Alisamento Exponencial UE 624 – Gráficos Comparativos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



| Modelo                       | Erro Acumulado Total | MAPE   |
|------------------------------|----------------------|--------|
| Alisamento SINGLE            | -6,11%               | 21,41% |
| PSAR - Lndet full - Ent. (C) | 4,19%                | 10,64% |

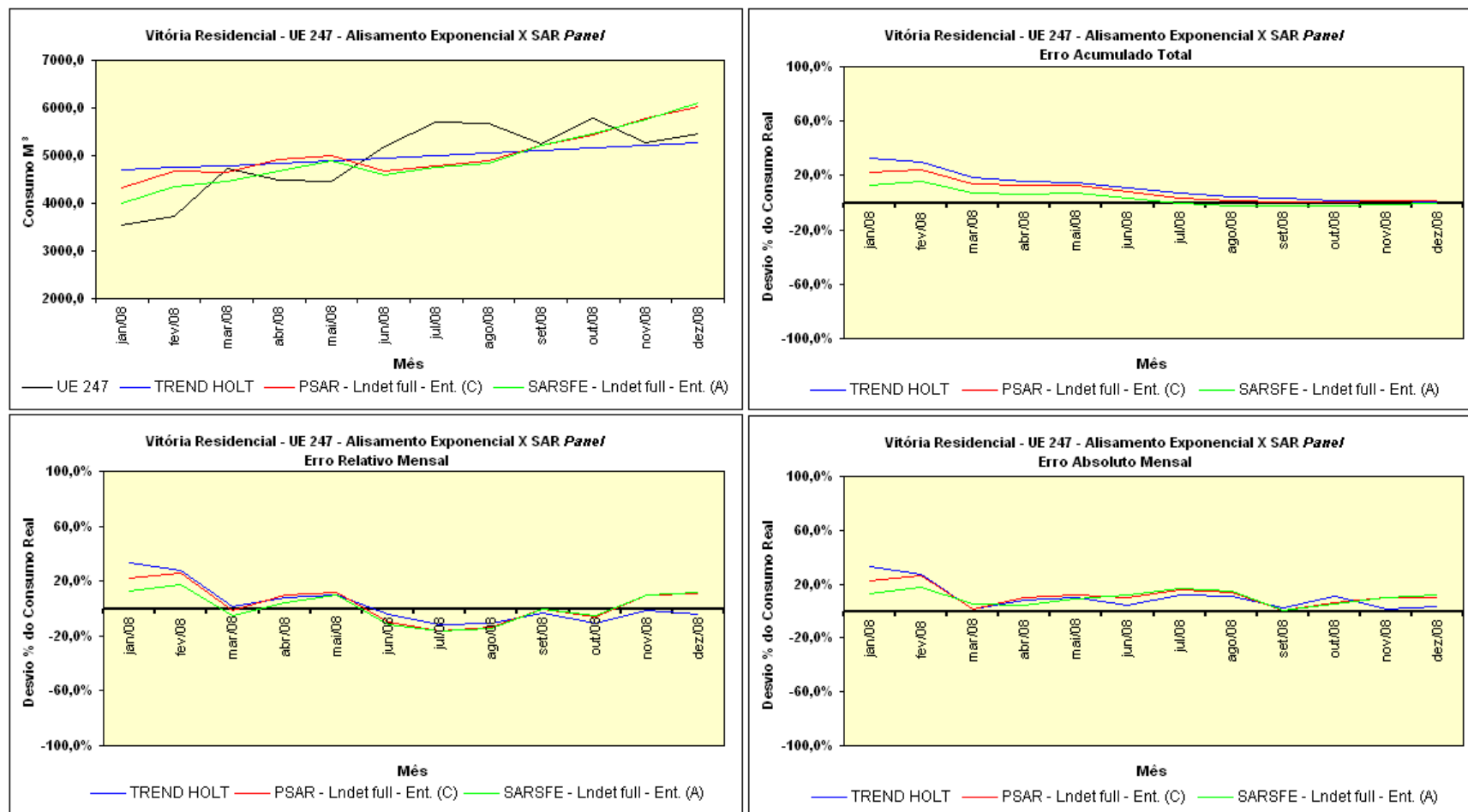
Figura K.7 – Previsão SAR Panel x Alisamento Exponencial UE 638 – Gráficos Comparativos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



| Modelo                     | Erro Acumulado Total | MAPE   |
|----------------------------|----------------------|--------|
| Alisamento SINGLE          | 1,65%                | 24,92% |
| PSAR - Lndet MC - Ent. (B) | 4,49%                | 25,99% |

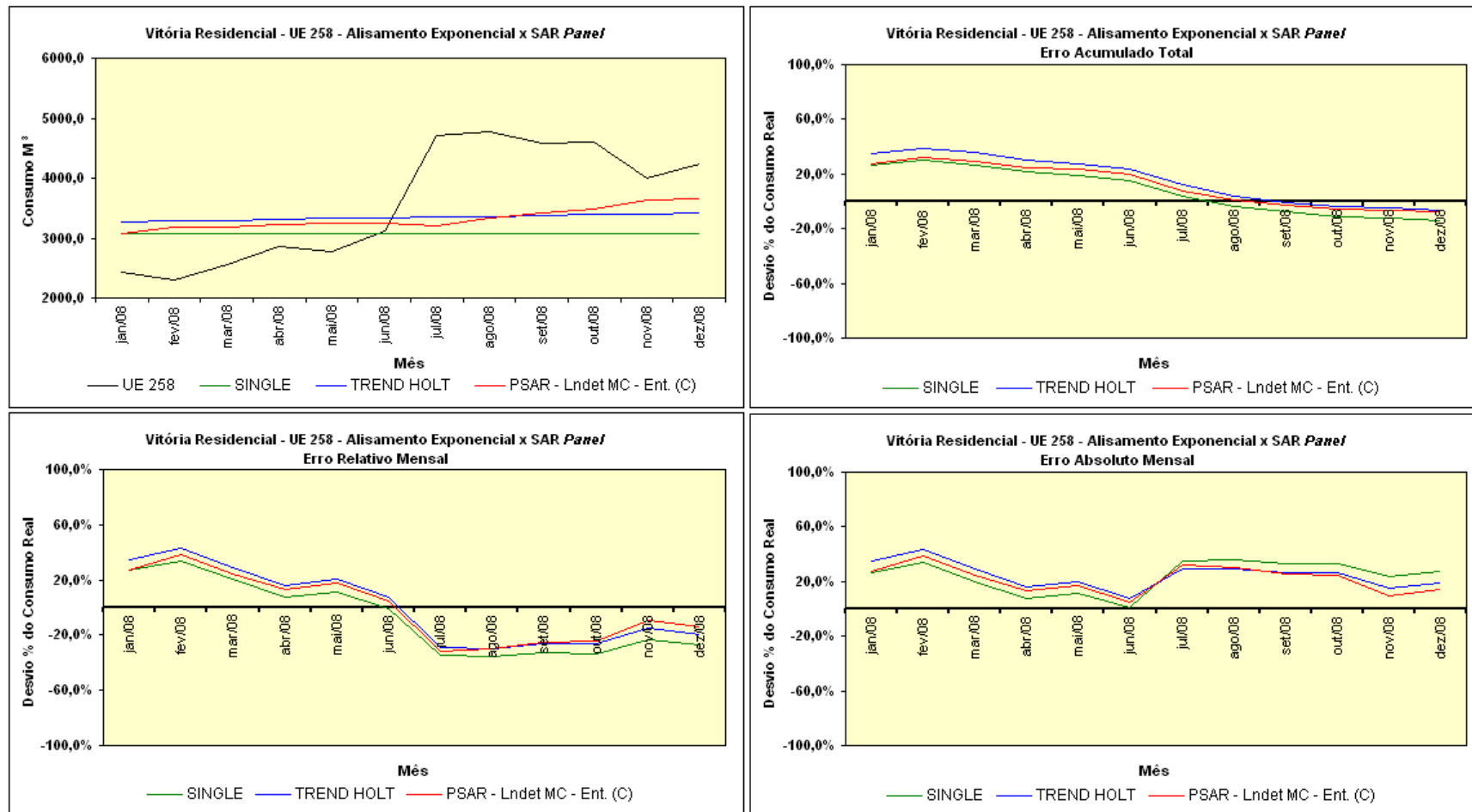
Figura K.8 – Previsão SAR Panel x Alisamento Exponencial UE 207 – Gráficos Comparativos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).





| Modelo                         | Erro Acumulado Total | MAPE   |
|--------------------------------|----------------------|--------|
| Alisamento TREND HOLT          | 0,80%                | 10,55% |
| PSAR - Lndet full - Ent. (C)   | 1,89%                | 11,59% |
| SARSFE - Lndet full - Ent. (A) | -0,28%               | 10,08% |

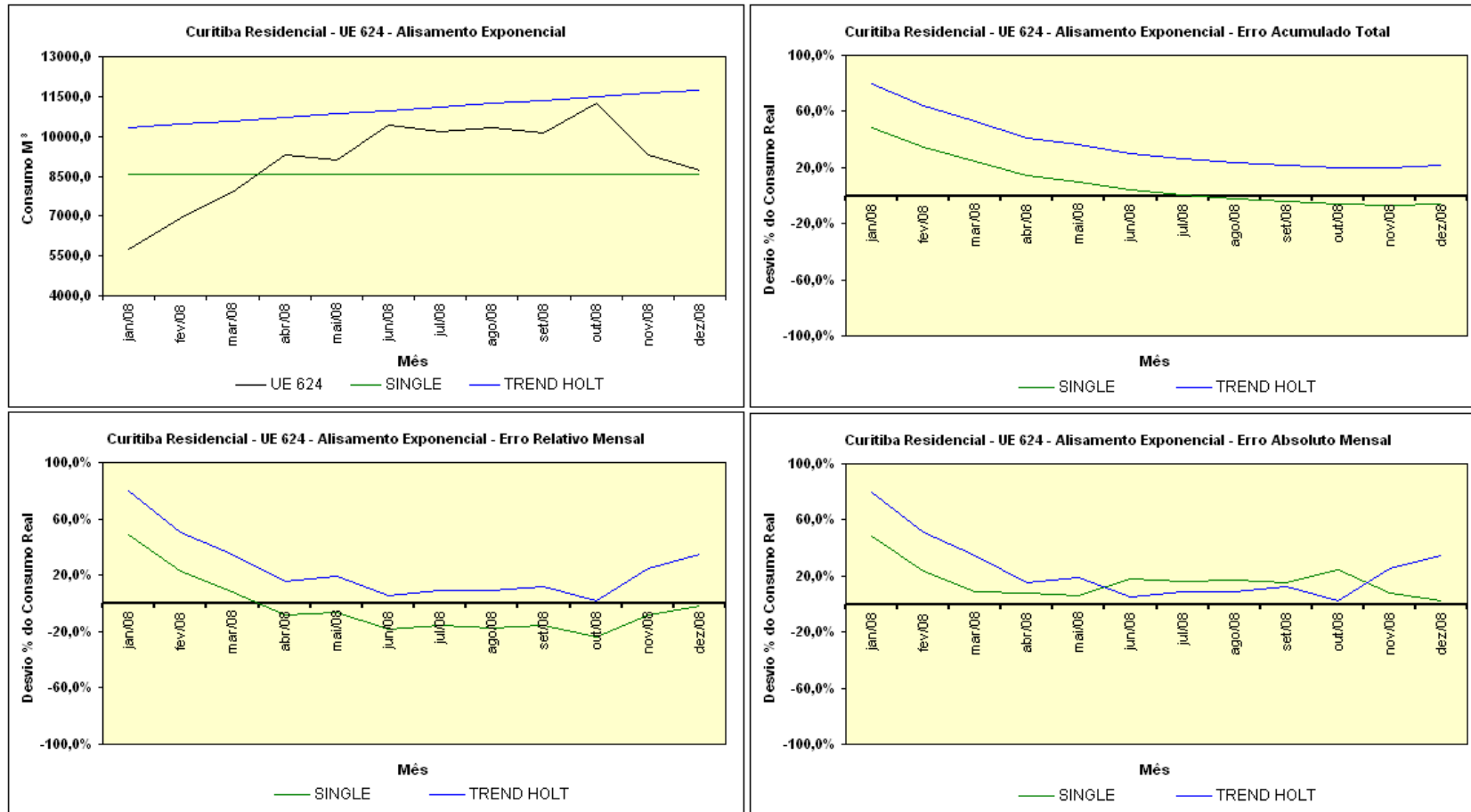
Figura K.9 – Previsão SAR Panel x Alisamento Exponencial UE 247 – Gráficos Comparativos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



| Modelo                     | Erro Acumulado Total | MAPE   |
|----------------------------|----------------------|--------|
| Alisamento SINGLE          | -14,00%              | 23,95% |
| Alisamento TREND HOLT      | -6,48%               | 24,65% |
| PSAR - Lndet MC - Ent. (C) | -7,13%               | 21,60% |

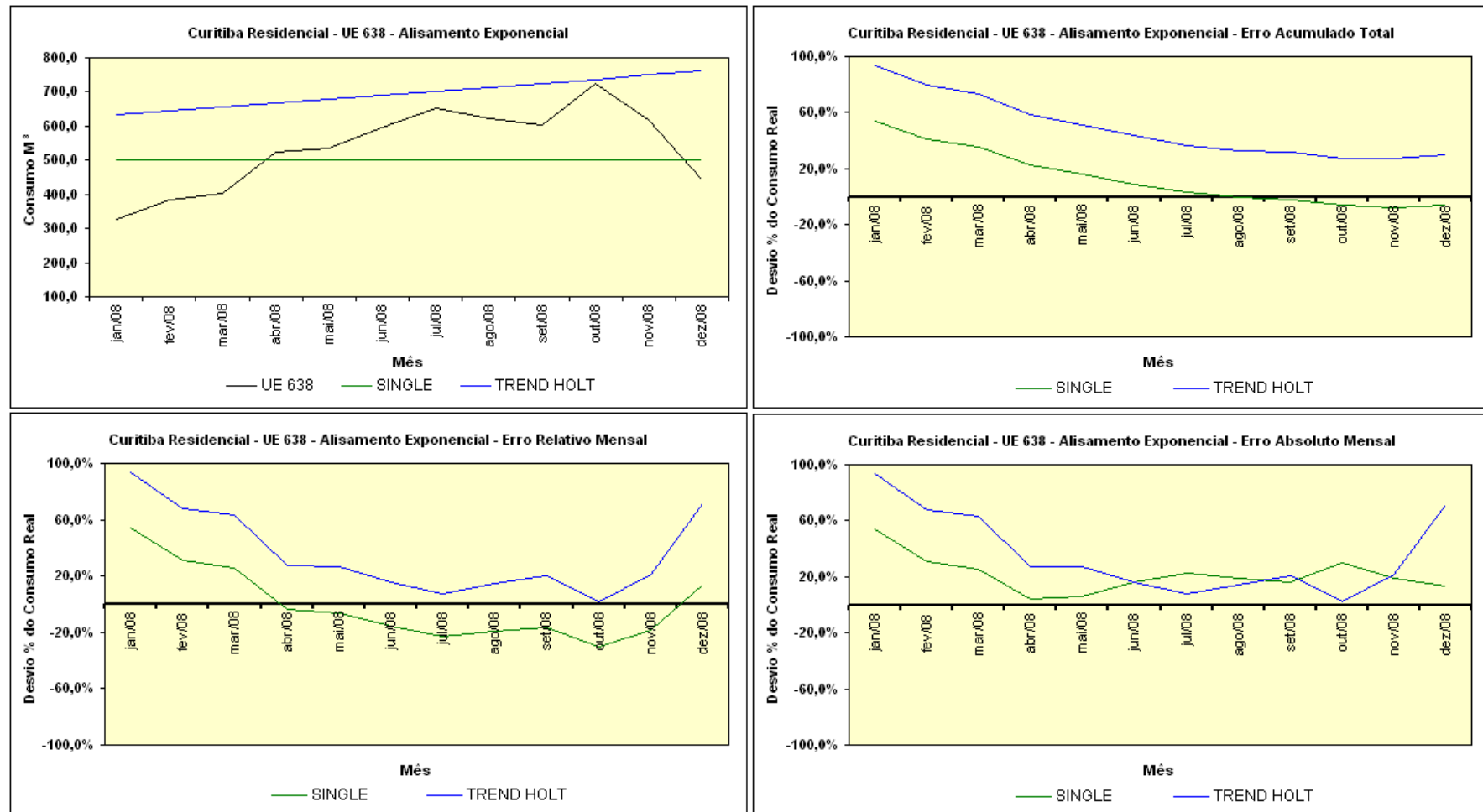
Figura K.10 – Previsão SAR Panel x Alisamento Exponencial UE 258 – Gráficos Comparativos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).

## APÊNDICE (K) - PREVISÃO ALISAMENTO EXPONENCIAL POR UE



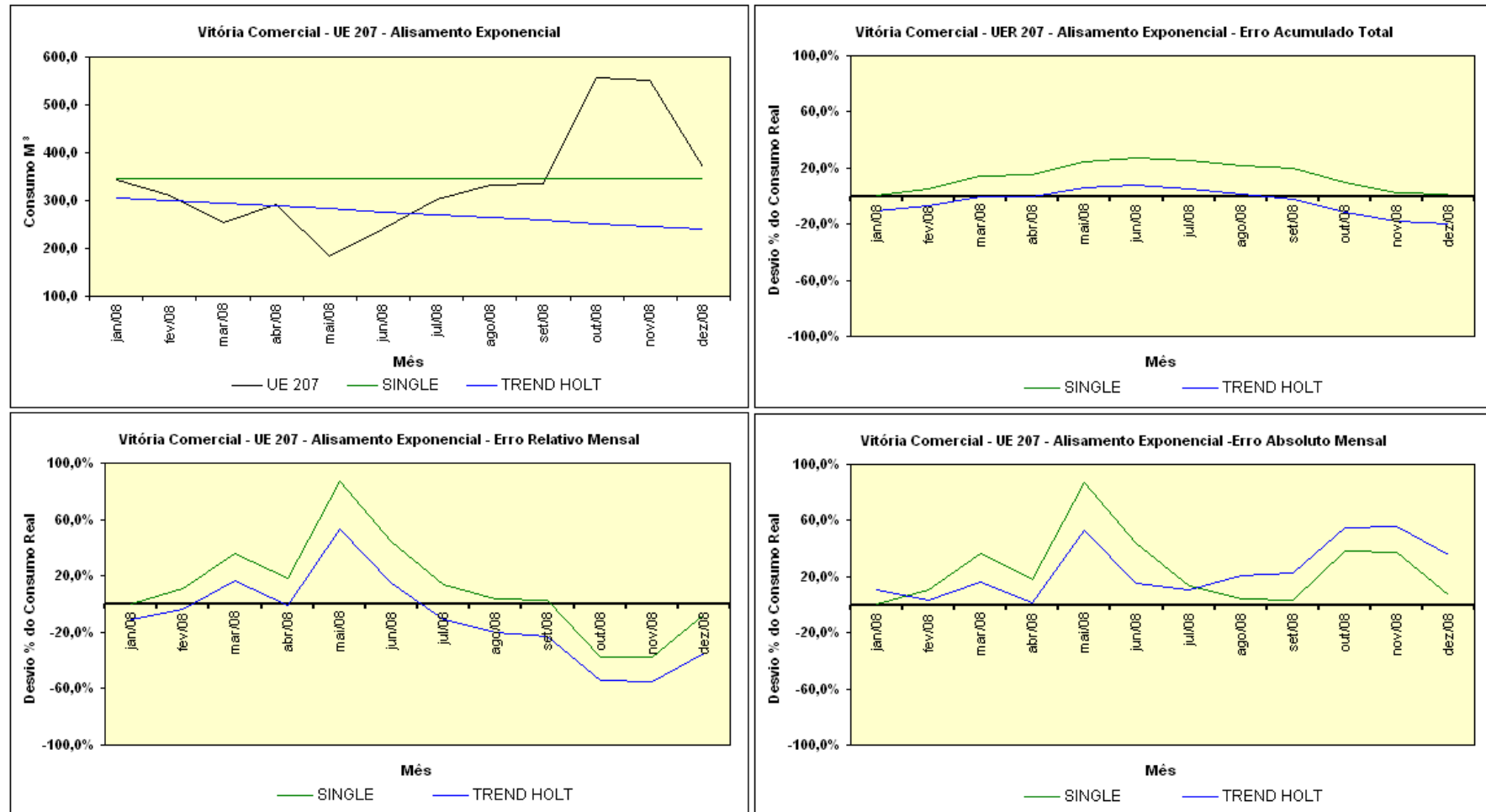
| Alisamento | Erro Médio | Erro Médio Absoluto | Soma dos Quadrados | Média Quadrada | Erro Médio Percentual | S0     | T0     | Alfa   | Gama   | Erro Acumulado Total | MAPE  |
|------------|------------|---------------------|--------------------|----------------|-----------------------|--------|--------|--------|--------|----------------------|-------|
| SINGLE     | 0,1534     | 0,5604              | 24,4696            | 0,4706         | 64,6358               | 0,0286 |        | 0,1000 | 0,1000 | -6,12%               | 16,3% |
| TREND HOLT | 0,0320     | 0,4725              | 18,5062            | 0,3559         | 30,3888               | 0,6380 | 0,0241 | 0,1000 | 0,1000 | 21,27%               | 24,8% |

Figura L.1 – Previsão Alisamento Exponencial UE 624 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



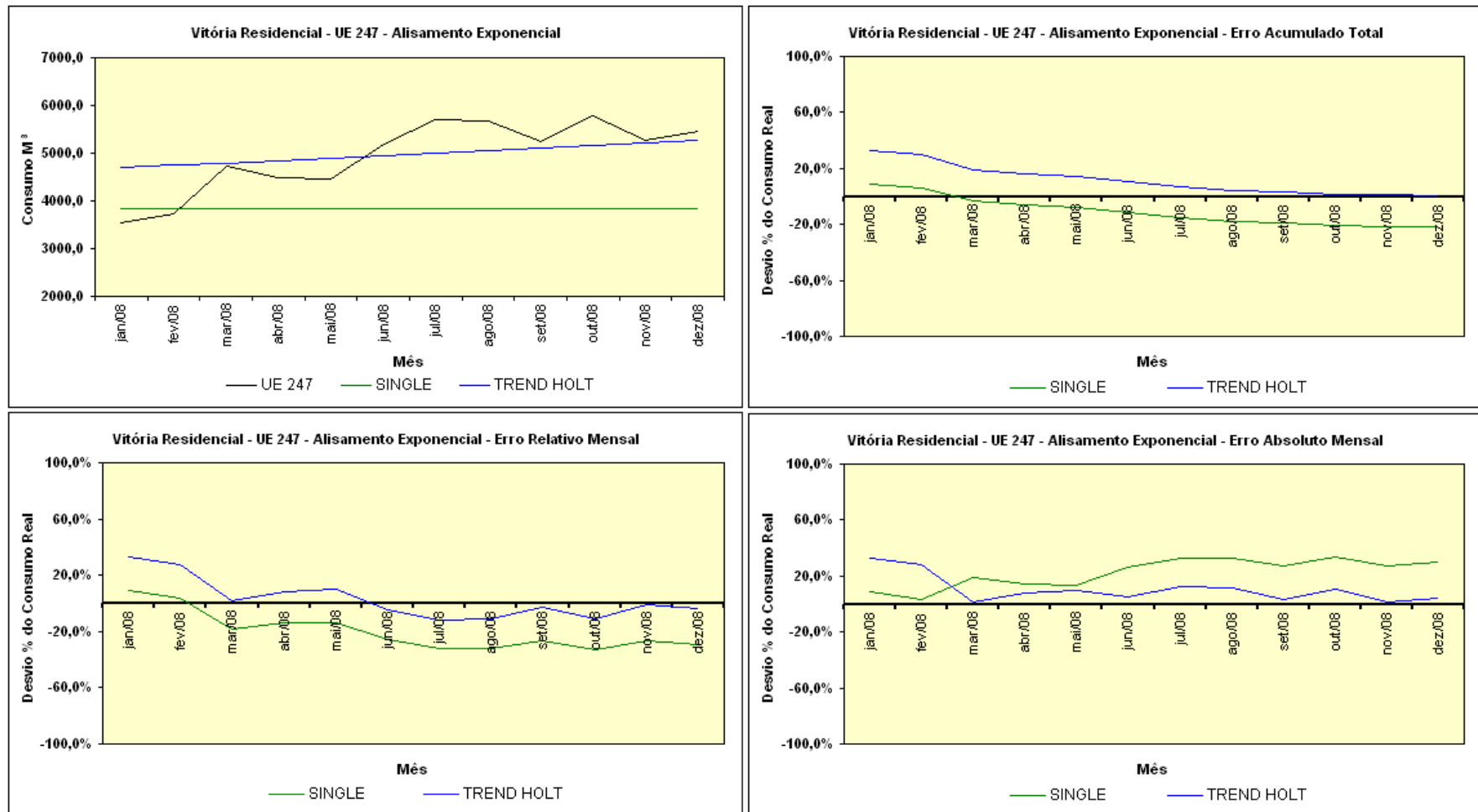
| Alisamento | Erro Médio | Erro Médio Absoluto | Soma dos Quadrados | Média Quadrada | Erro Médio Percentual | S0      | T0     | Alfa   | Gama   | Erro Acumulado Total | MAPE  |
|------------|------------|---------------------|--------------------|----------------|-----------------------|---------|--------|--------|--------|----------------------|-------|
| SINGLE     | 0,1663     | 0,5675              | 25,1022            | 0,4827         | 16,7206               | 0,0210  |        | 0,1000 | 0,1000 | -6,11%               | 21,4% |
| TREND HOLT | 0,0139     | 0,4025              | 14,6136            | 0,2810         | -51,5992              | -1,1300 | 0,0391 | 0,1000 | 0,1000 | 29,99%               | 36,0% |

Figura L.2 – Previsão Alisamento Exponencial UE 638 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



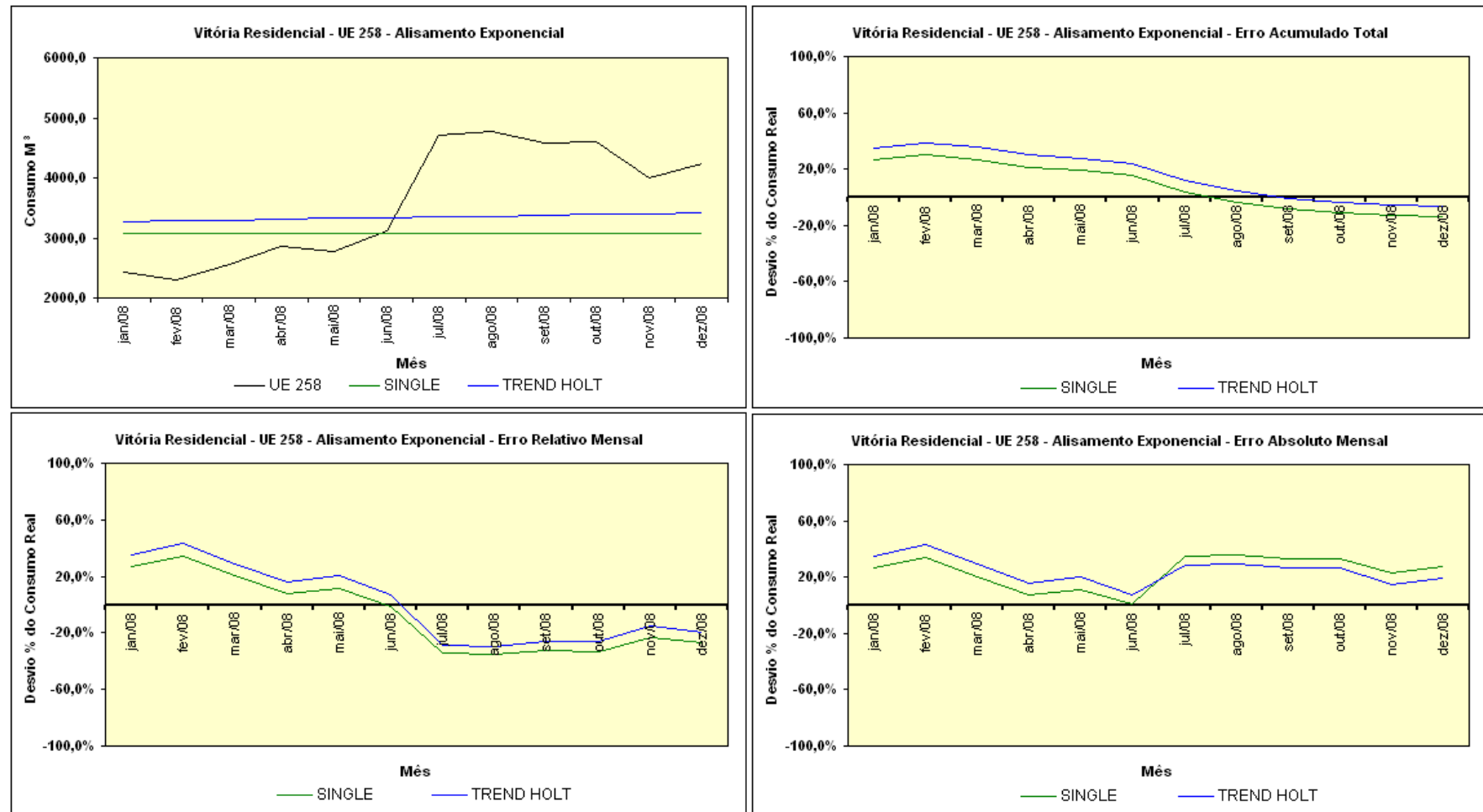
| Alisamento | Erro Médio | Erro Médio Absoluto | Soma dos Quadrados | Média Quadrada | Erro Médio Percentual | S0     | T0      | Alfa   | Gama   | Erro Acumulado Total | MAPE  |
|------------|------------|---------------------|--------------------|----------------|-----------------------|--------|---------|--------|--------|----------------------|-------|
| SINGLE     | -0,0845    | 0,3827              | 10,3707            | 0,2161         | 192,3540              | 0,3950 |         | 0,1000 | 0,1000 | 1,65%                | 24,9% |
| TREND HOLT | -0,0302    | 0,3554              | 9,6650             | 0,2014         | 105,4792              | 0,8066 | -0,0260 | 0,1000 | 0,1000 | -19,47%              | 24,9% |

Figura L.3 – Previsão Alisamento Exponencial UE 207 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



| Alisamento | Erro Médio | Erro Médio Absoluto | Soma dos Quadrados | Média Quadrada | Erro Médio Percentual | S0     | T0     | Alfa   | Gama   | Erro Acumulado Total | MAPE  |
|------------|------------|---------------------|--------------------|----------------|-----------------------|--------|--------|--------|--------|----------------------|-------|
| SINGLE     | 0,1287     | 0,4854              | 15,4095            | 0,3424         | 284,0727              | 0,0544 |        | 0,1000 | 0,1000 | -22,04%              | 22,2% |
| TREND HOLT | -0,0468    | 0,2859              | 6,3778             | 0,1417         | 334,9475              | 1,4400 | 0,0483 | 0,1000 | 0,1000 | 0,80%                | 10,6% |

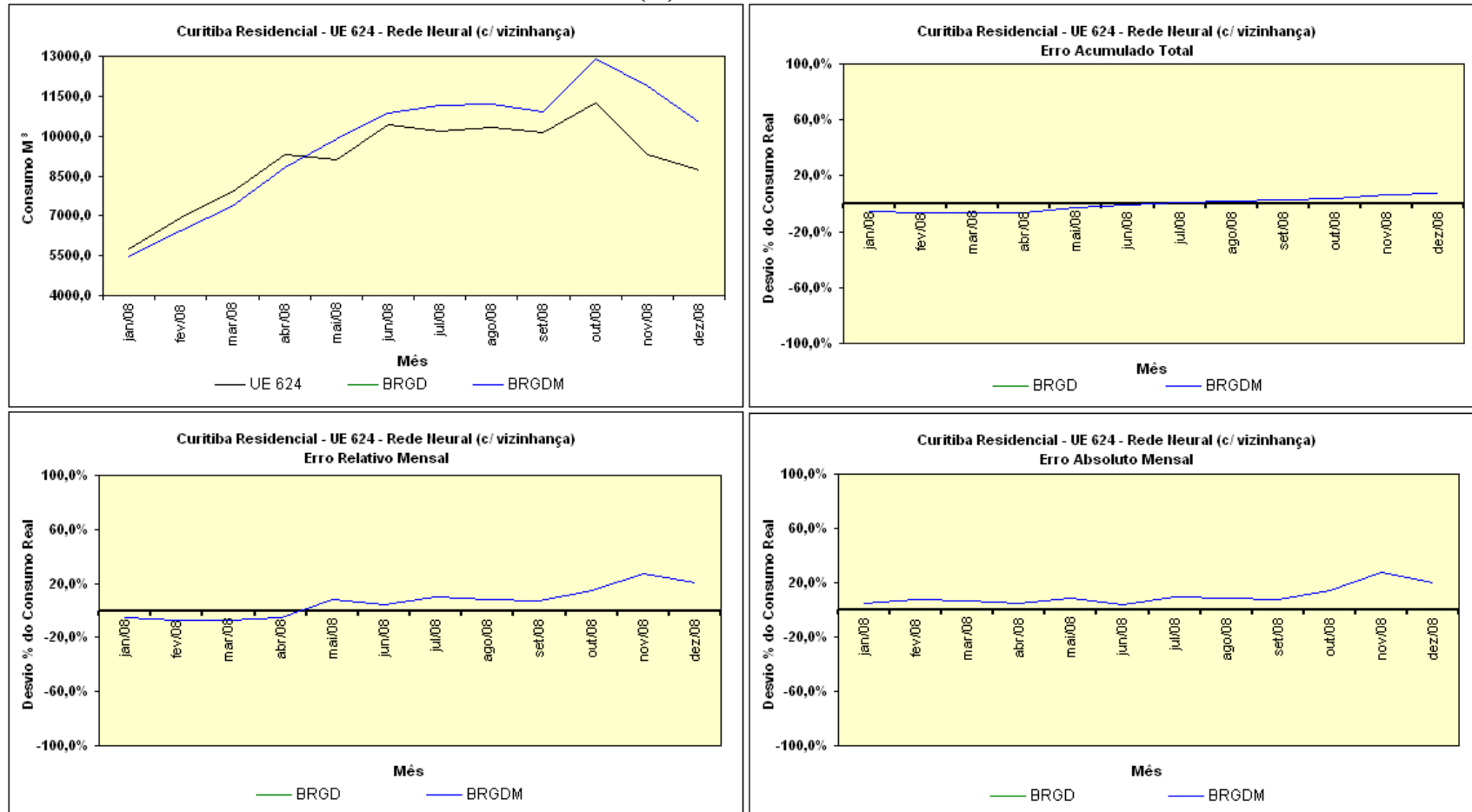
Figura L.4 – Previsão Alisamento Exponencial UE 247 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



| Alisamento | Erro Médio | Erro Médio Absoluto | Soma dos Quadrados | Média Quadrada | Erro Médio Percentual | S0     | T0     | Alfa   | Gama   | Erro Acumulado Total | MAPE  |
|------------|------------|---------------------|--------------------|----------------|-----------------------|--------|--------|--------|--------|----------------------|-------|
| SINGLE     | 0,0563     | 0,3608              | 9,0351             | 0,2008         | 94,6640               | 0,2129 |        | 0,1000 | 0,1000 | -14,00%              | 23,9% |
| TREND HOLT | -0,0269    | 0,3170              | 8,3854             | 0,1863         | 6,1944                | 0,6990 | 0,0234 | 0,1000 | 0,1000 | -6,48%               | 24,6% |

Figura L.5 – Previsão Alisamento Exponencial UE 258 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).

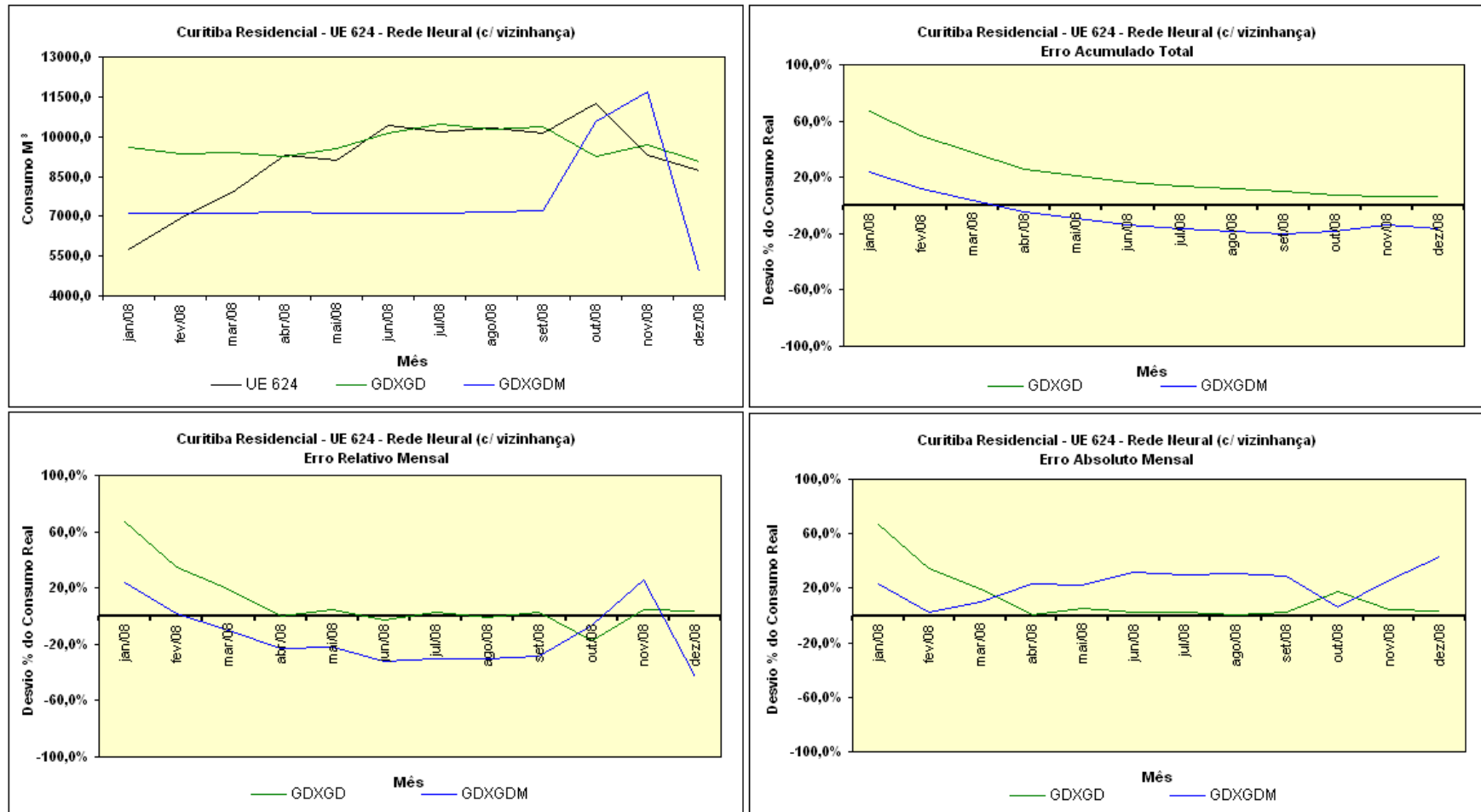
## APÊNDICE (L) - PREVISÃO RN POR UE



| Redes Neurais (c/ vizinhança) | Erro Acumulado Total | MAPE  |
|-------------------------------|----------------------|-------|
| BRGD                          | 7,30%                | 10,5% |
| BRGDM                         | 7,30%                | 10,5% |

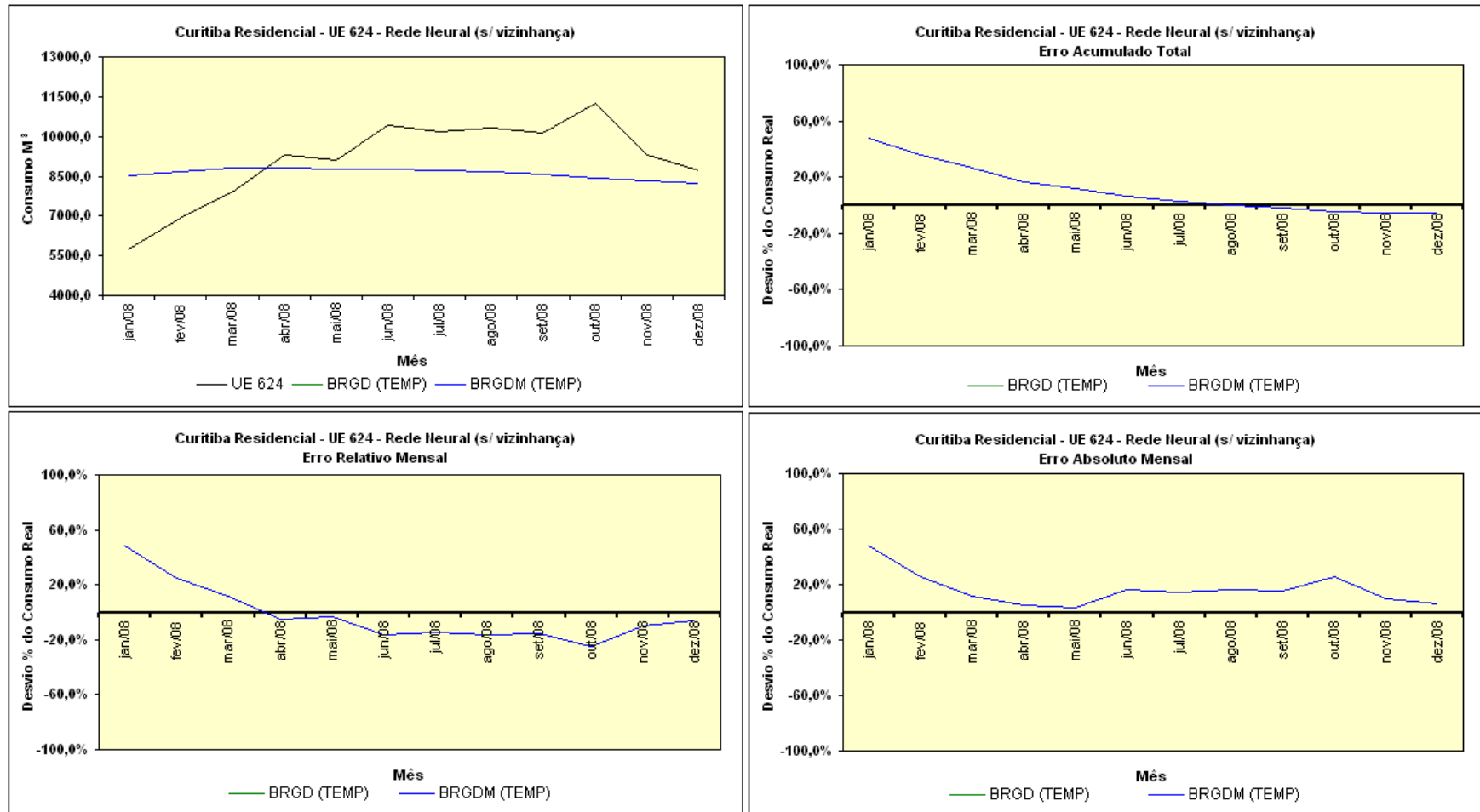
Figura M.1 – Previsão RN Método TRAIN BR c/ dados da Vizinhança UE 624 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).





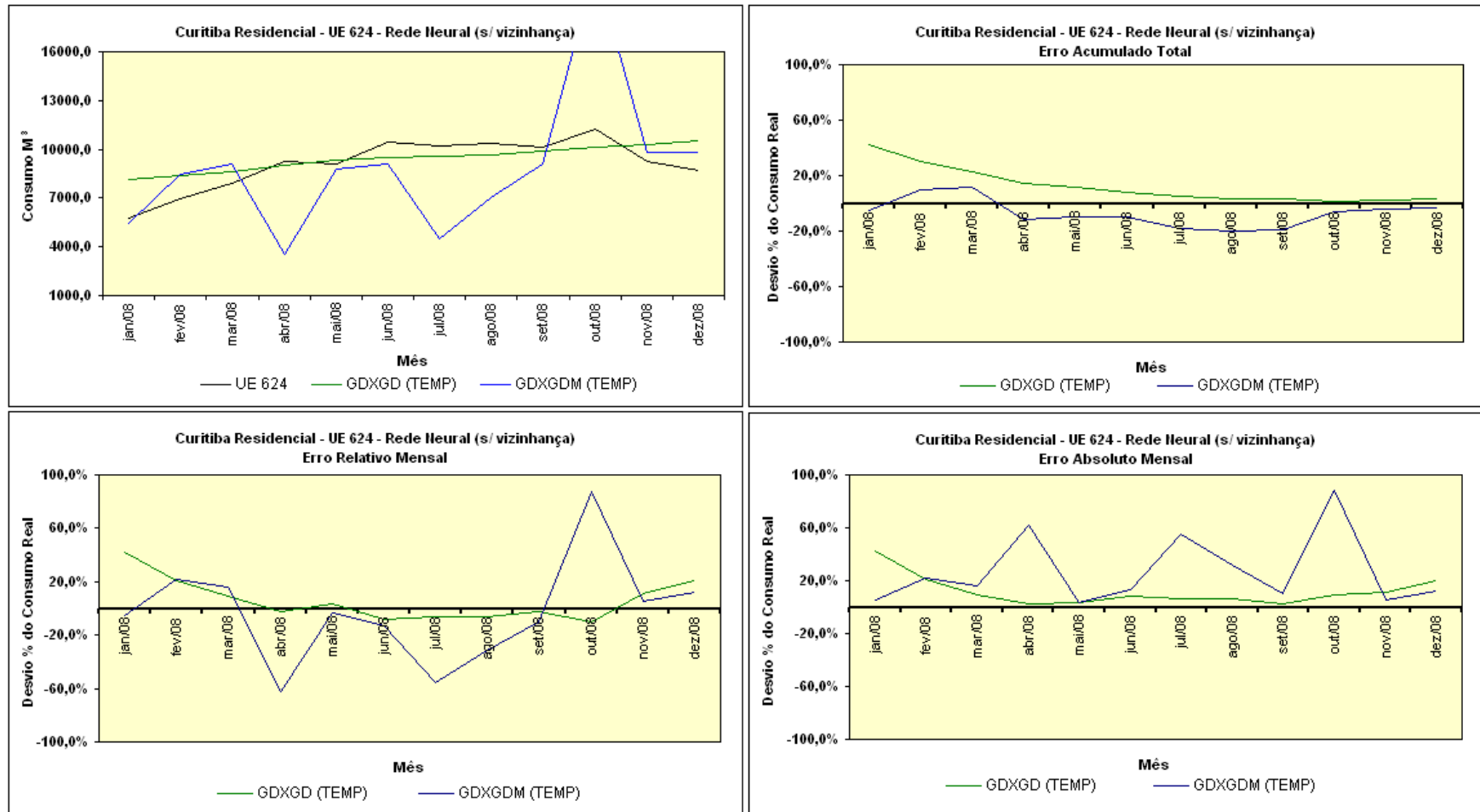
| Redes Neurais (c/ vizinhança) | Erro Acumulado Total | MAPE  |
|-------------------------------|----------------------|-------|
| GDXGD                         | 6,56%                | 13,4% |
| GDXGDM                        | -16,42%              | 23,1% |

Figura M.2 – Previsão RN Método TRAIN GDX c/ dados da Vizinhança UE 624 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



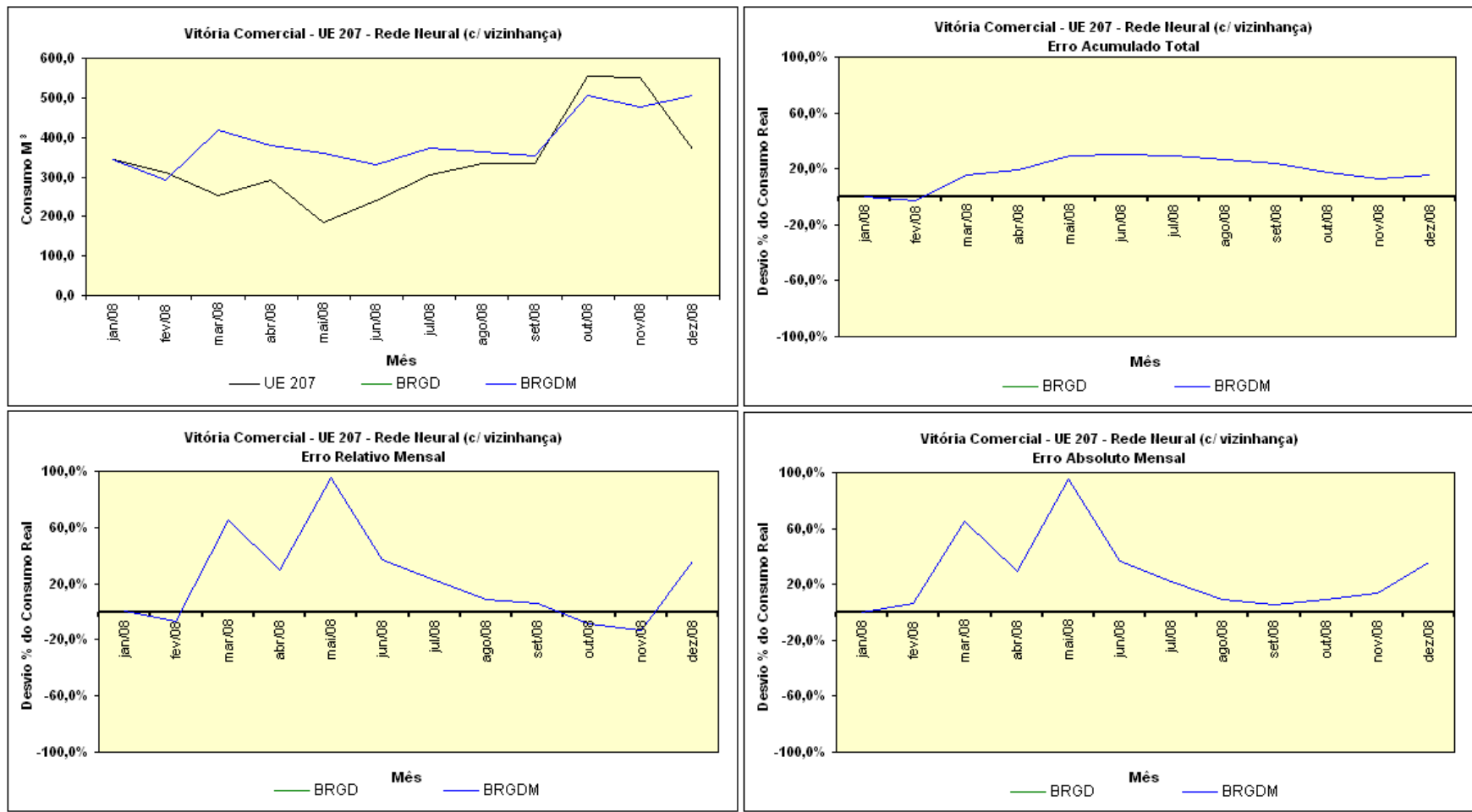
| Redes Neurais (s/ vizinhança) | Erro Acumulado Total | MAPE  |
|-------------------------------|----------------------|-------|
| BRGD (TEMP)                   | -5,46%               | 16,4% |
| BRGDM (TEMP)                  | -5,46%               | 16,4% |

Figura M.3 – Previsão RN Método TRAIN BR s/ dados da Vizinhança UE 624 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



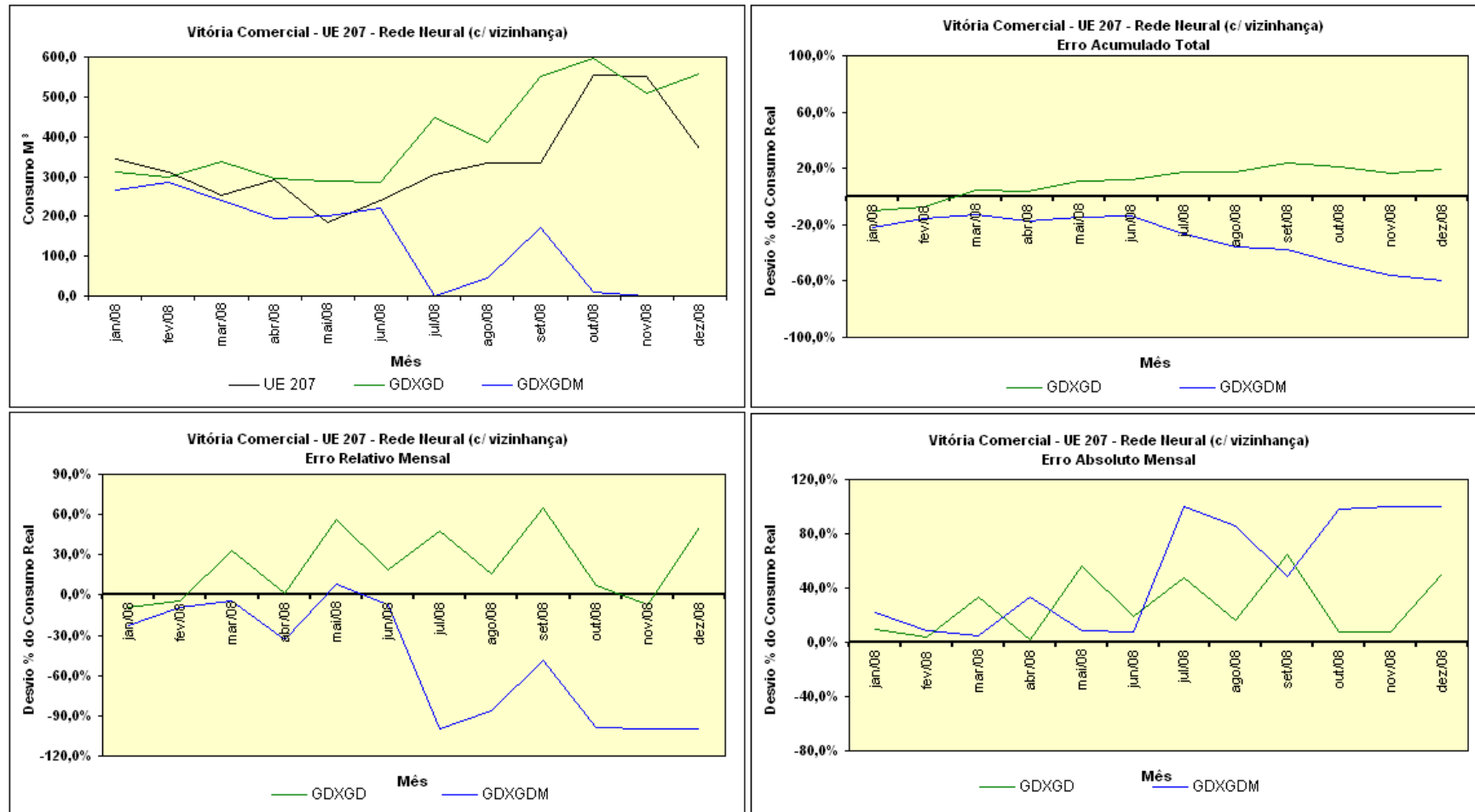
| Redes Neurais (s/ vizinhança) | Erro Acumulado Total | MAPE  |
|-------------------------------|----------------------|-------|
| GDXGD (TEMP)                  | 3,60%                | 11,9% |
| GDXGDM (TEMP)                 | -3,20%               | 27,1% |

Figura M.4 – Previsão RN Método TRAIN GDX s/ dados da Vizinhança UE 624 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



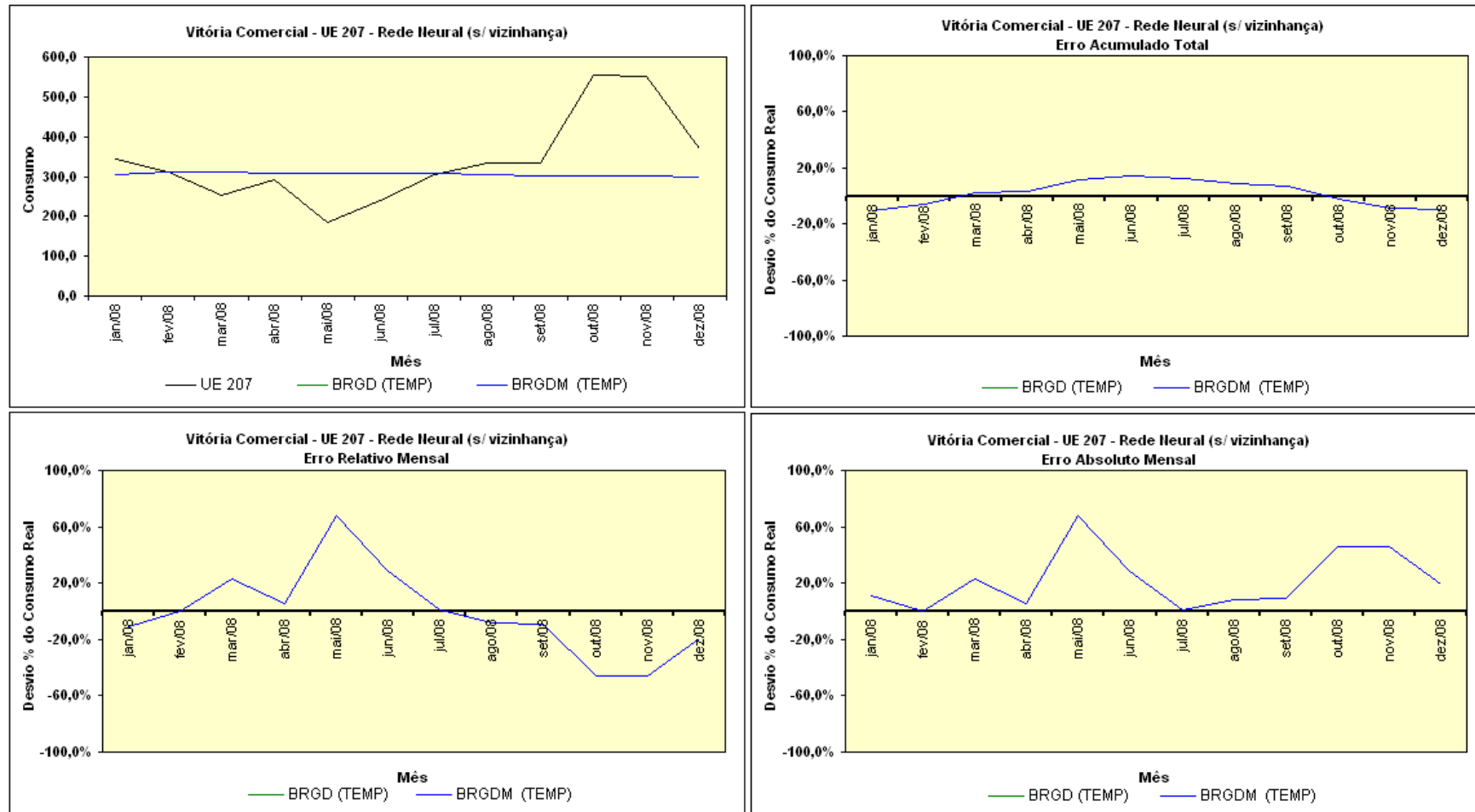
| Redes Neurais (c/ vizinhança) | Erro Acumulado Total | MAPE  |
|-------------------------------|----------------------|-------|
| BRGD                          | 15,30%               | 27,5% |
| BRGDM                         | 15,30%               | 27,5% |

Figura M.5 – Previsão RN Método TRAIN BR c/ dados da Vizinhança UE 207 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



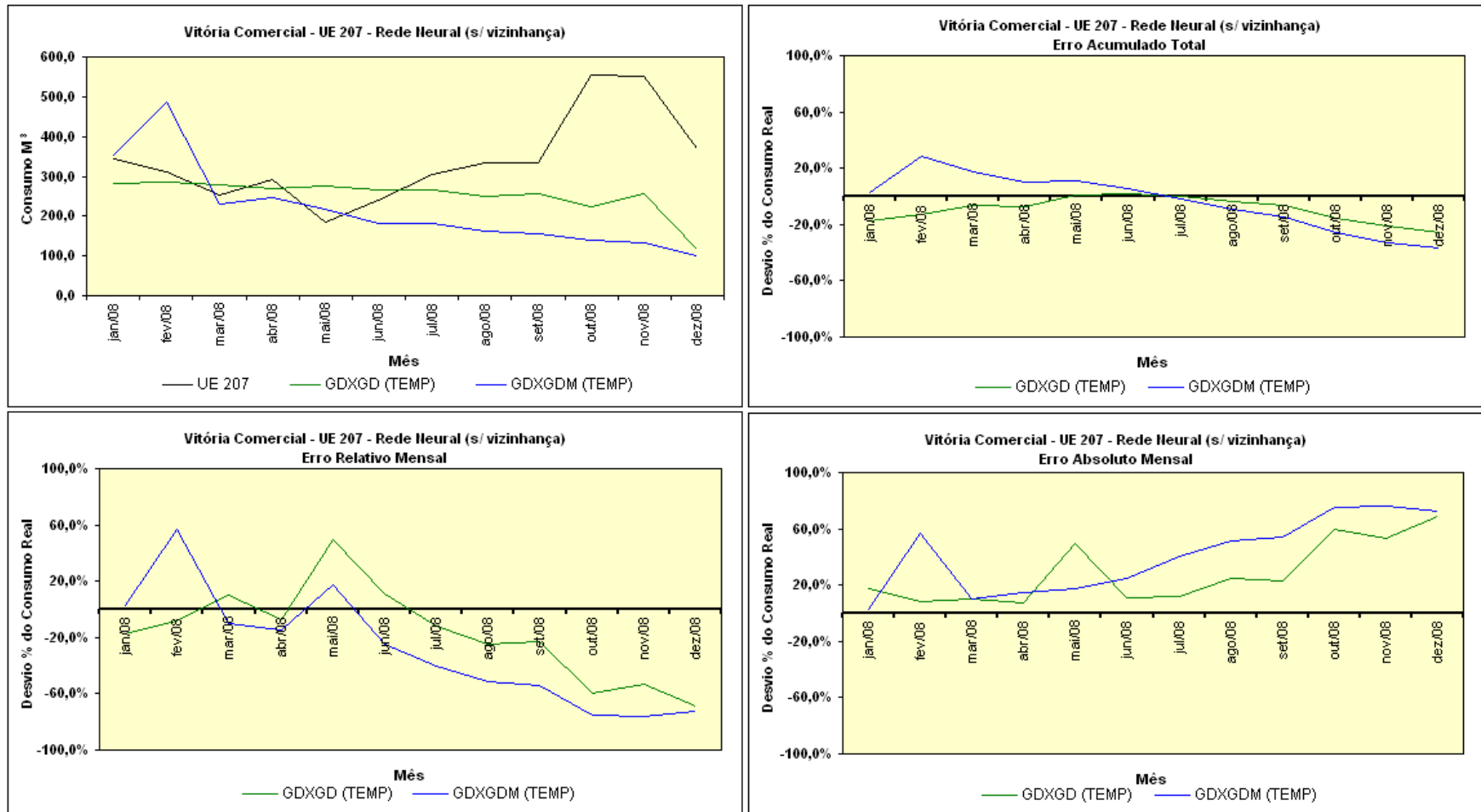
| Redes Neurais (c/ vizinhança) | Erro Acumulado Total | MAPE  |
|-------------------------------|----------------------|-------|
| GDXGD                         | 19,50%               | 26,5% |
| GDXGDM                        | -59,83%              | 51,5% |

Figura M.6 – Previsão RN Método TRAIN GDX c/ dados da Vizinhança UE 207 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



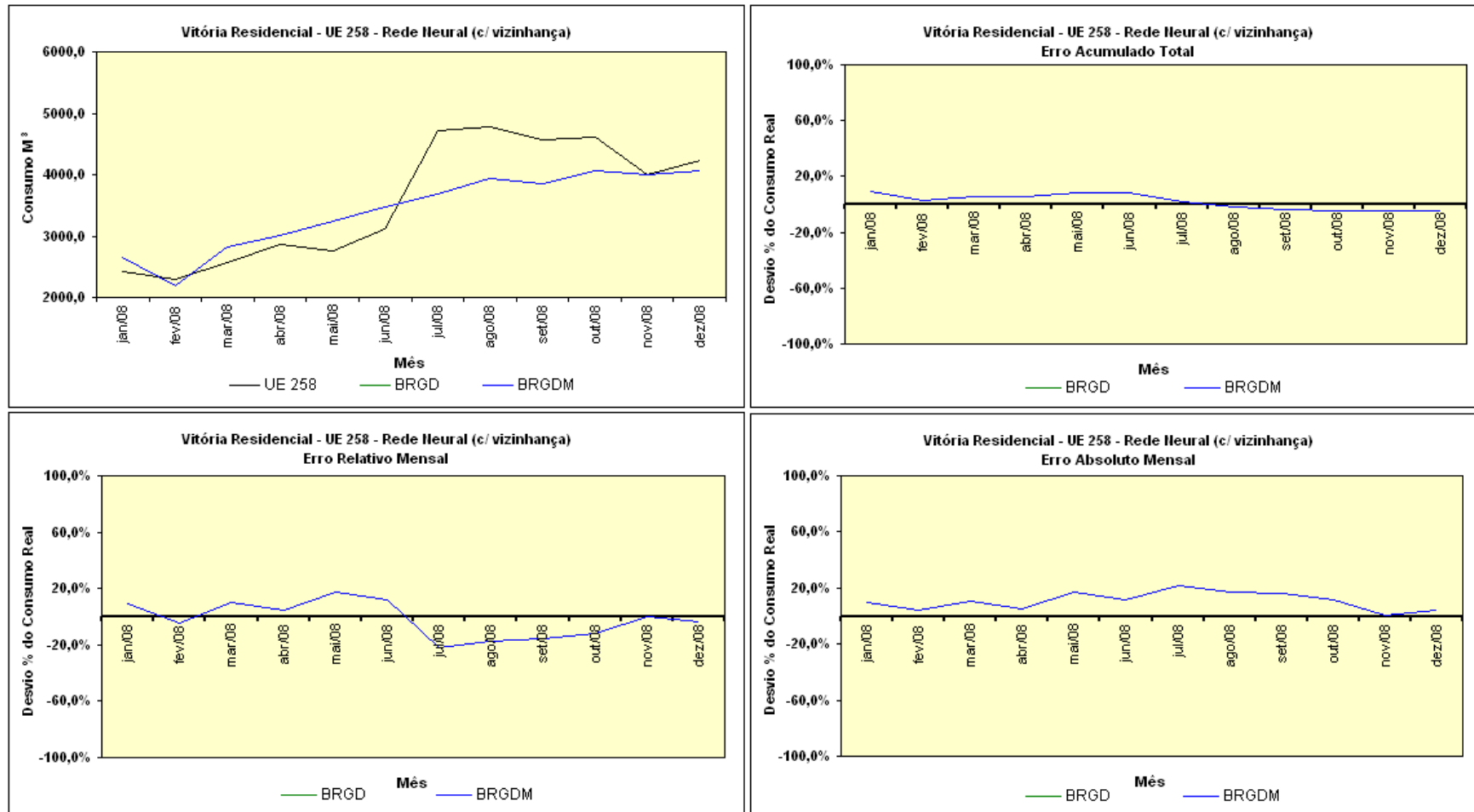
| Redes Neurais (s/ vizinhança) | Erro Acumulado Total | MAPE  |
|-------------------------------|----------------------|-------|
| BRGD (TEMP)                   | -9,92%               | 22,1% |
| BRGDM (TEMP)                  | -9,92%               | 22,1% |

Figura M.7 – Previsão RN Método TRAIN BR s/ dados da Vizinhança UE 207 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



| Redes Neurais (s/ vizinhança) | Erro Acumulado Total | MAPE  |
|-------------------------------|----------------------|-------|
| GDXGD (TEMP)                  | -25,68%              | 28,8% |
| GDXGDM (TEMP)                 | -36,55%              | 41,3% |

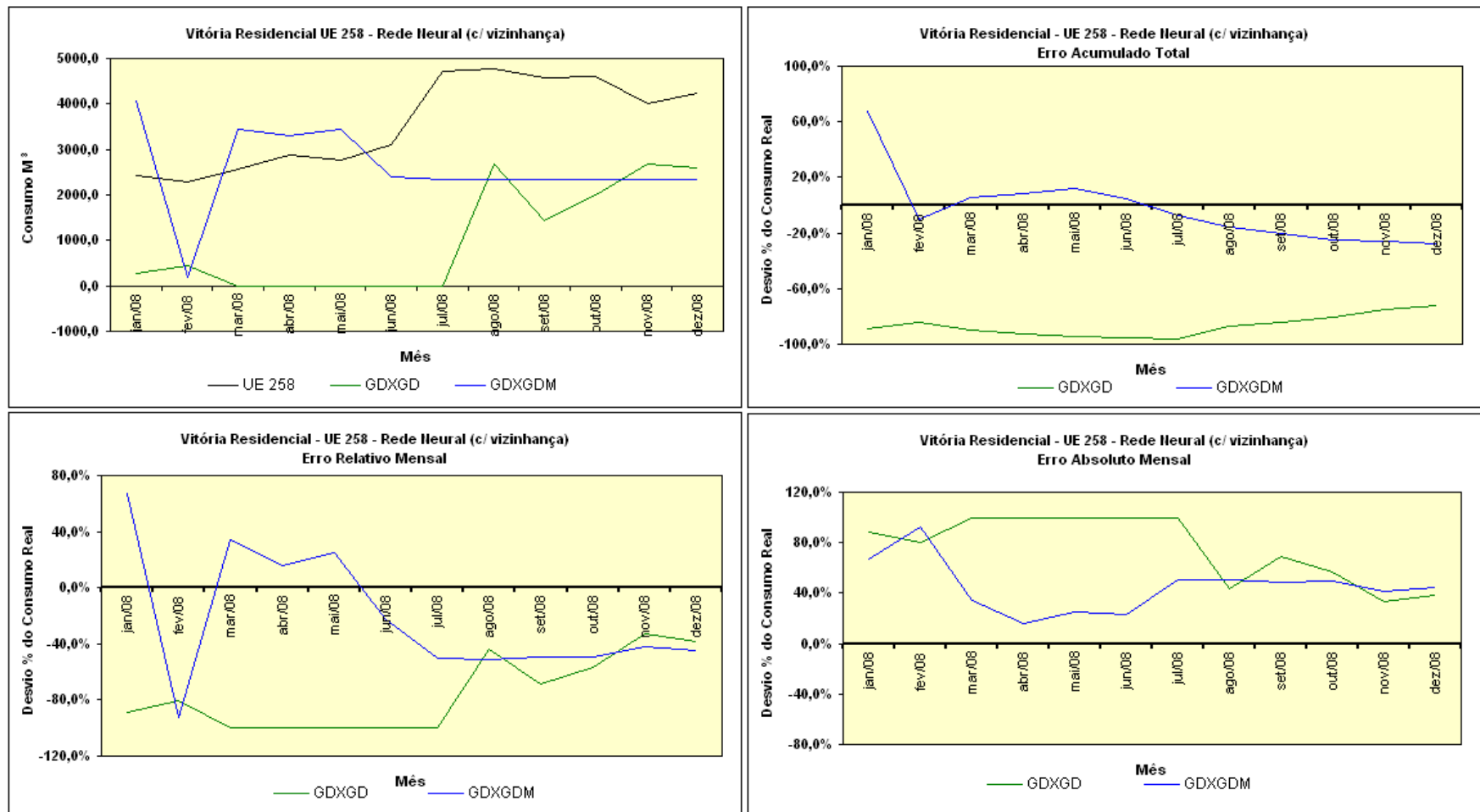
Figura M.8 – Previsão RN Método TRAIN GDX s/ dados da Vizinhança UE 207 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



| Redes Neurais (c/ vizinhança) | Erro Acumulado Total | MAPE  |
|-------------------------------|----------------------|-------|
| BRGD                          | -4,46%               | 10,8% |
| BRGDM                         | -4,46%               | 10,8% |

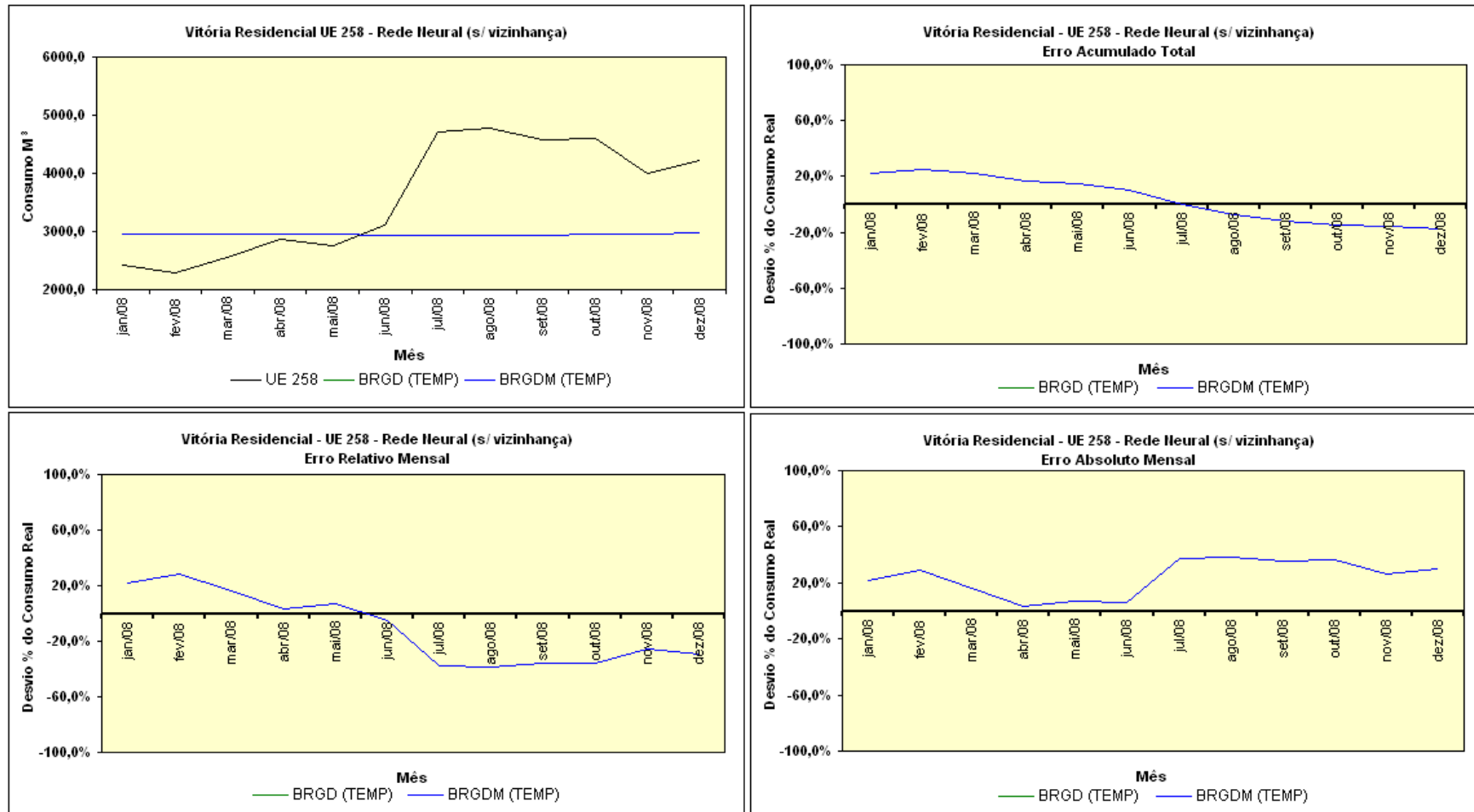
Figura M.9 – Previsão RN Método TRAIN BR c/ dados da Vizinhança UE 258 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).





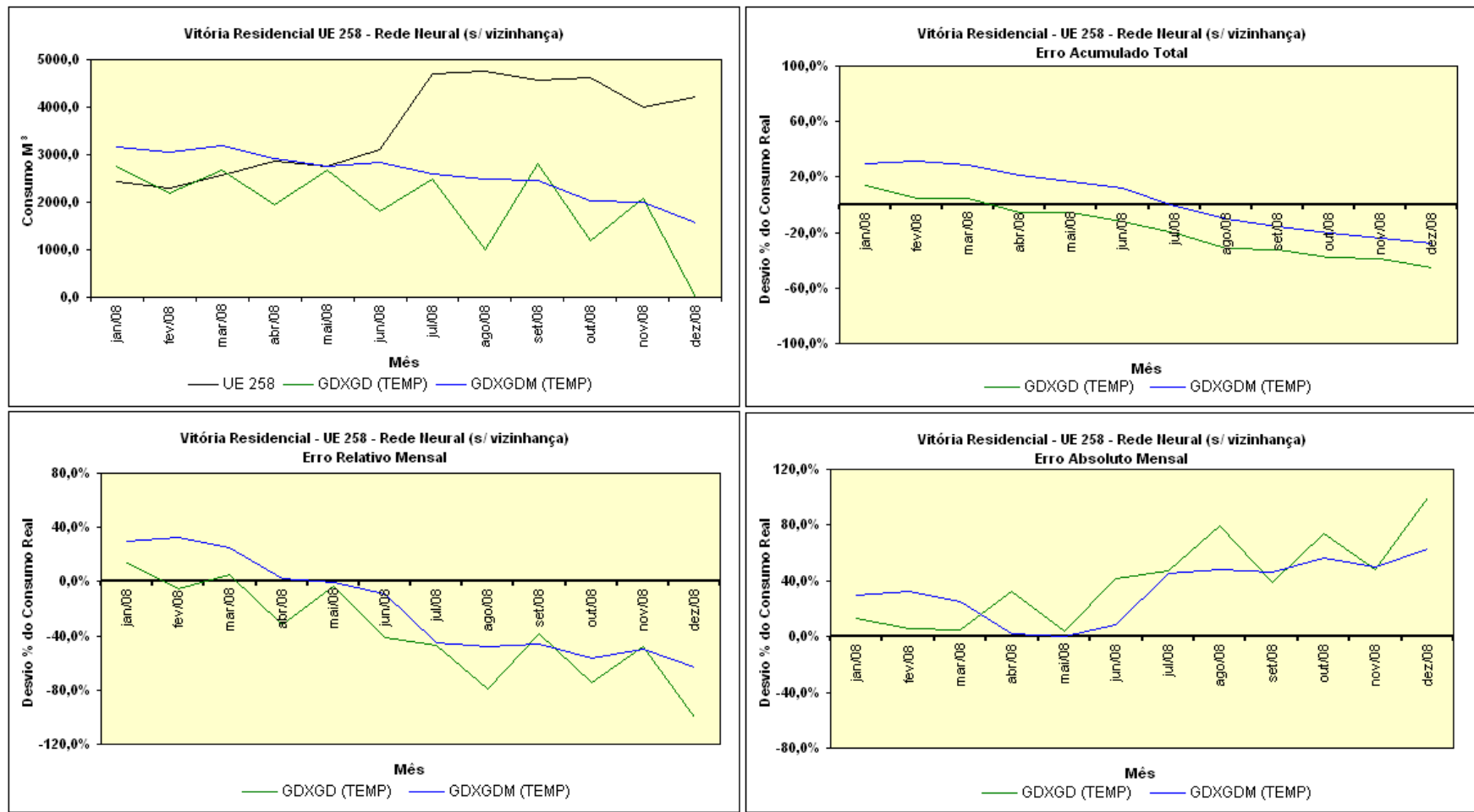
| Redes Neurais (c/ vizinhança) | Erro Acumulado Total | MAPE  |
|-------------------------------|----------------------|-------|
| GDXGD                         | -71,78%              | 75,8% |
| GDXGDM                        | -28,15%              | 45,3% |

Figura M.10 – Previsão RN Método TRAIN GDX c/ dados da Vizinhança UE 258 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



| Redes Neurais (s/ vizinhança) | Erro Acumulado Total | MAPE  |
|-------------------------------|----------------------|-------|
| BRGD (TEMP)                   | -17,51%              | 23,8% |
| BRGDM (TEMP)                  | -17,51%              | 23,8% |

Figura M.11 – Previsão RN Método TRAIN BR s/ dados da Vizinhança UE 258 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



| Redes Neurais (s/ vizinhança) | Erro Acumulado Total | MAPE  |
|-------------------------------|----------------------|-------|
| GDXGD (TEMP)                  | -44,99%              | 40,7% |
| GDXGDM (TEMP)                 | -27,67%              | 33,9% |

Figura M.12 – Previsão RN Método TRAIN GDX s/ dados da Vizinhança UE 258 – Gráficos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).

## APÊNDICE (M) - PREVISÕES SAR *Panel* X RN POR UE

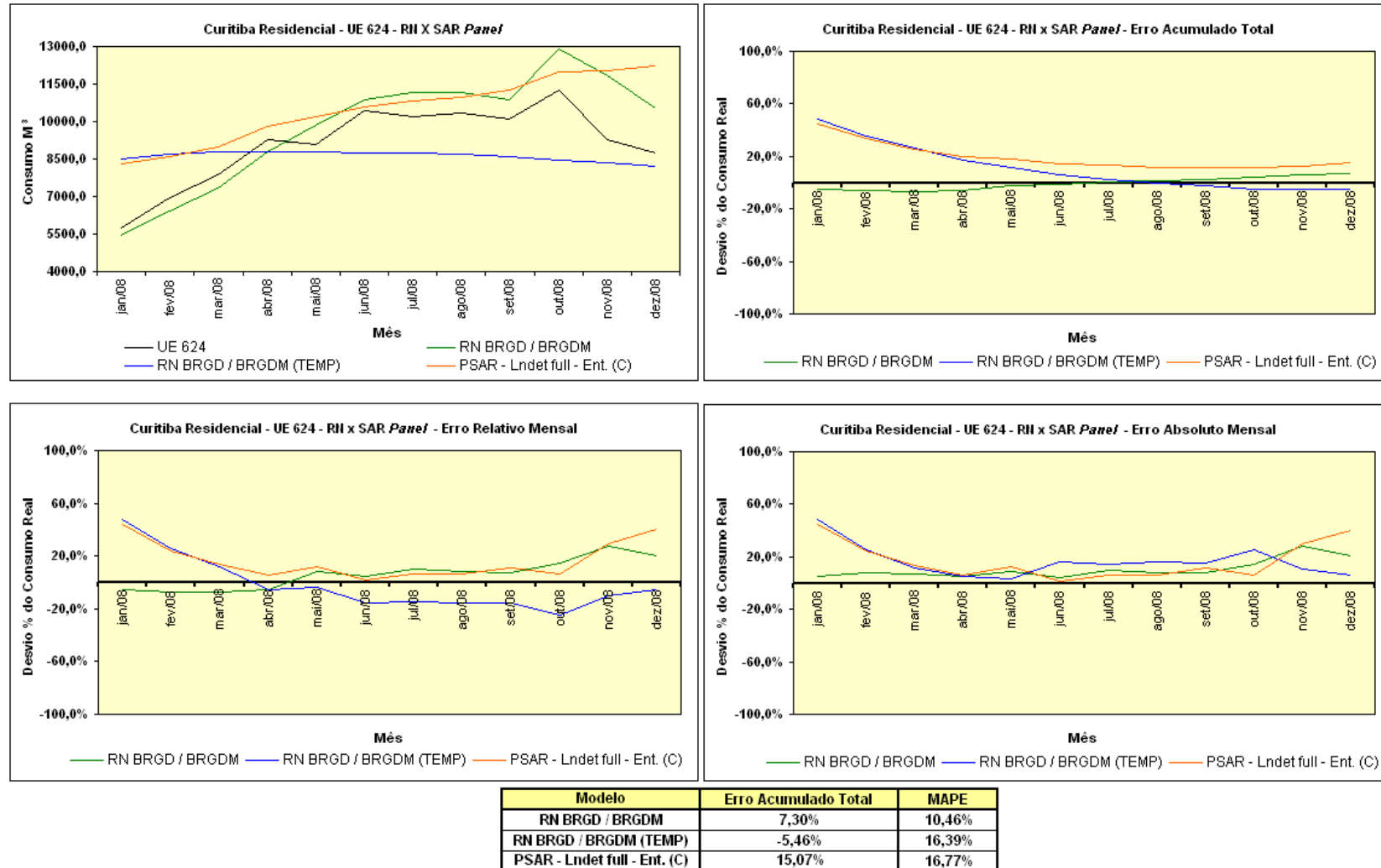
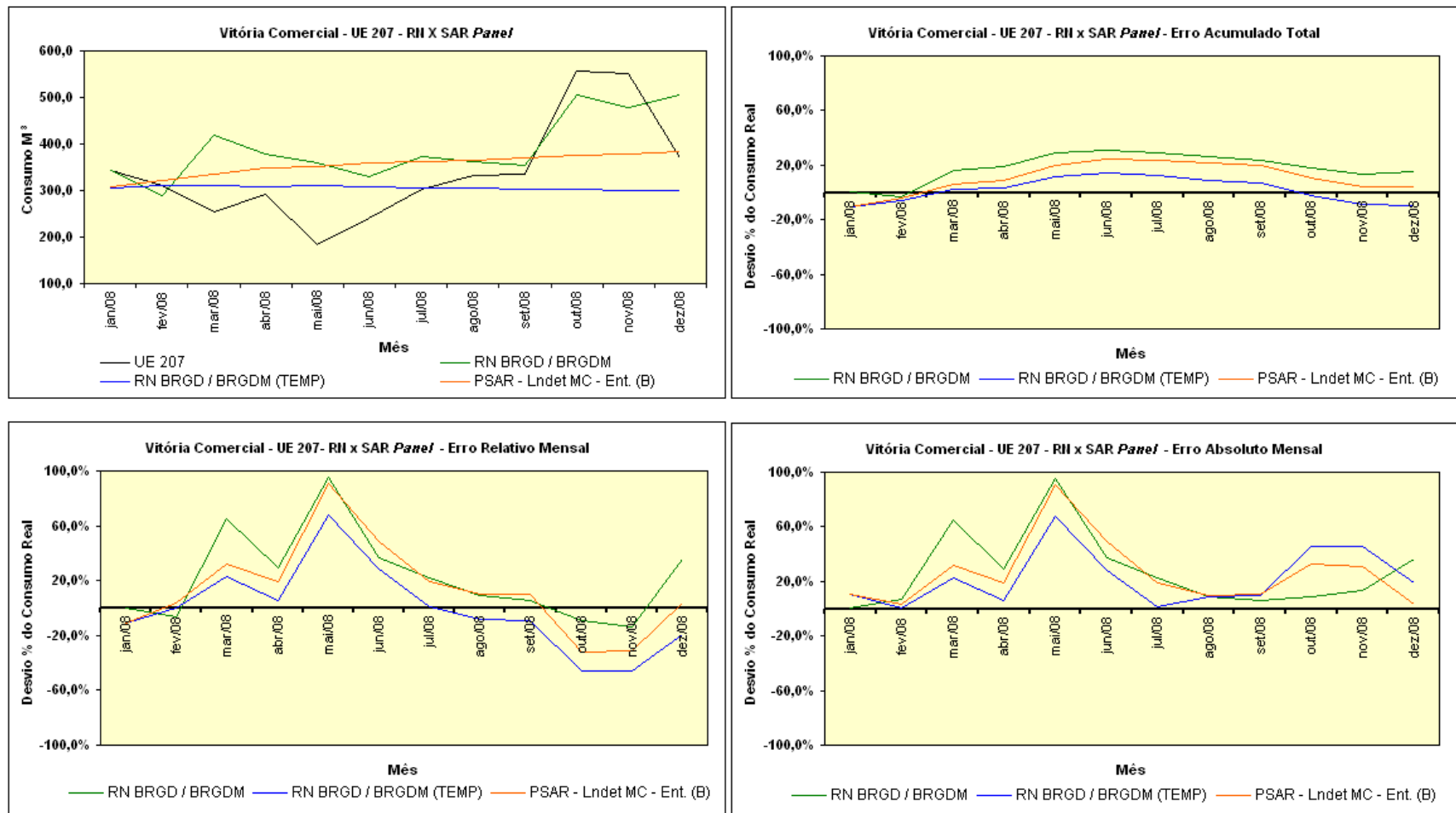
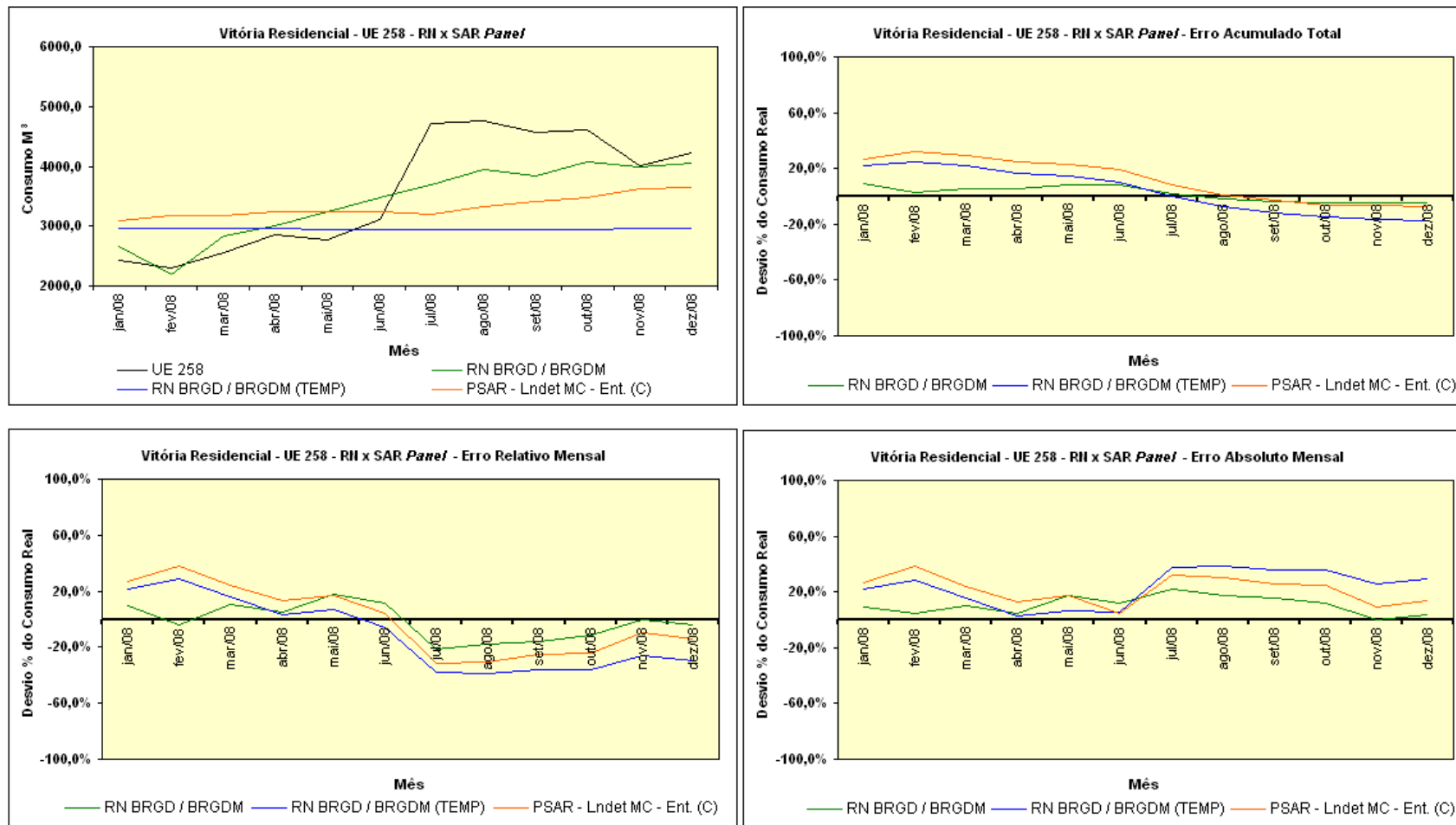


Figura N.1 – Previsão SAR *Panel* x RN Método TRAIN BR UE 624 – Gráficos Comparativos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



| Modelo                     | Erro Acumulado Total | MAPE   |
|----------------------------|----------------------|--------|
| RN BRGD / BRGDM            | 15,30%               | 27,46% |
| RN BRGD / BRGDM (TEMP)     | -9,92%               | 22,07% |
| PSAR - Lndet MC - Ent. (B) | 4,49%                | 25,99% |

Figura N.2 – Previsão SAR Panel x RN Método TRAIN BR UE 207 – Gráficos Comparativos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).



| Modelo                     | Erro Acumulado Total | MAPE   |
|----------------------------|----------------------|--------|
| RN BRGD / BRGDM            | -4,46%               | 10,79% |
| RN BRGD / BRGDM (TEMP)     | -17,51%              | 23,75% |
| PSAR - Lndet MC - Ent. (C) | -7,13%               | 21,60% |

Figura N.3 – Previsão SAR Panel x RN Método TRAIN BR UE 258 – Gráficos Comparativos de Previsão e Erros (Relativo Mensal / Acumulado Total / Absoluto Mensal).

## APÊNDICE (N) - PRODUÇÃO TEXTUAL

CASTRO, J. H., DA SILVA, A. P. A., 2010, “O emprego de modelos auto-regressivos espaciais na previsão de demanda para gás natural”. In: **Anais de VII Congresso Brasileiro de Planejamento Energético (CBPE)**, São Paulo. 08 – 10 Outubro.

CASTRO, J. H., DA SILVA, A. P. A., 2010, “A perspectiva espaço-temporal empregada na projeção de demanda de gás natural. In: **Anais de VII Congresso Brasileiro de Planejamento Energético (CBPE)**, São Paulo. 08 – 10 Outubro.