



Universidade Federal da Paraíba

Centro de Tecnologia

**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
URBANA E AMBIENTAL
– MESTRADO –**

**AVALIAÇÃO DA DISPONIBILIDADE DE ÁGUA EM AQUÍFEROS
POR MEIO DE ANÁLISES ESPAÇO-TEMPORAIS**

Por

Romildo Toscano de Brito Neto

*Dissertação de Mestrado apresentada à Universidade Federal da Paraíba
para obtenção do grau de Mestre*



Universidade Federal da Paraíba

Centro de Tecnologia

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA

URBANA E AMBIENTAL

– MESTRADO –

**AVALIAÇÃO DA DISPONIBILIDADE DE ÁGUA EM AQUÍFEROS
POR MEIO DE ANÁLISES ESPAÇO-TEMPORAIS**

Dissertação submetida ao Programa de Pós-Graduação em Engenharia Urbana e Ambiental da Universidade Federal da Paraíba, como parte dos requisitos para a obtenção do título de Mestre.

Romildo Toscano de Brito Neto

Orientador: Prof. Dr. Celso Augusto Guimarães Santos

João Pessoa – Paraíba

Dezembro de 2012

B862a Brito Neto, Romildo Toscano de.

Avaliação da disponibilidade de água em aquíferos por meio de análises espaço-temporais / Romildo Toscano de Brito Neto.- João Pessoa, 2012.

121f. : il.

Orientador: Celso Augusto Guimarães Santos

Dissertação (Mestrado) - UFPB/CT

1. Engenharia urbana e ambiental. 2. Águas subterrâneas. 3. Interpolação espacial. 4. Análise de cluster. 5. Teste de Mann-Kendall.

UFPB/BC

CDU: 62:711(043)

**“AVALIAÇÃO DA DISPONIBILIDADE DE ÁGUA EM AQUÍFEROS
POR MEIO DE ANÁLISES ESPAÇO-TEMPORAIS”**

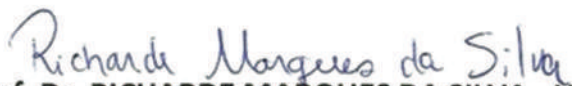
ROMILDO TOSCANO DE BRITO NETO

Dissertação aprovada em 21 de dezembro de 2012

Período Letivo: 2012.2



Prof^o. Dr. CELSO AUGUSTO GUIMARÃES SANTOS – UFPB
Orientador



Prof. Dr. RICHARDE MARQUES DA SILVA - UFPB
Examinador Interno



Prof^a. Dra. LUCILENE ANTUNES CORREIA MARQUES DE SÁ – UFPE
Examinador Externo

João Pessoa-PB
2012

AGRADECIMENTOS

Agradeço primeiramente ao Prof. Dr. Celso A. G. Santos pela excelente orientação, pela paciência nos momentos em que não pude corresponder as expectativas e, sobretudo, por ser o principal colaborador do meu desenvolvimento acadêmico.

Aos meus pais pelo conforto que sempre foi oferecido e por investir em mim no momento que mais precisei.

A equipe do *Center for Geospatial Technology* da *Texas Tech University* por me receber de braços abertos, pelos dados fornecidos e pela prontidão para tirar dúvidas.

A Cooperação Internacional do Semiárido (CISA) por financiar o intercâmbio com a *Texas Tech University*, onde defini o tema a ser abordado no mestrado.

A todos os professores, funcionários e alunos que fazem parte do LARHENA (Laboratório de Recursos Hídricos e Engenharia Ambiental) pelos ensinamentos e companheirismo.

Ao Prof. Marcello Benigno por sempre estar disponível para tirar dúvidas e por me orientar na preparação dos dados de entrada.

Ao colega de profissão Diego Valdevino pelo apoio durante o processamento dos dados espaciais.

A Wanessa Weridiana e Allysson Oliveira (Departamento de Estatística – UFPB) pelo apoio durante o processamento das séries temporais.

À CAPES pela bolsa REUNI fornecida durante o mestrado.

LISTA DE ILUSTRAÇÕES

Figura 2.1: Mapa da área de cobertura do Aquífero Ogallala	4
Figura 2.2: Definição da área do Aquífero Ogallala em estudo	5
Figura 3.1: Ciclo Hidrológico.....	9
Figura 3.2: Tipos de aquíferos.....	10
Figura 3.3: Diferença nas superfícies geradas por IDW e <i>spline</i>	13
Figura 3.4: Uso do histograma para definir a transformação	15
Figura 3.5: Representação 3D do nível dos poços do aquífero Ogallala Identificação de tendências globais (observado no ano 2000).....	15
Figura 3.6: Identificação de outliers por <i>Cluster</i> e Entropia	16
Figura 3.7: Definição do modelo de semivariograma	17
Figura 3.8: Definição da vizinhança.....	18
Figura 3.9: Conceito da krigagem ordinária.....	20
Figura 3.10: Conceito da krigagem universal.....	21
Figura 3.11: Exemplo de Dendrograma	25
Figura 3.12: Exemplo de um <i>Heat map</i>	27
Figura 3.13: Exemplo de um clustergrama.....	27
Figura 4.1: Distribuição espacial dos poços a serem interpolados	33
Figura 4.2: Distribuição espacial dos poços utilizados para a análise de tendências e validação cruzada.....	35
Figura 4.3: Fluxograma de processamento dos dados para cálculo do volume	42
Figura 4.4: Criação de superfícies de nível da água	43
Figura 4.5: Criação da superfície de camada saturada	44
Figura 4.6: Cálculo de volume <i>pixel a pixel</i>	45
Figura 4.7: Gráficos de nível da água com todas as séries	46
Figura 4.8: Gráfico da variação do nível da água com todas as séries	46
Figura 5.1: Variação do volume total do aquífero.....	48
Figura 5.2: Correlação krigagem <i>versus spline</i>	49

Figura 5.3: Espessura da camada saturada em 1962 e 1972.....	51
Figura 5.4: Espessura da camada saturada em 1982 e 1992.....	52
Figura 5.5: Espessura da camada saturada em 2002 e 2012.....	53
Figura 5.6: Evolução das camadas saturadas.....	53
Figura 5.7: Clustergrama	54
Figura 5.8: Clustergrama e o conjunto de séries de cada <i>cluster</i>	55
Figura 5.9: Gráficos de séries temporais classificadas por <i>cluster</i> (Método Hierárquico e Método <i>k-means</i>)	56
Figura 5.10: Síntese da análise de tendências realizada pelo teste de Mann-Kendall.....	58
Figura 5.11: Distribuição dos pontos por classe de cluster obtida pelo método hierárquico com tendência decrescente	59
Figura 5.12: Variação no nível da camada saturada entre 1962 e 2012	60
Figura 5.13: Identificação dos <i>clusters</i> dos poços inseridos na área crítica e proximidades ..	61
Figura 5.14: Conjunto de séries contidas dentro do polígono da área crítica e em suas proximidades	61
Figura 5.15: Variação do nível médio da camada saturada na área crítica	62
Figura 5.16: Identificação de cultivos dentro da área crítica.....	63
Figura 5.17: Sobreposição das camadas de variação no nível da camada saturada (entre 1962 e 2012) e de áreas cultivadas, destacando áreas para análises mais aprofundadas	64
Figura 5.18: Variação da camada saturada média em áreas destacadas.....	65

LISTA DE ABREVIATURAS E SIGLAS

CISA	Cooperação Internacional do Semiárido
CS	Camada saturada
EM	Erro Médio
EUA	Estados Unidos da América
IDW	<i>Inverse Distance Weighted</i>
MMA	Ministério do Meio Ambiente
MRLC	<i>Multi-Resolution Land Characteristics Consortium</i>
PPGEUA	Programa de Pós-Graduação em Engenharia Urbana e Ambiental
RMSE	<i>Root Mean Square Error</i>
RNA	Redes Neurais Artificiais
SQL	<i>Structured Query Language</i>
TTU	<i>Texas Tech University</i>
TWDB	<i>Texas Water Delevopment Board</i>
UFPB	Universidade Federal da Paraíba
UNESCO	<i>United Nations Educational, Scientific And Cultural Organization</i>
USDA	<i>United States Department of Agriculture</i>

LISTA DE TABELAS

Tabela 4.1: Quantidade de poços interpolados para cada ano da série histórica.....	34
Tabela 4.2: Parâmetros de configuração IDW e validação cruzada	37
Tabela 4.3: Parâmetros de configuração <i>spline</i> e validação cruzada	38
Tabela 4.4: Síntese dos resultados obtidos pela validação cruzada.....	40
Tabela 5.1: Diferença percentual entre as estimativas de volume.....	49
Tabela 5.2: Comparação entre <i>clusters</i> por <i>k-means</i> e hierárquico.....	56

LISTA DE APÊNDICES

APÊNDICE A - Consultas SQL para a obtenção dos dados de nível da água.....	73
APÊNDICE B - Criação da rotina em Visual Basic para preenchimento de falhas.....	75
APÊNDICE C - Configurações das krigagens realizadas entre 1960 e 2012	77
APÊNDICE D - Fluxograma do processamento das imagens para a geração dos resultados	91
APÊNDICE E - Comandos utilizados para processar os dados no <i>software</i> R	93
APÊNDICE F - Tabela com os resultados do teste de Mann-Kendall e agrupamentos	97

RESUMO

O crescimento da exploração de águas subterrâneas para suprir demandas de abastecimento, agricultura e indústria faz com que a extração frequentemente exceda a recarga natural, resultando num declínio de seu volume, deterioração do solo e qualidade da água. O objetivo geral deste trabalho é avaliar a disponibilidade de água na porção do aquífero Ogallala contido no Estado do Texas (EUA), onde a exploração tem alcançado níveis acima dos valores de recarga. Para isto, inicialmente, foram utilizados dados de nível da água de poços com uma série de 53 anos (1960 a 2012), criando-se 53 superfícies por cada método de interpolação espacial (krigagem e *spline*). Em seguida, foi realizada uma comparação entre as técnicas de interpolação pelos métodos *leave-one-out* e *holdout* de validação cruzada, além de verificar as diferenças nas estimativas do volume total do aquífero. Os dois métodos de interpolação produziram resultados semelhantes e desempenhos satisfatórios; entretanto, recomenda-se o *spline* para processar múltiplas superfícies de nível da água de modo automatizado e a krigagem para quando se demandar maior acurácia dos resultados. De posse das superfícies de nível da água, foi estimada a camada saturada do aquífero e o volume *pixel* a *pixel*, para toda a série histórica, o que permitiu a partir dos resultados espaciais, realizar análises temporais. Em paralelo, foi selecionado um conjunto de 492 séries temporais de variação do nível da água para se analisar tendências (teste de Mann-Kendall) e agrupá-las por meio de análise de *cluster* (método hierárquico e *k-means*). A partir destas análises temporais, foi verificada a distribuição espacial dos resultados, constatando que o grupo de *clusters* com a maior tendência de decrescimento converge com áreas críticas identificadas. Além disso, foi verificada a influência de áreas cultivadas na variação do nível da água, mostrando que existe uma forte relação entre os dois fenômenos. Por fim, os resultados mostram que houve uma redução de 33,9% do volume total em 53 anos, 74,39% das séries analisadas indicaram tendência decrescente e observou-se que o nível da camada saturada vem reduzindo gradativamente.

PALAVRAS-CHAVE: Águas subterrâneas, interpolação espacial, análise de *cluster*, teste de Mann-Kendall.

ABSTRACT

Increasing groundwater extraction for domestic, agricultural, and industrial supply frequently makes exploitation exceed the natural recharge rate. This results in decreasing volumes, deteriorating water quality, and soil degradation. The goal of this study is to assess water availability in the Texas State (USA) portion of the Ogallala aquifer, where extraction has reached levels that are above the recharge values. Fifty three years of water level data from wells between 1960 and 2012 were used. Fifty three surfaces each were created using both (kriging and spline techniques) of spatial interpolation, and cross-validation methods (leave-one-out, and holdout) were used to compare the predicted water level surfaces with the well data. The predictions were similar and the results were satisfactory, however we found spline technique simpler, and easier for automatic processing of multiple surfaces, and we recommend it. However, kriging technique presents higher accuracy. After processing the water level surfaces; the saturated thickness and volume (pixel by pixel), for the whole time series were estimated. This allowed temporal analysis to be performed from the results. A set of 492 water level variation time series (to analyze trends using the Mann-Kendall test) were selected and grouped by cluster analysis (using both hierarchical and k-means methods). From the temporal analyses, the spatial distribution of the results was verified, and it was observed that the clusters most likely to decrease converge with the critical areas identified. Furthermore, the influence of agricultural activities on water level variance was established, showing a strong relationship between the two. Finally, the results showed that over 53 years, the total water volume had been reduced by 33.9%. A full 74.39% of the time series analyzed also showed a decreasing trend in water levels, and it was observed that the saturated thickness is gradually being reduced as well.

KEYWORDS: Groundwater, Spatial Interpolation, Cluster analysis, Mann-Kendall test.

SUMÁRIO

LISTA DE ILUSTRAÇÕES	v
LISTA DE ABREVIATURAS E SIGLAS	vii
LISTA DE TABELAS	viii
LISTA DE APÊNDICES	ix
RESUMO	x
ABSTRACT	xi
1. INTRODUÇÃO	1
1.1 OBJETIVOS	2
1.2 ESTRUTURA DA DISSERTAÇÃO	3
2. A ÁREA DE ESTUDO	4
2.1 LOCALIZAÇÃO	5
2.2 IMPORTÂNCIA ECONÔMICA, SOCIAL E AMBIENTAL	5
2.3 CARACTERÍSTICAS HIDROGEOLÓGICAS	6
3. REFERENCIAL TEÓRICO	8
3.1 ÁGUAS SUBTERRÂNEAS	8
3.1.1 As Águas Subterrâneas e o ciclo hidrológico	8
3.1.2 Tipos de Aquíferos	9
3.1.3 Exploração de Aquíferos	11
3.2 MÉTODOS DE INTERPOLAÇÃO ESPACIAL	11
3.2.1 <i>Inverse Distance Weighted (IDW)</i>	12
3.2.2 <i>Spline</i>	13
3.2.3 Métodos Geoestatísticos	14
3.2.3.1 <i>Análise Exploratória</i>	14
3.2.3.2 <i>Modelagem do Semivariograma</i>	16
3.2.3.3 <i>Definição da Vizinhança</i>	18

3.2.3.4 Krigagem Ordinária	19
3.2.3.5 Krigagem Universal	20
3.3 VALIDAÇÃO CRUZADA	21
3.3.1 Método <i>Leave-one-out</i>	22
3.3.1 Método <i>Holdout</i>	22
3.5 ANÁLISE DE <i>CLUSTER</i>	23
3.6.1 Métodos por particionamento	24
3.6.2 Métodos hierárquicos	25
3.6 TESTE DE MANN-KENDALL	28
4. MATERIAIS E MÉTODOS	31
4.1 DADOS DE NÍVEL DA ÁGUA EM POÇOS	31
4.1.1 Preenchimento de falhas.....	35
4.2 DETERMINAÇÃO DAS SUPERFÍCIES DE NÍVEL DE ÁGUA	37
4.3 COMPARAÇÃO ENTRE AS SUPERFÍCIES INTERPOLADAS	39
4.4 ESTIMATIVAS DE VOLUME DO AQUÍFERO	41
4.5 IDENTIFICAÇÃO DE TENDÊNCIAS EM SÉRIES HISTÓRICAS	45
5. RESULTADOS E DISCUSSÕES	48
5.1 ESTIMATIVAS DE VOLUME	48
5.2 EVOLUÇÃO DAS ESPESSURAS DE CAMADA SATURADA.....	51
5.3 ANÁLISES DE TENDÊNCIAS ESPAÇO-TEMPORAIS	54
5.4 IDENTIFICAÇÃO DE ÁREAS CRÍTICAS.....	59
5.5 INFLUÊNCIA DAS ÁREAS CULTIVADAS	62
6. CONSIDERAÇÕES FINAIS	66
REFERÊNCIAS BIBLIOGRÁFICAS	68
APÊNDICES	72

1. INTRODUÇÃO

A escassez de água de boa qualidade é um problema global que se agrava a cada dia. O crescimento populacional e as mudanças climáticas são fatores que afetam indiretamente sua qualidade e disponibilidade. Sem mudanças na gestão e no uso da água, os problemas de escassez e qualidade chegarão, inevitavelmente, a níveis críticos com o passar do tempo (GRAFTON e HUSSEY, 2011).

As águas subterrâneas representam a maior fonte de água fresca no ciclo hidrológico (CHOW, 1964). Segundo Tilahun e Merkel (2009), o crescimento da exploração de águas subterrâneas para suprir demandas de abastecimento, agricultura e indústria, em algumas localidades do planeta, está ligado à carência de águas superficiais. A extração de águas subterrâneas frequentemente excede a recarga natural, resultando num declínio de seu volume, deterioração do solo e qualidade da água. Assim, o desenvolvimento sustentável é ameaçado pela excessiva exploração desses corpos hídricos, tornando-os vulneráveis ao esvaziamento de suas reservas (HERMANS, 2008).

Dentre as consequências da depleção de aquíferos estão o racionamento de água local, esvaziamento de poços, mudanças na qualidade da água e mudanças no fluxo da água. O aquífero Ogallala, localizado no Texas, EUA, onde está inserida a área de estudo desta pesquisa, possui grande parte de sua extensão localizada em regiões semiáridas e encontra-se com o nível de água diminuindo e sua qualidade deteriorando. Atividades como usinas hidrelétricas, irrigação por águas superficiais e, principalmente, por poços, resultam em um desequilíbrio do aquífero, ou seja, a descarga não é igual à recarga em muitas áreas (GURU e HORNE, 2000).

Uma das maneiras de facilitar os gerenciadores dos recursos hídricos a induzirem o uso racional das águas subterrâneas é obtendo tendências do nível da água. Gray (2007) destaca entre os métodos mais comumente aplicados para detecção de tendência a regressão linear, testes não paramétricos, *e.g.*, Mann-Kendall, regressão suavizada por mínimos quadrados, teste da razão de verossimilhança, entre outros.

Porém, muitos desses métodos baseiam-se em análises pontuais, o que torna a espacialização das informações uma limitação. Para Kumar e Ramadevi (2006), o sucesso da gestão dos recursos hídricos subterrâneos requer o conhecimento da distribuição espacial do nível da água, parâmetros do aquífero e outros dados. Métodos de interpolação espacial têm a

capacidade de produzir uma superfície a partir de dados pontuais e medir a qualidade do resultado por meio de validação cruzada.

Portanto, o presente estudo visa quantificar a disponibilidade de água subterrânea, identificar áreas críticas de redução do nível da água e, à medida do possível, relacioná-las com atividades antrópicas (extração da água). O que se busca é indicar que tal recurso tende a se tornar escasso, ajudando a persuadir usuários da água a implementar medidas de conservação ou ao uso sustentável. Para isto, serão geradas superfícies que representam o nível da água por meio de métodos de interpolação espacial. Em seguida serão estimados volumes de água para as áreas do aquífero e, por fim, serão realizadas as análises de tendências e espaço-temporais.

A escolha deste tema tem o propósito de consolidar uma parceria entre a *Texas Tech University* (TTU) e a UFPB, através do projeto de Cooperação Internacional do Semiárido (CISA), que conta com a participação de pesquisadores do Programa de Pós-Graduação em Engenharia Urbana e Ambiental (PPGEUA).

A razão para a escolha da área de estudo é a densa base de dados existente e de livre acesso. Por outro lado, não existe base de dados de poços no Nordeste brasileiro com séries históricas e densidade espacial que viabilize a aplicação da metodologia deste trabalho.

1.1 OBJETIVOS

Diante do exposto, o objetivo geral deste trabalho é avaliar a disponibilidade de água na porção do aquífero Ogallala contido no Texas mediante análises espaço-temporais, utilizando métodos de interpolação espacial e análise de tendência em séries temporais.

São objetivos específicos:

- a) Comparar o desempenho de métodos de interpolação espacial (Krigagem, IDW e Spline) para a geração de superfícies de nível da água;
- b) Estimar o volume total do aquífero na área em estudo para todos os anos analisados;
- c) **Detectar tendências** nas séries temporais através do teste de Mann-Kendall; e
- d) Identificar áreas críticas onde ocorre maior redução do nível da água por meio de agrupamento de séries temporais e avaliação da evolução das espessuras de camada saturada.

1.2 ESTRUTURA DA DISSERTAÇÃO

Esta dissertação está dividida da seguinte forma: (a) o Capítulo 1 contém uma visão geral a cerca do problema e os objetivos da dissertação; (b) o Capítulo 2 descreve a área de estudo, enfatizando suas características hidrogeológicas, além da importância, social, ambiental e econômica; (c) o Capítulo 3 mostra a fundamentação teórica sobre as águas subterrâneas, métodos de interpolação espacial e sua validação, métodos de análise de *cluster* (para agrupamento de séries temporais) e análise de tendência (teste de Mann-Kendall; d) o Capítulo 4 descreve a metodologia adotada; (e) o Capítulo 5 apresenta os resultados e uma análise acerca dos mesmos; e (f) o Capítulo 6 contém as conclusões e recomendações para futuros trabalhos.

2. A ÁREA DE ESTUDO

O Aquífero Ogallala (Figura 2.1), também conhecido como *High Plains*, cobre uma região de aproximadamente 440 mil quilômetros quadrados e está localizado em oito estados norte americanos (Colorado, Kansas, Nebraska, New Mexico, Oklahoma, South Dakota, Texas e Wyoming)(GUTENTAG *et al.*, 1984).

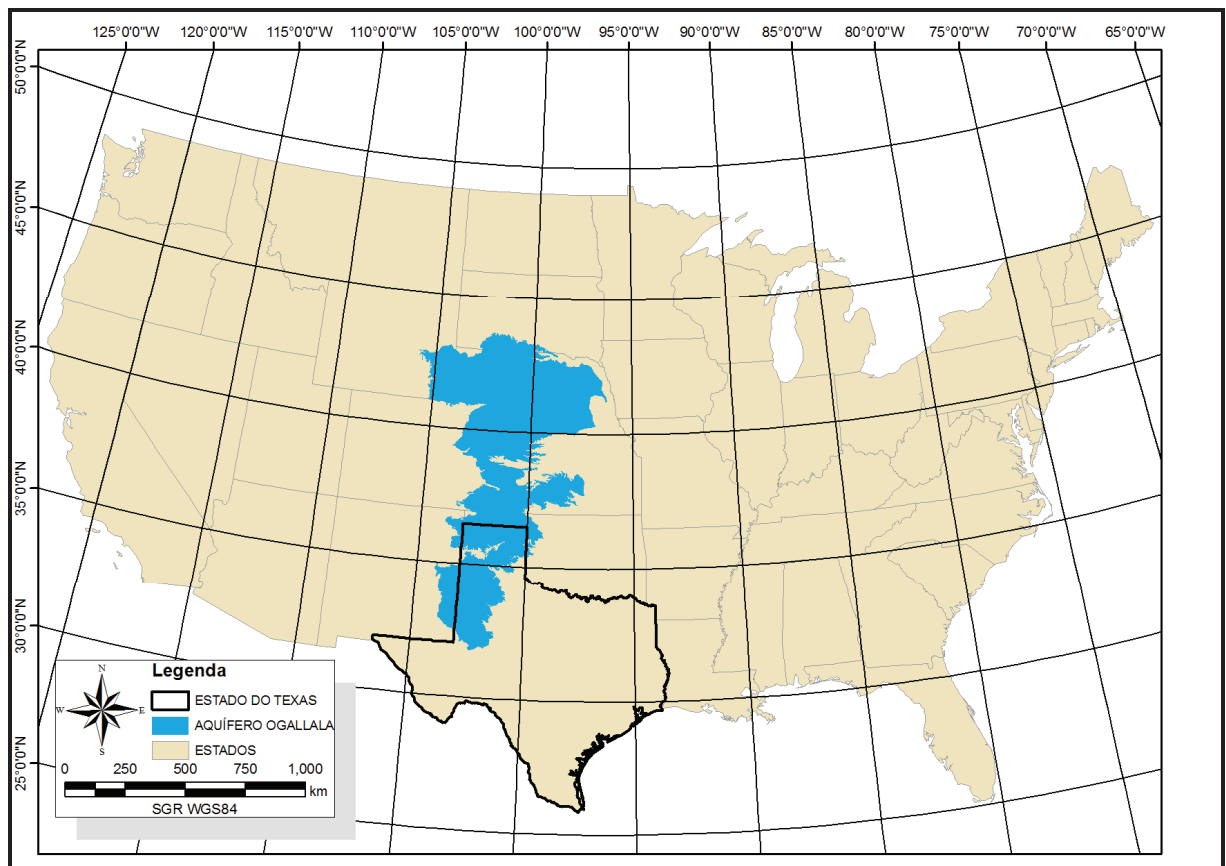


Figura 2.1: Mapa da área de cobertura do Aquífero Ogallala

Segundo Guru e Horne (2000), cerca de 90% da água extraída do Aquífero Ogallala é responsável por irrigar pelo menos um quinto de todas as áreas agricultáveis dos Estados Unidos. Miller (2012) afirma que o Aquífero Ogallala é o mais intensivamente explorado do território americano e que as retiradas chegam a ser 35 vezes maiores que a recarga em algumas regiões.

2.1 LOCALIZAÇÃO

A área de estudo deste trabalho compreende a porção do aquífero Ogallala que intersecta o Estado do Texas (Figura 2.2), possuindo aproximadamente 90 mil quilômetros quadrados. A região estudada possui características que diferem dos encontrados na região semiárida do Nordeste brasileiro. Apesar de não existirem rios perenes e volumosos, a agricultura irrigada é uma prática comum, tendo o aquífero local (Ogallala) como a principal fonte hídrica da região.

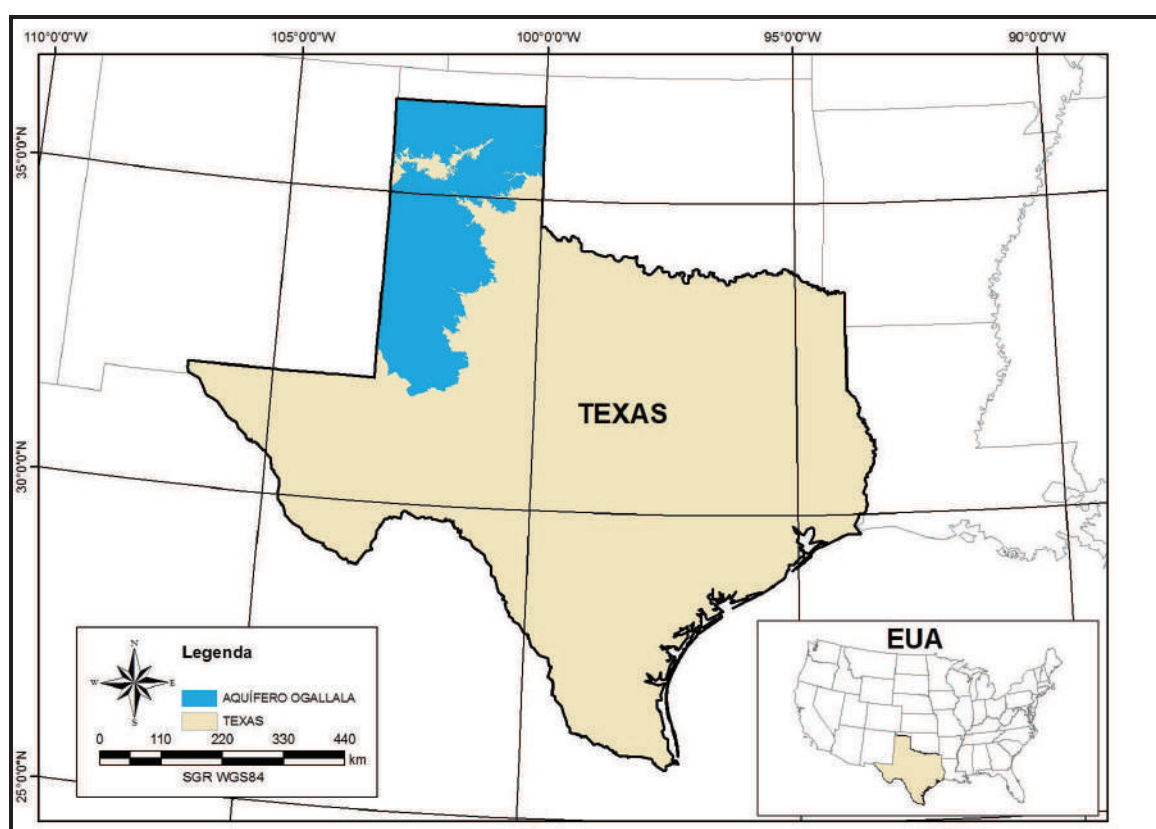


Figura 2.2: Definição da área do Aquífero Ogallala em estudo

2.2 IMPORTÂNCIA ECONÔMICA, SOCIAL E AMBIENTAL

Praticamente toda a população rural que vive nas áreas sobrejacentes ao aquífero utiliza sua água para uso doméstico. A economia da região depende fortemente da agroindústria como fator de desenvolvimento. Aproximadamente 1,01 milhão de hectares são

ocupados por fazendas que utilizam a água da irrigação para o plantio predominantemente de milho, trigo, soja e algodão (USDA, 1999).

A agricultura da região é dependente da água do aquífero Ogallala. O desenvolvimento da agricultura teve início na década de 1940 e se desenvolveu rapidamente (TERRELL *et al.*, 2002). Em 1949 havia uma exploração de $4,92 \times 10^9 \text{ m}^3$ de águas, aumentando para $6,97 \times 10^9 \text{ m}^3$ em 1979 e diminuindo para $5,77 \times 10^9 \text{ m}^3$ em 1989 (TWDB, 1993).

A forte extração de água do aquífero tem alcançado níveis acima dos valores de recarga. O relatório do *Texas Water Development Board* (TWDB, 1997) afirma que em 1994 a recarga do aquífero foi de aproximadamente $0,54 \times 10^9 \text{ m}^3$, o que representa apenas 7,4% do volume explotado ($7,26 \times 10^9 \text{ m}^3$).

As leis que regem o uso dos recursos hídricos do aquífero Ogallala diferem de Estado para Estado. O uso da água subterrânea é, em geral, outorgado sem restrições de vazão para o proprietário do terreno no qual a fonte hídrica se encontra. Porém, tem-se criado uma preocupação em torno das consequências desta exploração insustentável, gerando a adoção de métodos mais eficientes de irrigação, como o pivô central e investimentos em recarga (BRINEY, 2012; TERRELL *et al.*, 2002).

2.3 CARACTERÍSTICAS HIDROGEOLÓGICAS

O aquífero Ogallala foi formado há cerca de 10 milhões de anos, quando a água fluiu por camadas altamente permeáveis de areia e cascalho das planícies de geleiras e córregos próximas as *Rocky Mountains*. Devido às alterações causadas pela erosão e pela falta de água de degelo glacial, hoje o aquífero Ogallala não está mais recebendo recarga das *Rocky Mountains* (BRINEY, 2012).

O fluxo da água subterrânea geralmente é de oeste para leste, com uma velocidade média de aproximadamente 30cm por dia, com descargas para a rede de drenagem e nascentes, além de diretamente para a atmosfera por evapotranspiração. A precipitação (300–600 mm por ano) é a principal fonte de recarga do aquífero, estima-se que a recarga em algumas partes do Texas está em torno de 0,062 cm por ano. Tipicamente, estimativas de recarga são maiores para solos arenosos, já que se trata de um aquífero livre. Cerca de 66% da

água armazenada no aquífero se encontra no estado de Nebraska e 12% no Texas, possuindo a segunda maior porção (GUTENTAG *et al.*, 1984).

A condutividade hidráulica e a produção específica do aquífero dependem dos tipos de sedimentos encontrados, que variam bastante tanto horizontal quanto verticalmente. Por consequência, a condutividade hidráulica e a produção específica também são altamente variáveis (GUTENTAG *et al.*, 1984).

A base do aquífero na porção do Texas é formada por leitos rochosos compostos de rochas permianas, jurássica, triássicas e cretáceas inferiores. As rochas permianas em contato com o aquífero provocam grandes concentrações de sólidos dissolvidos, tornando estas áreas inadequadas para o plantio. As rochas jurássicas e triássicas foram classificadas juntas devido as suas características hidrológicas semelhantes, e possuem permeabilidade adequada para produzir quantidades suficientes de água para irrigação. As rochas cretáceas inferiores geralmente contém água de qualidade para irrigação, porém possuem baixa permeabilidade e capacidade de armazenamento, com baixas na produtividade do bombeamento e interferência entre poços distantes. Como resultado desta estrutura geológica, a água pode mover-se no aquífero sobre os leitos rochosos que estão em contato (GUTENTAG *et al.*, 1984).

3. REFERENCIAL TEÓRIO

3.1 ÁGUAS SUBTERRÂNEAS

As águas subterrâneas podem ser consideradas umas das mais importantes fontes de água limpa no planeta (CHOW, 1964). Apenas 2,5% de toda a água do planeta é considerada fresca. Desconsiderando o gelo polar, 96% da água fresca disponível é subterrânea (UNESCO, 2006).

O fato de ser acessível a um grande número de usuários, a possibilidade de extração de forma individualizada e a não dependência de grandes projetos hidráulicos para abastecimento são algumas das vantagens que elas oferecem ao desenvolvimento humano. Comparado às águas superficiais, aquíferos oferecem mais segurança contra secas, pois variações no nível da água devido a mudanças na recarga ou extração de poços só ocorrem após um longo tempo (HERMANS, 2008). Em nível global, 60% da extração de águas subterrâneas são destinados à agricultura e a parcela restante é dividida quase que igualmente entre os setores industrial e doméstico (UNESCO, 2006).

3.1.1 As Águas Subterrâneas e o ciclo hidrológico

O ciclo hidrológico representa o processo contínuo de movimentação da água entre o continente, os oceanos e a atmosfera. Esse movimento ocorre devido à força da gravidade e a energia solar. A Figura 3.1 mostra o esquema do ciclo hidrológico. A água que evapora dos oceanos e da superfície terrestre se torna parte da atmosfera, formando nuvens que se precipitam em forma de chuva, granizo, neve ou orvalho. A água precipitada pode escoar superficialmente, infiltrar e percolar por espaços vazios dos solos e das rochas, congelar ou evaporar. A água infiltrada pode penetrar profundamente e ficar armazenada, recarregando aquíferos e formando um escoamento subterrâneo que posteriormente terá sua descarga em fontes ou rios e córregos, retornando para o oceano e reiniciando o ciclo (MAIDMENT, 1993).

As águas subterrâneas são aquelas que se encontram sob a superfície da Terra. O caminho percorrido pela água ao infiltrar passa por uma zona não saturada, caracterizada pela

presença de vazios entre os grãos dos solos, rochas e fissuras, e posteriormente atinge uma zona saturada, onde todos os interstícios são preenchidos com água sob pressão hidrostática. O nível da água ou lençol freático é definido pelo limite entre essas duas zonas (TODD & MAYS, 2005).

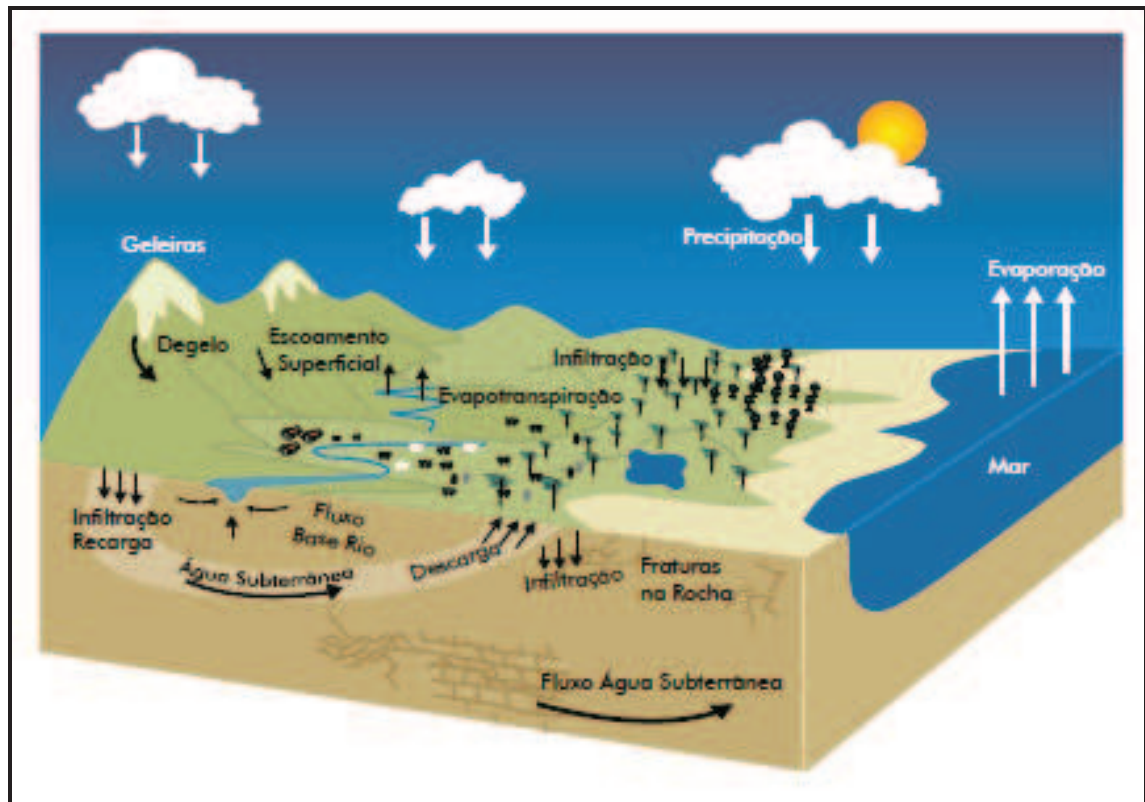


Figura 3.1: Ciclo Hidrológico
Fonte: MMA (2007)

3.1.2 Tipos de Aquíferos

Os aquíferos têm como características fundamentais a capacidade de armazenamento e capacidade de escoamento da água subterrânea (FOSTER, 2003). Eles podem ser classificados de acordo com as características hidrodinâmicas ou geológicas.

De acordo com as características hidrodinâmicas, os aquíferos são classificados em livres, confinados ou semiconfinados (Figura 3.2). Os aquíferos livres ou não confinados são aqueles que estão submetidos à pressão atmosférica e seu limite superior é o lençol freático, possuindo, geralmente, a área do aquífero como área de recarga. Aquíferos confinados são caracterizados pelo escoamento subterrâneo entre duas camadas impermeáveis (não drenante) ou semipermeáveis (drenante) e sua pressão no topo é maior que a pressão atmosférica, ou

seja, havendo uma perfuração o nível d'água se eleva até a superfície piezométrica. Os aquíferos semiconfinados são parcialmente livres e confinados (TUCCI & CABRAL, 2003).

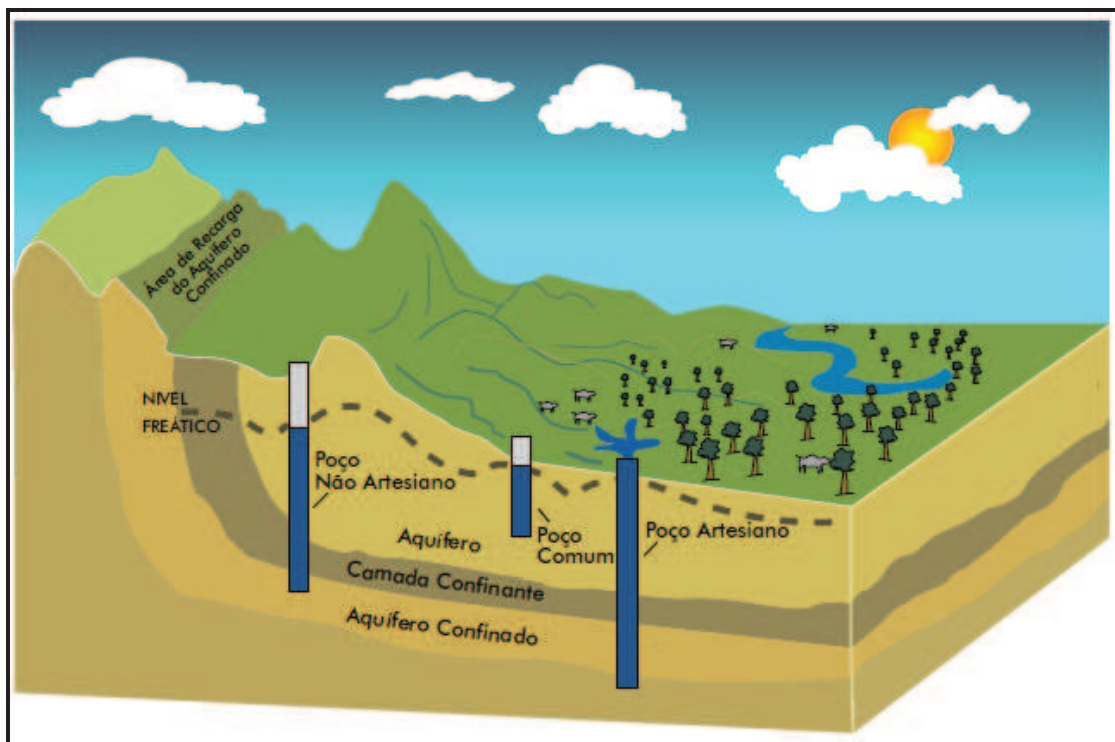


Figura 3.2: Tipos de aquíferos
Fonte: MMA (2007)

Ao analisar as características geológicas dos aquíferos, pode-se classificá-los quanto ao tipo de espaços vazios em: poroso, fissural ou cárstico (MMA, 2007). Os aquíferos podem ocorrer em camadas sedimentares ou em rochas ígneas ou metamórficas. De acordo com Tucci e Cabral (2003) os tipos de camadas mais comuns são:

- Aluvião: sedimentos não consolidados formados pelos sedimentos erodidos e transportados nas chuvas torrenciais, localizam-se nos locais favoráveis à recarga.
- Rocha sedimentar (arenito): formados pela compactação e cimentação das areias, possuindo boa condutividade hidráulica e boas condições de armazenamento.
- Rochas carbonáticas: ocorrem nas formas de calcário e dolomitas, sua significativa condutividade hidráulica é consequência de fraturas resultantes de movimentos tectônicos. Ao longo do tempo a água subterrânea vai dissolvendo o calcário e as dolomitas, formando um aquífero cárstico.

- Cristalino: formado por rochas ígneas e metamórficas, conseqüentemente, possuem uma porosidade muito baixa, provendo o acúmulo de água em regiões com rocha fraturada (aquífero fissural).

3.1.3 Exploração de Aquíferos

Existe dificuldade para definir o limiar entre exploração intensiva e superexploração das águas de um aquífero, considerando que existem muitos elementos envolvidos (*e.g.*, balanço hídrico, qualidade da água, econômicos, ambientais, administrativos e sociais). Quando o volume de exploração é menor, porém próximo ao de recarga, tem-se uma exploração intensiva. Esse modo pode ser sustentável se outras questões não reduzirem essa diferença. A superexploração considera que existem mais efeitos negativos e assume que a exploração é muito próxima ou excede a recarga. Quando a exploração excede demasiadamente a recarga, pode ser classificado como mineração da água (UNESCO, 2004).

Quando a exploração da água subterrânea é maior que a recarga, torna-se difícil atingir uma situação estável sem reduzir o bombeamento. A continuidade desse processo pode causar a diminuição do fluxo de água e dos níveis da água até cessar. Geralmente os fatores que fazem diminuir a o bombeamento de uma região são o limite físico para a depleção, aumento da salinidade ou indesejáveis problemas químicos (UNESCO, 2004).

Portanto, observa-se que a exploração diminui apenas quando as características locais inviabilizam a retirada da água, ou seja, as questões ambientais e conseqüências futuras são deixadas de lado. Uma solução para este tipo de prática é a intervenção pública, a fim de impor limites e mostrar diretrizes para a exploração em cada área a ser analisada.

3.2 MÉTODOS DE INTERPOLAÇÃO ESPACIAL

A interpolação espacial permite criar, a partir de pontos amostrais de um fenômeno, uma superfície que representa sua variação espacial em uma determinada área. O desafio dos modelos de interpolação é gerar a superfície mais acurada a partir de um grupo de dados amostrais.

As técnicas de interpolação espacial podem ser divididas em dois grupos: determinísticos e geoestatísticos. Os métodos determinísticos são baseados em funções

matemáticas que levam em consideração a vizinhança das amostras (*e.g.*, *Inverse Distance Weighted*) ou o grau de suavização da superfície (*e.g.*, *Spline* e *Global polynomial*). Os métodos geoestatísticos (*e.g.*, *Krigagem*) se baseiam no princípio da autocorrelação espacial, ou seja, consideram o grau de dependência entre vizinhos próximos e distantes (JOHNSTON *et al.*, 2001).

3.2.1 *Inverse Distance Weighted* (IDW)

O método IDW considera que pontos amostrais próximos tem mais peso que outros mais distantes. Para prever o valor de um local desconhecido, o IDW utiliza os valores amostrais da vizinhança, ponderando-os pela distância. Sua fórmula geral é dada por:

$$\hat{Z}(s_0) = \sum_{i=1}^N \lambda_i Z(s_i)$$

Onde:

$\hat{Z}(s_0)$ é o valor a ser predito no local s_0 ,

N é o número de valores amostrais conhecidos utilizados para a predição,

λ_i são os pesos de cada ponto, e

$Z(s_i)$ são os valores observados no ponto s_i .

Os pesos são determinados pela seguinte fórmula:

$$\lambda_i = \frac{d_{i0}^{-p}}{\sum_{i=1}^N d_{i0}^{-p}} \quad \sum_{i=1}^N \lambda_i = 1$$

Onde:

d_{i0}^{-p} é a distância entre o local de predição (s_0) e cada ponto observado (s_i), e

p é o parâmetro potência.

O parâmetro potência (p) é o principal fator que influencia na ponderação da locação a ser mensurada e, consequentemente, em sua acurácia (ISAACS e SRIVASTAVA, 1989). Os pesos são proporcionais ao inverso da distância, ou seja, quanto maior o parâmetro p mais rápido será o decrescimento do peso em função do aumento da distância. Os parâmetros potência 1, 2, 3, 4, 5 e 6 são os mais utilizados na literatura (KRAVCHENCO & BULLOCK, 1999; BURROUGH & MACDONNELL, 1998).

O IDW é geralmente utilizado na interpolação de dados hidrometeorológicos, como aplicou Collischonn (2001) e Dobesch *et al.* (2007). Porém, o fato de o IDW assumir que a superfície sofre uma variação local e passa por todos os pontos amostrais, torna-o sensível a presença de *outliers* de *clusters*, criando “*bulls eyes*” que são picos na superfície criada.

3.2.2 Spline

O *Spline* também é um método determinístico exato, ou seja, a superfície criada passa pelos pontos amostrais. Diferente do IDW, o *Spline* procura suavizar as superfícies e não considera a distância entre os pares amostrais. A Figura 3.3 demonstra a diferença na geração das duas superfícies.

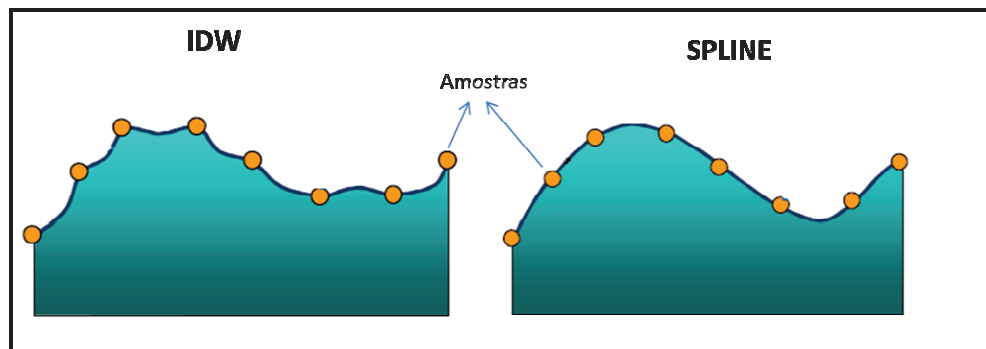


Figura 3.3: Diferença nas superfícies geradas por IDW e *spline*
Fonte: Adaptado de Johnston *et al.* (2001)

Existem cinco tipos diferentes de *splines*: *thin-plate spline*, *spline with tension*, *completely regularized spline*, *multiquadric function*, e *inverse multiquadric spline*. O *spline* deve ser utilizado para calcular superfícies suaves e com um grande número de amostras, como o nível da água. Particularmente, o *thin-plate spline* vem possuindo bons desempenhos ao interpolar dados pluviométricos e de elevação. Porém, aplicação do *spline* é inapropriada quando se tem pontos amostrais muito próximos e superfícies muito inclinadas (JOHNSTON *et al.*, 2001).

3.2.3 Métodos Geoestatísticos

A krigagem é um método geoestatístico baseado no princípio da autocorrelação espacial, ou seja, considera o grau de dependência entre vizinhos próximos e distantes. A autocorrelação espacial é verificada pelo ajuste de semivariogramas teóricos representados por uma curva (*e.g.*, Gauss, exponencial, circular, esférica etc.) (ISAACS e SRIVASTAVA, 1989). A vantagem de usar esta técnica é que suas superfícies tendem a ser mais suaves que as criadas por métodos determinísticos de interpolação, além de permitir ter um controle maior dos resultados e, conseqüentemente, dos erros gerados.

O processo para a krigagem pode ser dividido em três passos:

- Análise exploratória dos dados
- Modelagem do semivariograma
- Definição da vizinhança

3.2.3.1 Análise Exploratória

A análise exploratória permite investigar os dados de diferentes maneiras antes de gerar uma superfície. Ela possibilita um conhecimento mais profundo a cerca do fenômeno estudado, facilitando a tomada de decisão ao modelar a estrutura do variograma. Suas etapas consistem em: analisar a distribuição de frequência dos dados, identificar discrepâncias globais ou locais (*outliers*) e identificar tendências globais.

A krigagem, como um modelo de predição, tem melhor desempenho com dados que tenham uma distribuição normal. Ela também se baseia no pressuposto que todos os erros aleatórios são estacionários de segunda ordem, ou seja, têm média igual a zero e a covariância entre quaisquer dois erros aleatórios depende apenas da distância e direção que os separa. Portanto, o uso de transformações e remoção de tendências corrobora com a suposição de que os dados estão normalizados e são considerados estacionários (JOHNSTON *et al.*, 2001).

A partir da análise do histograma dos dados em estudo, é possível definir a transformação a ser aplicada de modo que sua distribuição tenda a se normalizar. Os métodos de krigagem universal e ordinária utilizam transformações do tipo Box-Cox e Lognormal. A Figura 3.4 apresenta histogramas de um mesmo conjunto de dados com e sem transformação. Pode-se notar que o histograma com a transformação Box-cox possui uma distribuição de

frequências que se aproxima mais de uma distribuição normal, diferentemente do histograma sem transformação.

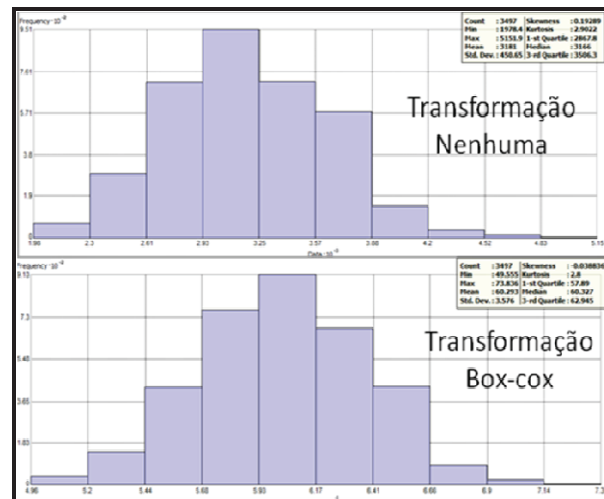


Figura 3.4: Uso do histograma para definir a transformação

A identificação de tendências globais é feita a partir da perspectiva tridimensional dos pontos amostrais, projetando-os em dois planos para visualizar as curvas de tendência (Figura 3.5). Em análises geoestatísticas, é necessária a remoção dessas curvas através da utilização de uma função polinomial de certo grau, a fim de minimizar os desvios quadráticos em relação aos valores observados e admitir a suposição de que os dados são estacionários.

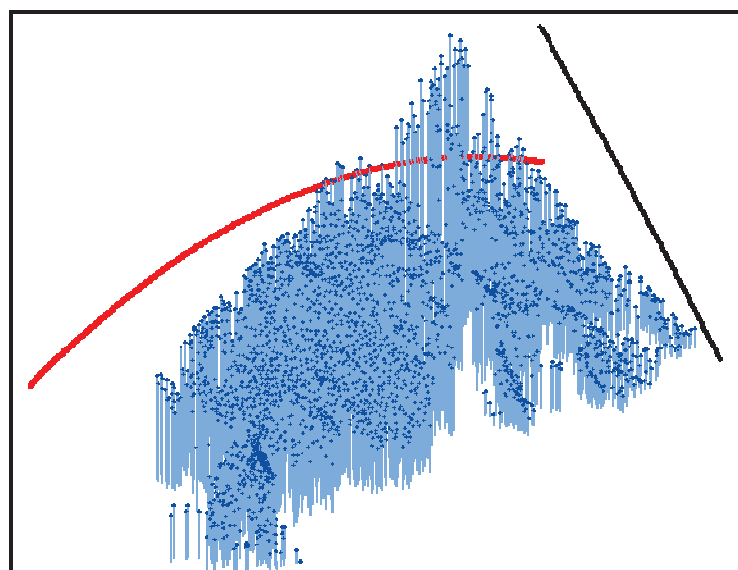


Figura 3.5: Representação 3D do nível dos poços do aquífero Ogallala Identificação de tendências globais (observado no ano 2000)

Por fim, o último passo da análise exploratória é procurar discrepâncias (*outliers*) globais ou locais. Ao se identificar a presença de *outliers* é necessário investigar se o valor é decorrente de um erro de implantação ou se o determinado fenômeno pode mesmo ter ocorrido. Os *outliers* podem ser encontrados analisando o histograma ou o semivariograma, porém, para facilitar essa busca, utilizam-se polígonos de Thiessen, como pode ser visto na Figura 3.6, classificando as amostras de acordo com seus valores amostrais, *e.g.*, entropia, *cluster*, desvio padrão.

Os polígonos de Thiessen com amostras de entropia e *cluster* são os métodos utilizados para identificar possíveis *outliers*. A entropia proporciona uma medida da dissimilaridade entre células vizinhas, considerando que na natureza se espera que as células mais próximas sejam mais semelhantes que células mais distantes, valores com alta entropia podem significar um *outlier*. O *cluster* analisa os cinco vizinhos da célula, espera-se que pelo menos um vizinho tenha características similares, caso contrário será um possível *outlier* (JOHNSTON *et al.*, 2001).

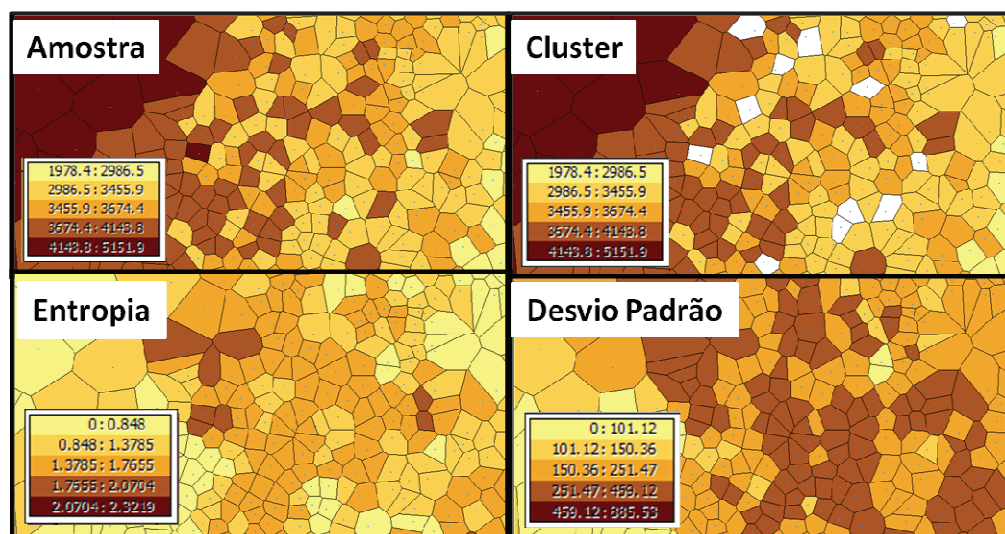


Figura 3.6: Identificação de outliers por *Cluster* e Entropia

3.2.3.2 Modelagem do Semivariograma

De posse dos resultados da análise exploratória, pode-se inferi-los na configuração da Krigagem. A definição do semivariograma e a escolha da vizinhança são as etapas mais sensíveis e importantes em interpolações geoestatísticas. O semivariograma é uma ferramenta que verifica a autocorrelação espacial dos dados, ou seja, ele é utilizado para investigar a

correlação entre pares de amostras de acordo com a direção e distância a serem arbitrados. O semivariograma clássico (MATHERON, 1962) é estimado pela seguinte equação:

$$\gamma(h) = \frac{1}{2N} \sum_{i=1}^N [Z(t_i + h) - Z(t_i)]^2$$

Sendo:

$\gamma(h)$ – Função semivariograma,

N – Número de pares das amostras, e

$Z(t_i + h)$, $Z(t_i)$ – Valores do por de amostras separados por um vetor h .

O gráfico com os pontos da função semivariograma plotados é chamado de semivariograma empírico (Figura 3.7). Ele representa a autocorrelação espacial e seu formato está diretamente ligado a fatores como o tamanho e o número de *lags* (distância em que os pares de amostras do semivariograma poderão ser agrupados), a direção de busca dos pares amostrais ou a anisotropia. No exemplo do gráfico da Figura 3.7, o semivariograma tem início com pontos bem próximos, o que indica que existe razoável autocorrelação espacial a pequenas distâncias e à medida que a distância aumenta a autocorrelação vai diminuindo, ou seja, os pontos vão se distanciando uns dos outros. Esse comportamento é de se esperar, já que na natureza se tem o pressuposto que fenômenos próximos tendem a ser mais semelhantes que os distantes.

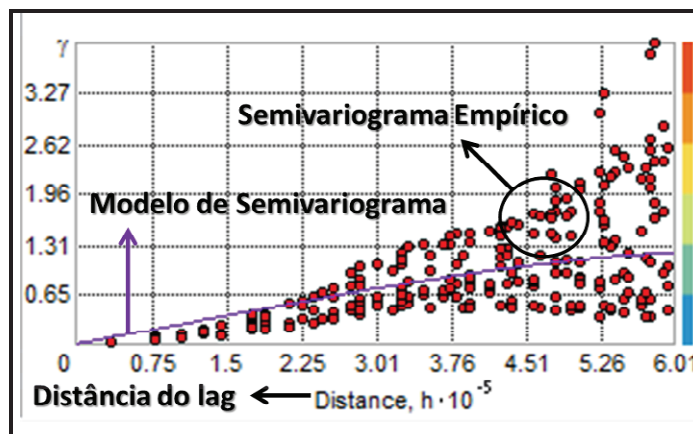


Figura 3.7: Definição do modelo de semivariograma

O tamanho do *lag* representa um limite de distâncias em que os pares amostrais podem ser agrupados, reduzindo o número de combinações possíveis. Reduzir o tamanho do *lag* significa diminuir o raio de busca por pares amostrais e, conseqüentemente, definir um

modelo com características de autocorrelação mais localizada. Portanto, tamanho do *lag* e o número de *lags* tem um papel importante no ajuste, por exemplo, valores muito altos podem mascarar pequenas variações na autocorrelação, enquanto que valores muito baixos podem causar picos inapropriados na superfície ou uma distância insuficiente entre pares de amostras, o que resulta em um semivariograma sem valores plotados.

A partir da criação do semivariograma empírico é possível definir o modelo de semivariograma (Figura 3.7). Entre os diversos modelos de semivariograma, destacam-se: Circular, Esférico, Tetraesférico, Pentaesférico, Exponencial, Gaussiano, Quadrático Racional, *Hole Effect*, *K-Bessel*, *J-Bessel* e *Stable* (JOHNSTON *et al.*, 2001). A definição do modelo de semivariograma se baseia em definir qual curva se adapta melhor ao semivariograma empírico e produz melhores resultados na validação cruzada.

Algumas vezes os dados amostrais possuem uma influência direcional que pode ser quantificada estatisticamente, mesmo aparentemente não existindo razões físicas para isso. Essa influência direcional é um fenômeno conhecido como anisotropia e pode melhorar o desempenho de alguns modelos de semivariograma (JOHNSTON *et al.*, 2001).

Durante a definição do modelo de semivariograma também pode ser necessário ajustar o efeito pepita (*Nugget*). Ele representa o valor que o modelo de semivariograma deve interceptar o eixo Y do gráfico. Esse efeito é comumente adotado como sendo zero (*e.g.*, $\gamma(h=0)=0$), porém, às vezes para uma distância infinitesimalmente pequena a diferença entre amostras não tende para zero, sendo, portanto, necessário o ajuste.

3.2.3.3 Definição da Vizinhança

A definição da vizinhança consiste em definir um polígono, geralmente de forma circular ou elipsoidal, onde estarão contidos os dados que poderão ser utilizados na interpolação. O tamanho do polígono é proporcional à distância do *lag*, porém é possível definir os setores em que serão procurados os vizinhos (Figura 3.8).

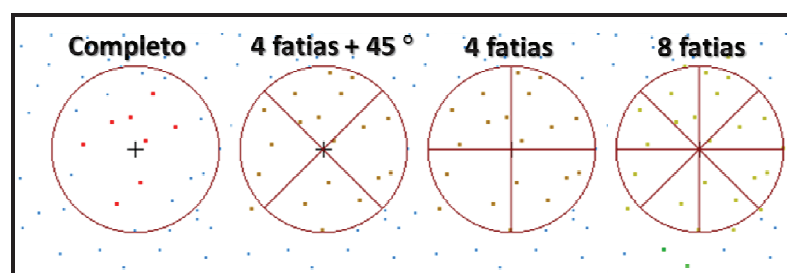


Figura 3.8: Definição da vizinhança

A escolha dos setores tem sua importância por ser um parâmetro bastante sensível e definir ângulos de visada nos quais serão procurados os vizinhos. Aliado a isto, deve ser definida a quantidade de vizinhos máxima e mínima que será utilizada em cada fatia do setor definido. Portanto, ao escolher o setor do tipo quatro fatias e definir que serão procurados no máximo oito vizinhos e no mínimo duas em cada fatia, ou seja, no total poderão ser escolhidos no máximo 32 vizinhos e no mínimo oito para interpolar uma célula. Avaliando a mesma situação para o setor do tipo completo seriam procurados no máximo oito vizinhos e no mínimo duas para toda a região definida pelo polígono.

3.2.3.4 Krigagem Ordinária

De acordo com Matheron (1971), a krigagem ordinária se baseia na seguinte suposição:

$$Z(s) = \mu + \varepsilon(s)$$

Onde:

$Z(s)$ é o valor a ser predito,

μ é uma constante desconhecida, e

$\varepsilon(s)$ são erros aleatórios.

Uma das principais questões sobre krigagem ordinária é se o pressuposto de uma média constante é razoável. Porém, como um método de predição geoestatística simples, ele tem uma flexibilidade notável (CRESSIE, 1993). A Figura 3.9 mostra a aplicação do conceito da krigagem ordinária em uma dimensão, onde a linha pontilhada indica a posição da constante μ de valor desconhecido e $\varepsilon(s)$ os erros aleatórios que possuem média igual a zero.

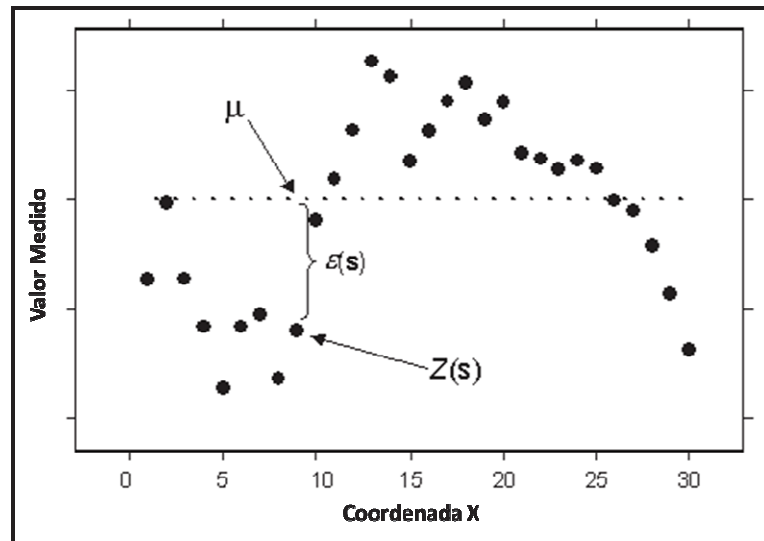


Figura 3.9: Conceito da krigagem ordinária
 Fonte: Adaptado de Johnston *et al.* (2001)

A krigagem ordinária utiliza semivariogramas ou covariâncias para expressar a correlação entre os dados. Além disso, podem ser utilizadas transformações e remoção de tendências (JOHNSTON *et al.*, 2001).

3.2.3.5 Krigagem Universal

A krigagem universal se baseia na mesma suposição vista na krigagem ordinária, porém μ passa a ser uma função determinística, sendo assim:

$$Z(s) = \mu(s) + \varepsilon(s)$$

Onde:

$Z(s)$ é o valor a ser predito,

$\mu(s)$ é uma função determinística, e

$\varepsilon(s)$ são erros aleatórios.

A Figura 3.10 indica a aplicação do conceito da krigagem universal em uma dimensão, onde a linha pontilhada representa a função $\mu(s)$ que é um polinômio de segunda ordem. Ao subtrair os valores medidos pela função $\mu(s)$, tem-se os erros, que são tidos como aleatórios e possuem média igual à zero.

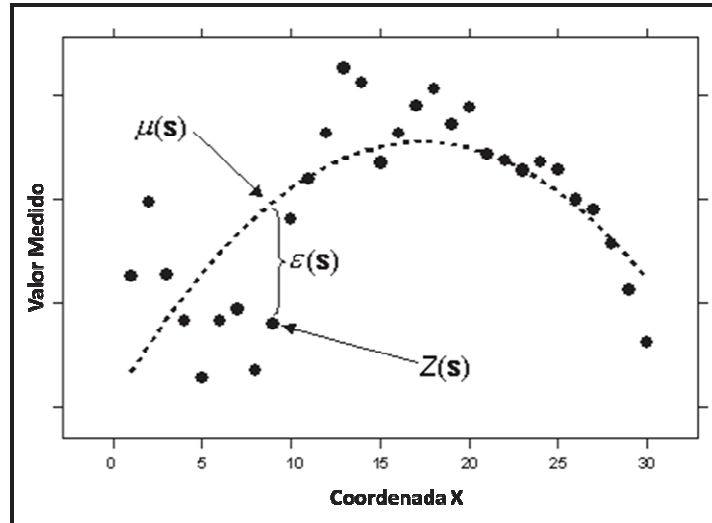


Figura 3.10: Conceito da krigagem universal
Fonte: Adaptado de Johnston *et al.* (2001)

Conceitualmente a autocorrelação é modelada a partir dos erros aleatórios $\varepsilon(s)$ (CRESSIE, 1993). A krigagem universal também utiliza semivariogramas ou covariâncias para expressar a correlação entre os dados. Além disso, podem ser utilizadas transformações e remoção de tendências (JOHNSTON *et al.*, 2001).

3.3 VALIDAÇÃO CRUZADA

A validação cruzada tem o objetivo de fornecer resultados quantitativos para que se possa avaliar qualitativamente o resultado de uma interpolação. Em outras palavras, ela permite avaliar o quanto a superfície criada se aproxima da realidade ou o quanto o modelo prediz valores desconhecidos. Portanto, a validação cruzada é útil para definir qual a melhor configuração dos parâmetros de um modelo para a predição de uma superfície (*e.g.*, krigagem) ou qual método de interpolação se adapta melhor a um determinado conjunto de dados.

De um modo geral, a validação cruzada consiste em análises estatísticas geradas a partir do erro entre amostras e a superfície modelada. De acordo com Webster e Oliver (2007), o diagnóstico estatístico pode ser obtido pelo resultado das seguintes expressões:

- Erro Médio (EM)

$$EM = \frac{1}{N} \sum_{i=1}^N [z(x_i) - \hat{Z}(x_i)]$$

- Raiz do erro médio quadrático (RMSE - *Root Mean Square Error*)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N [z(x_i) - \hat{Z}(x_i)]^2}$$

Sendo:

N é o número de amostras,

$z(x_i)$ é o valor da amostra i , e

$\hat{Z}(x_i)$ é o valor predito pela superfície correspondente a amostra i .

O valor ideal para o erro médio deve ser zero, enquanto o RMSE deve ser o menor valor possível.

Existem diversas formas de realizar a validação cruzada dos dados, sendo três métodos os mais utilizados: *holdout*, *k-fold* e *leave-one-out* (KOHAVI, 1995).

3.3.1 Método *Leave-one-out*

O método *leave-one-out* é um caso específico do *k-fold*. Nesse método, para todos os pontos da validação cruzada sequencialmente se omite um ponto, calcula seu valor considerando os pontos de sua vizinhança e depois compara o valor predito com o amostral, gerando um erro. Esse processo é repetido para todos os outros pontos e, por fim, são aplicadas as equações de análise estatística (RENNIE, 2003).

3.3.1 Método *Holdout*

O método *holdout* consiste em validar a superfície gerada utilizando um conjunto de pontos que não foram empregados na interpolação. Portanto, para a aplicação deste método é necessário dois conjuntos de dados: dados de treinamento (usados para interpolar) e dados de teste (usados na validação). O processo da validação cruzada verifica o erro que cada ponto dos dados de teste possui em relação à superfície gerada, então são aplicadas as equações de análise estatística (KOHAVI, 1995).

3.5 ANÁLISE DE *CLUSTER*

A mineração de dados (*Data Mining*), onde está inserida a análise de *cluster*, baseia-se na análise de base de dados (frequentemente grandes) para encontrar padrões e agrupá-los em grupos mais úteis para o usuário (HAND, 2001). A análise de *cluster* é uma técnica que busca agrupar um conjunto de dados se baseando na similaridade entre eles. Ela tem por objetivo proporcionar partições em uma população de dados heterogênea, formando subgrupos mais homogêneos.

O objetivo desse método é ajudar o usuário a entender a estrutura natural em um conjunto de dados. A análise de *cluster* é um processo de classificação não supervisionado que é fundamental para a mineração de dados, já que ela analisa como os objetos são agrupados e quais podem ser considerados distantes de um grupo natural (ZAIANE et al., 2003). Segundo Zaiane *et al.* (2003), uma análise *cluster* criteriosa deve possuir as seguintes características:

- **Adaptável a várias dimensões:** A análise *cluster* deve possuir a habilidade de trabalhar com dados multidimensionais, ou seja, com muitas variáveis.
- **Versatilidade:** A análise *cluster* pode se adequar a diferentes tipos de dados, *e.g.*, numéricos, booleanos ou por categorias.
- **Robusto a presença de ruídos:** Uma boa técnica de *cluster* deve possuir um bom desempenho mesmo com a presença de ruídos.
- **Indiferente ao ordenamento dos dados de entrada:** A análise *cluster* consegue produzir o mesmo resultado independente da ordem que os dados são apresentados.

Uma premissa básica dos métodos de análise de *cluster* é que os objetos pertencentes ao mesmo *cluster* são mais próximos que objetos de outros grupos. Portanto, a maioria das técnicas de análise de *cluster* é baseada em uma medida de similaridade entre os elementos a serem agrupados. Normalmente, para dados contínuos, o grau de similaridade entre objetos é medido por uma função distância, onde a mais comumente utilizada é a distância Euclidiana (YAN, 2005).

As medidas de similaridade para dados quantitativos multivariados (numéricos, booleanos ou por categorias) são habitualmente calculadas pela distância d_{ij} entre os indivíduos i e j , construídas a partir de valores de p variáveis, traduzidos nos vetores $x(i)$ e

$x(j)$. Estes vetores correspondem a linhas de uma matriz $X_{n \times p}$ de dados, mas seguindo a convenção habitual para vetores, serão considerados vetores coluna. As distâncias, Euclidiana, de Minkowski e de Manhattan são exemplos típicos de medida de distância entre indivíduos (CADIMA, 2010).

A distância Euclidiana usual em $\mathfrak{R}^p$ é dada por:

$$d_{ij} = \|x_{(i)} - x_{(j)}\| = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

Onde:

$x_{(i)}$ é o vetor do i -ésimo indivíduo

$x_{(j)}$ é o vetor do j -ésimo indivíduo

p é o número de variáveis ou dimensões

Existem dois grandes grupos de métodos de análise de *cluster*: métodos hierárquicos e por particionamento.

3.6.1 Métodos por particionamento

Os métodos por particionamento, ou não hierárquicos, foram desenvolvidos para agrupar n elementos em K grupos definidos previamente, onde cada grupo representa um *cluster* contendo pelo menos um elemento e, considerando $K \leq n$. A quantidade de grupos pode ser dividida, a priori, por meio de um conhecimento que se tenha dos dados ou, a posteriori, com base nos resultados da análise (BARROSO & ARTES, 2003).

O método de particionamento mais amplamente utilizado é o *k-means* criado por MacQueen (1967), porém outros métodos também implementam a técnica de *cluster* por particionamento, *e.g.*, *k-medoids*, *fuzzy c-means* e *fuzzy c-medoids* (LIAO, 2005).

A principal idéia por trás do método *k-means* é a minimização de uma função objetivo, dada pela distância total entre todos os elementos de um grupo e seu respectivo centroide (soma dos quadrados). A solução se baseia em um esquema iterativo, que começa com um número arbitrário de elementos em cada k grupo. A distribuição dos elementos vai se alternando entre os grupos, alterando também o centroide dos novos agrupamentos. O processo iterativo acaba quando o valor da função objetivo não consegue mais ser reduzido (MacQUEEN, 1967).

3.6.2 Métodos hierárquicos

O método hierárquico de *cluster* consiste em uma série de sucessivos agrupamentos geralmente representados por um diagrama bidimensional chamado de dendrograma ou diagrama de árvore. No dendrograma (Figura 3.11), as amostras (ramos) representam os elementos a serem agrupados e a raiz representa o agrupamento de todos os elementos.

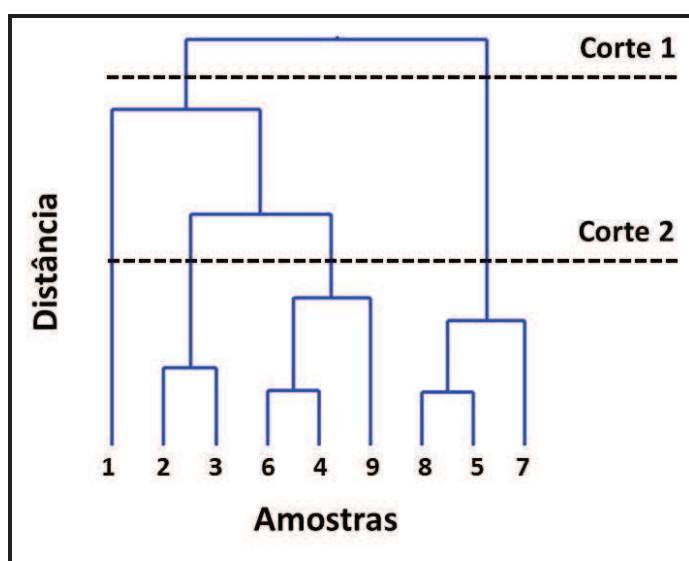


Figura 3.11: Exemplo de Dendrograma

Por meio da análise do dendrograma e do conhecimento prévio da distribuição dos dados, deve-se determinar uma distância de corte, definindo a formação dos grupos. Essa decisão depende do objetivo da análise e da quantidade de grupos que se deseja formar, sendo, portanto, subjetiva. Sabe-se que quanto maior a distância do corte, maior será a dissimilaridade entre os elementos agrupados. No dendrograma da Figura 3.11, pode-se verificar a presença de dois diferentes cortes. No corte 1 são formados dois grupos: (1,2,3,6,4,9) e (8,7,5). No corte 2, o número de grupos aumenta para quatro, sendo: (1), (2,3), (6,4,9) e (8,5,7).

Os métodos hierárquicos são subdivididos em métodos aglomerativos e divisivos. No método aglomerativo, os elementos iniciam-se separados e são agrupados de acordo com sua similaridade, por sua vez os grupos são fundidos pela similaridade entre classes, ocorrendo este procedimento sucessivamente até formar um grupo único com todos os elementos. No

método divisivo, ocorre de maneira análoga, mas no sentido inverso, *e.g.*, começa-se pela classe com a totalidade de indivíduos em um único grupo, desagregando-os em subgrupos considerados mais homogêneos até se obter todos os elementos dispostos separadamente (NORUŠIS, 2011). Segundo Cadima (2010), o método divisivo não é comumente utilizado devido a maior complexidade computacional do procedimento.

O método aglomerativo ou de agregação de classes mede a semelhança entre grupos de forma a realizar ou não fusões. Cadima (2010) descreve alguns métodos de agregação entre classes:

- **Vizinho mais próximo (*nearest neighbour* ou *single linkage*):** Considera que a distância entre dois subgrupos é a menor distância entre um elemento de um grupo e um elemento do outro grupo.
- **Vizinho mais distante (*furthest neighbour* ou *complete linkage*):** Considera que a distância entre dois subgrupos é a maior distância entre um elemento de um grupo e um elemento do outro grupo.
- **Método das Distâncias Médias entre Grupos (*group average* ou *average linkage*):** Considera que a distância entre duas classes é a média de todas as distâncias entre pares de elementos.
- **Método da Inércia Mínima (Ward):** Considera a soma de quadrados das diferenças entre cada indivíduo e a média dos indivíduos de uma classe. A distância entre classes é o aumento na soma total das inércias provocado pela fusão entre elas.
- **Método dos Centroides:** Toma-se a distância entre duas classes como sendo a distância entre seus os centroides.

Diante do processo de hierarquização dos dados por *cluster*, foi criado um gráfico que representa as células da matriz por cores, que são baseadas na similaridade de seus valores, calculada por meio de uma função distância normalizada. O objetivo deste produto é facilitar a inspeção do reordenamento de linhas e colunas em matrizes extensas (centenas de linhas e colunas). Este gráfico é chamado de *heat map* (Figura 3.12).

O *heat map* é mais comumente aplicado na área de biologia e biomedicina. Segundo Weinstein (2008), cerca de dez anos após a primeira publicação nessas áreas, apareceram mais de 4000 trabalhos com análises por *cluster heat maps*.

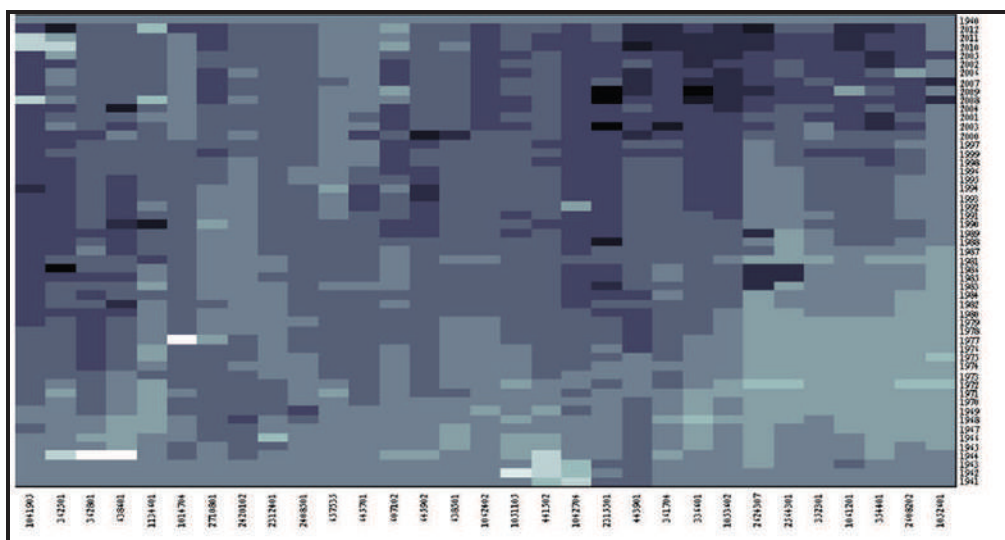


Figura 3.12: Exemplo de um *Heat map*

Os *heat maps* são tipicamente usados na biologia molecular para representar o nível de expressão de muitos genes (multivariáveis) sobre amostras a serem comparadas. Uma analogia a este uso pode ser feita para aplicar em séries temporais de nível da água, os genes são representados pelos anos da série histórica enquanto as amostras são os poços em análise.

O *cluster heat map* ou clustergrama é uma representação visual que simultaneamente representa um *heat map* e um dendrograma (Figura 3.13). Nesse tipo de gráfico, as relações do método de *cluster* hierárquico são indicadas nos dendrogramas que ordenam linhas e colunas, gerando um gráfico que permite extrair informações de grandes quantidades de dados em um pequeno espaço.

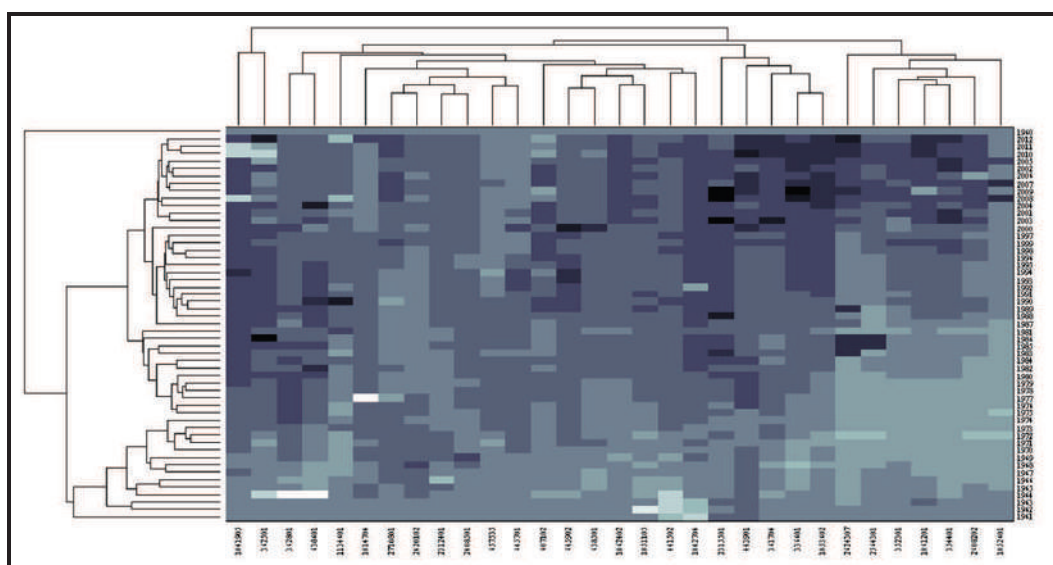


Figura 3.13: Exemplo de um clustergrama

3.6 TESTE DE MANN-KENDALL

O teste de Mann-Kendall é um teste não paramétrico para detecção de tendência. De acordo com Makridakis *et al.* (1998), existe um padrão de tendência em uma série quando ocorre uma variação do nível médio no longo prazo. O teste não paramétrico é útil em situações em que se deseja menosprezar suposições básicas (parâmetros) requeridas para testar uma hipótese, *e.g.*, despreza-se a influência de variáveis externas no comportamento da série. Esse teste assume que os dados possuem uma distribuição independente, ou seja, não existem parâmetros a serem relacionados com ela. O teste não paramétrico foi desenvolvido para ser utilizado na avaliação de impactos ambientais, já que os dados disponíveis são geralmente mal-arranjados para se adotar procedimentos paramétricos. Sua única restrição é que as amostras devem vir da mesma população básica (HIPEL & McLEOD, 1994).

Testes estatísticos, sejam paramétricos ou não paramétricos, devem ser baseados em testes de hipóteses e de significância. Esses testes servem para certificar se um conjunto de dados possui uma determinada característica, *e.g.*, para assegurar a existência ou não de tendência em uma série histórica de nível da água de um aquífero. A teoria do teste de hipóteses foi originalmente desenvolvida por Neyman e Pearson (1928, 1933) e o teste de significância é, em grande parte, devido a Fisher (1973). Cox e Hinkley (1974) apresentaram descrições detalhadas de vários tipos de testes de hipóteses e de significância, enquanto Hipel e McLeod (1994) contextualizaram sua aplicação no teste de Mann-Kendall.

O teste de hipóteses analisa duas hipóteses. A hipótese nula (H_0), que considera que a população de um conjunto de amostras não possui uma propriedade específica, como uma tendência, e a hipótese alternativa (H_1) que indica a presença da característica específica em análise (HIPEL & McLEOD, 1994).

O teste de Mann-Kendall não sazonal determina se uma série temporal possui tendência ao longo do tempo. O teste de tendência não paramétrico apresentado por Mann (1945), avaliava a hipótese de membros de certa sequência de variáveis aleatórias possuírem distribuição idêntica e independente umas das outras. Isto é um caso particular da aplicação do teste de Kendall (1975) para correlação, mais conhecido como teste de Mann-Kendall. Mann (1945) propôs testar a hipótese nula (H_0) para dados de uma população em que variáveis aleatórias são independente umas das outras. A hipótese alternativa (H_1) é que os dados seguem uma tendência ao longo do tempo.

A escolha do nível de significância define a probabilidade das amostras em estudo falharem devido à distribuição dos dados, tornando a hipótese H_0 aceita, *e.g.*, caso o p-valor obtido pelo teste de Mann-Kendall não atenda o valor do nível de significância (0,10, 0,05 ou 0,01). Se o p-valor for menor que o nível de significância adotado, é rejeitada a hipótese H_0 e adotada a hipótese H_1 , havendo, portanto, tendência.

Para identificar se a tendência em uma série é crescente ou decrescente, é analisado o coeficiente τ de Kendall. Este coeficiente é obtido pela razão:

$$\tau = \frac{S}{D}$$

Onde:

S é um valor assintótico normalmente distribuído,

D é o valor máximo provável de S .

O valor estatístico S é uma contagem do número de vezes que x_j excede x_k , para $j > k$, mais do que x_k excede x_j , dada uma série $x_1, x_2, \dots, x_n$, como é descrito na equação de S :

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sgn}(x_j - x_k)$$

Onde:

$$\text{sgn}(x) = \begin{cases} +1, x > 0 \\ 0, x = 0 \\ -1, x < 0 \end{cases}$$

Logo, valores positivos de S indicam uma tendência crescente, enquanto os valores negativos, decrescente. O valor máximo provável de S ocorre quando $x_1 < x_2 < \dots < x_n$. Este valor é chamado de D e foi definido por Kendall (1975) como sendo:

$$D = \left[\frac{1}{2}n(n-1) - \frac{1}{2} \sum_{j=1}^p t_j(t_j-1) \right]^{1/2} \left[\frac{1}{2}n(n-1) \right]^{1/2}$$

Onde:

n é o número de variáveis da série,

p é um número de grupos conectados dentro da série,

t_j é número de dados do j -ésimo grupo.

Portanto, τ possui valores que variam entre -1 e 1. Valores de τ negativos indicam que há uma tendência decrescente, já os valores positivos de tau indicam uma tendência crescente e valores próximos de zero indicam que não há tendência. O critério para definir que valores são estes é baseado no resultado do p-valor e do nível de significância adotado, como foi descrito anteriormente.

4. MATERIAIS E MÉTODOS

4.1 DADOS DE NÍVEL DA ÁGUA EM POÇOS

O nível de água dos poços foi obtido a partir do acesso a base de dados da *Texas Water Development Board* (TWDB, 2012). A base de dados é disponibilizada em formato Microsoft Access (.mdb), porém foi necessário converter a base para o formato nativo do *software* PostgreSQL, que é livre e possui maior robustez para realizar consultas SQL (*Structured Query Language*).

Inicialmente foram descartados os dados com o *status* de não publicáveis. Em seguida, foram extraídas da base de dados duas tabelas. A primeira (nomeada *ogallalawells*) tinha como dados: código dos poços, coordenadas, código do aquífero, nível do terreno etc. Esta tabela possuía 20.840 registros, que representavam cada poço pertencente ao Texas e ao aquífero Ogallala. A segunda tabela (nomeada *waterelev*) tinha como dados: código dos poços, profundidade para o nível da água, ano de coleta da amostra, mês de coleta da amostra etc. Esta tabela possuía 791.689 registros que entre elas existiam códigos de poços repetidos, porém, cada linha com datas de coletas diferentes.

O que se desejava era criar uma tabela com linhas que representassem apenas um poço com suas respectivas coordenadas e que a série histórica dos dados de nível da água fosse representada nas colunas, onde cada coluna seria um ano e as células seria o nível da água. Para isto, foram realizadas consultas SQL até chegar ao resultado final. Todas as consultas SQL realizadas no tratamento dos dados estão disponibilizadas no Apêndice A.

Dessa forma, o primeiro passo foi criar uma tabela (nomeada *waterelev_join*) por meio da junção tabular entre as tabelas *ogallalawells* e *waterelev*, utilizando como chave o código dos poços. Porém, fez-se necessário realizar alguns tratamentos nos dados da tabela, tais como: apagar linhas que possuíam valores de profundidades dos poços nulos, adicionar a coluna “*wte*” (nível da água) e preencher suas células correspondentes subtraindo o nível do terreno pela profundidade do poço. Por fim, a tabela passou a ter 368.254 registros e as colunas de saída ficaram com o código dos poços (com repetição de valores em mais de um registros), elevação do terreno, código do aquífero, profundidade do poço, mês de coleta, dia de coleta, ano de coleta e nível da água.

O passo seguinte foi definir os meses de coleta dos dados que seriam utilizados como nível de referência anual. De acordo com Guru e Horne (2000), durante os meses de novembro a fevereiro ocorre uma redução nas atividades agrícolas devido à chegada do inverno, recomendando-se, portanto, obter coletas de nível da água nesse período. Seguindo este princípio, foi criada uma nova tabela, excluindo os dados coletados entre os meses de março a outubro. Além disso, foi atribuída aos níveis de água coletados nos meses de novembro e dezembro a data do ano seguinte. O motivo desta alteração foi considerar o intervalo entre os meses de novembro a janeiro como sendo pertencentes ao mesmo ano, visto que, se não fosse adotado este procedimento, poderia haver até 12 meses de defasagem entre um dado coletado e outro no mesmo ano. Portanto, a nova tabela criada (`waterelev_join_exclude_3_to_10`) passou a ter 147.200 registros.

A última filtragem dos dados consistiu em obter uma tabela (nomeada `wells_timeseries`) contendo apenas três colunas: códigos dos poços, ano de coleta e nível da água. Porém, existiam em alguns poços, situações em que para o mesmo ano havia mais de um valor coletado para o nível da água, ou seja, houve coleta do nível da água mais de uma vez entre os meses de novembro a agosto. A solução adotada foi calcular o nível da água médio entre os valores coletados quando houvesse este conflito. Por fim, a tabela passou a ter 138.854 registros.

O passo final consistiu em criar uma tabela (nomeada `wells`) com cada linha representando apenas um poço e a série histórica dos dados de nível da água representados pelas colunas. A tabela passou a ter 9.292 registros (quantidade de códigos de poços distintos na tabela `wells_timeseries`) e 98 colunas (série histórica de 1914 a 2012). Por fim, valores das células foram preenchidos por meio de consultas SQL.

Após esta etapa inicial de preparação dos dados de entrada, foram exportados os dados para formatos de edição tabular (`.xlsx`) e de Sistemas de Informações Geográficas (SIG) (`.shp`). A partir da análise da distribuição espacial dos dados e da série histórica, optou-se por trabalhar com uma série histórica mais curta (1960 a 2012), porém com maior densidade de dados coletados e menos falhas, já que entre 1914 a 1959 haviam apenas 13.300 células preenchidas de um total possível de 427.432, ou seja, 3,1% de preenchimento.

Portanto, a tabela a ser utilizada para a interpolação dos dados contém 9.292 linhas (código dos poços), 53 colunas (anos da série histórica), formando uma matriz de 492.476 células. Desse valor, 126.553 células estão preenchidas com dados de nível da água, o que equivale a apenas 25,7% de células preenchidas. Por isso, fez-se necessário selecionar os

poços com as séries históricas com menos falhas para então realizar a análises de tendências dos dados. Dessa forma, foram selecionados 492 poços do universo anterior (9.292 poços), formando uma tabela com 492 linhas, 53 colunas e 23.427 células preenchidas, o que equivale a 89,84% do total de células. Portanto, estes foram os dados selecionados para realizar análises de tendências após serem preenchidas suas falhas.

A fim de qualificar e comparar o resultado das interpolações, foi necessário separar um conjunto de dados para fazer a validação cruzada pelo método *holdout*. A seleção destes dados baseou-se em uma amostragem aleatória de 200 poços contidos na tabela de análise de tendências (492 poços). Os 200 poços definidos para a validação cruzada foram retirados da tabela de interpolação, deixando-a com 9.092 poços.

A Figura 4.1 mostra a distribuição espacial dos 9.092 poços utilizados para a interpolação da série histórica e a Tabela 4.1 informa a quantidade de pontos interpolados para cada ano da série histórica.

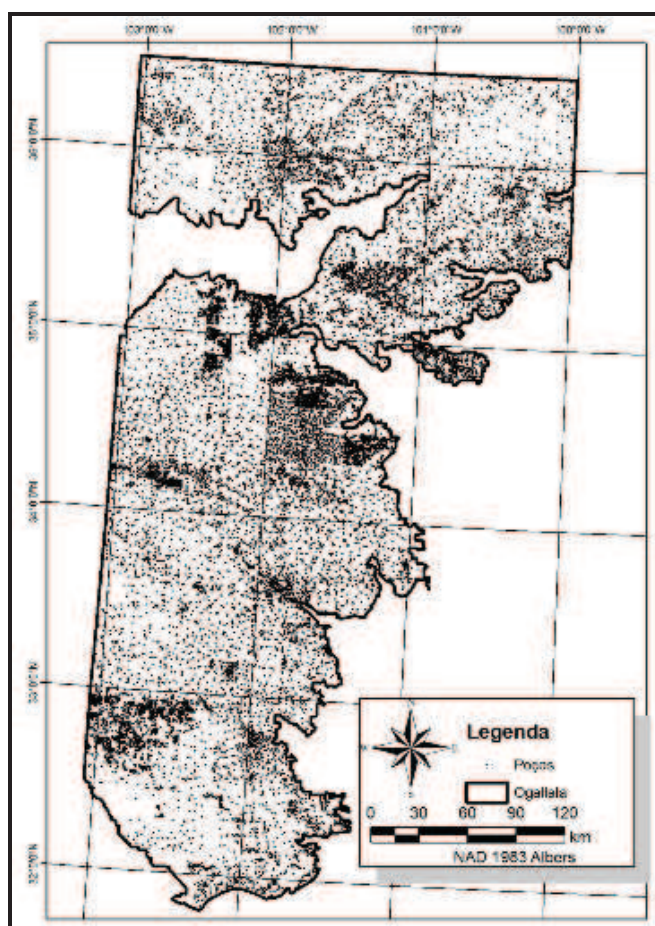


Figura 4.1: Distribuição espacial dos poços a serem interpolados

Tabela 4.1: Quantidade de poços interpolados para cada ano da série histórica

Ano	Nº de poços	Ano	Nº de poços	Ano	Nº de poços
1960	1099	1980	2951	2000	2665
1961	1418	1981	1985	2001	2257
1962	1437	1982	2353	2002	2724
1963	1574	1983	1825	2003	2414
1964	1879	1984	2183	2004	2712
1965	1672	1985	2280	2005	2690
1966	1646	1986	2076	2006	2844
1967	1715	1987	2571	2007	2671
1968	1745	1988	2536	2008	2744
1969	1554	1989	2758	2009	2822
1970	1650	1990	2323	2010	2539
1971	1685	1991	2603	2011	2780
1972	1719	1992	2649	2012	2526
1973	1727	1993	2488		
1974	1687	1994	2489		
1975	1849	1995	2490		
1976	2126	1996	2657		
1977	2254	1997	2452		
1978	1921	1998	2455		
1979	1990	1999	2158		

A Figura 4.2 mostra a distribuição espacial dos 492 poços utilizados para a análise de tendências e dos 200 poços aplicados na validação cruzada.

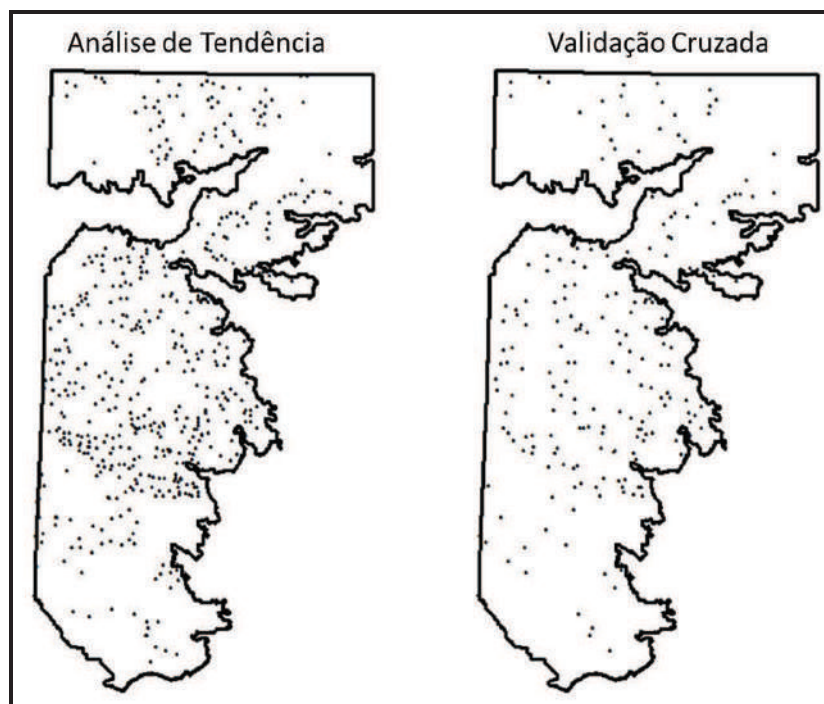


Figura 4.2: Distribuição espacial dos poços utilizados para a análise de tendências e validação cruzada

4.1.1 Preenchimento de falhas

Para a realização da análise de tendências, os 492 poços selecionados deveriam estar com séries temporais contínuas, ou seja, sem apresentar falhas. Havia um total de 2.649 falhas em uma matriz com 26.076 células (492 poços $\times$ 53 anos), o que equivale a 10,16% de falhas.

A metodologia aplicada para o preenchimento das falhas existentes se baseou no método de ponderação regional descrito por Tucci e Silveira (2004).

$$P_x = \frac{1}{3} \left(\frac{P_A}{M_A} + \frac{P_B}{M_B} + \frac{P_C}{M_C} \right) M_x$$

Onde:

P_x = Falha a ser corrigida

P_A, P_B, P_C = Níveis da água dos poços A, B, C, respectivamente

M_A, M_B, M_C = Média dos poços A, B, C, respectivamente

M_x = Média do poço a ser corrigido

Esse é um método simplificado normalmente utilizado para o preenchimento de séries mensais ou anuais de precipitações, visando sua análise estatística. Utilizam-se postos

vizinhos pertencentes a uma região climatológica semelhante ao posto a ser preenchido (TUCCI & SILVEIRA, 2004).

Dessa forma, para o preenchimento das falhas dos poços aplicou-se esta metodologia, utilizando poços semelhantes. A escolha dos poços (semelhantes) a serem utilizados na ponderação teve como princípio a análise da correlação entre os dados do poço a ser preenchido e os restantes, sendo escolhidos, a princípio, os três poços com maior correlação.

O método usualmente conhecido para medir a correlação entre duas variáveis é o coeficiente de correlação linear de Pearson, que é dado por:

$$\rho = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y}$$

Onde:

ρ = Coeficiente de correlação (varia entre -1 e 1)

$\text{cov}(X,Y)$ = Covariância das variáveis X e Y

σ_x, σ_y = Desvio padrão das variáveis X e Y, respectivamente.

Portanto, a partir da expressão acima, foi montada uma matriz com a correlação entre todos os pares de poços, para tornar possível selecionar os três poços com a melhor correlação entre os demais. A fim de automatizar o preenchimento dos dados, foi criada uma rotina em *Visual Basic*, desenvolvida por meio da criação de Macros no *software* de edição tabular Microsoft Excel (Apêndice B).

A rotina procura identificar primeiramente se a célula que representa o nível da água de algum dos poços de maior correlação está vazia, caso esteja, é escolhido o poço com a quarta melhor correlação e descartado o de valor vazio. Esta metodologia de seleção dos poços permite que existam até três poços vazios, ou seja, havendo necessidade, a rotina é capaz de selecionar até o poço com a sexta melhor correlação. Além disso, foram criados filtros para evitar o preenchimento com valores absurdos, limitando a faixa de valores que a célula poderia ser preenchida e apagando aquelas que não se enquadravam.

O processamento da rotina procura preencher primeiramente todas as células referentes ao primeiro ano da série (primeira coluna), em seguida são apagados os valores inconsistentes. Esse processo é repetido até que não haja evolução no preenchimento das falhas, ou seja, o número de células vazias no processo anterior seja o mesmo do processo atual. Em seguida é repetida essa sequência para todos os anos da série (todas as colunas).

Após finalizar o preenchimento de forma automatizada, existiam alguns valores não preenchidos. Esses últimos foram preenchidos por regressão polinomial de sexta ordem, utilizando valores de anos anteriores e posteriores a célula vazia de um determinado poço.

4.2 DETERMINAÇÃO DAS SUPERFÍCIES DE NÍVEL DE ÁGUA

A superfície de nível da água pode ser determinada a partir da interpolação de dados pontuais. Nesse trabalho, os dados pontuais representavam a localização dos poços e seus respectivos níveis de água. As superfícies foram criadas a partir de três métodos de interpolação: IDW, *spline* e krigagem. Foram geradas 53 (número de anos da série histórica) superfícies de nível da água em cada método de interpolação, totalizando 159 superfícies.

A definição dos parâmetros a serem utilizados em cada método foi determinada analisando a validação cruzada oriunda de diferentes configurações de cada modelo. Para testar as diferentes configurações, foram escolhidos os dados de nível da água referentes ao ano de 1960, que contém menos pontos para a interpolação e, por isso, seus resultados tendem a possuir um desempenho mais fraco em relação aos anos com maior quantidade de dados. A vantagem de escolher esse ano é analisar resultados em um cenário pessimista, ou seja, obter um resultado razoável deixa confortável por saber que eles tendem a melhorar com o tempo ou com o aumento do número de dados de entrada.

A definição dos parâmetros a serem utilizados no método IDW teve como base a escolha da potência, definição da vizinhança e no número de vizinhos (Tabela 4.2).

Tabela 4 2: Parâmetros de configuração IDW e validação cruzada

Nº de vizinhos	Potência	Setor de procura de vizinhos	<i>Leave-one-out</i>		<i>Holdout</i>	
			Erro médio	RMSE	Erro médio	RMSE
15	1	Completo	0.31	27.15	-11.11	122.1
15	2	Completo	-0.12	22.04	-13.82	119.1
15	3	Completo	-0.25	20.25	-15.03	117.2
15	4	Completo	-0.27	20.26	-15.56	115.7
15	5	Completo	-0.27	20.06	-15.79	114.5
15	6	Completo	-0.30	20.82	-15.87	113.4
15	10	Completo	-0.24	22.25	-15.85	110.4
15	15	Completo	-0.22	23.17	-15.81	108.3

Tabela 4.2: Continuação

Nº de vizinhos	Potência	Setor de procura de vizinhos	<i>Leave-one-out</i>		<i>Holdout</i>	
			Erro médio	RMSE	Erro médio	RMSE
20	20	Completo	-0.21	23.68	-15.81	107.3
15	2	4 fatias	-0.42	23.42	-14.27	123.3
15	6	4 fatias	-0.26	20.81	-16.06	114.3

Verificou-se, a partir da validação cruzada (método *leave-one-out* e *holdout*), que não houve uma configuração com melhor desempenho em todos indicadores estatísticos. Porém, foi definida para ser utilizada em toda a série histórica a configuração com potência dois, setor de procura de vizinhos do tipo completo (representação dos setores se encontra na Figura 3.8) e incluindo 15 vizinhos. O motivo desta escolha foi, além do bom desempenho, a consideração que para superfícies de nível da água não se espera o aparecimento de picos, portanto a escolha da potência dois visou minimizar esse fenômeno.

A definição dos parâmetros a serem utilizados no método *spline* teve como base a escolha do tipo de *spline* (*Thin Plate Spline*, *Spline with Tension* e *Completely Regularized Spline*), a definição da vizinhança e o número de vizinhos (Tabela 4.3).

Tabela 4.3: Parâmetros de configuração *spline* e validação cruzada

Tipo	Nº de vizinhos	Setor de procura de vizinhos	<i>Leave-one-out</i>		<i>Holdout</i>	
			Erro médio	RMSE	Erro médio	RMSE
<i>Thin Plate Spline</i>	15	Completo	-0.15	14.81	-6.14	61.86
<i>Thin Plate Spline</i>	15	4 fatias	-0.26	14.23	-4.77	55.29
<i>Spline with Tension</i>	15	Completo	-0.16	14.76	-7.66	71.83
<i>Spline with Tension</i>	15	4 fatias	-0.27	14.25	-5.43	59.61
<i>Completely Regularized Spline</i>	15	4 fatias	-0.10	17.88	-12.67	114
<i>Thin Plate Spline</i>	8	4 fatias	-0.30	14.33	-5.13	59.36
<i>Thin Plate Spline</i>	8	8 fatias	-0.29	14.24	5.65	59.59

Verificou-se, a partir da validação cruzada (método *leave-one-out* e *holdout*), que a configuração com melhor desempenho em todos indicadores estatísticos foi o tipo *Thin Plate Spline* com 15 vizinhos e setor de procura de vizinhos em oito fatias. Portanto, esta foi a configuração do método *spline* adotada para a criação das superfícies em toda a série histórica.

Já para o método da krigagem foi adotado um procedimento diferente dos demais métodos, uma vez que a krigagem exige uma configuração individualizada para cada grupo de pontos a serem interpolados. Por isso, na criação das superfícies de nível da água por krigagem, cada ano da série histórica possui uma configuração que melhor se adapta aos dados em estudo.

O processo de configuração da krigagem tem início com a análise exploratória dos dados, onde é verificada a transformação a ser aplicada de modo que sua distribuição de frequência dos dados tenda a se normalizar além de identificar tendências globais. Após a análise exploratória é feita a modelagem do semivariograma teórico, que consiste em ajustar o modelo de semivariograma ao semivariograma empírico, visando atingir níveis razoáveis de autocorrelação espacial. O ajuste do modelo de semivariograma é feito por meio da configuração de parâmetros, tais como: tamanho do *lag*, número de *lags*, consideração da anisotropia, efeito pepita e, por fim, o tipo de curva do modelo. A última etapa do processo de configuração se refere à definição da vizinhança, onde é determinado o número de vizinhos a serem incluídos e o setor de procura.

A configuração da krigagem é feita por um processo de tentativa e erro em que se busca a configuração que possui o melhor desempenho, que é avaliado por meio da validação cruzada. Ao final da configuração das krigagens de cada ano da série histórica, foi gerado um relatório com os valores de cada parâmetro utilizado, os quais podem ser vistos no Apêndice C.

4.3 COMPARAÇÃO ENTRE AS SUPERFÍCIES INTERPOLADAS

A principal metodologia utilizada nessa etapa para analisar o desempenho das interpolações foi a validação cruzada. Foram empregados os métodos *leave-one-out* e *holdout* de validação cruzada, descritos anteriormente. Portanto, a primeira análise comparativa entre as superfícies interpoladas consistiu em verificar o desempenho de cada uma, ano a ano. A Tabela 4.4 sintetiza todos os resultados obtidos pela validação cruzada.

Tabela 4.4: Síntese dos resultados obtidos pela validação cruzada

Ano	<i>Leave-one-out</i>						<i>Holdout</i>					
	Erro médio			RMSE			Erro médio			RMSE		
	Krig.	<i>Spline</i>	IDW	Krig.	<i>Spline</i>	IDW	Krig.	<i>Spline</i>	IDW	Krig.	<i>Spline</i>	IDW
1960	-0.07	-0.26	-0.12	13.31	14.23	22.04	-4.73	-4.77	-13.82	51.08	55.29	119.10
1961	-0.10	-0.06	0.10	19.24	18.00	23.62	2.09	2.65	3.59	26.21	27.68	32.46
1962	-0.22	-0.07	-0.40	18.90	20.69	23.84	-0.18	3.28	3.70	21.08	27.36	31.16
1963	-0.15	-0.02	-0.28	18.51	18.93	24.19	0.24	1.18	3.17	15.37	15.49	28.89
1964	0.07	-0.03	-0.18	15.87	16.39	21.19	-0.50	-1.46	1.80	19.61	19.71	30.48
1965	-0.06	-0.37	-0.54	18.84	21.67	21.97	0.72	3.13	2.55	19.77	23.78	25.06
1966	-0.15	-0.27	-0.28	20.71	18.67	24.37	-0.96	-0.02	2.69	20.69	19.33	29.42
1967	-0.23	-0.13	-0.76	23.09	25.31	26.58	0.27	0.01	1.00	20.83	28.05	24.23
1968	-0.15	-0.22	-0.77	19.70	20.50	24.23	-0.03	-0.05	0.80	19.49	20.21	25.91
1969	-0.51	-0.42	-0.72	19.42	20.92	24.90	1.38	1.78	3.00	20.34	20.41	26.78
1970	-0.43	-0.37	-0.70	18.43	20.14	23.25	1.16	1.73	3.04	19.77	19.84	31.03
1971	-0.49	-0.55	-0.84	18.84	21.01	25.05	1.44	1.40	2.02	20.39	18.86	29.28
1972	-0.16	-0.17	-0.82	20.75	21.34	25.01	-0.16	-0.21	1.35	18.01	19.03	26.98
1973	-0.18	-0.20	-0.79	20.11	21.11	24.26	-0.18	-0.18	1.00	20.88	21.38	29.36
1974	-0.59	-0.36	-1.19	20.37	21.75	24.07	0.26	-0.13	0.60	23.60	24.75	30.91
1975	-0.22	-0.28	-1.46	21.28	23.89	25.62	0.48	-0.16	1.38	20.04	20.28	29.25
1976	-0.17	-0.22	-1.84	21.18	22.78	25.54	1.33	0.80	0.44	19.64	19.77	29.77
1977	-0.18	-0.20	-1.06	21.15	22.24	25.76	0.26	-0.25	-0.83	20.04	20.05	24.38
1978	-0.07	-0.05	-1.13	21.60	22.79	25.88	0.87	-0.95	1.20	25.99	32.36	29.27
1979	-0.11	-0.02	-0.92	22.75	24.42	27.26	-0.65	-2.18	-0.01	21.10	23.30	26.29
1980	-0.01	0.03	-0.68	22.39	23.85	26.35	-0.46	-1.12	-0.65	18.16	18.18	22.97
1981	-0.30	-0.26	-1.78	24.04	25.74	29.97	0.12	-0.30	-0.94	25.41	25.45	30.91
1982	-0.05	0.09	-0.80	21.35	22.11	25.95	0.85	0.31	0.24	24.29	25.34	27.19
1983	-0.07	0.04	-1.31	22.19	23.96	28.16	4.10	3.79	4.07	38.73	43.97	52.83
1984	-0.12	-0.50	-1.14	22.40	23.59	27.06	1.08	0.53	-0.35	25.01	25.40	28.10
1985	-0.19	-0.15	-1.32	22.80	25.06	27.62	0.35	0.26	-1.60	21.63	21.35	27.15
1986	-0.11	-0.90	-1.10	23.32	23.18	29.17	-0.84	-1.15	-1.54	24.93	24.94	29.59
1987	-0.12	-0.14	-1.11	22.28	23.73	26.68	0.07	-0.25	-1.57	21.16	20.66	26.29
1988	-0.34	-0.35	-1.07	21.80	22.82	26.41	-0.46	-0.25	-1.36	19.11	19.94	22.49
1989	-0.14	-0.95	-0.93	21.37	22.55	26.23	0.70	0.77	-0.62	17.92	17.69	23.20
1990	-0.13	-0.19	-0.90	22.26	23.24	27.08	-0.20	-1.62	0.22	20.29	21.89	27.82
1991	-0.26	-0.20	-0.98	21.52	22.76	26.11	1.32	1.31	1.56	20.32	20.88	28.85
1992	-0.17	-0.20	-1.15	22.68	23.88	27.15	-1.20	-1.38	-2.56	20.32	20.51	24.43
1993	-0.03	0.04	-0.71	22.82	24.20	27.24	-0.55	-0.60	-0.90	20.52	20.67	25.12
1994	-0.20	-0.23	-0.70	21.84	23.03	26.67	0.16	0.32	0.43	22.53	21.98	27.85
1995	-0.06	-0.10	-0.80	22.04	23.11	26.69	1.25	1.33	0.27	22.03	22.24	26.01
1996	-0.15	-0.14	-0.83	22.26	23.30	26.43	0.69	0.79	0.58	19.53	19.62	21.25
1997	-0.05	0.02	-0.98	23.01	24.47	27.47	1.72	1.33	2.77	22.49	23.86	27.57
1998	-0.09	-0.03	-0.54	23.33	24.71	27.58	0.28	-0.01	0.21	21.34	22.28	22.62
1999	-0.19	0.27	-0.76	23.14	24.64	28.77	4.20	3.96	4.74	24.63	25.63	49.85

Tabela 4.4: Continuação

Ano	<i>Leave-one-out</i>						<i>Holdout</i>					
	Erro médio			RMSE			Erro médio			RMSE		
	Krig.	<i>Spline</i>	IDW	Krig.	<i>Spline</i>	IDW	Krig.	<i>Spline</i>	IDW	Krig.	<i>Spline</i>	IDW
2000	-0.16	-0.3	-0.66	23.94	24.43	28.17	1.17	1.56	0.34	21	21.4	24.9
2001	-0.2	-0.2	-0.75	23.97	25.93	29.14	2.052	0.95	5.33	27.38	28.31	49.12
2002	-0.12	-0.26	-0.87	23.66	26.31	28.92	-0.46	-0.96	-0.35	20.88	21.04	25.83
2003	-0.03	-0.24	-0.46	25.42	26.76	29.6	1.71	1.33	7.45	23.86	24.08	36.03
2004	-0.33	-0.45	-1.05	24.24	25.17	28.68	-1.3	-1.62	-1.57	24.36	28.45	25.27
2005	-0.18	-0.45	-1.35	24.62	26.02	29.07	-0.71	-1.43	-3.24	23.6	24.96	36.28
2006	-0.24	-0.18	-1.56	27.02	25.63	29.4	-5.26	-2.48	-1.81	27.91	27.91	25.8
2007	-0.17	-0.16	-1.47	24.68	25.72	29.61	0.57	-0.2	0.9	25.68	26.47	27.97
2008	-0.18	-0.24	-1.53	25.34	26.34	30.06	-1.87	-2.45	-2.41	25.63	25.98	27.63
2009	-0.18	-0.02	-1.44	25.9	27.65	30.6	-1.63	-1.89	-1.54	25.1	25.74	25.87
2010	-0.22	-0.2	-1.3	24.63	25.58	30.06	-1.73	-2.57	-2.46	28.48	30.79	28.64
2011	-0.17	-0.3	-1.42	25.59	27.48	30.35	-0.99	-1.09	-1.26	27.53	28.96	26.7
2012	-0.21	-0.43	-1.3	27.02	30.33	32.07	1.5	1.86	0.89	27.99	28.44	31.22

Na Tabela 4.4 foram destacados em negrito os valores que representam o melhor resultado. A partir da análise da Tabela 4.4, percebe-se o fraco desempenho do método IDW em relação aos demais, por isso ele foi descartado de análises posteriores. Além disso, é fácil ver que a krigagem possui o melhor desempenho entre os métodos, porém seus resultados estão bastante próximos do *spline*, o que leva a realizar mais análises comparativas entre esses dois métodos.

4.4 ESTIMATIVAS DE VOLUME DO AQUÍFERO

A estimativa de volume do aquífero é obtida a partir de uma série de processamentos de imagens, como pode ser visto, detalhadamente, no fluxograma do Apêndice D. A Figura 4.3 mostra o fluxograma de modo resumido.

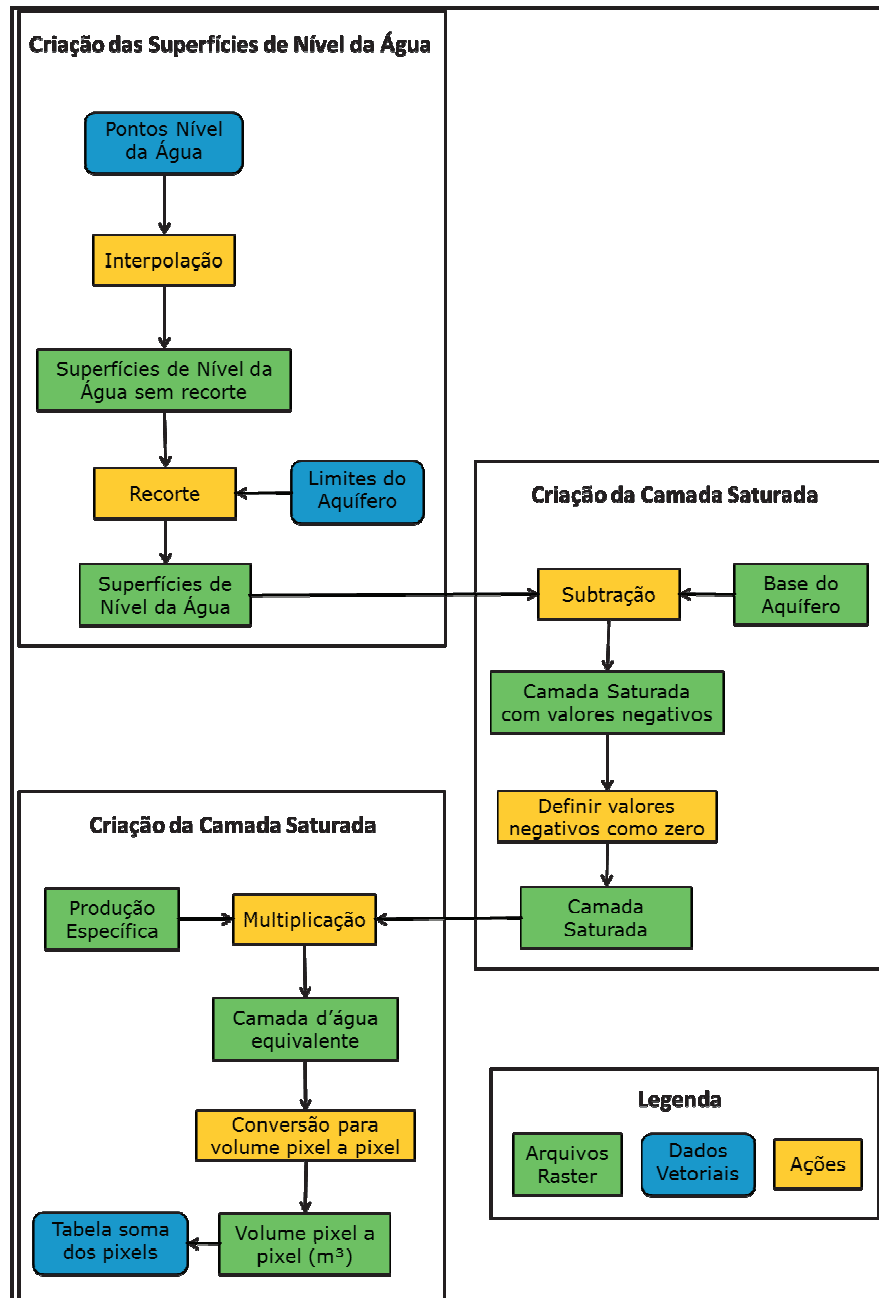


Figura 4.3: Fluxograma de processamento dos dados para cálculo do volume

No primeiro momento são criadas as superfícies de nível da água por meio de interpolações utilizando dados pontuais, como descritos anteriormente. Foram geradas 53 superfícies referentes pelo método da krigagem e 53 por *spline*, totalizando 106 superfícies que deveriam passar por todas as etapas do processamento descritas na Figura 4.3. As superfícies criadas extrapolavam os limites da área em estudo, por isso foi necessário recortá-las de acordo com os limites estabelecidos. A Figura 4.4 ilustra o processamento para a criação de superfícies de nível da água.

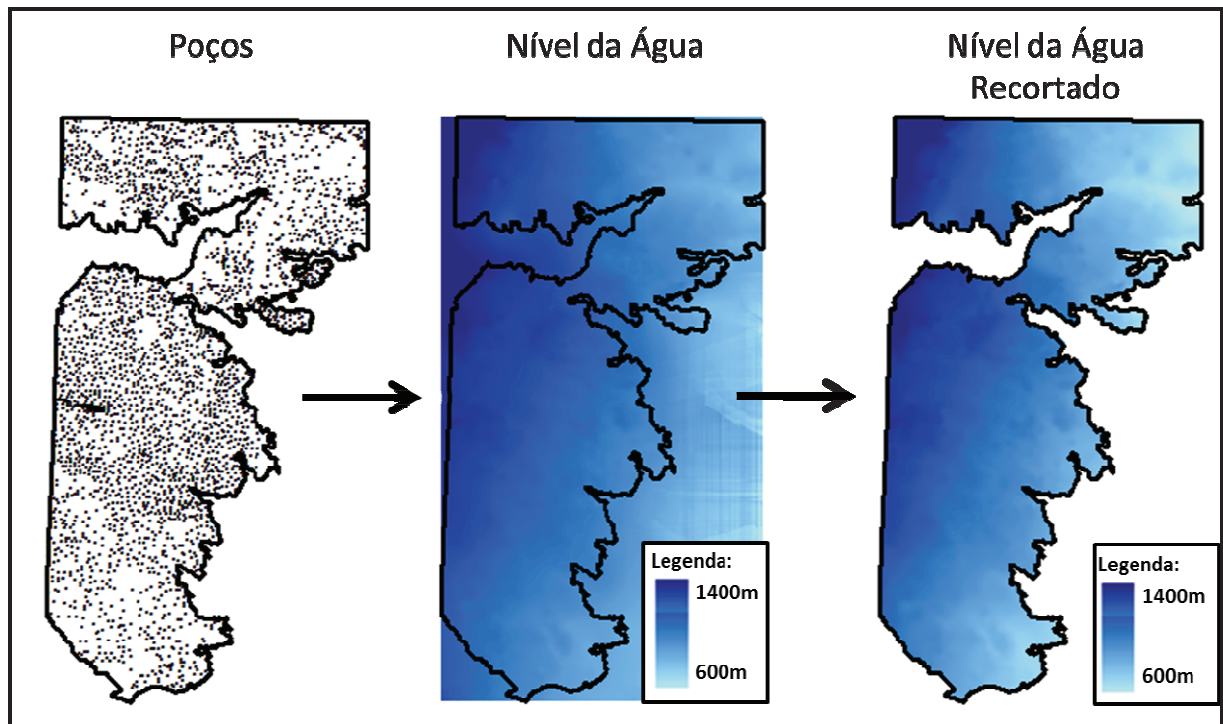


Figura 4.4: Criação de superfícies de nível da água

Após a criação das superfícies, é definida a espessura da camada saturada, ou seja, a altura da lâmina d'água em relação à base do aquífero, como pode ser visto na Figura 4.5.

Durante o cálculo da camada saturada, ocorre a aparição de valores negativos, que resultam de inconsistências entre a base do aquífero¹ e os níveis da água gerados, devendo ser eliminados. Portanto, os valores negativos da camada saturada são definidos como sendo zero aplicando as operações algébricas definidas a seguir:

$$CS = \frac{CS_{NEG} + |CS_{NEG}|}{2}$$

Onde:

CS = Camada Saturada

CS_{NEG} = Camada Saturada com valores negativos

¹ A Base do Aquífero foi fornecida pelo Center for Geospatial Technology da Texas Tech University (TTU)

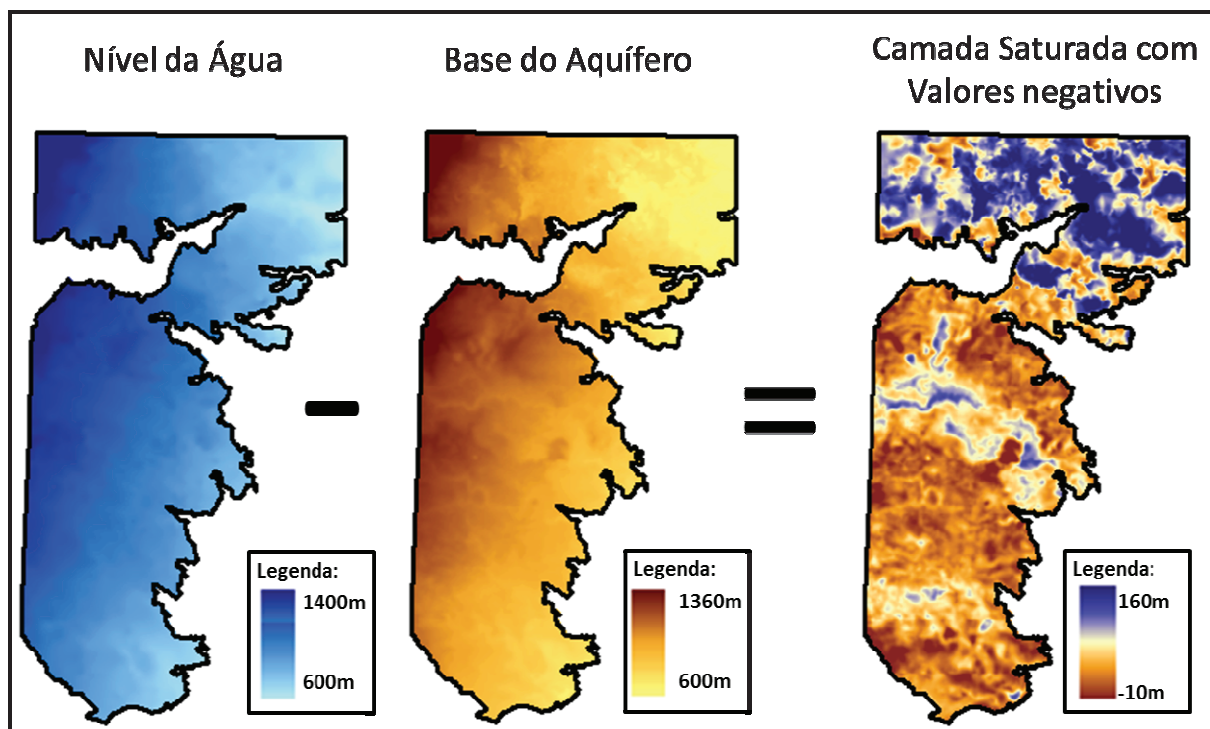


Figura 4.5: Criação da superfície de camada saturada

A estimativa do volume total do aquífero, para um determinado ano da série histórica, foi calculada *pixel a pixel* e, posteriormente, somados todos os valores. Para realizar esse cálculo foi necessário calcular a camada d'água equivalente, ou seja, retirar a parcela referente às partículas sólidas presentes na camada saturada, deixando apenas a altura da lâmina d'água equivalente. Esse resultado é obtido pelo produto entre a produção específica² e a camada saturada. Porém, o que se desejava era converter o valor da lâmina d'água equivalente de cada *pixel* para unidade de volume (m³), como a resolução espacial da imagem (528 × 528) e a camada d'água equivalente estavam na unidade de comprimento pés³, foi necessário multiplicar cada *pixel* de lâmina d'água equivalente (medidos em pés) por uma constante de conversão (conv_{pixel}), calculada a seguir:

$$\text{conv}_{\text{pixel}} = 528 \times 528 \times (0,3048)^3 = 7.894,28376 \text{ m}^3/\text{pé}$$

A Figura 4.6 ilustra o processo para o cálculo de volume *pixel a pixel*.

² A Produção Específica foi fornecida pelo Center for Geospatial Technology da TTU

³ 1 pé = 0,3048m

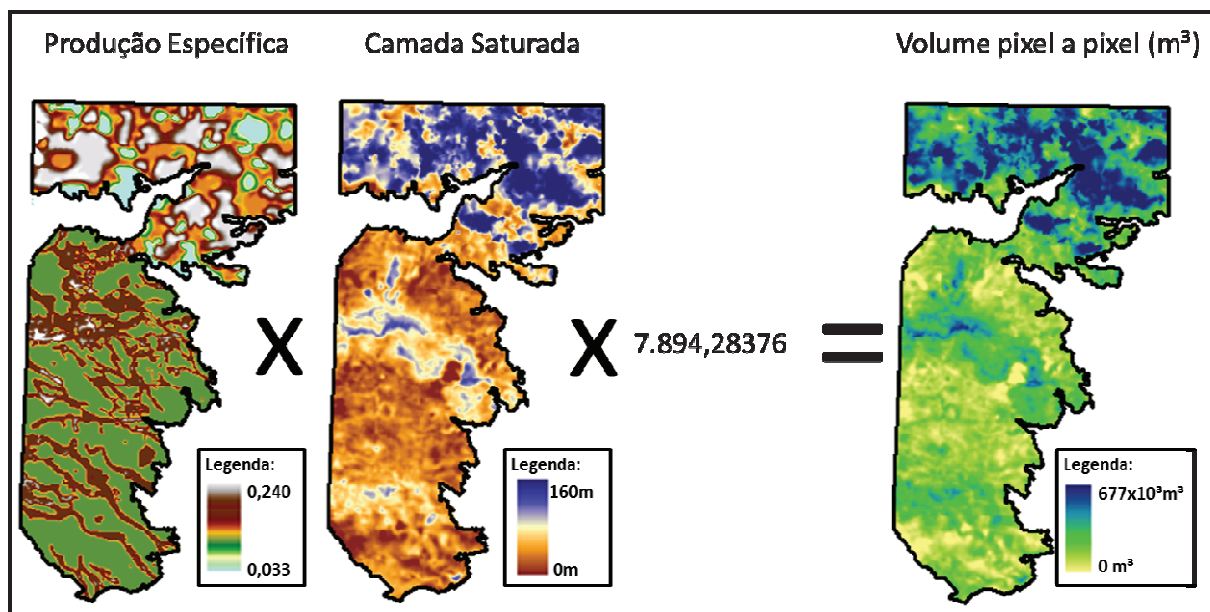


Figura 4.6: Cálculo de volume *pixel a pixel*

4.5 IDENTIFICAÇÃO DE TENDÊNCIAS EM SÉRIES HISTÓRICAS

Antes de analisar a presença de tendência nas séries definidas para este estudo, foi necessário investigar seu comportamento. Por se tratar de dados anuais de nível da água, não foi considerada a presença de sazonalidade nas séries. Esta hipótese é facilmente confirmada quando se observa a curva dos gráficos. Outra hipótese era de que as séries não seriam estacionárias, *e.g.*, haveria uma tendência ao longo do tempo.

Por se tratar de muitas séries (492), fez-se necessário realizar uma análise exploratória do conjunto de dados para poder definir que técnicas seriam mais apropriadas para avaliá-los.

A princípio foi analisado o gráfico de nível da água para todas as séries, como pode ser visto na Figura 4.7(a). O gráfico mostra o nível da água em relação ao nível médio dos mares. As séries são de dados anuais e contínuos (houve o preenchimento das falhas). É possível notar que, devido a enorme quantidade de dados, é difícil definir algum padrão de comportamento das curvas. Por isso, decidiu-se agrupar as séries de acordo com o perfil das curvas, então foram utilizadas técnicas de análise de *cluster* (*k-means*) nas séries de nível da água (Figura 4.7(b)).

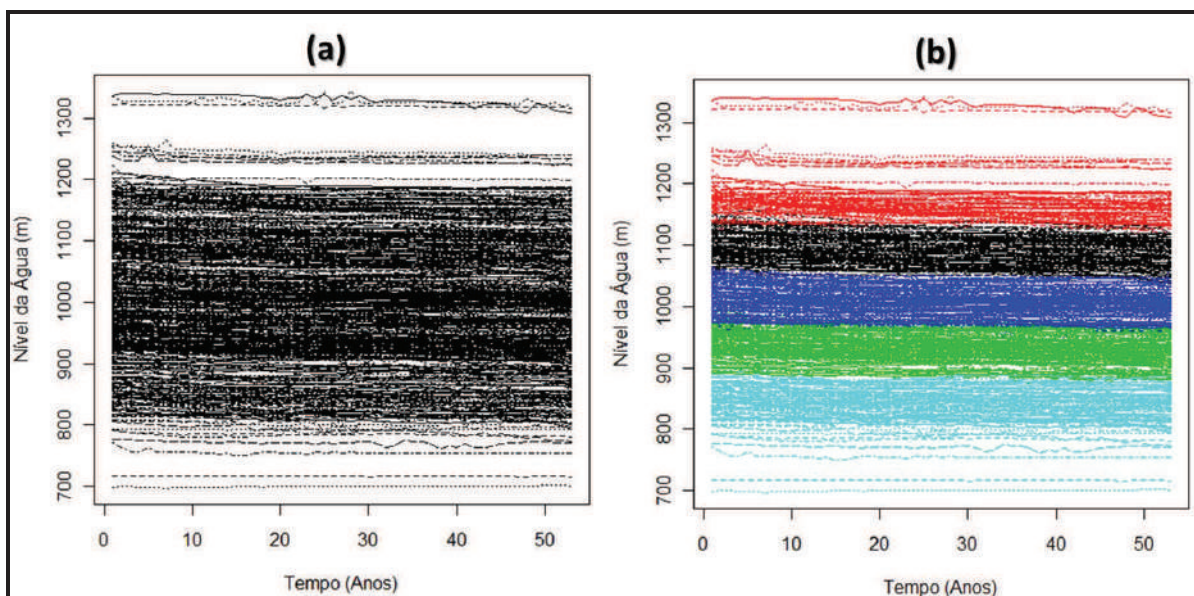


Figura 4.7: Gráficos de nível da água com todas as séries

Porém, constatou-se que este gráfico não é adequado para realizar análises de *cluster*, uma vez que os valores de níveis de água médios entre as séries predominam sobre a variação do nível da água entre elas. O que se esperava era um agrupamento que definissem classes baseado na variação do nível da água. Por isso, foi realizada uma transformação no conjunto de séries de nível da água para que estas representassem a variação do nível da água. Desse modo, os valores de cada série foram subtraídos do seu respectivo valor inicial (nível da água em 1960), assim as novas séries representariam a variação do nível da água em relação ao ano de 1960, como pode ser visto na Figura 4.8.

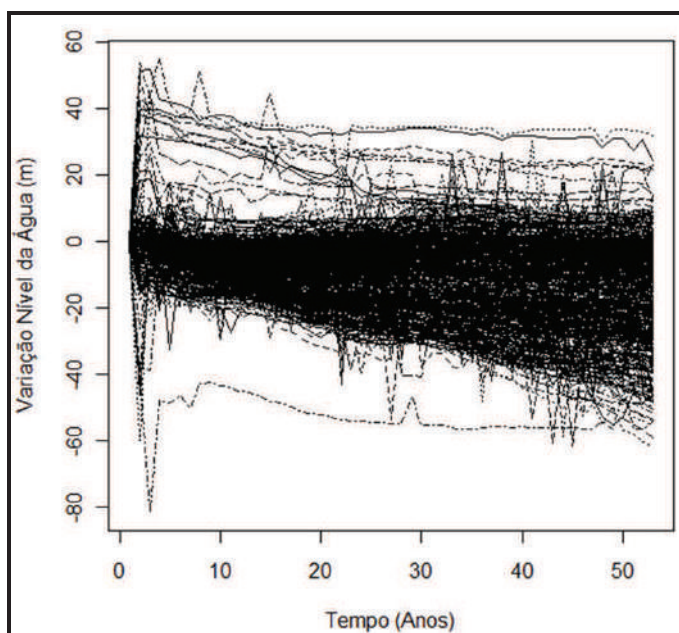


Figura 4.8: Gráfico da variação do nível da água com todas as séries

Ao analisar o gráfico da Figura 4.8, a presença de picos (ruído, em algumas séries, se dá devido ao preenchimento de falhas. Porém, considerando que uma boa técnica de *cluster* deve possuir robustez a presença de ruídos, não foi realizado tipo algum de tratamento nessas séries. Desse modo, espera-se avaliar o desempenho dos métodos de análise de *cluster* com a presença de ruídos.

Portanto, as séries foram agrupadas por dois métodos: *k-means* e hierárquico. Grande parte do processamento de dados estatísticos foi realizada no *software* R (os comandos utilizados se encontram no Apêndice E). No processamento pelo método *k-means*, foram adotadas oito classes K, formando assim oito grupos de *cluster*. Já para o método hierárquico, foi criado um clustergrama utilizando a função distância Euclidiana, padronização dos valores do *heat map* variando entre -3 e 3, método aglomerativo de classes do tipo *average linkage* e um método de otimização para organizar os ramos do dendrograma (*OptimalLeafOrderValue*).

Por fim, foi verificado em cada série histórica se havia tendência, além de testar para a média de cada *cluster*. O método utilizado para a análise de tendência foi o teste de Mann-Kendall. Um dos resultados do teste é o p-valor, onde pode ser testada a significância da hipótese H_0 , que considera não haver tendência. Neste trabalho, foi considerada uma significância de 0,01, ou seja, caso o p-valor tenha uma significância menor que 0,01, a hipótese H_0 é rejeitada e adotada a hipótese H_1 . Sob as condições citadas, ao se adotar a hipótese H_1 , considera-se com 99% de confiança que há tendência nos dados. Ainda é possível identificar se a tendência em uma série é crescente ou decrescente analisando o coeficiente τ de Kendall. Considerando que foi adotada a hipótese H_1 , valores negativos de tau indicam que há uma tendência decrescente, já valores positivos, crescente.

No Apêndice F, é mostrada uma tabela com o resultado do teste de Mann-Kendall para cada poço, onde é mostrado o p-valor, o coeficiente τ , a hipótese adotada, classificação da série por categoria (0 = não há tendência; 1 = tendência crescente; -1 = tendência decrescente), grupo pertencente de cada série pelo método *k-means* e pelo método hierárquico.

5. RESULTADOS E DISCUSSÕES

5.1 ESTIMATIVAS DE VOLUME

A partir da metodologia definida para a estimativa do volume *pixel a pixel* do aquífero, foi gerado um relatório em forma tabular contendo o volume total do aquífero. De posse do relatório contendo o volume total para cada ano da série histórica pelos métodos de krigagem e *spline*, foi possível gerar o gráfico que indica a variação do volume total do aquífero em função do tempo, como pode ser visto na Figura 5.1.

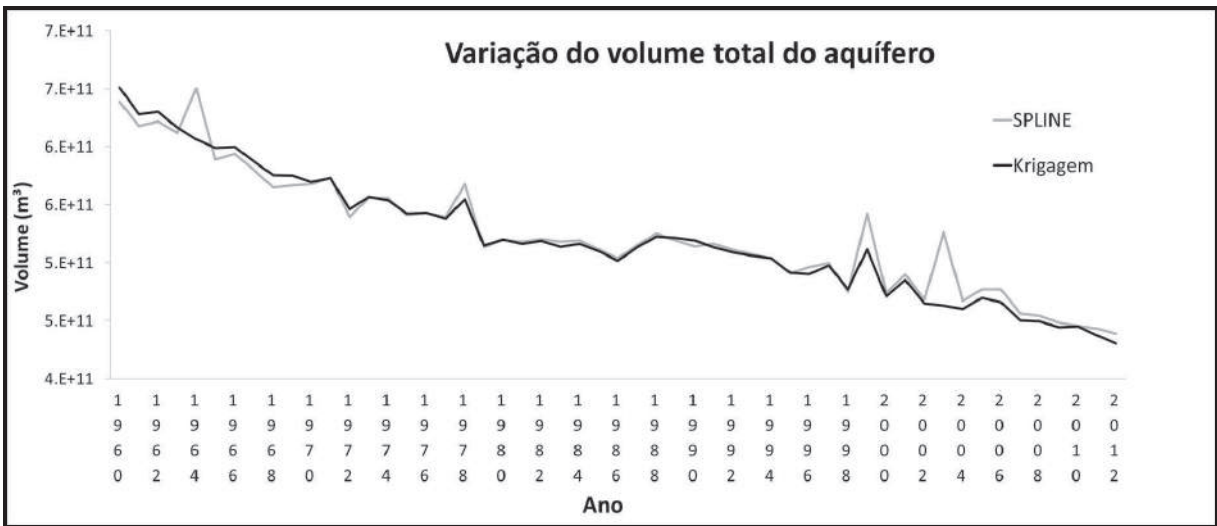


Figura 5.1: Variação do volume total do aquífero

Ao analisar as curvas de variação do volume total geradas utilizando os métodos *spline* e krigagem, percebe-se que as duas possuem comportamentos semelhantes, com exceção dos anos de 1964, 1999 e 2003, quando o *spline* gerou picos de volume que não foram previstos pela krigagem.

O gráfico também indica a redução constante do volume total da série histórica, o que leva a constatar o desequilíbrio entre descarga e recarga que está havendo nessa porção do aquífero.

A fim de ratificar o pressuposto visto na metodologia de que as superfícies de nível da água geradas por krigagem e *spline* possuem comportamentos semelhantes, foi realizada uma análise da correlação entre os resultados de volumes totais gerados por *spline* e krigagem,

além de comparar as diferenças percentuais entre os volumes totais calculados pelos dois métodos.

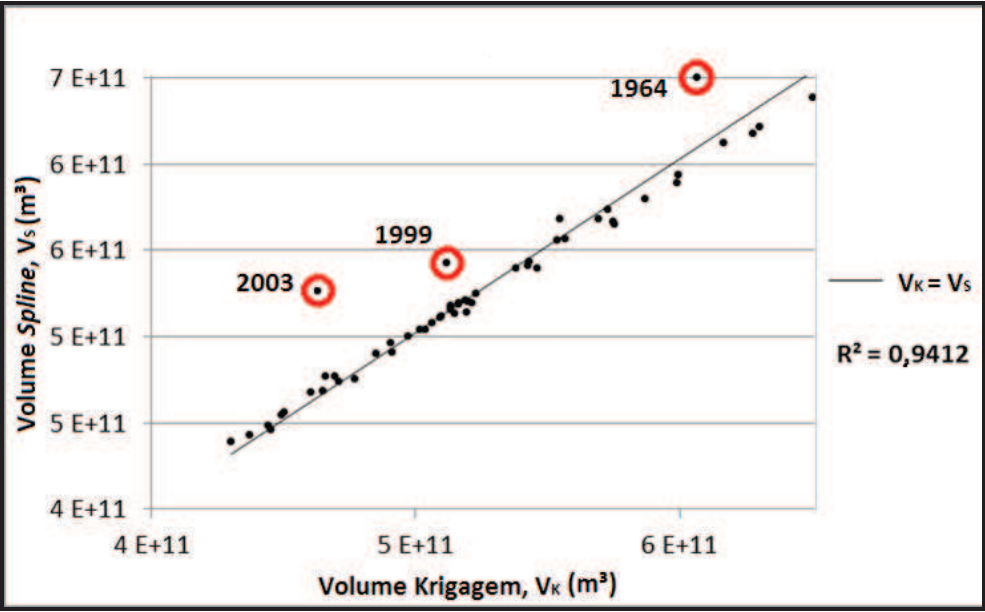


Figura 5.2: Correlação entre krigagem e spline

A partir da análise do gráfico de correlação entre Krigagem e Spline (Figura 5.2), percebe-se que a dispersão dos pontos que representa os volumes totais originados por cada método e em cada ano da série histórica, aproxima-se da curva $x = y$, ou seja, possuem uma alta correlação ($R^2 = 0,9412$).

Além disso, foi verificada a diferença percentual entre os volumes totais gerados por krigagem e spline. A Tabela 5.1 indica a diferença percentual para cada ano da série histórica, sendo os valores percentuais positivos indicando maior estimativa da krigagem e, consequentemente, os percentuais negativos indicando a menor estimativa da krigagem em relação ao spline (com destaque os três anos com as maiores diferenças percentuais).

Tabela 5.1: Diferença percentual entre as estimativas de volume

Ano	Krigagem (m³)	Spline (m³)	Diferença	Ano	Krigagem (m³)	Spline (m³)	Diferença
1960	6.51E+11	6.39E+11	1.87%	1987	5.14E+11	5.15E+11	-0.35%
1961	6.28E+11	6.18E+11	1.64%	1988	5.23E+11	5.25E+11	-0.44%
1962	6.30E+11	6.22E+11	1.34%	1989	5.22E+11	5.19E+11	0.39%
1963	6.17E+11	6.12E+11	0.77%	1990	5.19E+11	5.14E+11	1.00%
1964	6.07E+11	6.50E+11	-7.17%	1991	5.13E+11	5.16E+11	-0.57%
1965	5.99E+11	5.89E+11	1.68%	1992	5.09E+11	5.11E+11	-0.40%
1966	6.00E+11	5.94E+11	0.98%	1993	5.06E+11	5.08E+11	-0.31%

Tabela 5.1: Continuação

Ano	Krigagem (m ³)	Spline (m ³)	Diferença	Ano	Krigagem (m ³)	Spline (m ³)	Diferença
1967	5.87E+11	5.80E+11	1.23%	1994	5.04E+11	5.04E+11	-0.02%
1968	5.76E+11	5.65E+11	1.87%	1995	4.91E+11	4.91E+11	0.05%
1969	5.75E+11	5.67E+11	1.47%	1996	4.91E+11	4.96E+11	-1.17%
1970	5.70E+11	5.68E+11	0.30%	1997	4.98E+11	5.00E+11	-0.45%
1971	5.73E+11	5.73E+11	-0.08%	1998	4.77E+11	4.75E+11	0.45%
1972	5.46E+11	5.39E+11	1.20%	1999	5.12E+11	5.43E+11	-6.01%
1973	5.57E+11	5.56E+11	0.13%	2000	4.71E+11	4.74E+11	-0.57%
1974	5.54E+11	5.56E+11	-0.37%	2001	4.85E+11	4.90E+11	-1.01%
1975	5.42E+11	5.41E+11	0.20%	2002	4.65E+11	4.68E+11	-0.70%
1976	5.43E+11	5.43E+11	0.05%	2003	4.63E+11	5.26E+11	-13.63%
1977	5.38E+11	5.39E+11	-0.25%	2004	4.60E+11	4.67E+11	-1.51%
1978	5.55E+11	5.68E+11	-2.38%	2005	4.70E+11	4.77E+11	-1.58%
1979	5.15E+11	5.13E+11	0.34%	2006	4.66E+11	4.77E+11	-2.37%
1980	5.20E+11	5.20E+11	0.06%	2007	4.50E+11	4.56E+11	-1.30%
1981	5.16E+11	5.18E+11	-0.37%	2008	4.49E+11	4.54E+11	-1.06%
1982	5.19E+11	5.21E+11	-0.30%	2009	4.44E+11	4.48E+11	-0.95%
1983	5.14E+11	5.18E+11	-0.88%	2010	4.45E+11	4.46E+11	-0.09%
1984	5.16E+11	5.19E+11	-0.59%	2011	4.37E+11	4.43E+11	-1.33%
1985	5.10E+11	5.12E+11	-0.33%	2012	4.30E+11	4.39E+11	-2.05%
1986	5.02E+11	5.04E+11	-0.41%				

A partir da análise da Tabela 5.1, é possível afirmar que não houve grandes variações nas estimativas por krigagem e *spline*. A média aritmética entre as diferenças percentuais foi de -0,64%, o que indica que as interpolações por krigagem geraram estimativas inferiores em relação ao *spline*, porém extremamente próximas. O desvio padrão das diferenças percentuais foi de 2,43%, ratificando a semelhança entre as estimativas.

Por isso, conclui-se que os dois métodos de representação são satisfatórios, porém a krigagem mostrou um desempenho melhor tanto na validação cruzada quanto na variação do volume total, por não apresentar picos em suas previsões, considerando que não deveria haver uma recarga e uma descarga muito fora do padrão em apenas três anos isolados. Assim, a partir deste ponto, serão realizadas análises apenas com os resultados provenientes da krigagem.

5.2 EVOLUÇÃO DAS ESPESSURAS DE CAMADA SATURADA

A análise da variação das espessuras da camada saturada permite identificar áreas onde camada saturada encontra-se com valores críticos, porções do aquífero onde sua redução acontece em um processo mais acelerado, regiões que não estão sofrendo redução significativa entre outras possíveis análises. A Figura 5.3 mostra o resultado do cálculo da espessura da camada saturada para os anos de 1962 e 1972. Na Figura 5.3, a superfície foi dividida em três classes que representam a espessura da camada saturada, 0–10 m, 10–50 m e 50 – (valor máximo) m, também é mostrado um gráfico que representa o percentual de cobertura de cada classe.

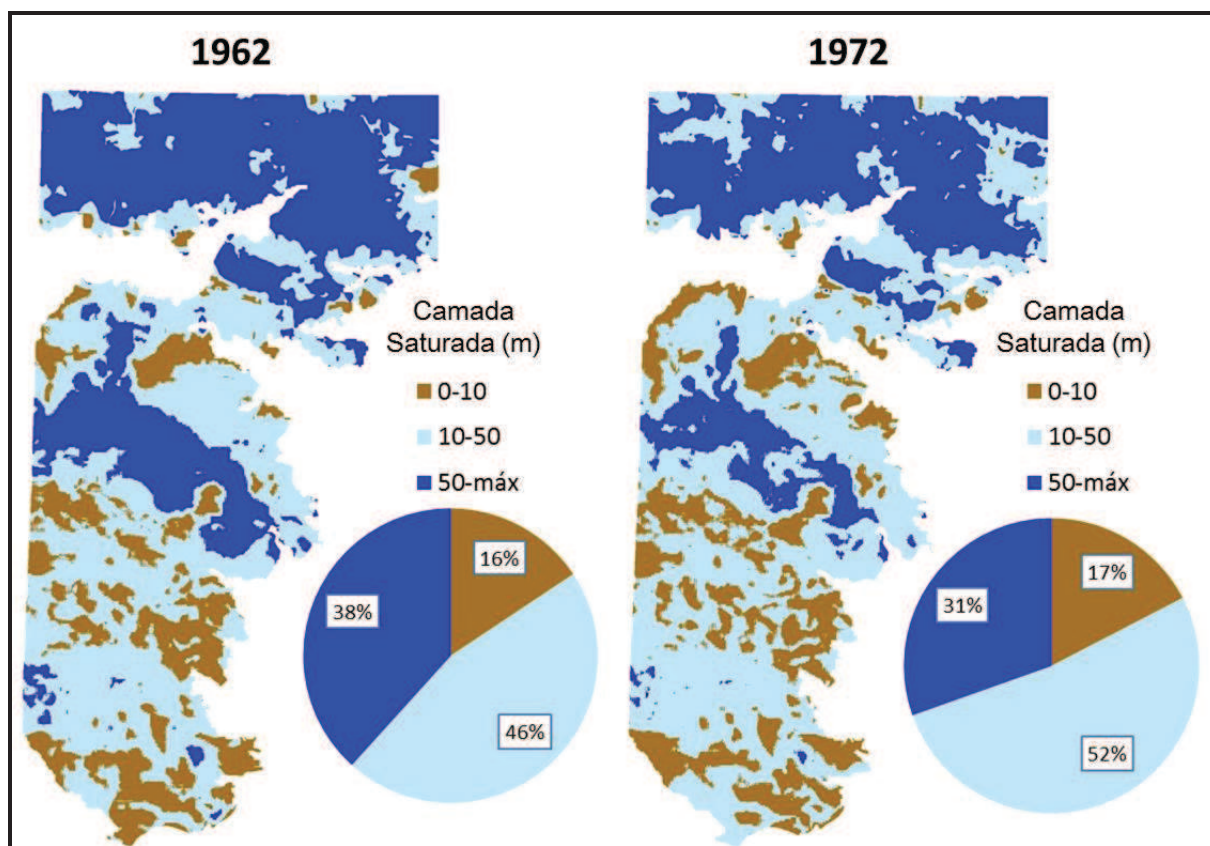


Figura 5.3: Espessura da camada saturada em 1962 e 1972

A partir da análise da Figura 5.3, percebe-se uma redução da classe que representa a camada mais espessa (50–máx.). Esse comportamento se repete ao longo dos anos da série histórica, como pode ser visto nas Figuras 5.4 e 5.5, que foram criadas de modo análogo à Figura 5.3. A Figura 5.4 mostra o resultado do cálculo da espessura camada saturada para os anos de 1982 e 1992, e a Figura 5.5 representa os anos 2000 e 2012.

Ao realizar uma análise espaço-temporal da variação das camadas saturadas, percebe-se que a região central do aquífero possuía, na década de 1960, grande parte ocupada pela camada classificada como mais espessa. Ao longo dos anos ela foi diminuindo gradativamente até um valor insignificativo de áreas com a camada saturada maior que 50m. Também foi possível identificar uma redução significativa na porção noroeste do aquífero.

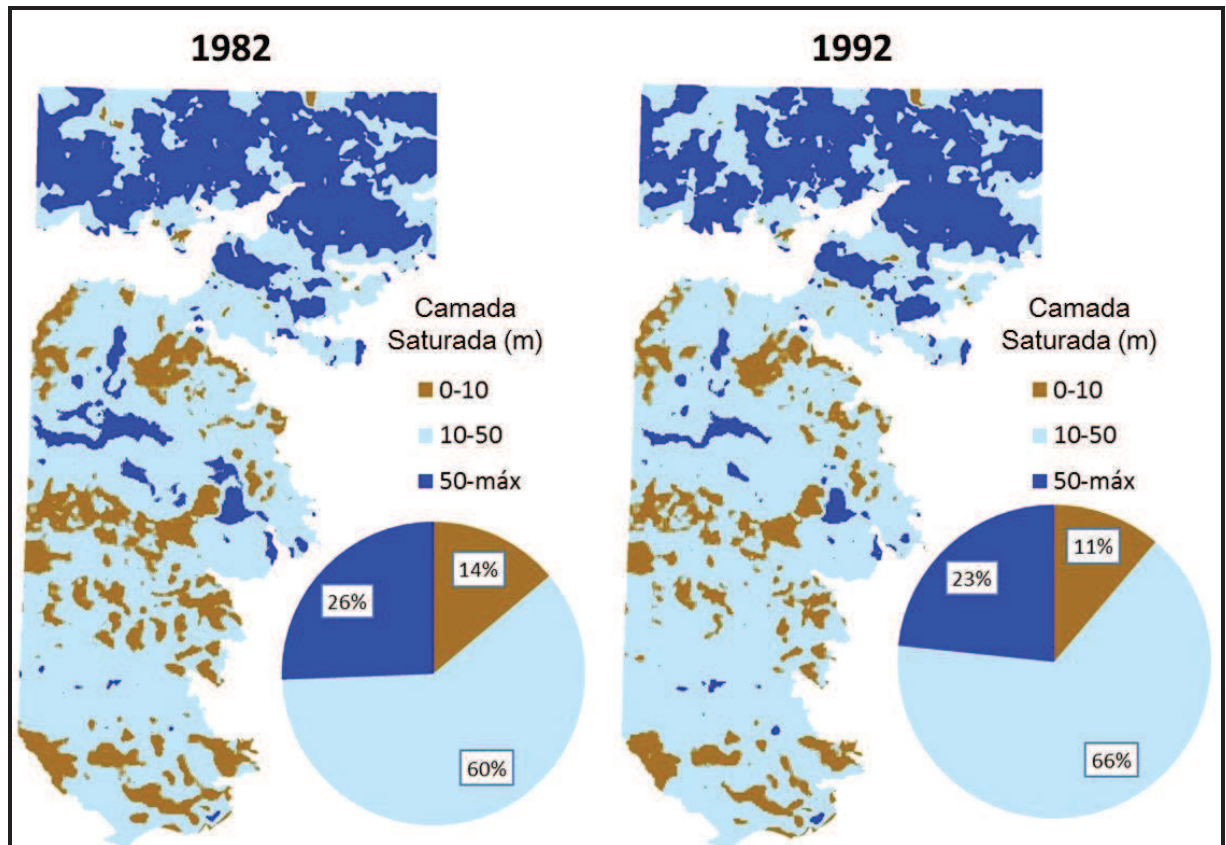


Figura 5.4: Espessura da camada saturada em 1982 e 1992

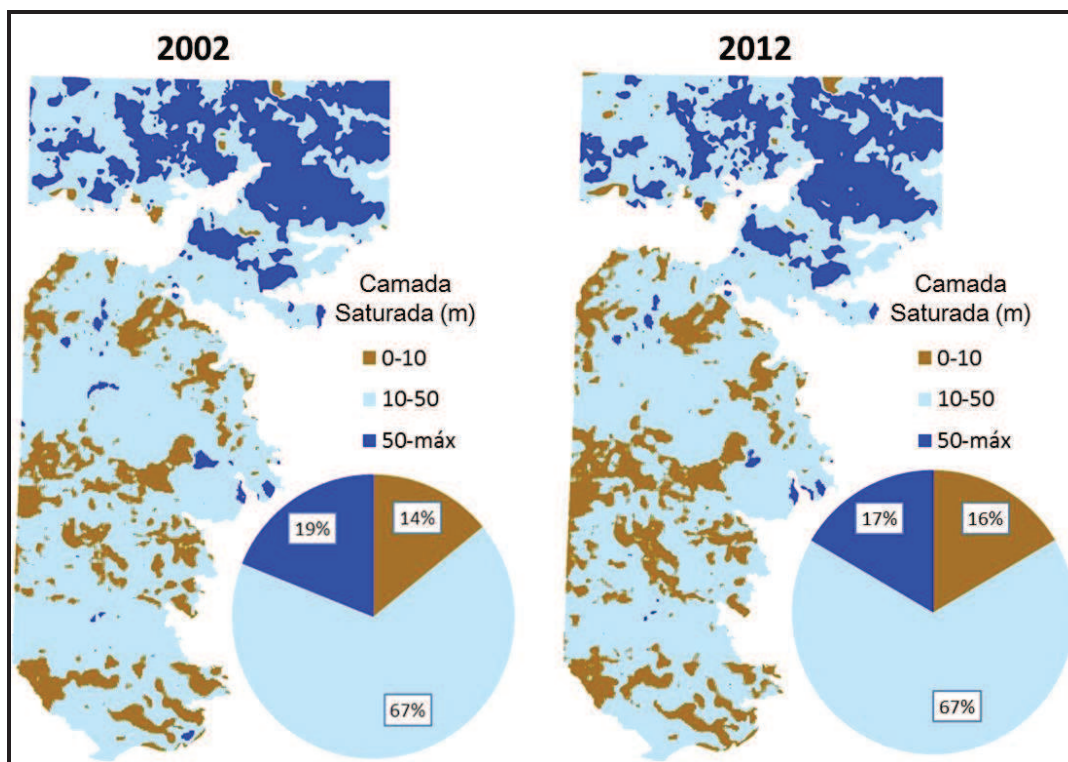


Figura 5.5: Espessura da camada saturada em 2002 e 2012

O que se observa ao analisar a evolução das espessuras é que, de um modo geral, a classe que representa as espessuras críticas (0–10m) tem uma flutuação em torno de 15% da cobertura total, enquanto as espessuras médias (10–50m) têm um aumento constante, ganhando 21% do percentual de cobertura total em 50. Esse percentual (21%) é o mesmo que foi perdido pela camada mais espessa (50–máx). A Figura 5.6 mostra um gráfico que representa as mudanças nos percentuais de cada classe de camada saturada ao longo de 50 anos com um passo de 10 anos.

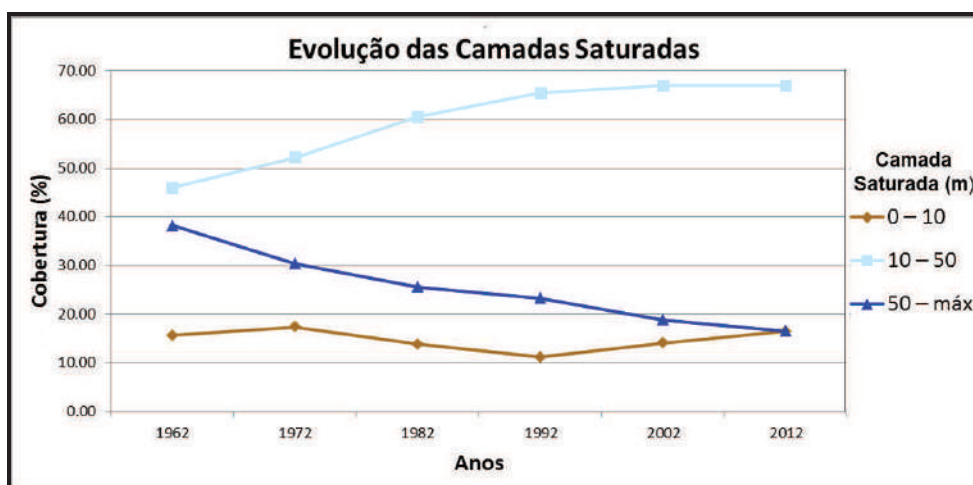


Figura 5.6: Evolução das camadas saturadas

5.3 ANÁLISES DE TENDÊNCIAS ESPAÇO-TEMPORAIS

A metodologia para a análise de tendência teve como base a avaliação de dois tipos de resultados: as análises de *cluster* e o teste de Mann-Kendall. Inicialmente foram analisados os *clusters* obtidos pelo método hierárquico e de agrupamento. Pelo método hierárquico foi criado um clustergrama (Figura 5.7) para agrupar linhas (dendrograma dos anos) e colunas (dendrograma de poços) de acordo com a similaridade entre as séries. As cores do *heat map* presente na Figura 5.7 representam valores da distância Euclidiana normalizada e permite verificar a similaridade entre grupos.

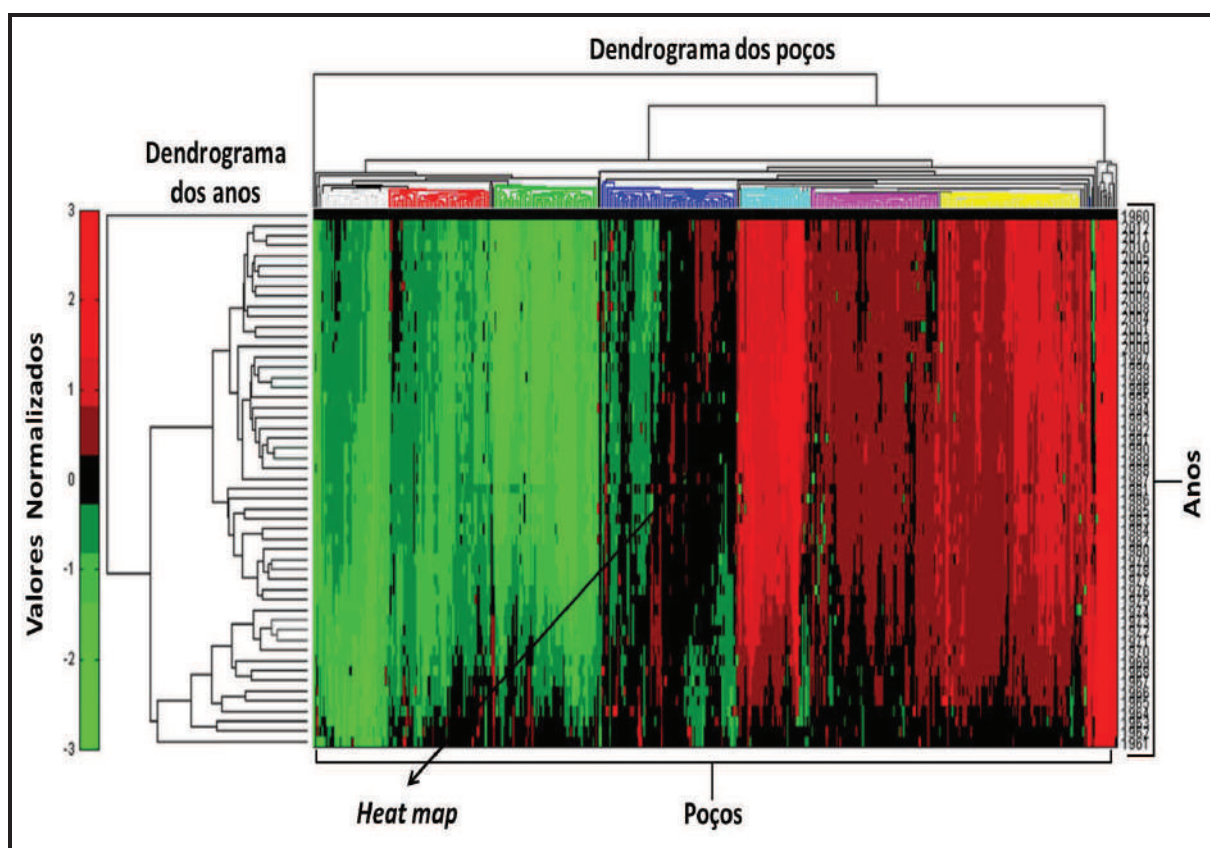


Figura 5.7: Clustergrama

Após definir o corte a ser realizado no dendrograma dos poços, foram serapados os *clusters*, como pode ser visto na Figura 5.8.

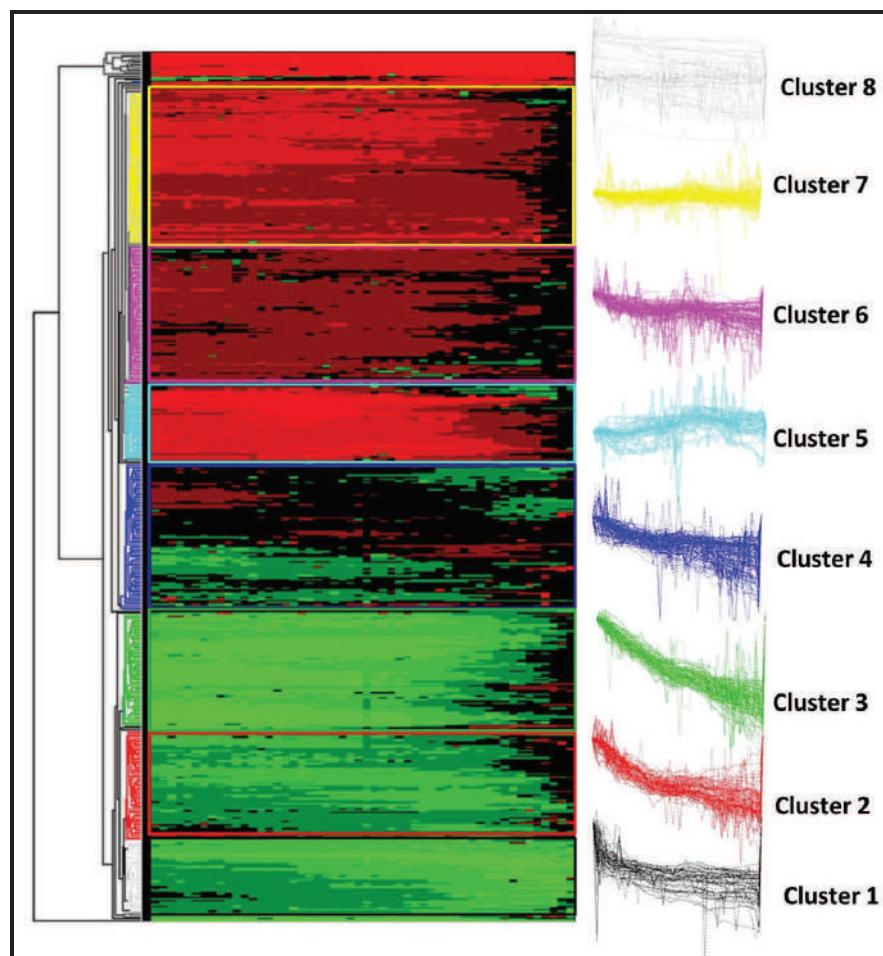


Figura 5.8: Clustergrama e o conjunto de séries de cada *cluster*

Ao analisar os *clusters* criados, pode-se concluir que cada grupo de séries possui um comportamento semelhante, como é esperado. A exceção ocorre apenas no *cluster* 8, que foi formado pelas séries que não foram agrupadas em nenhum dos outros sete *clusters*. Outra característica marcante entre os *clusters* é que se pode ter uma idéia visual do comportamento de cada um; *e.g.*, os *clusters* 1, 2, 3, 4 e 6 devem possuir uma tendência decrescente, o *cluster* 5 uma tendência crescente, enquanto o *cluster* 7 aparentemente não possui tendência. A fim de verificar estas hipóteses, foi utilizado o teste de Mann-Kendall em todas as séries de cada grupo para verificar a presença de tendência, como apresentado na Tabela 5.2.

A fim de comparar o desempenho entre métodos de agrupamento, também foi realizada a análise de *cluster* pelo método *k-means*. Para facilitar a comparação, foram definidas oito classes para o agrupamento *k-means*, *e.g.*, a mesma quantidade de grupos definida pelo método hierárquico. A Figura 5.9 mostra dois gráficos (método hierárquico e método *k-means*) contendo todas as séries temporais de variação do nível da água, coloridas de acordo com o tipo de *cluster*.

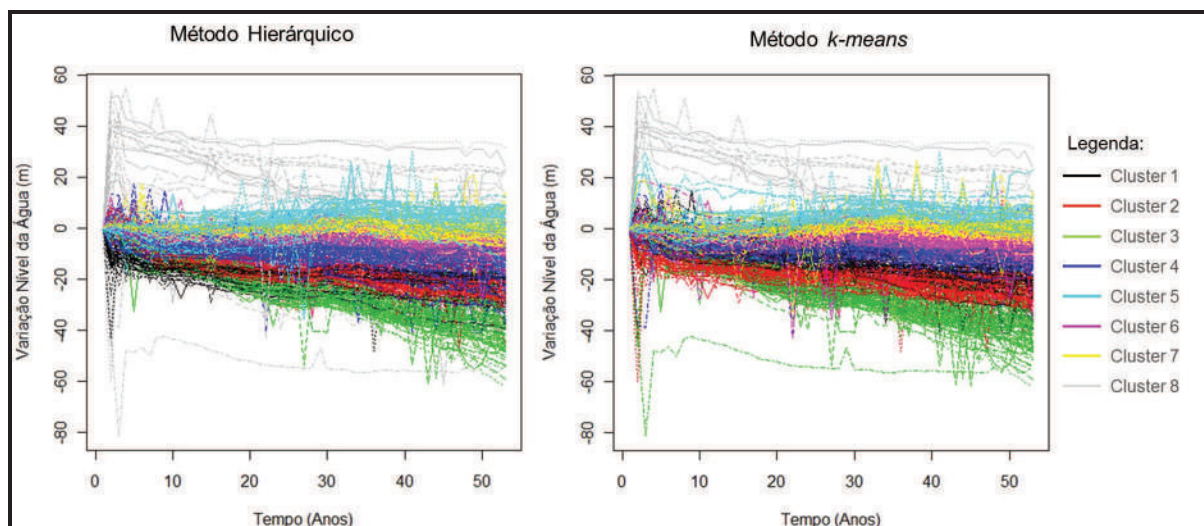


Figura 5.9: Gráficos de séries temporais classificadas por *cluster* (Método Hierárquico e Método *k-means*)

Apesar da visível semelhança entre a disposição dos *clusters* pelos dois métodos de agrupamento, deve-se ressaltar que cada *cluster* possui uma quantidade diferente de séries e que as séries agrupadas em um mesmo *cluster* não necessariamente são as mesmas. O que se deseja é identificar qual método possui melhor desempenho ao agrupar séries de acordo com a tendência, *e.g.*, espera-se que os *clusters* agrupem séries com o mesmo tipo de tendência (crescente, decrescente ou sem tendência) e, além disso, que os grupos possuam curvas semelhantes. A Tabela 5.2 compara o resultado dos dois métodos por meio de estatísticas de cada *cluster*, *e.g.*, define a quantidade e o percentual de séries em cada *cluster* de acordo com o tipo de tendência. Esta tabela teve como base os resultados do teste de Mann-Kendall com 90% confiança. No Apêndice F é mostrada uma tabela com o teste de Mann-Kendall para cada série (poço) e seu respectivo *cluster* pelo método hierárquico e de agrupamento.

Tabela 5.2: Comparação entre *clusters* por *k-means* e hierárquico

Cluster	Método	Tendência					
		Decrescente		Não há		Crescente	
		Quant.	%	Quant.	%	Quant.	%
Cluster 1	Hierárquico	41	100	0	0	0	0
	<i>k-means</i>	51	100	0	0	0	0
Cluster 2	Hierárquico	60	96,77	2	3,23	0	0
	<i>k-means</i>	88	100	0	0	0	0
Cluster 3	Hierárquico	63	100	0	0	0	0
	<i>k-means</i>	69	100	0	0	0	0
Cluster 4	Hierárquico	81	97,59	2	2,41	0	0
	<i>k-means</i>	52	94,55	2	3,64	1	1,82

Tabela 5.2: Continuação

Cluster	Método	Tendência					
		Decrescente		Não há		Crescente	
		Quant.	%	Quant.	%	Quant.	%
Cluster 5	Hierárquico	0	0	8	19,05	34	80,95
	<i>k-means</i>	3	7,32	6	14,63	32	78,05
Cluster 6	Hierárquico	64	82,05	3	3,85	11	14,1
	<i>k-means</i>	74	80,43	4	15,22	14	4,35
Cluster 7	Hierárquico	46	54,12	21	24,71	18	21,18
	<i>k-means</i>	40	47,06	24	28,24	21	24,71
Cluster 8	Hierárquico	32	84,21	4	10,53	2	5,26
	<i>k-means</i>	11	100	0	0	0	0

Ao analisar a Tabela 5.2, é possível notar que os métodos apresentaram resultados bastante próximos e desempenhos satisfatórios para o estudo proposto. Porém, fez-se necessário definir qual método obteve o melhor desempenho para que se pudessem realizar análises posteriores de identificação de áreas críticas. Os *clusters* 1, 2, 3 e 4 agruparam quase que em sua totalidade apenas séries com tendência decrescente; apenas o *cluster* 4 do agrupamento *k-means* apresentou apenas uma série com tendência crescente. No *cluster* 5, o método hierárquico teve um desempenho muito superior, uma vez que não agrupou séries de tendência crescente e decrescente, como ocorreu no método *k-means*. Os *clusters* seis e sete, que possuem séries de pequenas variações do nível da água, apresentaram resultados heterogêneos em ambos os métodos. Já o *cluster* oito não foi analisado, pois, pelo método hierárquico, ele representa séries que não foram agrupadas. Por fim, considerando o objeto de análise deste estudo (agrupar séries com comportamento e tendências semelhantes), o método de análise de *cluster* hierárquico apresentou um resultado levemente superior, sendo, portanto, o método utilizado nas análises seguintes.

A Figura 5.10 apresenta uma síntese do resultando da análise de tendências realizada pelo teste de Mann-Kendall. Neste estudo, foi considerado um nível de 99% de confiança nos testes de hipóteses, visando-se obter o resultado mais confiável possível. Ao ver a Figura 5.10, destaca-se o grande percentual de poços com tendência decrescente (74,39%) e sua localização espacial (predominância de poços com tendência decrescente na metade superior do aquífero). Porém, o gráfico não indica que grupos de poços possuem séries de decrescimento mais acentuadas, como pode ser visto nos resultados do agrupamento.

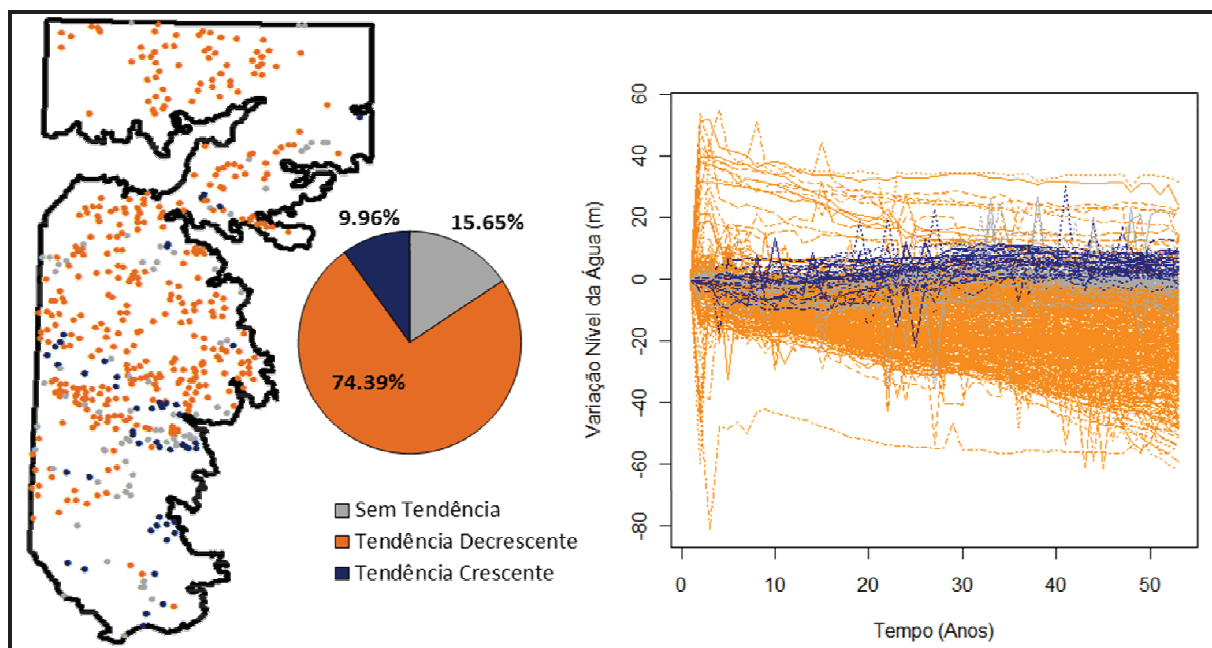


Figura 5.10: Síntese da análise de tendências realizada pelo teste de Mann-Kendall

5.4 IDENTIFICAÇÃO DE ÁREAS CRÍTICAS

O primeiro procedimento para identificar as áreas críticas com redução do nível da água foi por meio da análise dos resultados de evolução da camada saturada (Figuras 5.3, 5.4 e 5.5). De posse do resultado de agrupamento de *clusters*, foi possível localizar geograficamente os pontos pertencentes às séries com decrescimento mais acentuado. A Figura 5.11 mostra a distribuição espacial dos pontos em que foi realizada a análise de *cluster*. Na figura estão representadas as classes de *cluster* obtidas pelo método hierárquico com tendência decrescente.

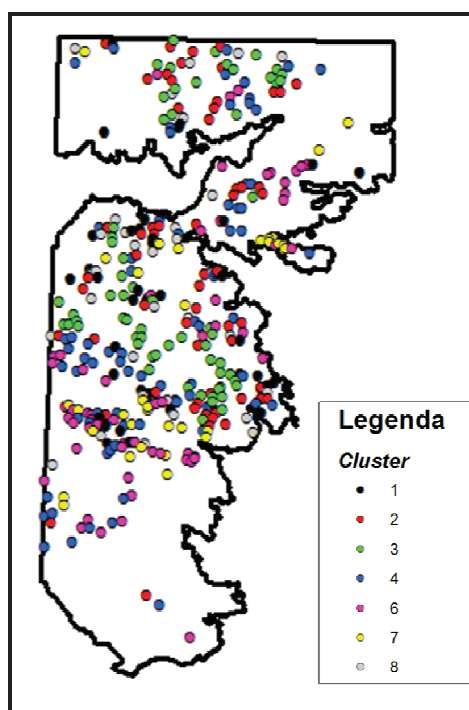


Figura 5.11: Distribuição dos pontos por classe de cluster obtida pelo método hierárquico com tendência decrescente

Na Figura 5.11 é possível notar que a distribuição espacial dos *clusters* é, de modo geral, aleatória. Porém, destaca-se um aglomerado de pontos representados pelo *cluster* 3 próximo ao centro da área de estudo e, ao verificar a Figura 5.9, percebe-se que estes pontos representam o grupo de séries com tendência decrescente mais acentuada. Por isso, considerando estes pontos agrupados pelo *cluster* 3 e a área onde houve maior redução da camada saturada foi definida uma região crítica dentro do aquífero. A Figura 5.12 representa a mudança no nível da camada saturada entre 1962 e 2012, destacando a área crítica com uma borda mais espessa.

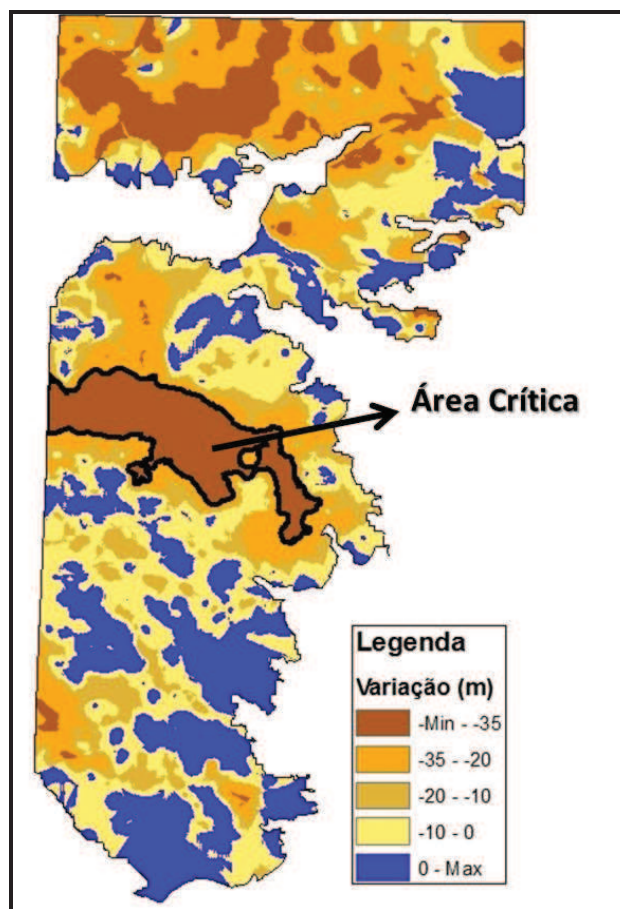


Figura 5.12: Variação no nível da camada saturada entre 1962 e 2012

A Figura 5.12 mostra que houve redução no nível da água em grande parte do aquífero, destacando-se na cor vermelha as áreas com redução maior que 35 metros em 50 anos. Foi analisada como sendo uma área crítica apenas a área destacada na Figura 5.12 por conter uma grande densidade de poços (o que representa confiança nos resultados das interpolações), enquanto que em outras áreas em vermelho não existe tal densidade.

A Figura 5.13 mostra os poços tematizados por cores que representam o resultado do *cluster* hierárquico. É possível perceber um forte predomínio do *cluster* três (cor verde) dentro da área crítica, grupos quais representam as séries de curvas com decréscimo mais acentuado (Figura 5.9). A Figura 5.13 também mostra poços fora da área crítica (até 10 km além do limite da área crítica), a fim de verificar se o comportamento deles difere dos inseridos na área crítica. A princípio se evidencia a presença de poços agrupados por outros *clusters*, o que indica que a região não sofre um decréscimo tão acelerado.

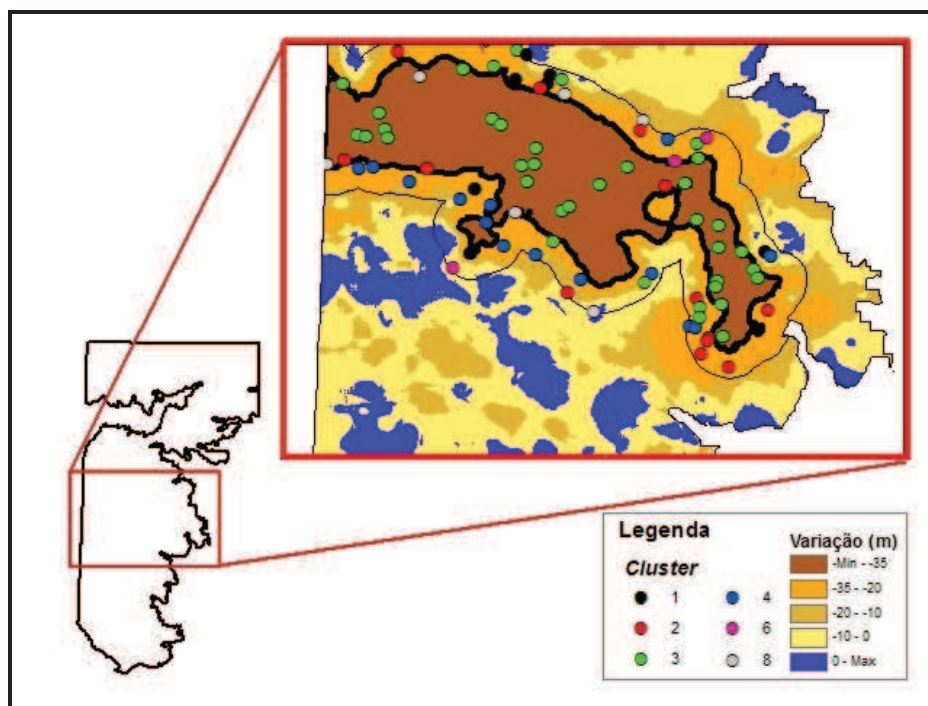


Figura 5.13: Identificação dos *clusters* dos poços inseridos na área crítica e proximidades

A Figura 5.14 mostra dois gráficos com o conjunto de séries contidas dentro do polígono da área crítica e nas proximidades, como foi discutido anteriormente. Nota-se que o padrão das curvas dentro da área crítica é mais homogêneo e de decréscimo mais acentuado, enquanto as séries das proximidades, como eram esperadas, possuem comportamentos distintos e possuem uma menor variação do nível da água de modo geral. Portanto, conclui-se que a área crítica escolhida possui boa representatividade.

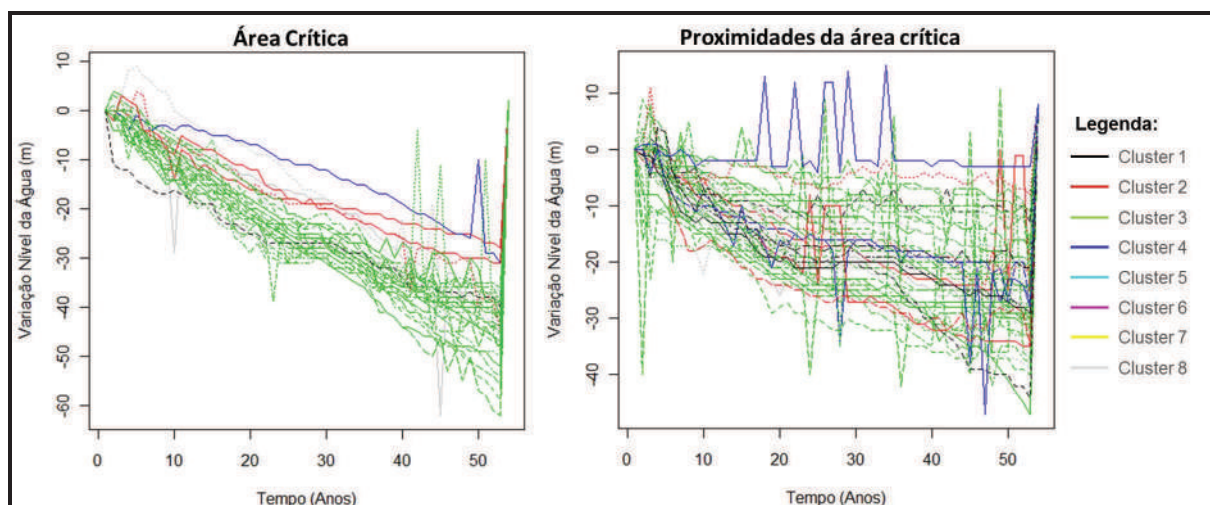


Figura 5.14: Conjunto de séries contidas dentro do polígono da área crítica e em suas proximidades

Por fim, a partir da análise da variação do nível médio da camada saturada na área crítica (Figura 5.15), pode-se perceber a alarmante diminuição dos valores medidos ao longo dos anos.

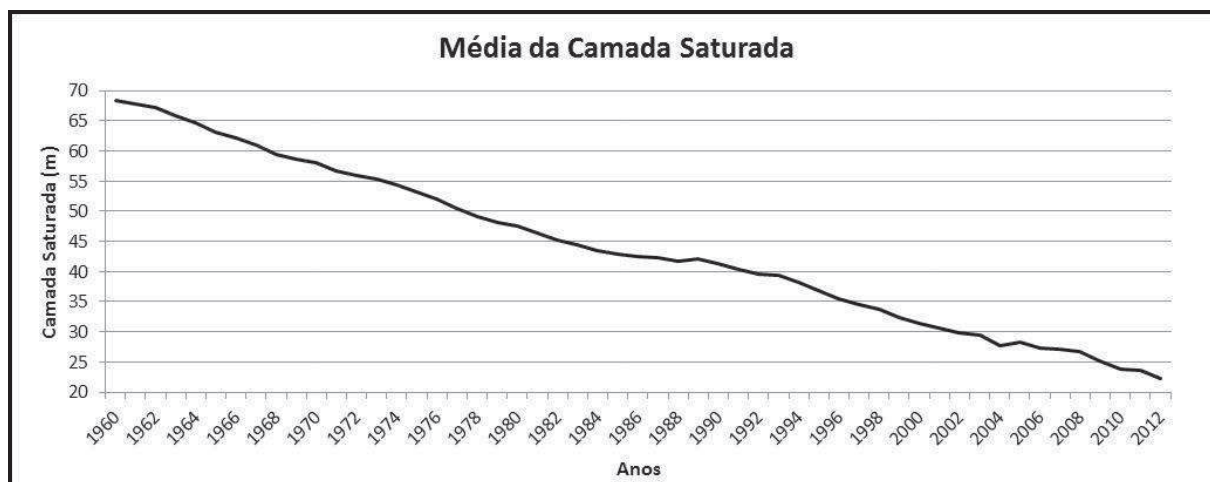


Figura 5.15: Variação do nível médio da camada saturada na área crítica

5.5 INFLUÊNCIA DAS ÁREAS CULTIVADAS

Após a identificação das áreas críticas e da constatação de que existe um declínio do nível da camada saturada em grande parte da área em estudo (Figura 5.12), decidiu-se aprofundar a análise e verificar se atividades antrópicas, como a presença de áreas cultivadas⁴, possuem relação com a variação do nível da água. A hipótese é que a presença de áreas densamente cultivadas exerce influência na variação do nível da água, pois se considera que elas são irrigadas.

Inicialmente, foi verificado se a área crítica identificada nesta pesquisa possui áreas cultivadas. A Figura 5.16 mostra que a área crítica possui uma alta densidade de áreas cultivadas, corroborando com a hipótese. Observa-se que a cor da legenda que representa as áreas cultivadas (Figura 5.16 e Figura 5.17) se altera devido à sobreposição do *layer* de variação no nível da camada saturada.

⁴ Dados de áreas cultivadas retirados da base de dados nacional norte americana de cobertura do solo (MRLC, 2012), disponível em: < <http://www.mrlc.gov/index.php>>.

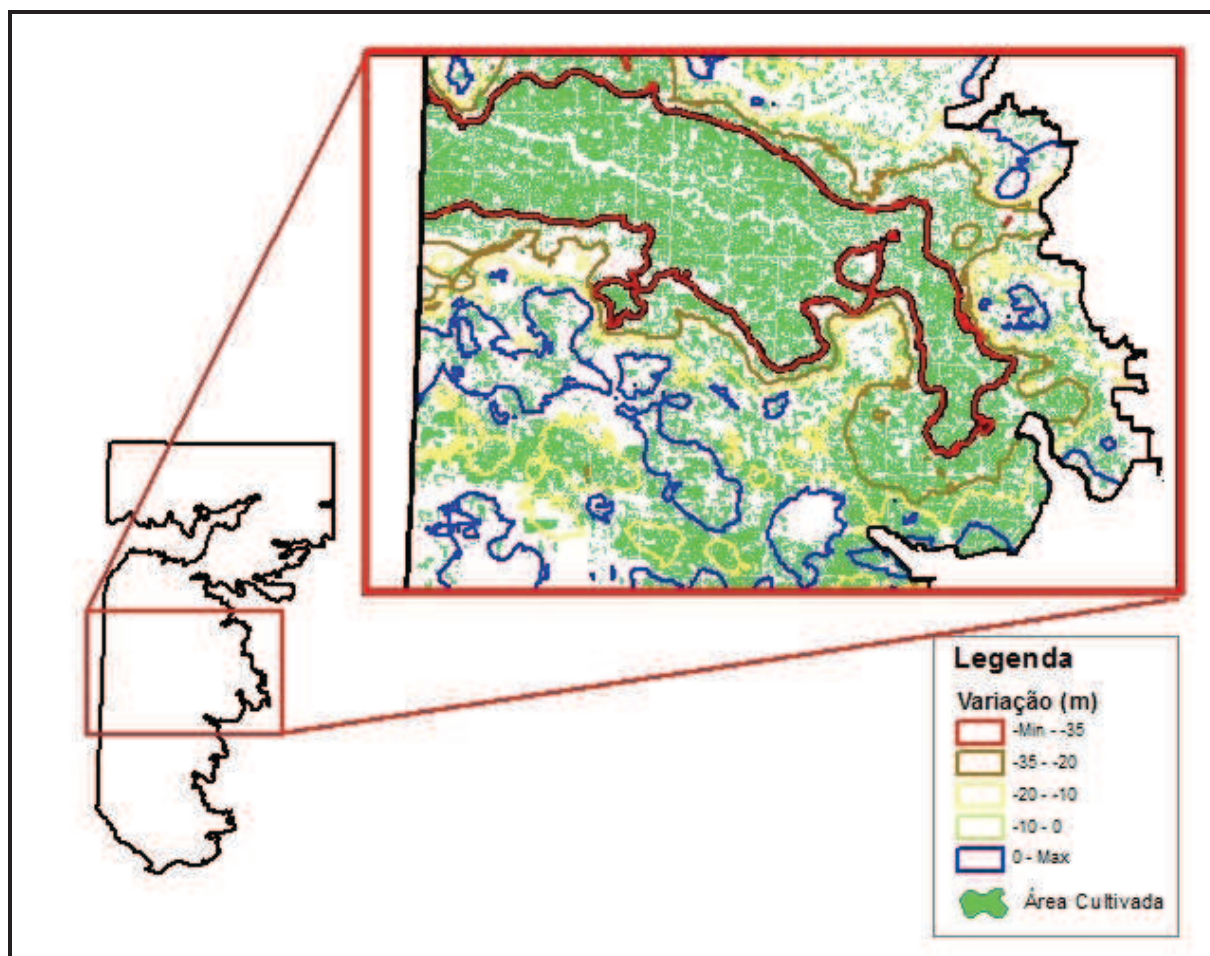


Figura 5.16: Identificação de cultivos dentro da área crítica

Além da área crítica, foram definidas outras áreas com características distintas para se avaliar a variação do nível da camada saturada e sua relação com a presença de áreas cultivadas ao longo dos anos em estudo (Figura 5.17).

A Figura 5.17 mostra a sobreposição das camadas de variação no nível da camada saturada (entre 1962 e 2012) e de áreas cultivadas. De modo geral, pode-se associar a presença de áreas cultivadas com a redução do nível da água e vice-versa. No entanto, também existem áreas onde ocorre o oposto, *e.g.*, variação negativa da camada saturada sem a presença de áreas cultivadas e variação positiva com a presença de áreas cultivadas. Nas poucas áreas sem a presença de cultivos onde ocorreu a redução do nível da camada saturada, considera-se que a interpolação nesta região específica não obteve bom desempenho devido à pequena quantidade de poços existente na área.

A Figura 5.17 destaca quatro áreas onde foram realizadas análises mais aprofundadas. A Área 01 mostra uma região que indica grande diminuição no nível da camada saturada, porém lá não existe uma boa densidade de poços, de modo que os resultados da interpolação

podem oscilar dependendo do ano em estudo. O gráfico da Figura 5.18 confirma esta expectativa, pois entre os anos de 1998 e 2005 houve grandes variações no nível médio. Contudo, o gráfico indica um decrescimento contínuo, mostrando que o resultado obtido, apesar de existirem oscilações, é confiável. Logo, associa-se novamente a presença de áreas cultivadas com a redução do nível da água.

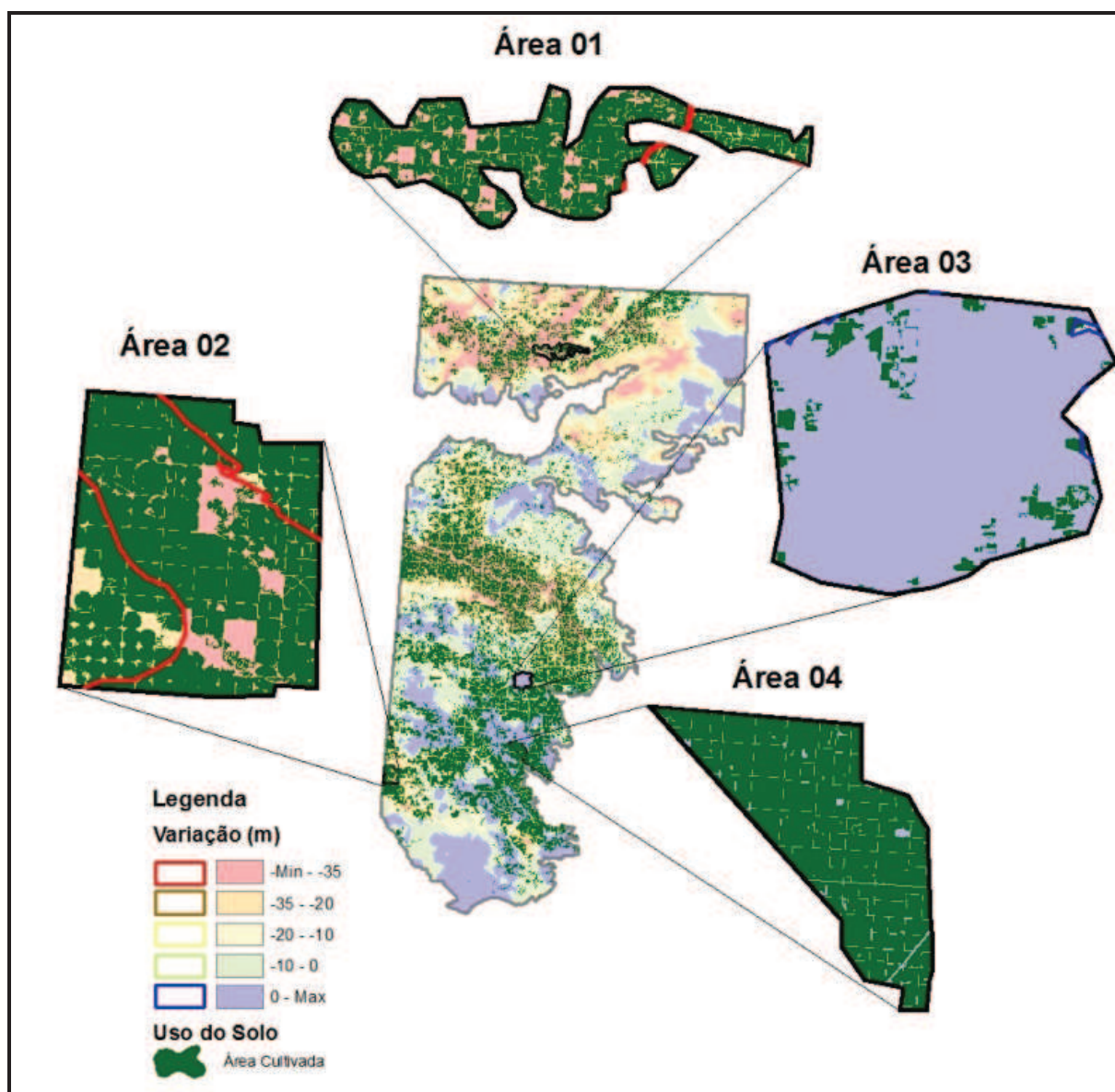


Figura 5.17: Sobreposição das camadas de variação no nível da camada saturada (entre 1962 e 2012) e de áreas cultivadas, destacando áreas para análises mais aprofundadas

A Área 02 (Figura 5.17) também indica uma região com grande densidade de áreas cultivadas, porém esta, com grande quantidade de poços, ou seja, com resultados teoricamente mais confiáveis. O gráfico da Figura 5.18 mostra que houve variações no nível médio da camada saturada e sem apresentar oscilações.

A Área 03 (Figura 5.17) verifica a hipótese, mas de modo inverso, *e.g.*, analisa-se uma área onde ocorre um aumento do nível da camada saturada, porém está sem cultivo. Trata-se de uma área urbana (cidade de Lubbock) e com boa quantidade de poços (provavelmente utilizados para abastecimento). O gráfico da Figura 5.18 apresenta um crescimento contínuo para sua respectiva curva, como era esperado.

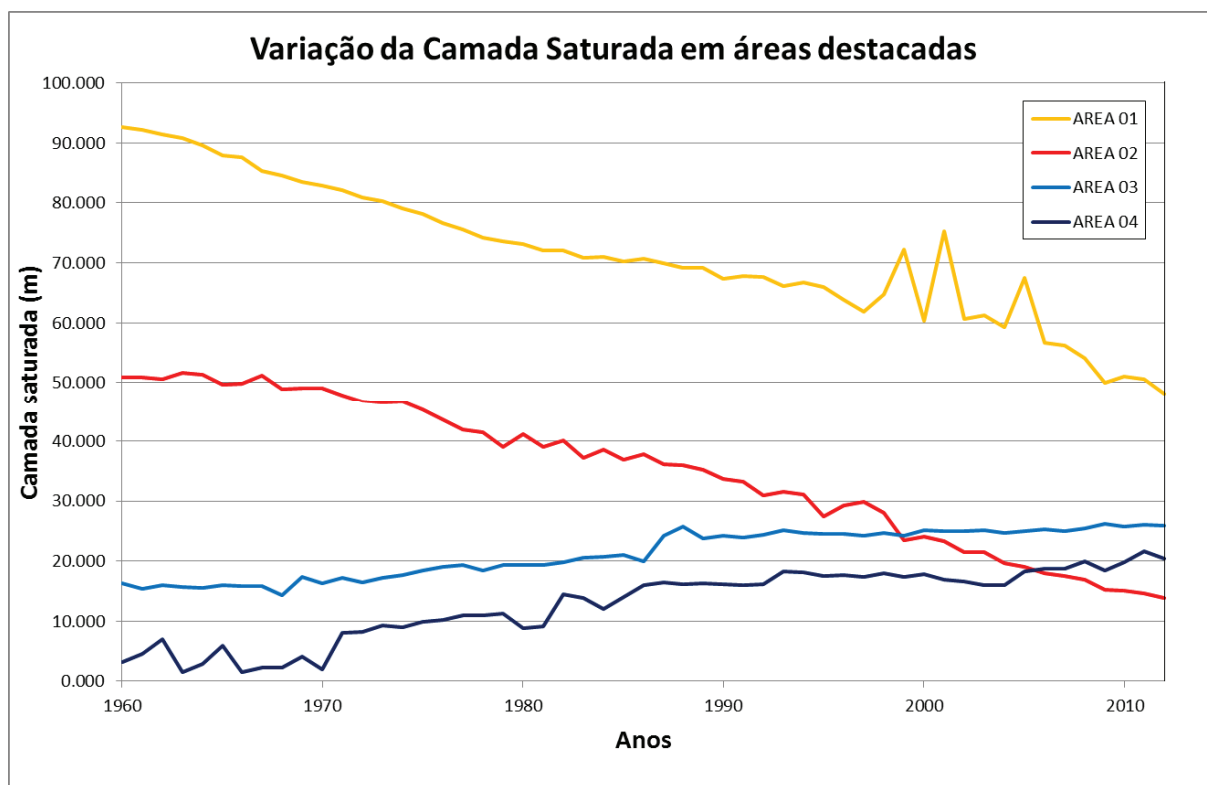


Figura 5.18: Variação da camada saturada média em áreas destacadas

A Área 04 (Figura 5.17) mostra um exemplo contrário à hipótese. Nesta área existe uma densa área cultivada e está aumentando o nível médio da camada saturada (Figura 5.18). Além disso, a região apresenta uma quantidade razoável de poços, o que dá confiabilidade aos resultados. Para identificar o fenômeno que está ocorrendo, é necessário fazer uma investigação mais aprofundada, analisando, por exemplo, vazão de exploração dos poços da área, características hidrogeológicas locais, tipos de cultivos etc. Porém, esse tipo de análise foge do escopo dessa dissertação, sendo recomendado para trabalhos futuros.

6. CONSIDERAÇÕES FINAIS

Ao relembrar os objetivos deste trabalho, elencados no Capítulo 1, e de posse dos resultados do Capítulo 5, pode-se afirmar que o objetivo geral foi atendido, uma vez que foi possível avaliar a disponibilidade da água do aquífero de diversas formas (*e.g.*, estimativas do volume total ao longo da série histórica, análises de variação do nível da água e da camada saturada).

Este trabalho buscou desenvolver uma metodologia capaz de processar grande quantidade de dados de modo rápido, eficaz e aplicável em outras áreas de estudo. De outra maneira, não seria viável realizar tais análises devido ao consumo de tempo. Ressalta-se, por exemplo, a metodologia abordada no item 4.4 (Estimativas de volume do aquífero) que possuía 106 superfícies de entrada (nível da água) e 636 superfícies de saída, sendo processada em minutos, enquanto um processamento manual poderia durar semanas.

A comparação entre os métodos de interpolação mostrou que o IDW não obteve bom desempenho para gerar as superfícies de nível da água. No entanto, os métodos de krigagem e *spline* tiveram resultados bastante semelhantes e desempenhos satisfatórios. Recomenda-se o uso do método *spline* quando se necessita processar múltiplas superfícies de modo automatizado, já a krigagem é mais aconselhável quando se demanda maior acurácia dos resultados.

As estimativas de volume mostraram que está havendo uma preocupante redução no total de água disponível (Figura 5.1), já que houve uma diminuição de 33,9% nos últimos 53 anos. A análise da evolução das camadas saturadas ajudou a ter uma idéia de quais áreas estavam sofrendo maior redução e quais estavam em situação crítica.

A análise temporal dos dados, realizada por meio da análise de *cluster* em conjunto com a análise de tendência, permitiu identificar os grupos de séries que possuíam tendência de decréscimo mais acelerado e, posteriormente, espacializá-los. Esta metodologia, junto à análise da variação da camada saturada entre 1962 e 2012, permitiu definir áreas críticas de redução do nível da água.

A verificação da influência de áreas cultivadas na variação do nível da água mostrou que existe uma forte relação entre os dois fenômenos, porém é necessário aprofundar as análises incluindo outros níveis de informação para se obter diagnósticos mais precisos.

Por fim, recomenda-se para trabalhos futuros:

- Testar outros métodos para o preenchimento de falhas em séries anuais de nível da água em poços, visando melhorar a qualidade dos dados de entrada da análise de tendência;
- Uma pesquisa com o objetivo de prever o nível da água utilizando *e.g.*: modelos de águas subterrâneas, Redes Neurais Artificiais (RNAs) ou modelos estocásticos; e
- Uma pesquisa mais detalhada sobre a influência de áreas cultivadas na variação do nível da água.

REFERÊNCIAS BIBLIOGRÁFICAS

- BRINEY, Amanda. **Aquifers and the Ogallala Aquifer**. Disponível em: <<http://geography.about.com/od/physicalgeography/a/aquifers.htm>>. Acesso em: 11 nov. 2012.
- BARROSO, L. P., ARTES, R. **Análise de Multivariada**. Lavras: UFLA, 2003. 157p.
- BURROUGH, P. A. & MACDONNELL, R. A. 1998. **Principles of Geographical Information Systems. Spatial Information Systems and Geostatistics**. Oxford University Press, New York, 333 p.
- CADIMA, Jorge. **Apontamentos de estatística multivariada**. Universidade Técnica de Lisboa, 2010, 194p.
- CHOW, Ven Te. **Handbook of Applied Hydrology**. New York: McGraw-Hill, 1964.
- CRESSIE, N. **Statistics for spatial data (Revised edition)**, New York: Wiley, 1993, 900p.
- COLLISCHONN, W. (2001). **Simulação hidrológica de grandes bacias**. Tese (Doutorado). Programa de Pós-graduação em Engenharia de Recursos Hídricos e Saneamento Ambiental, IPH, UFRGS, Porto Alegre, 270 p.
- COX, D. R.; HINKLEY, D. V. **Theoretical Statistics**. Chapman and Hall, London, 1974.
- DOBESCH, H. *et al.* **Spatial Interpolation for Climate Data: The Use of GIS in Climatology and Meteorology**. London: ISTE Ltd, 2007, 284p.
- FISHER, R. A. **Statistical Methods and Scientific Inference (3rd edition)**. Edinburg: Oliver and Boyd, 1973
- FOSTER, S. *et al.* **Groundwater Management Strategies: facets of the integrated approach**. GWMATE Briefing Note n.3. 2003
- GRAFTON, R. Q.; HUSSEY, K. **Water Resources Planning and Management**. New York: Cambridge, 2011, 777p.
- GRAY, K. L. **Comparison of Trend Detection Methods**. Tese (Doutorado) University of Montana, Missoula, Montana. 2007, 86 p.
- GURU, M. V.; HORNE, J. E. **The Ogalla Aquifer**. Poteau, Oklahoma: The Kerr Center for Sustainable Agriculture Poteau, 2000, 32p.
- GUTENTAG, E. D. *et al.* **Geohydrology of the High Plains Aquifer In Parts of Colorado, Kansas, Nebraska, New Mexico, Oklahoma, South Dakota, Texas, and Wyoming**. Geological Survey Professional Paper. Washington: United States Government Printing Office, 63p, 1984.

- HAND, D. *et al.* **Principles of Data Mining**. Massachusetts: The MIT Press, 2001, 546p.
- HERMANS, E. *et al.* **Groundwater Management and Exploration Package: State of the Art for Northern China**. FutureWater Report. v. 74, p. 1-52, 2008.
- HIPEL, K. W.; McLEOD, A.I. **Time Series Modelling of Water Resources and Environmental Systems**. Amsterdam: Elsevier, 1994, 1013 p.
- ISAAKS, E.; SRIVASTAVA, R. M. **An introduction to applied geostatistics**. New York: Oxford University Press, 1989, 561p.
- JOHNSTON, K. *et al.* **Using ArcGIS Geostatistical Analyst**. California: ESRI, 2001, 300p.
- KENDALL, M. G. **Rank correlation methods**. Charles Griffin, London, 4th ed., 1975.
- KOHAVI, R. **A study of cross-validation and bootstrap for accuracy estimation and model selection**. In: International joint Conference on artificial intelligence. v. 14, p. 1137–1145, 1995.
- KRAVCHENCO, A.; BULLOCK, D. G. 1999. **A comparative study of interpolation methods for mapping soil properties**. Agronomy Journal, 91: 393-400.
- KUMAR, V; RAMADEVI. **Kriging of groundwater levels – a case study**. Journal of Spatial Hydrology. v.6, no 1, p. 81–94, 2006.
- LIAO, W. T. **Clustering of time series data—a survey**. Journal of Pattern Recognition Society. v. 38, p. 1857-1874, 2005
- MacQUEEN, J. **Some methods for classification and analysis of multivariate observations**. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, p. 281-297, 1967.
- MAIDMENT, D. R. Hydrology cycle. In: **Handbook of Hydrology**. New York: McGraw-Hill, 1993, 1424p.
- MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. **Forecasting Methods and Applications** (Third Edition ed.): John Wiley & Sons, Inc, 1998.
- MANN, H. B. **Nonparametric tests against trend**. Econometrica, 13:245-259, 1945.
- MATHERON, G. **Traité de géostatistique appliquée**. Tome 1, Editions Technip, Paris, 1962, 334 p.
- MATHERON, G. **The theory of regionalized variables and its applications**. Paris, Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, 1971. 211p.
- MILLER, A. J. **Ground Water Atlas of The United States: unconsolidated sand and gravel aquifers**. Disponível em: <http://pubs.usgs.gov/ha/ha730/ch_a/A-text2.html>. Acesso em: 11 nov. 2012.

MINISTÉRIO DO MEIO AMBIENTE - MMA. **Águas Subterrâneas - um recurso a ser conhecido e protegido**. Brasília: Crio, 2007, 37p.

MRLC – Multi-Resolution Land Characteristics Consortium. **National Land Cover Database (NLCD) (2006)**. Disponível em: < <http://www.mrlc.gov/index.php> >. Acesso em: 04 dez. 2012.

NEYMAN, J.; PEARSON, E. S. **On the use and interpretation of certain test criteria for purposes of statistical inference**. Biometrika, Series A, 20:175-240, 263-294, 1928.

NEYMAN, J.; PEARSON, E. S. **On the problem of the most efficient test of statistical hypotheses**. Philosophical Transactions of the Royal Society, Series A, 231:289-337, 1933.

NORUŠIS, Marija J. Cluster Analysis. **In: IBM SPSS Statistics 19 Statistical Procedures Companion**. 2011, 672 p.

RENNIE, J. D. M. **On The Value of Leave-One-Out Cross-Validation Bounds**. 2003. Disponível em: <<http://www.ai.mit.edu/~jrennie/writing/loocv.ps.gz>> Acesso em 15 nov. 2012

TILAHUN, K.; MERKEL, B. J. **Estimation of groundwater recharge using a GIS-based distributed water balance model in Dire Dawa, Ethiopia**. Hydrogeology Journal, v.17, p.1443-1457, 2009.

TODD, D. K.; MAYS, L. W. **Groundwater Hydrology**. John Wiley & Sons, Inc., 2005, 636p.

TWDB. **The High Plains Aquifer System of Texas, 1980 to 1990 – Overview and Projections**. Texas Water Development Board, Report N. 341, 1993

TWDB. **Water for Texas – Technical Planning Appendix**. Texas Water Development Board, Report GP-6-2, 1997

TWDB. **Groundwater Database Reports**. Disponível em: <<http://www.twdb.state.tx.us/groundwater/data/gwdbbrpt.asp>>. Acesso em: 06 ago. 2012.

TERRELL, B. L. *et al.* **Ogallala Aquifer Depletion: Economic Impact on the Texas High Plains**. 26p. 2002.

TUCCI, C. E. M.; CABRAL, J. J. S. P. **Qualidade da Água Subterrânea - Documento Final**. Brasília: CGEE - Centro de Gestão e Estudos Estratégicos, 2003, 59p.

TUCCI, C. E. M.; SILVEIRA, A. L. L.; **Hidrologia: ciência e aplicação**. Porto Alegre: UFRGS/ABRH, 3ed., 2004, 943p.

United Nations Educational, Scientific And Cultural Organization – UNESCO. **Ground Water Resources of the World and their Use**. IHP-VI, Series on Groudwater No. 6, 2004, 346p.

_____. UNESCO. **Groundwater for Emergency Situations**. Proceedings of the International Workshop Tehran, Out, 2006, 128p.

USDA. **Census of Agriculture – Texas State and County Data**. U. S. Department of Agriculture. Geographic Area Series Part 43A, v. 1, 1999.

WEBSTER, R.; OLIVER, M. A. **Geostatistics for Environmental Scientists** - Second Edition. England: John Wiley & Sons, Ltd, 2007, 315p.

WEINSTEIN, J. **A Postgenomic Visual Icon**. Science, 319, 1772{1773, 2008

YAN, Mingjin. **Methods of Determining the Number of Clusters in a Data Set and a New Clustering Criterion**. Blacksburg: Virginia, 2005, 120 p. Tese de Doutorado (Estatística). Virginia Polytechnic Institute and State University, 2005.

ZAIANE, O. R. *et al.* **On data clustering analysis: scalability, constraints and validation**. Edmonton Alberta, University of Alberta, 2003.

APÊNDICES

APÊNDICE A – Consultas SQL para a obtenção dos dados de nível da água

```
-----
-- Criação da tabela "waterelev_join" por meio da junção tabular entre as tabelas
-- "ogallalawells" e "waterelev", utilizando como chave o código dos poços
-- Dados de saída: código dos poços (com repetição de valores em mais de uma
-- linha), elevação do terreno, código do aquífero, profundidade do poço, mês de
-- coleta, dia de coleta e ano de coleta

CREATE TABLE waterelev_join AS
SELECT o.state_well, o.elev_of_ls, o.aquifer_co, w.depth_from, w.mm_date,
w.dd_date, w.yy_date
FROM ogallalawells o, waterelev w
WHERE o.state_well = w.state_well

-- Tratamento dos dados da tabela "waterelev_join"
-- Apagar linhas que possuem valores de profundidades dos poços nulos

DELETE FROM waterelev_join WHERE depth_from is null;

-- Adicionar a coluna "wte"

ALTER TABLE waterelev_join ADD COLUMN wte numeric(6,2);

-- Preencher as células correspondentes a coluna "wte" subtraindo o nível do
-- terreno pela profundidade do poço

UPDATE waterelev_join SET wte = (elev_of_ls + depth_from);

-----
-- Criação da tabela "waterelev_join_exclude_3_to_10" excluindo os dados coletados
-- entre os meses de março a outubro

CREATE TABLE waterelev_join_exclude_3_to_10 AS
SELECT state_well, elev_of_ls, aquifer_co, depth_from, mm_date, dd_date, yy_date,
wte
FROM waterelev_join
WHERE mm_date not in (0,3,4,5,6,7,8,9,10)
ORDER BY state_well, yy_date;

-- Criar coluna com o nível de água anual

ALTER TABLE waterelev_join_exclude_3_to_10 ADD COLUMN year numeric (4,0);

-- Atribuir aos meses de novembro e dezembro a data do ano seguinte

UPDATE waterelev_join_exclude_3_to_10 SET year=year+1
WHERE mm_date = 12 or mm_date = 11
```

```
-----  
  
-- Criação da tabela "wells_timeseries" contendo apenas três colunas: códigos dos  
-- poços, ano de coleta e nível da água, onde o nível da água será o valor médio  
-- entre os níveis da água quando houver mais de uma coleta no ano
```

```
CREATE TABLE wells_timeseries AS  
SELECT state_well, year, avg(wte) as wte  
from waterelev_join_exclude_3_to_10  
group by state_well, year;
```

```
-----  
  
-- Criação da tabela "wells"
```

```
CREATE TABLE wells AS  
SELECT DISTINCT state_well  
FROM wells_timeseries;
```

```
-- Criação das colunas da tabela "wells"
```

```
ALTER TABLE wells ADD COLUMN wte_1914 NUMERIC  
ALTER TABLE wells ADD COLUMN wte_1915 NUMERIC  
ALTER TABLE wells ADD COLUMN wte_1916 NUMERIC  
  
...  
ALTER TABLE wells ADD COLUMN wte_2011 NUMERIC  
ALTER TABLE wells ADD COLUMN wte_2012 NUMERIC
```

```
-- Preenchimento das células da tabela "wells"
```

```
UPDATE wells SET wte_1914 = wells_timeseries.wte  
FROM wells_timeseries  
WHERE wells_timeseries.state_well = wells.state_well AND wells_timeseries.year =  
1914  
UPDATE wells SET wte_1915 = wells_timeseries.wte  
FROM wells_timeseries  
WHERE wells_timeseries.state_well = wells.state_well AND wells_timeseries.year =  
1915  
  
...  
UPDATE wells SET wte_2011 = wells_timeseries.wte  
FROM wells_timeseries  
WHERE wells_timeseries.state_well = wells.state_well AND wells_timeseries.year =  
2011  
UPDATE wells SET wte_2012 = wells_timeseries.wte  
FROM wells_timeseries  
WHERE wells_timeseries.state_well = wells.state_well AND wells_timeseries.year =  
2012
```

APÊNDICE B – Criação da rotina em Visual Basic para preenchimento de falhas

```
Sub Preench_Falhas()
Dim MRow As Double
Dim Ma As Double
Dim Mb As Double
Dim Mc As Double
Dim Pa As Double
Dim Pb As Double
Dim Pc As Double
Dim Row As Integer
Dim Col As Integer
MRow = 0
Ma = 0
Mb = 0
Mc = 0
Pa = 0
Pb = 0
Pc = 0
For Col = 4 To 56 Step 1
Row = 5
while Cells(498, Col) <> Cells(499, Col)
while Cells(Row, Col).Value <> 99999
MRow = Cells(Row, 59).Value
If Cells(Row, Col).Value = "" Then
Ma = Cells((Cells(Row, 71).Value), 59).Value
Pa = Cells((Cells(Row, 71).Value), Col).Value
Mb = Cells((Cells(Row, 73).Value), 59).Value
Pb = Cells((Cells(Row, 73).Value), Col).Value
Mc = Cells((Cells(Row, 75).Value), 59).Value
Pc = Cells((Cells(Row, 75).Value), Col).Value
If Pa = 0 Then
Ma = Cells((Cells(Row, 77).Value), 59).Value
Pa = Cells((Cells(Row, 77).Value), Col).Value
Else
End If
If Pb = 0 And Ma = Cells((Cells(Row, 77).Value), 59).Value Then
Pb = Cells((Cells(Row, 79).Value), Col).Value
Mb = Cells((Cells(Row, 79).Value), 59).Value
Else
If Pb = 0 Then
Pb = Cells((Cells(Row, 77).Value), Col).Value
Mb = Cells((Cells(Row, 77).Value), 59).Value
Else
End If
End If

If Pc = 0 And Ma = Cells((Cells(Row, 77).Value), 59).Value And Mb =
Cells((Cells(Row, 79).Value), 59).Value Then
```

```

        Pc = Cells((Cells(Row, 81).Value), Col).Value
        Mc = Cells((Cells(Row, 81).Value), 59).Value
    Else
        If Pc = 0 And (Ma = Cells((Cells(Row, 77).Value), 59).Value Or Mb =
Cells((Cells(Row, 77).Value), 59).Value) Then
            Pc = Cells((Cells(Row, 79).Value), Col).Value
            Mc = Cells((Cells(Row, 79).Value), 59).Value
        Else
            If Pc = 0 Then
                Pc = Cells((Cells(Row, 77).Value), Col).Value
                Mc = Cells((Cells(Row, 77).Value), 59).Value
            Else
                End If
            End If
        End If
        Cells(Row, Col).Value = ((MRow * Pa / Ma) + (MRow * Pb / Mb) + (MRow *
Pc / Mc)) / 3
    Else
        End If
    Row = Row + 1
wend
Row = 5
while Cells(Row, Col).Value <> 99999
    If Cells(Row, Col).Value < Cells(2, Col).Value Then
        Cells(Row, Col).Value = ""
    Else
        End If
    If Cells(Row, Col).Value < 0.96 * (Cells(Row, 59).Value) Then
        Cells(Row, Col).Value = ""
    Else
        End If
    If Cells(Row, Col).Value > 1.032 * (Cells(Row, 59).Value) Then
        Cells(Row, Col).Value = ""
    Else
        End If
    Row = Row + 1
wend
    Cells(498, Col).Select
    Selection.Copy
    Selection.Offset(1, 0).Select
    Selection.PasteSpecial Paste:=xlPasteValues, Operation:=xlNone, SkipBlanks
-
        :=False, Transpose:=False
    Application.CutCopyMode = False
wend
Next
End Sub

```

APÊNDICE C – Configurações das krigagens realizadas entre 1960 e 2012

Year – 1960	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	7
Include at least:	2
Sector type:	Four
Semivariogram	
Type:	Circular
Number of lags:	12
Lag Size:	50084
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year – 1962	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,1
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Full
Semivariogram	
Type:	Gaussian
Number of lags:	12
Lag Size:	50000
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1961	
Method	
Type:	Universal Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	6
Include at least:	2
Sector type:	Full
Semivariogram	
Type:	Circular
Number of lags:	12
Lag Size:	50000
Nugget	0.00E+00
Anisotropy	Yes
Direction	303.2621059

Year - 1963	
Method	
Type:	Universal Kriging
Exploratory Analysis	
Transformation:	BoxCox 0.4
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	8
Include at least:	2
Sector type:	Full
Semivariogram	
Type:	Circular
Number of lags:	12
Lag Size:	50084
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1964	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	8
Include at least:	2
Sector type:	Full
Semivariogram	
Type:	Circular
Number of lags:	12
Lag Size:	51000
Nugget	0.00E+00
Anisotropy	Yes
Direction	284.1794729

Year - 1966	
Method	
Type:	Universal Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,7
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	6
Include at least:	2
Sector type:	Full
Semivariogram	
Type:	Spherical
Number of lags:	8
Lag Size:	60000
Nugget	3.12E+00
Anisotropy	No
Direction	No

Year - 1965	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,6
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Full
Semivariogram	
Type:	Gaussian
Number of lags:	10
Lag Size:	50000
Nugget	4.78E-01
Anisotropy	Yes
Direction	294.5295372

Year - 1967	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,5
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Full
Semivariogram	
Type:	Circular
Number of lags:	11
Lag Size:	80000
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1968	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Full
Semivariogram	
Type:	Spherical
Number of lags:	11
Lag Size:	80000
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1970	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,9
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	7
Include at least:	2
Sector type:	Full
Semivariogram	
Type:	Circular
Number of lags:	10
Lag Size:	80580
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1969	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,6
Trend Type:	Second
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Full
Semivariogram	
Type:	Spherical
Number of lags:	7
Lag Size:	68820
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1971	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,01
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	7
Include at least:	2
Sector type:	Eight
Semivariogram	
Type:	Circular
Number of lags:	11
Lag Size:	100820
Nugget	0.00E+00
Anisotropy	Yes
Direction	290.029

Year - 1972	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Full
Semivariogram	
Type:	Circular
Number of lags:	10
Lag Size:	58410
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1974	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Full
Semivariogram	
Type:	Circular
Number of lags:	10
Lag Size:	88820
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1973	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	6
Include at least:	2
Sector type:	Full
Semivariogram	
Type:	Circular
Number of lags:	10
Lag Size:	58820
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1975	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	7
Include at least:	3
Sector type:	Four
Semivariogram	
Type:	Spherical
Number of lags:	10
Lag Size:	80760
Nugget	0.00E+00
Anisotropy	Yes
Direction	292.469305

Year - 1976	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	6
Include at least:	3
Sector type:	Four
Semivariogram	
Type:	Spherical
Number of lags:	10
Lag Size:	80820
Nugget	0.00E+00
Anisotropy	Yes
Direction	291.1302204

Year - 1978	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	6
Include at least:	3
Sector type:	Four and 45 degree
Semivariogram	
Type:	Spherical
Number of lags:	10
Lag Size:	86820
Nugget	0.00E+00
Anisotropy	Yes
Direction	287.1106339

Year - 1977	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	8
Include at least:	3
Sector type:	Eight
Semivariogram	
Type:	Spherical
Number of lags:	10
Lag Size:	85820
Nugget	0.00E+00
Anisotropy	Yes
Direction	289.3877811

Year - 1979	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	None
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	6
Include at least:	3
Sector type:	Four and 45 degree
Semivariogram	
Type:	Spherical
Number of lags:	10
Lag Size:	88480
Nugget	0.00E+00
Anisotropy	Yes
Direction	278.7255716

Year - 1980	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 1
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Four and 45 degree
Semivariogram	
Type:	Spherical
Number of lags:	12
Lag Size:	142510
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1982	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Four and 45 Degree
Semivariogram	
Type:	Spherical
Number of lags:	11
Lag Size:	142160
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1981	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,9
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Four
Semivariogram	
Type:	Spherical
Number of lags:	11
Lag Size:	139280
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1983	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Four and 45 degree
Semivariogram	
Type:	Sphercial
Number of lags:	11
Lag Size:	140130
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1984	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Four and 45 degree
Semivariogram	
Type:	Circular
Number of lags:	11
Lag Size:	135960
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1986	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,7
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	9
Include at least:	3
Sector type:	Full
Semivariogram	
Type:	Circular
Number of lags:	10
Lag Size:	11542
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1985	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,5
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	8
Include at least:	3
Sector type:	Four
Semivariogram	
Type:	Spherical
Number of lags:	10
Lag Size:	80760
Nugget	0.00E+00
Anisotropy	Yes
Direction	278.7544584

Year - 1987	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	None
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	3
Sector type:	Four and 45 degree
Semivariogram	
Type:	Tetraspherical
Number of lags:	9
Lag Size:	92510
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1988	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	7
Include at least:	3
Sector type:	Four
Semivariogram	
Type:	Circular
Number of lags:	11
Lag Size:	139870
Nugget	0.00E+00
Anisotropy	Yes
Direction	285.5030222

Year - 1990	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,4
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Four and 45 degree
Semivariogram	
Type:	Spherical
Number of lags:	12
Lag Size:	142510
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1989	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,1
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	9
Include at least:	3
Sector type:	Four
Semivariogram	
Type:	Spherical
Number of lags:	10
Lag Size:	131010
Nugget	0.00E+00
Anisotropy	Yes
Direction	287.58

Year - 1991	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,9
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	8
Include at least:	3
Sector type:	Eight
Semivariogram	
Type:	Circular
Number of lags:	11
Lag Size:	122510
Nugget	0.00E+00
Anisotropy	Yes
Direction	281.068819

Year - 1992	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,3
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Four and 45 Degree
Semivariogram	
Type:	Spherical
Number of lags:	11
Lag Size:	142510
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1994	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,1
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	8
Include at least:	3
Sector type:	Four and 45 degree
Semivariogram	
Type:	Spherical
Number of lags:	10
Lag Size:	142510
Nugget	0.00E+00
Anisotropy	Yes
Direction	289.67

Year - 1993	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Four and 45 degree
Semivariogram	
Type:	Spherical
Number of lags:	11
Lag Size:	142510
Nugget	0.00E+00
Anisotropy	Yes
Direction	288.393013

Year - 1995	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Four and 45 degree
Semivariogram	
Type:	Spherical
Number of lags:	11
Lag Size:	142510
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1996	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	9
Include at least:	3
Sector type:	Eight
Semivariogram	
Type:	Circular
Number of lags:	12
Lag Size:	102510
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1998	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	8
Include at least:	3
Sector type:	Four and 45 degree
Semivariogram	
Type:	Circular
Number of lags:	10
Lag Size:	142510
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1997	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	8
Include at least:	3
Sector type:	Four and 45 degree
Semivariogram	
Type:	Circular
Number of lags:	11
Lag Size:	142510
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 1999	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	8
Include at least:	3
Sector type:	Eight
Semivariogram	
Type:	Exponential
Number of lags:	10
Lag Size:	142510
Nugget	0.00E+00
Anisotropy	Yes
Direction	314.7230272

Year - 2000	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,4
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	10
Include at least:	2
Sector type:	Four
Semivariogram	
Type:	Gaussian
Number of lags:	11
Lag Size:	4510
Nugget	1.07E-02
Anisotropy	No
Direction	No

Year - 2002	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,3
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	10
Include at least:	2
Sector type:	Four and 45 Degree
Semivariogram	
Type:	Spherical
Number of lags:	10
Lag Size:	49210
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 2001	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,7
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	8
Include at least:	2
Sector type:	Four
Semivariogram	
Type:	Spherical
Number of lags:	8
Lag Size:	42510
Nugget	3.97E-01
Anisotropy	No
Direction	No

Year - 2003	
Method	
Type:	Universal Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,4
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	10
Include at least:	2
Sector type:	Four and 45 degree
Semivariogram	
Type:	Spherical
Number of lags:	10
Lag Size:	15860
Nugget	7.98E-03
Anisotropy	No
Direction	No

Year - 2004	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	BoxCox 0,4
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	8
Include at least:	2
Sector type:	Four and 45 degree
Semivariogram	
Type:	Spherical
Number of lags:	12
Lag Size:	38860
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 2006	
Method	
Type:	Universal Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	10
Include at least:	2
Sector type:	Four
Semivariogram	
Type:	Tetraspherical
Number of lags:	10
Lag Size:	21430
Nugget	0.00E+00
Anisotropy	Yes
Direction	277.24

Year - 2005	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Four and 45 degree
Semivariogram	
Type:	Spherical
Number of lags:	12
Lag Size:	38940
Nugget	1.84E-05
Anisotropy	No
Direction	No

Year - 2007	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Four and 45 degree
Semivariogram	
Type:	Spherical
Number of lags:	12
Lag Size:	38940
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 2008	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	5
Include at least:	2
Sector type:	Four and 45 degree
Semivariogram	
Type:	Circular
Number of lags:	12
Lag Size:	38940
Nugget	0.00E+00
Anisotropy	No
Direction	No

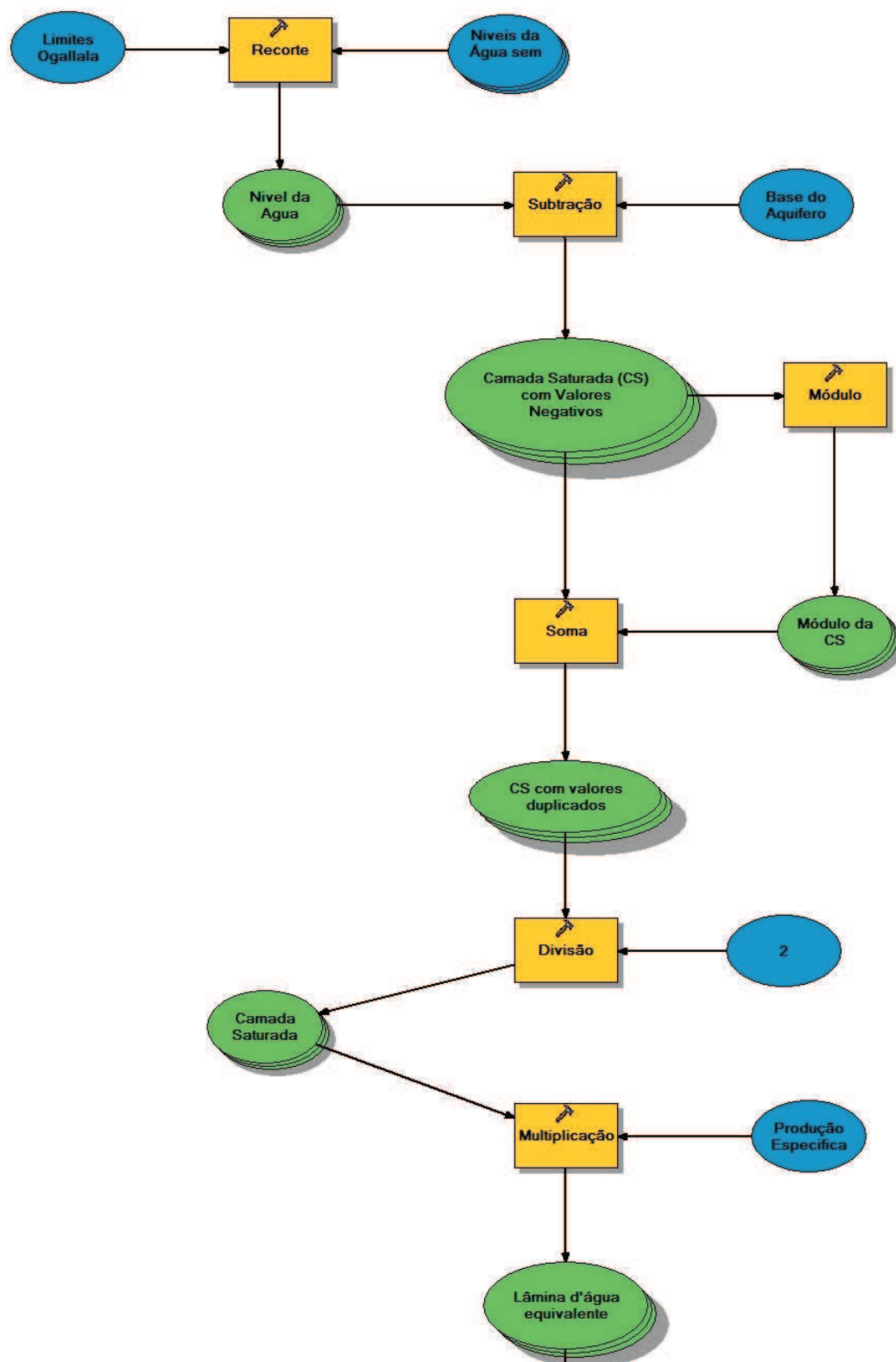
Year - 2010	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	7
Include at least:	2
Sector type:	Four and 45 degree
Semivariogram	
Type:	Spherical
Number of lags:	12
Lag Size:	34100
Nugget	0.00E+00
Anisotropy	No
Direction	No

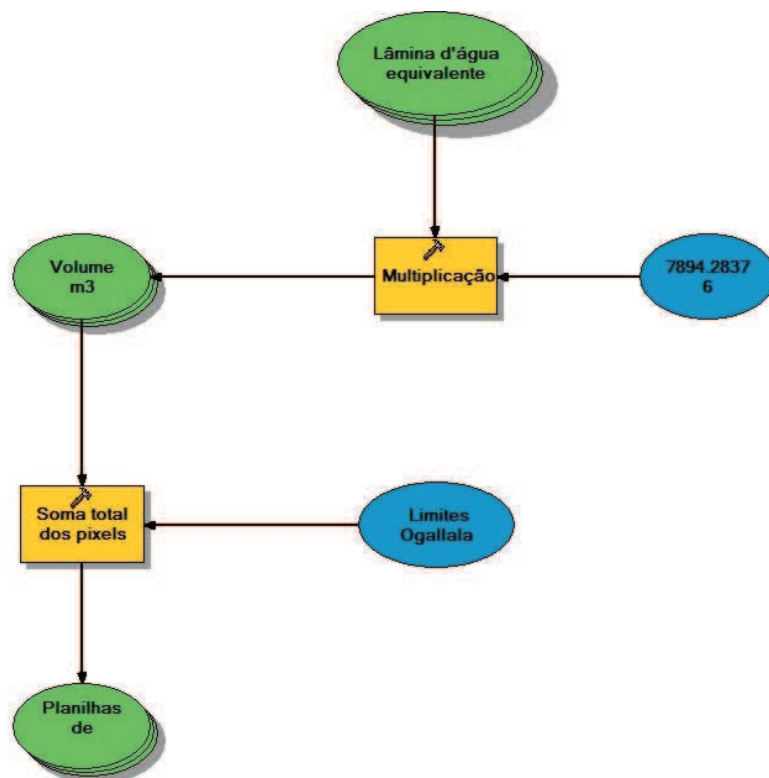
Year - 2009	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Log
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	8
Include at least:	2
Sector type:	Eight
Semivariogram	
Type:	Pentaspherical
Number of lags:	10
Lag Size:	38940
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 2011	
Method	
Type:	Ordinary Kriging
Exploratory Analysis	
Transformation:	Box Cox 0,3
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	8
Include at least:	1
Sector type:	Four and 45 degree
Semivariogram	
Type:	Spherical
Number of lags:	10
Lag Size:	40000
Nugget	0.00E+00
Anisotropy	No
Direction	No

Year - 2012	
Method	
Type:	Universal Kriging
Exploratory Analysis	
Transformation:	Box Cox 0.1
Trend Type:	First
Searching Neighborhood	
Type:	Standard
Neighbors to Include:	10
Include at least:	2
Sector type:	Four
Semivariogram	
Type:	Tetraspherical
Number of lags:	10
Lag Size:	32700
Nugget	2.49E-05
Anisotropy	Yes
Direction	277.27

APÊNDICE D – Fluxograma do processamento das imagens para a geração dos resultados





APÊNDICE E – Comandos utilizados para processar os dados no *software* R.

```
# Definindo Diretório
setwd("C:\\...\\Mestrado\\ogallala\\Analises de Tendencia\\Processamentos R")

##### Método k-means #####

# Leitura da tabela de dados
dados = read.table("banco.txt", header = T, sep = "\\t", dec = ",")
x = t(dados[, -c(1,2,3)])

# Gráfico multidimensional
matplot(x, type = "l")

# Definindo numero de clusters
fit1 = kmeans(t(x), 8)

#Definindo arquivo de saida
z = rbind(x, fit1$cluster)

# Salvar em csv
write.csv2(z, 'C:\\...\\Mestrado\\ogallala\\Analises de Tendencia\\Processamentos
R\\k-means.csv')

##### Configuracao dos graficos #####

# Leitura da tabela de dados
trend<-read.csv("Trend_Analysis.csv", as.is=TRUE)
x = t(trend[, -c(1,2,3,4,5,6,7,8,9,10)])

# Metodo K-means #

#selecionando coluna com grupos de cluster
kmeans=trend$kmeans

#Grafico multidimensional colorido por grupo de cluster
matplot(x, type = "l", col = kmeans, ylab="Variação Nível da Água (m)", xlab = "Tempo
(Anos)")

#Definindo matrizes com as series de cada cluster
zk = rbind(x, kmeans)
kmeans1 = zk[, kmeans==1]
kmeans2 = zk[, kmeans==2]
kmeans3 = zk[, kmeans==3]
kmeans4 = zk[, kmeans==4]
kmeans5 = zk[, kmeans==5]
kmeans5 = zk[, kmeans==6]
kmeans5 = zk[, kmeans==7]
```

```
kmeans5 = zk[,kmeans==8]
```

```
#Grafico de cada cluster com suas respectivas cores
matplot(kmeans1, type = "l", col = 1 ,ylab="Variação Nível da Água (m)",xlab =
"Tempo (Anos)")
matplot(kmeans2, type = "l", col = 2 ,ylab="Variação Nível da Água (m)",xlab =
"Tempo (Anos)")
matplot(kmeans3, type = "l", col = 3,ylab="Variação Nível da Água (m)",xlab =
"Tempo (Anos)")
matplot(kmeans4, type = "l", col = 4,ylab="Variação Nível da Água (m)",xlab =
"Tempo (Anos)")
matplot(kmeans5, type = "l", col = 5,ylab="Variação Nível da Água (m)",xlab =
"Tempo (Anos)")
matplot(kmeans5, type = "l", col = 6,ylab="Variação Nível da Água (m)",xlab =
"Tempo (Anos)")
matplot(kmeans5, type = "l", col = 7,ylab="Variação Nível da Água (m)",xlab =
"Tempo (Anos)")
matplot(kmeans5, type = "l", col = 8,ylab="Variação Nível da Água (m)",xlab =
"Tempo (Anos)")
```

```
# Metodo Hierarquico #
```

```
#Selecionando coluna com grupos de cluster
Hcluster=trend$Hcluster
```

```
#Grafico multidimensional colorido por grupo de cluster
matplot(x, type = "l", col = Hcluster, ylab="Variação Nível da Água (m)",xlab =
"Tempo (Anos)")
zh = rbind(x,Hcluster)
```

```
#Definindo matrizes com as series de cada cluster
```

```
Hcluster1 = zk[,Hcluster==1]
Hcluster2 = zk[,Hcluster==2]
Hcluster3 = zk[,Hcluster==3]
Hcluster4 = zk[,Hcluster==4]
Hcluster5 = zk[,Hcluster==5]
Hcluster6 = zk[,Hcluster==6]
Hcluster7 = zk[,Hcluster==7]
Hcluster8 = zk[,Hcluster==8]
```

```
#Grafico de cada cluster com suas respectivas cores
```

```
matplot(Hcluster1, type = "l", col = "black",ylab="Variação Nível da Água (m)",xlab =
"Tempo (Anos)")
matplot(Hcluster2, type = "l", col = "red",ylab="Variação Nível da Água (m)",xlab =
"Tempo (Anos)")
matplot(Hcluster3, type = "l", col = "green3",ylab="Variação Nível da Água
(m)",xlab = "Tempo (Anos)")
matplot(Hcluster4, type = "l", col = "blue",ylab="Variação Nível da Água (m)",xlab
= "Tempo (Anos)")
```

```

matplot(Hcluster5, type = "l", col = "cyan",ylab="Variação Nível da Água (m)",xlab
= "Tempo (Anos)")
matplot(Hcluster6, type = "l", col = "magenta",ylab="Variação Nível da Água
(m)",xlab = "Tempo (Anos)")
matplot(Hcluster7, type = "l", col = "yellow",ylab="Variação Nível da Água
(m)",xlab = "Tempo (Anos)")
matplot(Hcluster8, type = "l", col = "grey",ylab="Variação Nível da Água (m)",xlab
= "Tempo (Anos)")

```

```
##### Teste de Mann-Kendall #####
```

```

# Leitura da tabela de dados
trend<-read.csv("Trend_Analysis.csv",as.is=TRUE)
dados1 = read.table("BancoTeste.txt", header = T, sep = "\t", dec = ",")

```

```

#carregando biblioteca
library(kendall)

```

```
### Teste de Mann- Kendall por série ###
```

```

# p-valor do teste
mann = matrix(0,492,1,byrow=TRUE)
for(i in 1:492){
mann[i] = MannKendall(dados1[i,])$s1
}
mann

```

```

#salvando a saída com o p-valor
write.csv2(mann, 'C:\\...\\Mestrado\\ogallala\\Analises de
Tendencia\\Processamentos R\\saidasp_valor.csv')

```

```

# Valor do tau
tau = matrix(0,492,1,byrow=TRUE)
for(i in 1:492){
tau[i] = MannKendall(dados1[i,])$tau
}
tau

```

```

#salvando a saída com o valor do tau
write.csv2(mann, 'C:\\...\\Mestrado\\ogallala\\Analises de
Tendencia\\Processamentos R\\saidas_tau.csv')

```

```
### Teste de Mann- Kendall por grupos (metodo k-means) ###
```

```

#selecionando as series de cada cluster
kserie1 = trend[kmeans==1,]
kserie1 = kserie1 [,-c(1,2,3,4,5,6,7,8,9,10)]

```

```

kserie2 = trend[kmeans==2,]
kserie2 = kserie2 [,-c(1,2,3,4,5,6,7,8,9,10)]
kserie3 = trend[kmeans==3,]
kserie3 = kserie3 [,-c(1,2,3,4,5,6,7,8,9,10)]
kserie4 = trend[kmeans==4,]
kserie4 = kserie4 [,-c(1,2,3,4,5,6,7,8,9,10)]
kserie5 = trend[kmeans==5,]
kserie5 = kserie5 [,-c(1,2,3,4,5,6,7,8,9,10)]
kserie6 = trend[kmeans==6,]
kserie6 = kserie6 [,-c(1,2,3,4,5,6,7,8,9,10)]
kserie7 = trend[kmeans==7,]
kserie7 = kserie7 [,-c(1,2,3,4,5,6,7,8,9,10)]
kserie8 = trend[kmeans==8,]
kserie8 = kserie8 [,-c(1,2,3,4,5,6,7,8,9,10)]

```

Teste de Mann- Kendall por grupos (Metodo Hierarquico)

```

#selecionando as series de cada cluster
hserie1 = trend[Hcluster==1,]
hserie1 = hserie1 [,-c(1,2,3,4,5,6,7,8,9,10)]
hserie2 = trend[Hcluster==2,]
hserie2 = hserie2 [,-c(1,2,3,4,5,6,7,8,9,10)]
hserie3 = trend[Hcluster==3,]
hserie3 = hserie3 [,-c(1,2,3,4,5,6,7,8,9,10)]
hserie4 = trend[Hcluster==4,]
hserie4 = hserie4 [,-c(1,2,3,4,5,6,7,8,9,10)]
hserie5 = trend[Hcluster==5,]
hserie5 = hserie5 [,-c(1,2,3,4,5,6,7,8,9,10)]
hserie6 = trend[Hcluster==6,]
hserie6 = hserie6 [,-c(1,2,3,4,5,6,7,8,9,10)]
hserie7 = trend[Hcluster==7,]
hserie7 = hserie7 [,-c(1,2,3,4,5,6,7,8,9,10)]
hserie8 = trend[Hcluster==8,]
hserie8 = hserie8 [,-c(1,2,3,4,5,6,7,8,9,10)]

```

APÊNDICE F – Tabela com os resultados do teste de Mann-Kendall e agrupamentos

ID	Código	p-valor	τ	99% de confiança		90% de confiança		Agrupamentos	
				Hipótese	Tendência	Hipótese	Tendência	<i>k-means</i>	Hierárquico
1	233914	4.25E-19	-0.837469	H1	-1	H1	-1	1	4
2	234804	5.96E-19	-0.833683	H1	-1	H1	-1	3	7
3	236102	2.07E-17	-0.795947	H1	-1	H1	-1	2	1
4	236501	3.96E-19	-0.838462	H1	-1	H1	-1	2	2
5	236601	1.59E-17	-0.799301	H1	-1	H1	-1	1	4
6	239101	1.04E-01	0.153040	H0	0	H0	0	3	5
7	240601	1.96E-22	-0.913986	H1	-1	H1	-1	2	2
8	241602	1.56E-07	-0.492136	H1	-1	H1	-1	1	4
9	247603	1.94E-20	-0.868623	H1	-1	H1	-1	4	3
10	248901	1.09E-22	-0.918938	H1	-1	H1	-1	4	3
11	255301	1.59E-17	-0.799301	H1	-1	H1	-1	1	4
12	255401	1.04E-09	-0.572327	H1	-1	H1	-1	2	2
13	256201	2.23E-20	-0.867226	H1	-1	H1	-1	2	1
14	256601	1.82E-17	-0.797345	H1	-1	H1	-1	2	2
15	264303	5.45E-21	-0.881202	H1	-1	H1	-1	2	1
16	264503	0.00E+00	-0.960867	H1	-1	H1	-1	4	3
17	333101	5.12E-20	-0.858840	H1	-1	H1	-1	4	3
18	334801	7.27E-19	-0.831877	H1	-1	H1	-1	2	2
19	336401	1.34E-19	-0.849057	H1	-1	H1	-1	4	3
20	336501	6.04E-23	-0.924528	H1	-1	H1	-1	2	2
21	336601	0.00E+00	-0.955276	H1	-1	H1	-1	2	1
22	337901	1.09E-07	-0.498253	H1	-1	H1	-1	4	3
23	338901	7.33E-22	-0.900769	H1	-1	H1	-1	4	3
24	340901	1.10E-08	-0.535989	H1	-1	H1	-1	1	6
25	341502	3.09E-14	-0.712089	H1	-1	H1	-1	2	1
26	342401	5.05E-17	-0.786164	H1	-1	H1	-1	5	8
27	344501	4.10E-22	-0.906359	H1	-1	H1	-1	2	2
28	344601	3.06E-22	-0.909154	H1	-1	H1	-1	4	3
29	345801	1.27E-20	-0.872816	H1	-1	H1	-1	4	3
30	348802	0.00E+00	-0.956309	H1	-1	H1	-1	4	3
31	348906	6.45E-18	-0.808526	H1	-1	H1	-1	4	3
32	349701	3.28E-21	-0.887024	H1	-1	H1	-1	4	3
33	350601	1.13E-21	-0.896576	H1	-1	H1	-1	4	3
34	352301	6.04E-23	-0.924528	H1	-1	H1	-1	2	1
35	353901	6.17E-05	-0.376092	H1	-1	H1	-1	1	6
36	354401	1.83E-23	-0.935709	H1	-1	H1	-1	2	1
37	356101	3.54E-22	-0.907757	H1	-1	H1	-1	4	3
38	356501	0.00E+00	-0.939902	H1	-1	H1	-1	4	3
39	356601	2.28E-22	-0.911950	H1	-1	H1	-1	4	3
40	356801	2.28E-10	-0.594689	H1	-1	H1	-1	2	1

ID	Código	p-valor	τ	99% de confiança		90% de confiança		Agrupamentos	
				Hipótese	Tendência	Hipótese	Tendência	k-means	Hierárquico
41	360401	3.87E-23	-0.92872	H1	-1	H1	-1	2	2
42	361404	5.48E-20	-0.85844	H1	-1	H1	-1	1	6
43	361704	3.28E-22	-0.90877	H1	-1	H1	-1	2	1
44	362501	1.54E-10	-0.60028	H1	-1	H1	-1	2	1
45	362801	1.33E-18	-0.82530	H1	-1	H1	-1	2	1
46	364301	3.04E-17	-0.79175	H1	-1	H1	-1	2	2
47	435201	6.38E-01	0.04474	H0	0	H0	0	3	7
48	435302	2.83E-01	-0.10140	H0	0	H0	0	3	6
49	441502	1.70E-22	-0.91474	H1	-1	H1	-1	2	1
50	443901	3.88E-20	-0.86164	H1	-1	H1	-1	2	1
51	449601	4.78E-16	-0.76101	H1	-1	H1	-1	2	2
52	505901	1.62E-07	-0.49230	H1	-1	H1	-1	3	7
53	516402	0.00E+00	0.72098	H1	1	H1	1	3	7
54	519601	6.59E-08	-0.50735	H1	-1	H1	-1	3	7
55	525904	2.21E-04	-0.34722	H1	-1	H1	-1	1	6
56	526301	1.91E-07	-0.48881	H1	-1	H1	-1	1	6
57	526501	2.62E-09	-0.55874	H1	-1	H1	-1	1	6
58	527102	7.82E-12	-0.64218	H1	-1	H1	-1	2	2
59	527901	1.38E-02	-0.23163	H0	0	H1	-1	1	6
60	528201	6.98E-01	-0.03711	H0	0	H0	0	3	6
61	528401	5.42E-02	-0.18112	H0	0	H1	-1	1	6
62	529201	1.73E-02	-0.22385	H0	0	H1	-1	1	6
63	529301	3.07E-01	-0.09667	H0	0	H0	0	3	7
64	530801	2.84E-07	-0.48249	H1	-1	H1	-1	2	2
65	533401	1.62E-10	-0.59979	H1	-1	H1	-1	1	6
66	541101	1.51E-19	-0.84944	H1	-1	H1	-1	1	6
67	541401	9.63E-03	-0.24327	H1	-1	H1	-1	1	6
68	541701	3.16E-02	-0.20281	H0	0	H1	-1	3	7
69	601201	4.60E-13	-0.67855	H1	-1	H1	-1	2	2
70	601401	4.49E-23	-0.92732	H1	-1	H1	-1	4	3
71	601603	1.62E-15	-0.74703	H1	-1	H1	-1	5	8
72	601803	9.43E-23	-0.92034	H1	-1	H1	-1	4	2
73	603803	3.04E-19	-0.84067	H1	-1	H1	-1	2	2
74	604102	3.24E-19	-0.84027	H1	-1	H1	-1	2	2
75	605601	3.48E-19	-0.83927	H1	-1	H1	-1	4	3
76	607102	3.29E-16	-0.76573	H1	-1	H1	-1	2	1
77	609202	3.97E-15	-0.73681	H1	-1	H1	-1	2	2
78	628601	3.05E-04	-0.33940	H1	-1	H1	-1	1	6
79	632701	1.51E-03	-0.29906	H1	-1	H1	-1	1	6
80	632702	8.39E-09	-0.54132	H1	-1	H1	-1	1	6

ID	Código	p-valor	τ	99% de confiança		90% de confiança		Agrupamentos	
				Hipótese	Tendência	Hipótese	Tendência	k-means	Hierárquico
81	637903	2.94E-18	-0.81691	H1	-1	H1	-1	2	2
82	638501	1.36E-22	-0.91716	H1	-1	H1	-1	2	1
83	638601	4.55E-19	-0.83648	H1	-1	H1	-1	2	1
84	638701	5.06E-21	-0.88221	H1	-1	H1	-1	2	2
85	639301	1.22E-02	-0.23616	H0	0	H1	-1	1	6
86	639501	1.96E-23	-0.93534	H1	-1	H1	-1	2	2
87	643301	2.45E-14	-0.71488	H1	-1	H1	-1	5	8
88	645201	0.00E+00	-0.95491	H1	-1	H1	-1	2	2
89	645701	5.36E-16	-0.76067	H1	-1	H1	-1	2	1
90	645902	1.77E-16	-0.77273	H1	-1	H1	-1	2	1
91	646302	2.84E-22	-0.91017	H1	-1	H1	-1	4	2
92	646701	4.25E-11	-0.61845	H1	-1	H1	-1	1	4
93	647205	2.64E-22	-0.91055	H1	-1	H1	-1	2	2
94	650901	8.63E-14	-0.69951	H1	-1	H1	-1	2	2
95	652301	2.22E-04	0.34673	H1	1	H1	1	3	7
96	652501	3.42E-04	0.33613	H1	1	H1	1	1	6
97	652801	2.31E-11	-0.62727	H1	-1	H1	-1	1	6
98	652902	7.49E-10	-0.57782	H1	-1	H1	-1	2	2
99	657202	1.99E-11	-0.62915	H1	-1	H1	-1	5	8
100	657535	1.08E-13	-0.69672	H1	-1	H1	-1	1	4
101	657802	1.39E-11	-0.63382	H1	-1	H1	-1	5	8
102	659101	2.19E-05	-0.39846	H1	-1	H1	-1	1	6
103	661101	6.44E-01	0.04403	H0	0	H0	0	3	7
104	661201	4.69E-03	-0.26564	H1	-1	H1	-1	1	4
105	661301	1.41E-03	0.30021	H1	1	H1	1	3	7
106	662101	1.61E-06	-0.45019	H1	-1	H1	-1	1	4
107	662501	2.05E-01	-0.11958	H0	0	H0	0	1	6
108	663801	1.05E-08	-0.53763	H1	-1	H1	-1	3	7
109	664501	3.23E-04	-0.33776	H1	-1	H1	-1	3	7
110	664701	3.76E-07	-0.47676	H1	-1	H1	-1	3	7
111	664702	7.85E-07	-0.46364	H1	-1	H1	-1	3	7
112	708802	0.00E+00	-0.97065	H1	-1	H1	-1	4	3
113	712401	3.30E-16	-0.76520	H1	-1	H1	-1	2	2
114	716302	1.45E-14	-0.72143	H1	-1	H1	-1	1	4
115	716602	2.07E-17	-0.79595	H1	-1	H1	-1	1	4
116	753701	3.88E-14	-0.70929	H1	-1	H1	-1	1	6
117	755701	2.84E-22	-0.91017	H1	-1	H1	-1	4	2
118	756501	1.98E-18	-0.82110	H1	-1	H1	-1	2	2
119	756702	2.39E-10	-0.59420	H1	-1	H1	-1	2	1
120	759301	9.78E-22	-0.89797	H1	-1	H1	-1	1	4

ID	Código	p-valor	τ	99% de confiança		90% de confiança		Agrupamentos	
				Hipótese	Tendência	Hipótese	Tendência	k-means	Hierárquico
161	1021201	2.00E-21	-0.89161	H1	-1	H1	-1	4	3
162	1022801	0.00E+00	-0.96506	H1	-1	H1	-1	4	2
163	1022903	2.06E-14	-0.71723	H1	-1	H1	-1	2	2
164	1024401	1.10E-01	-0.15046	H0	0	H0	0	1	6
165	1025301	2.57E-01	-0.10692	H0	0	H0	0	1	4
166	1025701	0.00E+00	-0.94410	H1	-1	H1	-1	4	3
167	1027501	0.00E+00	-0.94689	H1	-1	H1	-1	4	3
168	1028301	0.00E+00	-0.99022	H1	-1	H1	-1	4	3
169	1029302	5.21E-19	-0.83508	H1	-1	H1	-1	4	3
170	1030505	0.00E+00	-0.98463	H1	-1	H1	-1	4	2
171	1031201	1.27E-20	-0.87282	H1	-1	H1	-1	4	2
172	1031301	5.79E-05	-0.37749	H1	-1	H1	-1	1	6
173	1031501	4.39E-22	-0.90598	H1	-1	H1	-1	2	2
174	1031601	0.00E+00	-0.97484	H1	-1	H1	-1	2	2
175	1031701	1.13E-21	-0.89658	H1	-1	H1	-1	2	2
176	1032201	6.81E-19	-0.83229	H1	-1	H1	-1	4	2
177	1032703	2.67E-17	-0.79315	H1	-1	H1	-1	4	3
178	1033802	0.00E+00	-0.98043	H1	-1	H1	-1	4	3
179	1033902	0.00E+00	-0.98742	H1	-1	H1	-1	4	3
180	1034404	0.00E+00	-0.98602	H1	-1	H1	-1	4	3
181	1034802	0.00E+00	-0.99161	H1	-1	H1	-1	4	3
182	1037601	0.00E+00	-0.99581	H1	-1	H1	-1	4	3
183	1038401	0.00E+00	-0.99720	H1	-1	H1	-1	4	3
184	1041402	0.00E+00	-0.98183	H1	-1	H1	-1	2	2
185	1041905	5.73E-17	-0.78477	H1	-1	H1	-1	2	1
186	1042202	0.00E+00	-0.99825	H1	-1	H1	-1	4	3
187	1042704	0.00E+00	-0.94512	H1	-1	H1	-1	2	1
188	1043905	0.00E+00	-0.95947	H1	-1	H1	-1	4	3
189	1046601	0.00E+00	-0.97345	H1	-1	H1	-1	4	3
190	1047101	0.00E+00	-0.97764	H1	-1	H1	-1	4	3
191	1047401	0.00E+00	-0.97904	H1	-1	H1	-1	4	3
192	1049602	2.06E-15	-0.74423	H1	-1	H1	-1	2	1
193	1049801	0.00E+00	-0.96506	H1	-1	H1	-1	1	6
194	1050801	1.84E-07	0.48934	H1	1	H1	1	3	5
195	1051105	1.47E-22	-0.91614	H1	-1	H1	-1	2	1
196	1051703	1.36E-20	-0.87242	H1	-1	H1	-1	1	4
197	1052601	0.00E+00	-0.98043	H1	-1	H1	-1	1	4
198	1053101	0.00E+00	-0.97484	H1	-1	H1	-1	4	2
199	1053602	0.00E+00	-1.00000	H1	-1	H1	-1	2	1
200	1054301	0.00E+00	-0.96785	H1	-1	H1	-1	4	3

ID	Código	p-valor	τ	99% de confiança		90% de confiança		Agrupamentos	
				Hipótese	Tendência	Hipótese	Tendência	k-means	Hierárquico
201	1054801	0.00E+00	-0.97097	H1	-1	H1	-1	4	3
202	1055902	0.00E+00	-0.95211	H1	-1	H1	-1	4	3
203	1056403	0.00E+00	-0.95947	H1	-1	H1	-1	4	3
204	1057501	1.56E-03	-0.29720	H1	-1	H1	-1	1	4
205	1058502	1.93E-08	0.52709	H1	1	H1	1	3	5
206	1058701	0.00E+00	0.65898	H1	1	H1	1	3	5
207	1058801	9.49E-04	-0.31060	H1	-1	H1	-1	1	4
208	1059302	2.11E-13	-0.68881	H1	-1	H1	-1	1	4
209	1059401	1.44E-07	-0.49353	H1	-1	H1	-1	1	6
210	1061201	0.00E+00	-0.94130	H1	-1	H1	-1	2	1
211	1061701	0.00E+00	-0.95248	H1	-1	H1	-1	4	3
212	1062402	0.00E+00	-0.97484	H1	-1	H1	-1	2	1
213	1063601	0.00E+00	-0.97484	H1	-1	H1	-1	4	3
214	1063702	0.00E+00	-0.95631	H1	-1	H1	-1	1	4
215	1101507	3.45E-03	-0.27483	H1	-1	H1	-1	3	7
216	1108201	2.59E-14	-0.71444	H1	-1	H1	-1	1	6
217	1108308	1.08E-13	-0.69672	H1	-1	H1	-1	3	6
218	1109501	7.88E-01	-0.02586	H0	0	H0	0	1	4
219	1109601	1.42E-04	0.35709	H1	1	H1	1	3	7
220	1110402	2.09E-04	0.34813	H1	1	H1	1	3	7
221	1110802	2.33E-01	0.11251	H0	0	H0	0	1	6
222	1111801	2.21E-12	-0.65851	H1	-1	H1	-1	2	2
223	1112401	2.18E-09	-0.56193	H1	-1	H1	-1	2	2
224	1112522	3.67E-08	-0.51678	H1	-1	H1	-1	2	2
225	1112701	4.47E-11	-0.61797	H1	-1	H1	-1	5	8
226	1112702	1.55E-06	-0.45073	H1	-1	H1	-1	1	4
227	1112801	7.46E-02	-0.16777	H0	0	H1	-1	1	4
228	1112802	1.03E-03	-0.30818	H1	-1	H1	-1	1	4
229	1112803	3.09E-07	-0.48008	H1	-1	H1	-1	2	1
230	1112901	3.35E-06	-0.43652	H1	-1	H1	-1	2	1
231	1117401	1.00E+00	0.00070	H0	0	H0	0	1	6
232	1117701	2.99E-02	-0.20412	H0	0	H1	-1	1	4
233	1119401	1.54E-11	-0.63242	H1	-1	H1	-1	2	1
234	1121801	7.43E-01	-0.03145	H0	0	H0	0	3	7
235	1126701	1.46E-11	-0.63335	H1	-1	H1	-1	3	7
236	1126908	3.94E-05	-0.38588	H1	-1	H1	-1	2	1
237	1128501	1.06E-15	-0.75218	H1	-1	H1	-1	1	6
238	1134501	1.08E-13	-0.69672	H1	-1	H1	-1	2	2
239	1134601	1.62E-15	-0.74703	H1	-1	H1	-1	1	1
240	1134801	3.35E-18	-0.81551	H1	-1	H1	-1	2	2

ID	Código	p-valor	τ	99% de confiança		90% de confiança		Agrupamentos	
				Hipótese	Tendência	Hipótese	Tendência	k-means	Hierárquico
241	1135702	9.59E-08	-0.50052	H1	-1	H1	-1	1	4
242	1136708	3.48E-19	-0.83927	H1	-1	H1	-1	4	3
243	1136802	3.74E-16	-0.76380	H1	-1	H1	-1	1	6
244	1137601	1.51E-21	-0.89378	H1	-1	H1	-1	2	1
245	1137701	8.28E-20	-0.85425	H1	-1	H1	-1	2	2
246	1138402	1.57E-20	-0.87102	H1	-1	H1	-1	2	2
247	1139201	5.61E-01	-0.05521	H0	0	H0	0	1	6
248	1141602	0.00E+00	-0.98463	H1	-1	H1	-1	4	3
249	1143501	6.87E-18	-0.80811	H1	-1	H1	-1	1	6
250	1144401	2.02E-19	-0.84486	H1	-1	H1	-1	4	3
251	1145512	8.37E-18	-0.80573	H1	-1	H1	-1	5	8
252	1145802	1.79E-13	-0.69067	H1	-1	H1	-1	2	2
253	1146101	4.60E-13	-0.67855	H1	-1	H1	-1	1	6
254	1146701	8.91E-19	-0.82949	H1	-1	H1	-1	2	2
255	1149101	0.00E+00	-0.95807	H1	-1	H1	-1	4	3
256	1151102	0.00E+00	-0.97309	H1	-1	H1	-1	2	2
257	1151303	5.08E-22	-0.90458	H1	-1	H1	-1	4	3
258	1152703	0.00E+00	-0.98183	H1	-1	H1	-1	4	3
259	1155901	1.82E-14	-0.71913	H1	-1	H1	-1	1	4
260	1160302	7.79E-19	-0.83089	H1	-1	H1	-1	4	3
261	1160605	0.00E+00	-0.95491	H1	-1	H1	-1	4	3
262	1161801	7.53E-23	-0.92275	H1	-1	H1	-1	4	3
263	1162201	1.87E-01	0.12443	H0	0	H0	0	1	6
264	1162401	2.73E-01	0.10346	H0	0	H0	0	3	6
265	1162601	3.83E-04	0.33380	H1	1	H1	1	3	7
266	1162702	3.52E-15	-0.73820	H1	-1	H1	-1	2	2
267	1162801	3.48E-11	-0.62124	H1	-1	H1	-1	1	4
268	1163801	1.26E-15	-0.75035	H1	-1	H1	-1	2	2
269	1201101	7.25E-04	-0.31759	H1	-1	H1	-1	3	7
270	1201102	7.20E-05	-0.37325	H1	-1	H1	-1	3	7
271	1201301	1.61E-06	-0.45019	H1	-1	H1	-1	1	6
272	1202607	3.69E-13	-0.68134	H1	-1	H1	-1	1	4
273	2302302	4.85E-13	-0.67808	H1	-1	H1	-1	1	4
274	2302501	0.00E+00	-0.95528	H1	-1	H1	-1	4	3
275	2302801	2.95E-13	-0.68462	H1	-1	H1	-1	1	4
276	2303802	0.00E+00	-0.98742	H1	-1	H1	-1	4	3
277	2304602	0.00E+00	-0.96226	H1	-1	H1	-1	4	3
278	2304603	6.04E-23	-0.92453	H1	-1	H1	-1	4	3
279	2304701	0.00E+00	-0.96154	H1	-1	H1	-1	2	1
280	2305301	0.00E+00	-0.94689	H1	-1	H1	-1	4	3

ID	Código	p-valor	τ	99% de confiança		90% de confiança		Agrupamentos	
				Hipótese	Tendência	Hipótese	Tendência	k-means	Hierárquico
281	2305602	7.33E-22	-0.90077	H1	-1	H1	-1	4	3
282	2307501	1.68E-20	-0.87002	H1	-1	H1	-1	2	2
283	2307601	3.38E-20	-0.86303	H1	-1	H1	-1	2	2
284	2308201	1.88E-19	-0.84586	H1	-1	H1	-1	2	2
285	2308401	7.13E-13	-0.67296	H1	-1	H1	-1	2	2
286	2308701	5.45E-21	-0.88120	H1	-1	H1	-1	3	7
287	2309201	9.32E-13	-0.67016	H1	-1	H1	-1	1	6
288	2309501	1.21E-21	-0.89619	H1	-1	H1	-1	1	4
289	2309601	1.35E-12	-0.66690	H1	-1	H1	-1	1	6
290	2309701	2.20E-17	-0.79553	H1	-1	H1	-1	2	2
291	2311401	0.00E+00	-0.94410	H1	-1	H1	-1	2	2
292	2311601	0.00E+00	-0.97764	H1	-1	H1	-1	1	4
293	2311702	0.00E+00	-0.96785	H1	-1	H1	-1	2	2
294	2312102	0.00E+00	-0.97345	H1	-1	H1	-1	4	3
295	2312301	5.88E-20	-0.85744	H1	-1	H1	-1	4	3
296	2312401	0.00E+00	-0.96785	H1	-1	H1	-1	2	1
297	2312402	0.00E+00	-0.97904	H1	-1	H1	-1	4	3
298	2312801	9.43E-23	-0.92034	H1	-1	H1	-1	2	2
299	2312803	1.94E-20	-0.86862	H1	-1	H1	-1	2	2
300	2312902	1.21E-21	-0.89619	H1	-1	H1	-1	4	3
301	2314101	2.02E-19	-0.84486	H1	-1	H1	-1	2	1
302	2315204	6.75E-20	-0.85604	H1	-1	H1	-1	2	2
303	2315501	1.17E-19	-0.85045	H1	-1	H1	-1	2	1
304	2315801	3.04E-19	-0.84067	H1	-1	H1	-1	2	2
305	2316101	1.87E-21	-0.89200	H1	-1	H1	-1	2	2
306	2316501	7.75E-20	-0.85465	H1	-1	H1	-1	1	4
307	2317202	5.73E-12	-0.64593	H1	-1	H1	-1	5	8
308	2317704	4.73E-21	-0.88260	H1	-1	H1	-1	3	7
309	2317901	3.01E-06	0.43816	H1	1	H1	1	3	5
310	2318703	8.17E-01	-0.02239	H0	0	H0	0	3	7
311	2318704	3.21E-07	-0.47955	H1	-1	H1	-1	3	7
312	2319304	7.01E-23	-0.92313	H1	-1	H1	-1	2	2
313	2319501	2.28E-22	-0.91195	H1	-1	H1	-1	2	2
314	2320401	1.13E-21	-0.89658	H1	-1	H1	-1	2	2
315	2320701	1.52E-18	-0.82390	H1	-1	H1	-1	2	2
316	2321706	1.73E-18	-0.82250	H1	-1	H1	-1	2	2
317	2322204	1.94E-14	-0.71768	H1	-1	H1	-1	5	8
318	2325304	7.26E-04	0.31737	H1	1	H1	1	3	5
319	2325401	3.89E-03	0.27123	H1	1	H1	1	3	6
320	2326101	6.99E-05	0.37330	H1	1	H1	1	3	5

ID	Código	p-valor	τ	99% de confiança		90% de confiança		Agrupamentos	
				Hipótese	Tendência	Hipótese	Tendência	k-means	Hierárquico
321	2326301	3.98E-03	0.27044	H1	1	H1	1	3	7
322	2326604	3.95E-02	-0.19371	H0	0	H1	-1	3	7
323	2327207	2.35E-17	-0.79455	H1	-1	H1	-1	1	4
324	2327302	3.52E-04	-0.33555	H1	-1	H1	-1	3	7
325	2327402	8.31E-04	0.31377	H1	1	H1	1	3	5
326	2327601	4.02E-06	-0.43256	H1	-1	H1	-1	3	7
327	2328701	7.77E-01	-0.02725	H0	0	H0	0	3	7
328	2329302	1.03E-02	-0.24126	H0	0	H1	-1	1	4
329	2330502	6.98E-01	0.03704	H0	0	H0	0	1	6
330	2331101	4.07E-07	-0.47536	H1	-1	H1	-1	5	5
331	2333401	7.07E-09	-0.54298	H1	-1	H1	-1	3	7
332	2333601	1.46E-01	0.13711	H0	0	H0	0	3	7
333	2333801	1.84E-02	-0.22168	H0	0	H1	-1	3	7
334	2334502	4.36E-14	-0.70790	H1	-1	H1	-1	1	6
335	2334503	2.90E-10	-0.59140	H1	-1	H1	-1	1	6
336	2334801	1.23E-12	-0.66597	H1	-1	H1	-1	1	6
337	2334902	1.68E-20	-0.87002	H1	-1	H1	-1	1	6
338	2334903	1.02E-14	-0.72562	H1	-1	H1	-1	1	6
339	2335701	9.66E-12	-0.63894	H1	-1	H1	-1	1	6
340	2335801	6.82E-01	0.03917	H0	0	H0	0	3	7
341	2335901	2.38E-02	-0.21251	H0	0	H1	-1	3	7
342	2341401	4.17E-04	-0.33135	H1	-1	H1	-1	3	7
343	2341501	2.77E-03	0.28102	H1	1	H1	1	3	5
344	2341901	2.28E-04	0.34591	H1	1	H1	1	3	7
345	2342301	7.71E-01	0.02798	H0	0	H0	0	3	7
346	2342401	5.06E-02	0.18379	H0	0	H1	1	3	7
347	2342601	3.75E-01	0.08389	H0	0	H0	0	3	7
348	2342602	6.28E-01	0.04614	H0	0	H0	0	3	7
349	2342801	5.64E-03	0.26005	H1	1	H1	1	3	5
350	2343301	3.95E-01	0.08036	H0	0	H0	0	3	5
351	2343502	2.53E-07	0.48375	H1	1	H1	1	3	7
352	2343504	3.19E-10	0.59000	H1	1	H1	1	3	5
353	2343901	1.06E-06	0.45804	H1	1	H1	1	3	5
354	2344101	1.52E-01	0.13487	H0	0	H0	0	3	5
355	2344401	4.00E-08	0.51502	H1	1	H1	1	3	5
356	2344702	2.04E-05	0.39986	H1	1	H1	1	3	5
357	2401201	0.00E+00	0.69626	H1	1	H1	1	3	5
358	2402701	0.00E+00	0.71209	H1	1	H1	1	3	5
359	2404301	1.51E-04	-0.35570	H1	-1	H1	-1	1	6
360	2405502	0.00E+00	0.81230	H1	1	H1	1	3	5

ID	Código	p-valor	τ	99% de confiança		90% de confiança		Agrupamentos	
				Hipótese	Tendência	Hipótese	Tendência	k-means	Hierárquico
361	2406402	5.71E-02	-0.17908	H0	0	H1	-1	3	7
362	2407701	0.00E+00	-0.96506	H1	-1	H1	-1	2	2
363	2407901	6.10E-17	-0.78434	H1	-1	H1	-1	2	2
364	2408501	8.73E-23	-0.92136	H1	-1	H1	-1	2	1
365	2408701	0.00E+00	-0.95388	H1	-1	H1	-1	2	1
366	2409801	4.72E-02	0.18671	H0	0	H1	1	3	7
367	2409901	4.75E-15	-0.73445	H1	-1	H1	-1	1	6
368	2410201	1.44E-01	0.13767	H0	0	H0	0	3	7
369	2410702	7.07E-12	-0.64313	H1	-1	H1	-1	3	7
370	2412101	0.00E+00	0.68508	H1	1	H1	1	3	5
371	2414501	0.00E+00	0.73026	H1	1	H1	1	3	5
372	2415108	5.45E-12	-0.64640	H1	-1	H1	-1	1	6
373	2415504	1.13E-11	-0.63662	H1	-1	H1	-1	3	6
374	2415506	2.64E-04	-0.34254	H1	-1	H1	-1	3	7
375	2415507	3.41E-02	-0.19930	H0	0	H1	-1	1	4
376	2415605	1.17E-19	-0.85045	H1	-1	H1	-1	1	6
377	2415802	2.51E-03	-0.28382	H1	-1	H1	-1	3	7
378	2415901	1.71E-15	-0.74659	H1	-1	H1	-1	3	7
379	2416311	7.92E-05	-0.37050	H1	-1	H1	-1	3	7
380	2416405	6.72E-12	-0.64361	H1	-1	H1	-1	1	6
381	2416601	5.98E-05	-0.37666	H1	-1	H1	-1	1	6
382	2416901	5.33E-02	-0.18201	H0	0	H1	-1	3	7
383	2418201	3.88E-14	-0.70929	H1	-1	H1	-1	1	4
384	2418202	1.52E-13	-0.69252	H1	-1	H1	-1	1	6
385	2418301	2.45E-15	-0.74292	H1	-1	H1	-1	3	7
386	2418401	7.46E-12	-0.64221	H1	-1	H1	-1	1	6
387	2418601	1.22E-14	-0.72327	H1	-1	H1	-1	1	4
388	2418801	9.44E-11	-0.60727	H1	-1	H1	-1	1	4
389	2418901	1.12E-01	-0.14960	H0	0	H0	0	3	7
390	2419201	5.14E-13	-0.67715	H1	-1	H1	-1	1	6
391	2419301	2.07E-10	-0.59609	H1	-1	H1	-1	1	6
392	2419601	3.52E-15	-0.73820	H1	-1	H1	-1	1	6
393	2419801	1.31E-17	-0.80112	H1	-1	H1	-1	1	4
394	2420102	5.56E-19	-0.83467	H1	-1	H1	-1	1	1
395	2420202	2.02E-19	-0.84486	H1	-1	H1	-1	2	1
396	2420401	2.01E-21	-0.89099	H1	-1	H1	-1	2	2
397	2420701	2.84E-22	-0.91017	H1	-1	H1	-1	3	7
398	2420702	1.73E-14	-0.71908	H1	-1	H1	-1	1	6
399	2420901	2.47E-23	-0.93291	H1	-1	H1	-1	4	2
400	2421201	3.67E-13	-0.68230	H1	-1	H1	-1	1	4

ID	Código	p-valor	τ	99% de confiança		90% de confiança		Agrupamentos	
				Hipótese	Tendência	Hipótese	Tendência	k-means	Hierárquico
401	2421301	2.58E-16	-0.768	H1	-1	H1	-1	1	6
402	2421401	5.71E-17	-0.785	H1	-1	H1	-1	1	6
403	2421905	2.45E-14	-0.715	H1	-1	H1	-1	1	4
404	2422201	9.02E-03	0.245	H1	1	H1	1	3	5
405	2422802	6.33E-04	-0.321	H1	-1	H1	-1	3	7
406	2423701	2.41E-18	-0.819	H1	-1	H1	-1	1	4
407	2427201	8.65E-10	-0.575	H1	-1	H1	-1	1	6
408	2427301	3.61E-09	-0.554	H1	-1	H1	-1	3	7
409	2428103	8.48E-02	-0.162	H0	0	H1	-1	3	6
410	2428501	1.15E-09	-0.571	H1	-1	H1	-1	3	5
411	2428901	3.29E-09	-0.555	H1	-1	H1	-1	2	2
412	2429308	6.05E-12	-0.645	H1	-1	H1	-1	2	2
413	2429401	6.65E-01	-0.041	H0	0	H0	0	3	6
414	2429901	2.90E-06	-0.439	H1	-1	H1	-1	1	6
415	2430801	1.82E-17	-0.797	H1	-1	H1	-1	2	1
416	2430901	7.28E-07	-0.465	H1	-1	H1	-1	1	6
417	2431501	3.87E-01	0.082	H0	0	H0	0	3	7
418	2431601	1.66E-09	0.566	H1	1	H1	1	3	7
419	2431801	3.45E-09	-0.554	H1	-1	H1	-1	1	4
420	2434302	1.91E-04	-0.350	H1	-1	H1	-1	1	6
421	2436601	1.48E-03	-0.299	H1	-1	H1	-1	1	4
422	2437101	2.20E-08	-0.525	H1	-1	H1	-1	1	4
423	2437204	8.57E-15	-0.727	H1	-1	H1	-1	1	4
424	2437308	4.18E-12	-0.650	H1	-1	H1	-1	1	6
425	2438801	5.79E-05	-0.377	H1	-1	H1	-1	1	6
426	2439101	5.99E-07	-0.469	H1	-1	H1	-1	3	6
427	2439901	2.10E-08	0.526	H1	1	H1	1	3	7
428	2439904	9.87E-07	0.459	H1	1	H1	1	3	5
429	2440403	6.37E-15	-0.731	H1	-1	H1	-1	1	6
430	2440504	7.85E-17	-0.782	H1	-1	H1	-1	1	6
431	2445901	6.98E-02	-0.171	H0	0	H1	-1	3	7
432	2445902	7.71E-01	-0.028	H0	0	H0	0	3	7
433	2446801	1.49E-06	-0.452	H1	-1	H1	-1	1	6
434	2447201	2.05E-01	0.119	H0	0	H0	0	3	7
435	2447202	8.17E-01	0.022	H0	0	H0	0	3	7
436	2447402	1.56E-02	0.227	H0	0	H1	1	3	5
437	2448302	5.43E-02	-0.181	H0	0	H1	-1	3	7
438	2448401	2.25E-06	0.444	H1	1	H1	1	3	5
439	2454901	4.54E-08	-0.513	H1	-1	H1	-1	1	6
440	2455401	2.76E-01	-0.103	H0	0	H0	0	3	5

ID	Código	p-valor	τ	99% de confiança		90% de confiança		Agrupamentos	
				Hipótese	Tendência	Hipótese	Tendência	<i>k-means</i>	Hierárquico
441	2457301	2.34E-03	-0.28581	H1	-1	H1	-1	3	7
442	2457601	1.31E-08	-0.53319	H1	-1	H1	-1	3	7
443	2458101	2.70E-01	0.10420	H0	0	H0	0	3	7
444	2458301	3.05E-03	0.27861	H1	1	H1	1	3	7
445	2460501	5.67E-04	0.32355	H1	1	H1	1	3	5
446	2461702	1.89E-01	0.12369	H0	0	H0	0	3	7
447	2462103	1.15E-08	-0.53548	H1	-1	H1	-1	1	4
448	2463301	1.54E-01	-0.13422	H0	0	H0	0	3	7
449	2540601	4.25E-01	0.07550	H0	0	H0	0	3	7
450	2548301	1.08E-03	-0.30699	H1	-1	H1	-1	1	6
451	2548801	2.43E-02	-0.21174	H0	0	H1	-1	3	7
452	2556201	1.27E-20	-0.87282	H1	-1	H1	-1	1	4
453	2564301	3.56E-21	-0.88539	H1	-1	H1	-1	1	1
454	2608202	3.54E-22	-0.90776	H1	-1	H1	-1	2	1
455	2624307	1.17E-19	-0.85045	H1	-1	H1	-1	2	1
456	2701101	1.70E-13	-0.69113	H1	-1	H1	-1	2	1
457	2701701	3.06E-22	-0.90915	H1	-1	H1	-1	2	2
458	2703501	1.38E-06	-0.45331	H1	-1	H1	-1	1	6
459	2703701	7.80E-08	-0.50420	H1	-1	H1	-1	1	6
460	2703901	2.41E-16	-0.76950	H1	-1	H1	-1	1	4
461	2705401	2.92E-16	-0.76660	H1	-1	H1	-1	1	4
462	2705601	8.18E-05	-0.36967	H1	-1	H1	-1	1	6
463	2706901	8.58E-01	-0.01747	H0	0	H0	0	3	7
464	2707201	5.02E-01	0.06359	H0	0	H0	0	3	5
465	2707401	3.54E-02	0.19776	H0	0	H1	1	3	5
466	2707901	1.03E-02	0.24109	H0	0	H1	1	3	5
467	2710801	1.62E-18	-0.82349	H1	-1	H1	-1	2	1
468	2712201	2.50E-13	-0.68647	H1	-1	H1	-1	1	6
469	2719601	2.07E-01	0.11892	H0	0	H0	0	3	7
470	2735701	4.34E-05	0.38392	H1	1	H1	1	3	7
471	2738401	2.83E-02	-0.20644	H0	0	H1	-1	3	7
472	2739903	2.25E-07	-0.48567	H1	-1	H1	-1	2	1
473	2744201	9.14E-07	0.46084	H1	1	H1	1	3	7
474	2748602	3.57E-12	-0.65199	H1	-1	H1	-1	1	4
475	2748904	2.07E-02	0.21733	H0	0	H1	1	1	6
476	2751701	6.98E-02	-0.17057	H0	0	H1	-1	1	4
477	2756202	7.54E-01	0.03005	H0	0	H0	0	1	6
478	2756301	2.63E-01	-0.10552	H0	0	H0	0	1	6
479	2764901	9.85E-05	0.36548	H1	1	H1	1	3	7
480	2809901	8.45E-03	0.24747	H1	1	H1	1	3	5

ID	Código	p-valor	τ	99% de confiança		90% de confiança		Agrupramentos	
				Hipótese	Tendência	Hipótese	Tendência	<i>k-means</i>	Hierárquico
481	2817501	5.59E-08	0.50979	H1	1	H1	1	3	5
482	2817702	5.61E-05	0.37806	H1	1	H1	1	3	5
483	2818301	8.44E-03	0.24764	H1	1	H1	1	3	6
484	2818501	2.09E-09	0.56204	H1	1	H1	1	3	5
485	2818902	2.19E-05	0.39846	H1	1	H1	1	3	7
486	2826206	1.31E-08	0.53357	H1	1	H1	1	3	5
487	2841501	0.00E+00	0.66317	H1	1	H1	1	3	5
488	2849801	2.83E-01	0.10140	H0	0	H0	0	3	5
489	2858102	1.81E-09	0.56513	H1	1	H1	1	3	7
490	2858304	1.15E-05	-0.41160	H1	-1	H1	-1	1	6
491	4508602	8.65E-10	0.57512	H1	1	H1	1	3	5
492	4515201	2.03E-02	0.21811	H0	0	H1	1	3	7