UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS AGRONÔMICAS
CAMPUS DE BOTUCATU

UTILIZAÇÃO DA GEOESTATÍSTICA NA CONSTRUÇÃO DE MAPAS DE TEMPERATURA MÉDIA MENSAL PARA O ESTADO DO PARANÁ.

VANDERLI MARINO MELEM

Tese apresentada à Faculdade de Ciências Agronômicas da UNESP - Campus de Botucatu, para obtenção do título de Doutor em Agronomia - Área de Concentração em Energia na Agricultura.

BOTUCATU-SP
Abril – 2002

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE CIÊNCIAS AGRONÔMICAS
CAMPUS DE BOTUCATU

UTILIZAÇÃO DA GEOESTATÍSTICA NA CONSTRUÇÃO DE MAPAS DE TEMPERATURA MÉDIA MENSAL PARA O ESTADO DO PARANÁ.

VANDERLI MARINO MELEM

ORIENTADOR: Prof. Dr. CARLOS ROBERTO PADOVANI

Tese apresentada à Faculdade de Ciências Agronômicas da UNESP - Campus de Botucatu, para obtenção do título de Doutor em Agronomia - Área de Concentração em Energia na Agricultura.

BOTUCATU-SP
Abril – 2002

DEDICATÓRIA

Esta tese é especialmente dedicada ao *Cláudio Roberto Melem*, meu esposo, por ter sido paciente e amoroso, principalmente nesses últimos anos em que estive envolvida com o doutorado.

AGRADECIMENTOS

A *Deus*, pela oportunidade de estar aqui e ter esta família e todos estes amigos maravilhosos.

Ao *Prof. Dr. Carlos Roberto Padovani*, pela orientação, amizade e confiança, sempre constantes.

Ao Instituto Agronômico do Paraná – IAPAR, pelos dados fornecidos que permitiram a realização deste trabalho. Em especial aos engenheiros *João Henrique Caviglione*, *Edmirson Borrozino* e ao meteorologista *Dr. Paulo Henrique Caramori*.

Ao *Prof. Dr. Antônio Carlos A. Gonçalves*, da Universidade Estadual de Maringá – UEM, pelos ensinamentos e sugestões que contribuíram para a realização deste trabalho.

À Universidade Estadual de Londrina – UEL, pela oportunidade para realizar o curso e pelos recursos financeiros despendidos.

Aos amigos *Ana Vergínia Libos Messetti*, *Jacinta Ludovico Zamboti*, *Marie Oshiiwa*, *Rogério Mendonça Martins* e *Simone de Castro Queiroz*, pela companhia agradável nas viagens a Botucatu e pelo incentivo e carinho nos momentos difíceis.

Ao *Prof. Dr. Flávio Ferrari Aragon*, pelos ensinamentos e atenção sempre dedicados.

Aos familiares, em especial minha irmã *Susy* e aos pais do *Cláudio*, *Sr. Irineu* e *Dona Terezinha Melem*, pela paciência compreensão e carinho.

Ao amigo *Luiz Melo*, pelo auxílio na digitação da tese e à *Profa. Dr. Edina Regina Pugas Panichi*, do Departamento de Letras da Universidade Estadual de Londrina – UEL, pela atenção nas correções gramaticais.

LISTA DE QUADROS

Quadro **Página**

1 - Limites usados na construção dos mapas. 83

2 - Estatísticas descritivas das temperaturas médias (°C) dos meses de janeiro a dezembro no Estado do Paraná. 90

3 - Estatísticas descritivas: da altitude (A), em m e; das temperaturas médias dos meses de janeiro (TJA) e julho (TJU), em °C. 91

4 - Coeficientes de correlação de Pearson das temperaturas dos meses de janeiro a dezembro com as variáveis X, Y e A. 93

5 - Quadrado das diferenças entre valores observados e preditos, TJA. 106

6 - Quadrado das diferenças entre valores observados e preditos, TJU. 107

7 - Estatísticas dos valores de altitude (A) e valores preditos de temperatura média do mês de janeiro ($TJA_{preditos}$) 109

8 - Estatísticas dos valores de altitude (A) e valores preditos de temperatura média do mês de julho ($TJU_{preditos}$) 110

LISTA DE FIGURAS

Figura **Página**

1 - Exemplo de semivariograma de nuvem. 23

2 - Ilustração de um semivariograma ideal. 35

3 - (a) Semivariograma - Modelo efeito pepita puro com c_0=1; (b) Covariograma - Modelo efeito pepita puro com c_0=1. 38

4 - Semivariograma, modelo esférico com a=10 e c=1. 40

5 - Covariograma, modelo esférico com a=10 e c=1. 40

6 - Semivariograma, modelo exponencial com a=10 (“alcance prático”) e c=1. 42

7 - Covariograma, modelo exponencial com a=10 (“alcance prático”) e c=1. 42

8 - Semivariograma, modelo gaussiano com a=10 (“alcance prático”) e c=1. 44

9 - Covariograma, modelo gaussiano com a=10 (“alcance prático”) e c=1. 44

10 - Semivariograma, modelo potência, $0<w<2$. 45

11 - Localização das estações agrometeorológicas do IAPAR. 79

12 - Posição geográfica dos municípios extremos do Paraná. 82

13 - Posição geográfica das estações que se situam nos extremos do Estado. 83

14 - Histograma e curva normal esperada, altitude. 91

15 - Histograma e curva normal esperada, temperatura de janeiro. 92

16 - Histograma e curva normal esperada, temperatura de julho. 92

17 - Semivariograma dos dados de TJA, modelo esférico. 94

18 - Semivariograma dos dados de TJA, modelo gaussiano. 95

19 - Mapa de temperatura média do mês de janeiro (°C), modelo esférico, KO. 97

20 - Mapa de temperatura média do mês de janeiro (°C), modelo gaussiano, KO. 97

21 - Semivariograma dos dados de TJU, modelo esférico. 98

22 - Semivariograma dos dados de TJU, modelo gaussiano. 98

23 - Mapa de temperatura média do mês de julho (°C), modelo esférico, KO. 101

24 - Mapa de temperatura média do mês de julho (°C), modelo gaussiano, KO. 101

25 - Gráficos de dispersão das variáveis TJA e TJU em relação a X e Y. 102

26 - Mapa de temperatura média do mês de janeiro (°C), modelo esférico, KU – L. 104

27 - Mapa de temperatura média do mês de janeiro (°C), modelo gaussiano, KU – L. 104

28 - Mapa de temperatura média do mês de julho (°C), modelo esférico, KU - L. 105

29 - Mapa de temperatura média do mês de julho (°C), modelo gaussiano, KU - L. 105

30 - Gráfico de dispersão entre $TJA_{preditos}$ (°C) e altitude (m). 110

31 - Gráfico de dispersão entre $TJU_{preditos}$ (°C) e altitude (m). 111

32 - Mapa de temperatura média do mês de janeiro (°C) – TESE/2002. 112

33 - Mapa de temperatura média do mês de julho (°C) – TESE/2002. 112

34 - Mapa de temperatura média do mês de janeiro (°C) – IAPAR/2000. 114

35 - Mapa de temperatura média do mês de julho (°C) – IAPAR/2000. 114

36 - Temperatura média anual diminuindo com o aumento da altitude. 116

SUMÁRIO

LISTA DE QUADROS VI

LISTA DE FIGURAS VII

1 INTRODUÇÃO 1

2 REVISÃO DE LITERATURA 6

2.1 Introdução *6*

2.2 Princípios Básicos de Experimentação *7*

2.3 Princípios Básicos de Geoestatística *16*

2.3.1 Variáveis Regionalizadas 17
2.3.2 Estacionariedade 19
2.3.3 Semivariograma 21
2.3.4 Covariograma 30
2.3.5 Correlograma 32
2.3.6 Efeito Pepita 32
2.3.7 Alcance 34
2.3.8 Patamar 35
2.3.9 Modelos Básicos 36
2.3.9.1 Efeito Pepita ("Nugget Effect") 38
2.3.9.2 Esférico 39
2.3.9.3 Exponencial 41
2.3.9.4 Gaussiano 43
2.3.9.5 Potência 45

2.4 Krigagem *46*

2.4.1 Krigagem Simples (KS) 49
2.4.2 Krigagem Ordinária (KO) 55
2.4.3 Krigagem Universal (KU) 63

2.5 Validação Cruzada *75*

3 MATERIAL E MÉTODOS 78

3.1 Material *78*

3.2 Métodos *80*

4 RESULTADOS E DISCUSSÃO 90

4.1 Análise Exploratória dos Dados *90*

4.2 Aplicação dos Métodos de Predição *94*

4.2.1 Krigagem Ordinária (KO) 94
4.2.1.1 Dados de temperatura de janeiro 94
4.2.1.2 Dados de temperatura de julho 98

4.2.2 Krigagem Universal (KU) 102
4.2.2.1 Dados de temperatura de janeiro 104
4.2.2.2 Dados de temperatura de julho 105
4.2.3 Cartas Climáticas do Paraná 114
4.2.4 Sugestão para Futuros Trabalhos 115

5 CONCLUSÕES 117

REFERÊNCIAS BIBLIOGRÁFICAS 119

RESUMO

Sabe-se que a época de plantio de determinadas culturas é dependente da ocorrência de chuva e das condições de temperatura e, também, que estas épocas variam de região para região. Muitos estudos, utilizando procedimentos clássicos de estatística para analisar como a temperatura afeta o crescimento de plantas e a produção de grãos na agricultura, já foram realizados no Brasil. Este trabalho apresenta uma alternativa para analisar valores de temperatura média do ar envolvendo distribuição espacial dos dados. Assim, com o objetivo de construir o mapeamento climático do Estado do Paraná por meio de uma técnica eficiente, simples e acurada (Geoestatística), foram utilizados dados obtidos pelo Instituto Agronômico do Paraná – IAPAR, coletados diariamente em 33 estações agrometeorológicas.

A Geoestatística considera as associações entre amostras vizinhas através de sua geometria, ou seja, da localização das amostras no campo e detecta se existe influência de uma amostra sobre a outra, de acordo com a distância entre elas. As técnicas geoestatísticas de interpolação que se baseiam nessas associações entre amostras vizinhas, podem estimar com eficiência valores em locais não amostrados. A existência de dependência espacial é detectada pelo variograma (ou pelo covariograma).

As análises permitiram concluir pela existência de dependência espacial na temperatura média do ar para os meses de janeiro e julho e o modelo esférico foi o que melhor expressou esta dependência espacial. As estimativas de temperatura em pontos não amostrados (um quadriculado com 50 colunas e 35 linhas, totalizando 1750 pontos sobre o Estado) foram obtidas através da técnica de interpolação Krigagem Universal e os mapas, comparados com os existentes no IAPAR, apresentaram-se eficientes, mesmo não considerando a variável altitude como variável auxiliar.

THE USE 0F GEOSTATISTICS IN THE DEVELOPMENT OF MONTHLY MEAN TEMPERATURE MAPS FOR THE STATE OF PARANÁ, BRAZIL.

Thesis (Doutorado em Agronomia/Energia na Agricultura) – Faculdade de Ciências Agronômicas, Universidade Estadual Paulista.
Author: VANDERLI MARINO MELEM
Adviser: CARLOS ROBERTO PADOVANI

SUMMARY

It is well known that sowing periods for some crops depend on rain and temperature conditions, and that the harvesting period may vary from region to region. Several studies on the effect of temperature on plants and grain production have been carried out in Brazil, using classic statistical procedures. This study offers an alternative for the analysis of mean air temperature values involving spatial data distribution. Data from 33 agrometeorological stations, collected daily by IAPAR (Agronomical Institute of Paraná, Brazil), were used to generate climatic maps for the state, adopting an efficient, simple and accurate technique (Geostatistics). Geostatistics takes into consideration the associations among neighboring samples through its geometry, i.e., the location of samples in the field, and detects the influence of one sample over the other according to the distance between them. Interpolation geostatistical techniques based on these associations among neighboring samples allows to estimating values for non-sampled locations efficiently. The existence of spatial dependence is detected by the variogram or covariance function. Data from the geostatistical analysis showed spatial dependence in the atmospheric mean temperature for the months of January and July which was best expressed by the spherical model. Temperature estimates from non-sampled points (a square with 50 columns and 35 lines, totaling 1750 points around the state) were obtained by the Universal Kriging interpolation technique, and the mappings were more efficient compared to those from Caviglione *et al.* (2000), even when altitude was considered an auxiliary variable.

Key words: Temperature, spatial dependence, semivariogram, Universal and Ordinary Kriging.

1 INTRODUÇÃO

Na área agronômica, quase todos os trabalhos experimentais são analisados e os resultados comprovados, com certa margem de erro, quando se utilizam métodos estatísticos. Por exemplo: os delineamentos em blocos casualizados (DBC) são os mais requisitados, pois o controle das unidades experimentais (UE) por meio dos blocos (formando subgrupos de UE homogêneas dentro de cada bloco) e a aleatorização dos tratamentos às parcelas, são maneiras de minimizar a variabilidade associada à heterogeneidade das UE.

Assim, ao se realizar um experimento, o pesquisador deve controlar ao máximo tais variações para que a maior parte da variação existente entre as UE seja devida ao efeito dos tratamentos. A homogeneidade das UE (ou parcelas), antes de receberem os tratamentos, é condição imprescindível para a validade do experimento. Ademais, as parcelas, na estatística clássica, devem ser independentes para que os testes dos dados tenham validade. Esses princípios nem sempre estão presentes e, portanto, há que se pensar em alternativas mais acuradas de procedimentos quantitativos. A Geoestatística trabalha com os casos onde há

dependência entre as UE, complementando a estatística "clássica", não se tratando de uma "nova" estatística.

Como saber se a aleatorização foi suficiente para o "controle" mencionado anteriormente? O uso da Geoestatística pode responder a esta pergunta, estudando-se uma possível estrutura de dependência espacial nas UE. Se a existência de tal dependência não for constatada, o uso do princípio da aleatorização nos experimentos atingirá seu objetivo que é o de controlar a variabilidade associada à heterogeneidade das parcelas.

Com a disponibilidade no mercado de computadores cada vez mais potentes e softwares estatísticos e geoestatísticos cada vez mais acessíveis, todo experimento pode ser analisado quanto a sua variabilidade espacial, pois diante de evidências de que a variabilidade do solo não é aleatória, as pressuposições básicas da análise de variância (independência entre as EU, erros independentes e homocedásticos e aditividade dos efeitos) estariam sendo violadas, podendo levar a interpretações errôneas dos resultados experimentais.

Deixar de considerar a correlação espacial, não apenas transgride os pressupostos, mas também significa deixar de observar importantes aspectos que determinam a ocorrência do fenômeno estudado (Ribeiro Jr., 1995).

Segundo Guerra (1988) e Isaaks & Srivastava (1989), na natureza (em Ciências da Terra) quase sempre existe alguma dependência espacial, muitas vezes não sendo evidente nas amostras disponíveis por serem em número insuficiente, erro amostral, valores erráticos ou possíveis valores discrepantes.

Uma análise geoestatística da dependência espacial dos atributos físicos e químicos do solo pode ajudar na correção do mesmo, antes da instalação do experimento, evitando-se que falsos efeitos sejam atribuídos aos tratamentos testados. Torna-

se oportuna a observação de que muitas instituições de pesquisa, por insuficiência de área experimental, vêm reutilizando áreas usadas em experimentação, na maioria das vezes com objetivos completamente diferentes dos anteriores, sem a devida correção do solo.

No Brasil, ainda existe uma carência considerável de dados georreferenciáveis para que a dependência espacial (se existir), seja detectada e analisada adequadamente. É sempre mais fácil aleatorizar e considerar amostras independentes. Nas ciências agronômicas encontramos alguns trabalhos onde a dependência espacial foi detectada, utilizando-se técnicas geoestatísticas como em: Reichard *et al.* (1986), Souza (1992), Queiroz (1995), Vieira (1997), Pereira *et al.* (1999) e Timm *et al.* (2000).

A época de semeadura de culturas anuais é dependente da ocorrência de chuva e das condições de temperatura e estas épocas podem ser bastante distintas entre regiões. Têm sido realizados no Brasil muitos estudos de como a temperatura afeta o crescimento de plantas e a produção de grãos na agricultura, como por exemplo: Berlato *et al.* (1984), Pezzi & Cavalcanti (1988), Câmara *et al*. (1996), Martorano *et al.* (1999), Gadioli *et al*. (2000) e outros.

A temperatura é o elemento climático mais importante para predizer os eventos fenológicos da cultura, desde que não haja deficiência hídrica (Gadioli *et al.*, 2000).

Portanto, um melhor entendimento do comportamento da temperatura média mensal do ar é de muita importância na área agrícola, assim como no gerenciamento de água e energia elétrica, principalmente nos dias atuais em que o racionamento de energia em algumas regiões do Brasil está sendo necessário.

O uso de temperatura na modelagem do desenvolvimento de algumas culturas, envolvendo o conhecimento do comportamento médio através dos meses em

determinadas regiões para, por exemplo, precisar a melhor época de plantio e qual tipo de cultura melhor se adapta, torna-se imprescindível para uma resposta de qualidade do fenômeno biológico.

Em Geoestatística, cada observação amostral é descrita não apenas pelo seu valor, mas também por informações de sua posição expressa num sistema de coordenadas $(z(\mathbf{u}_\alpha)$ onde $\mathbf{u}_\alpha = (x_a, y_a)$, $a = 1,..., n$, identificam as posição das amostras no campo). Isso possibilita verificar se existe dependência entre amostras vizinhas, considerando que observações mais próximas geograficamente tendem a ter valores mais similares, e se utiliza medida de associação para detectar a estrutura de dependência espacial através das funções de semivariância (Seção 2.3.3), covariância (Seção 2.3.4) e correlação (Seção 2.3.5).

A Geoestatística considera as associações entre amostras vizinhas através de sua geometria, ou seja, da localização das amostras no campo e detecta se existe influência de uma amostra sobre a outra, de acordo com a distância entre elas. A distância na qual as amostras mantêm uma dependência espacial é chamada de *alcance* e corresponde à idéia de zona de influência de uma amostra (Seção 2.3.7).

As técnicas geoestatísticas de interpolação que se baseiam em associações entre amostras vizinhas, podem predizer valores de variáveis em locais não amostrados (de forma que as condições de não tendência e de variância mínima do preditor sejam satisfeitas), se existir dependência espacial nos dados. Na literatura encontram-se muitos métodos de interpolação como, por exemplo, Krigagem Ordinária e Krigagem Universal.

Para a discussão da aplicação de método geoestatístico são considerados dados de clima das 33 estações agrometeorológicas do Instituto Agronômico do

Paraná – IAPAR, que representam os pontos amostrais. Assim, interpolações são necessárias para gerar predições para pontos não amostrados, possibilitando o melhor entendimento dos fatores climáticos e a construção de mapas de temperatura mensal mais precisos. As técnicas de interpolação espacial como Krigagem Ordinária e Krigagem Universal (Seção 2.4), são aplicadas nos dados de temperatura média do ar para os meses de janeiro e julho, objetivando predizer valores onde não existem dados disponíveis.

Sob estas considerações, este trabalho tem como objetivo principal apresentar a Geoestatística como um referencial que pode contribuir na melhoria da precisão das estimativas de temperatura média mensal, utilizadas na construção das cartas climáticas do Estado do Paraná, que servem atualmente como orientação a agricultores e pesquisadores.

2 REVISÃO DE LITERATURA

2.1 Introdução

A variabilidade espacial de propriedades do solo é uma preocupação de pesquisadores, praticamente desde o início do século. Durante um longo tempo ela foi "esquecida" devido à adoção de técnicas como Casualização e Repetição e o melhor conhecimento de funções de distribuição, que levam à adoção de amostragem ao acaso, desprezando assim as coordenadas geográficas do ponto amostrado. Este procedimento, somado à distribuição normal de freqüência, era e ainda é usado para assumir independência entre as amostras e assim garantir a validade do uso da média e do desvio-padrão para representar o fenômeno.

Por outro lado, a distribuição normal não garante, de maneira alguma, a independência entre amostras, a qual pode ser verificada através da autocorrelação. A principal razão para isso é que o cálculo da freqüência de distribuição não leva em conta a distância na qual as amostras foram coletadas no campo.

A presença de dependência espacial requer o uso de um tipo de estatística chamada Geoestatística, a qual surgiu na África do Sul, quando Krige em 1951, trabalhando com dados de concentração de ouro, concluiu que não conseguia encontrar sentido nas variâncias se não levasse em conta a distância entre as amostras. Matheron em 1963 e 1971, citado por Guerra (1988), baseado nestas observações, desenvolveu uma teoria a qual ele chamou de Teoria das Variáveis Regionalizadas, que contém os fundamentos da Geoestatística (veja Seção 2.3).

2.2 Princípios Básicos de Experimentação

Na realização de um experimento, alguns procedimentos devem ser obedecidos para que os objetivos sejam alcançados e as análises estatísticas realizadas com certa segurança, principalmente quanto aos pressupostos exigidos em cada análise estatística.

Nos delineamentos experimentais mais usuais, estatísticas obtidas de dados amostrais como estimativas da variância, coeficiente de variação, entre outras, são consideradas como pertencentes à Estatística Clássica e são encontradas em livros de estatística experimental como: Fisher (1956); Kempthorne (1967); Kronka & Banzatto (1989); Hoffmann & Vieira (1989); Pimentel Gomes (1990); Fisher (1993); Hinkelmann & Kempthorne (1994); Mischan & Pinho (1996) e vários outros.

A experimentação inclui em seu processo planejamento, execução, análise dos dados obtidos e interpretação dos resultados. A experimentação, formalizada através da Estatística, está presente em diversas áreas como: agronomia, medicina, biologia, engenharia, psicologia e outras. Sua origem, porém, se deve à área agronômica onde boa parte da formalização experimental existente, hoje, se deve a sir Ronald Aylmer Fisher (1890-

1962), um estatístico que trabalhou na Estação Experimental de Agricultura de Rothamstead, na Inglaterra. A origem agrícola da experimentação explica o uso de vários termos técnicos, dados a seguir:

Unidade Experimental (UE) ou Parcela designa o elemento experimental indivisível, no qual será aplicado o tratamento. É a menor divisão do material experimental, tal que duas quaisquer UE possam receber diferentes tratamentos.

Tratamentos são procedimentos que estão sendo comparados no experimento. Cada UE receberá um único tratamento. Em alguns experimentos, algumas UE não recebem tratamento, ou recebem um tratamento já conhecido, constituindo o chamado Grupo Controle (Testemunha). A resposta deste grupo será comparada com as respostas dos grupos tratados.

Dados Experimentais (ou Respostas) são as medidas de características do objeto em estudo. Exemplos: 1- Dois suplementos vitamínicos (os tratamentos) são adicionados à ração de cada animal de uma ninhada de porcos (a UE é um animal). O efeito dos tratamentos pode ser avaliado pelo ganho de peso dos animais (resposta); 2- Algumas variedades de uma certa cultura (os tratamentos) são dispostos em faixas homogêneas de terra (as UE). A medida do efeito do tratamento pode ser a resistência a uma doença, avaliada através da intensidade de lesão causada (resposta).

Erro Experimental (ou Resíduo). Mesmo em condições experimentais aparentemente idênticas, existe uma variação "natural" entre as UE, variações estas que não podem ser controladas pelo pesquisador e que fornecem respostas diferentes. A esta(s) fonte(s) de variação não controlada(s) pelo pesquisador denomina-se erro experimental (ou resíduo). Por exemplo: um experimento com plantas que recebem os tratamentos,

possíveis diferenças entre o material genético, e/ou às condições físicas do experimento, como diferenças não controláveis de temperatura, luz, água do solo. “Erros humanos” (intrínsecos ou de conduta do pesquisador), não detectados na instalação do experimento, aplicação dos tratamentos e coleta dos resultados, também são fontes de variação incontroláveis.

Assim, ao se realizar um experimento, o pesquisador deve controlar ao máximo tais fontes de variação para que a maior parte da variação existente entre as UE seja devida ao efeito dos tratamentos. A homogeneidade das UE (ou parcelas), antes de receberem os tratamentos, é condição imprescindível para a validade do experimento.

Princípio da Aleatorização. A variabilidade associada à heterogeneidade das unidades experimentais é chamada de erro casual, “plot error” ou “unit error”. Esta é a variabilidade mais importante do ponto de vista experimental e o seu controle será efetuado pela aleatorização. Assim, uma das maiores funções da aleatorização é controlar o erro das UE. Uma outra função não menos importante da aleatorização, é a construção das distribuições de probabilidades que vão servir como base para a inferência.

Supondo que a aleatorização não seja suficiente para o “controle” acima, ou seja, que existe uma heterogeneidade nas UE cujas causas não são conhecidas, o uso da Geoestatística torna-se fundamental para estudar uma possível dependência espacial na estrutura das UE. A Geoestatística usa esta dependência espacial como sua principal ferramenta de trabalho.

Na literatura, a importância da aleatorização é sentida nas frases de alguns importantes autores:

? **FISHER (1947)**: “Pode-se dizer que a simples precaução da aleatorização é suficiente para garantir a validade do teste de significância, pelo qual o resultado do experimento é julgado”.

? **ANDERSON & BANCROFT (1952)**: “Em experimentos de campo, é conhecido que os erros das UE adjacentes são usualmente positivamente correlacionadas. A aleatorização é usada para obviar este fato”.

? **KEMPTHORNE (1975)**: “O experimentador deve decidir entre as várias causas que produzem variação e quais podem ser controladas experimentalmente. Aquelas que não podem sofrer este processo devem ser controladas por aleatorização”.

A aleatorização consiste no sorteio dos tratamentos às unidades experimentais, com a finalidade de propiciar a todos os tratamentos a mesma probabilidade de serem designados a qualquer das unidades experimentais. Para cada repetição do experimento é realizado um sorteio do tratamento. Com esse processo se evita que um tratamento seja sistematicamente favorecido ou prejudicado em relação a outro. Qualquer método de sorteio pode ser empregado, desde o popular “cara ou coroa”, para o caso de dois tratamentos, até o uso de tabelas de dígitos pseudo-aleatórios (para conhecer mais sobre técnicas de amostragem ver Bussab & Bolfarine, 1994).

O sorteio não vai evitar completamente que um tratamento seja mais ou menos favorecido que outro, mas pelo menos isto acontecerá em uma extensão suficientemente pequena, de modo a não prejudicar a validade dos testes de significância e o estabelecimento de intervalos de confiança.

Repetição. O erro experimental (ou resíduo) das UE pode ser mais bem controlado quando, além do princípio da aleatorização, utiliza-se o princípio da repetição. As UE que recebem o mesmo tratamento são denominadas repetições.

A repetição consiste na reprodução do experimento básico e tem por finalidade propiciar melhor estimativa do erro experimental que mede a variação entre respostas de UE que recebem tratamentos iguais, portanto, somente repetindo o tratamento será possível avaliar o erro.

Quantas repetições devem ser feitas no experimento? Se o pesquisador já dispuser de dados de outros experimentos semelhantes ou de um experimento piloto, algumas fórmulas matemáticas podem ajudar a responder a esta pergunta. Se não, deve-se utilizar o máximo de repetições, dentro de seus recursos materiais, financeiros e éticos. Ressalta-se que, quanto mais heterogêneo for o material experimental, mais repetições devem ser necessárias para que os resultados sejam confiáveis.

Fica intuitivamente claro que a avaliação da diferença real entre os efeitos de dois tratamentos deve melhorar com o aumento do número de repetições do experimento.

Até aqui se abordou o Delineamento Inteiramente Casualizado (DIC), onde as UE são homogêneas e os tratamentos são designados às parcelas de maneira aleatória, casual.

Controle Local. O erro devido à heterogeneidade das unidades experimentais pode ser reduzido, incluindo-se blocos no experimento. Assim temos o Delineamento em Blocos Casualizados (DBC). Este é o princípio do Controle Local, onde as UE são agrupadas em subgrupos homogêneos. A estes subgrupos chamamos de blocos e

dentro de cada bloco os tratamentos são (aleatorizados) sorteados às UE como no DIC. O número de unidades em cada um dos blocos é igual ao número de tratamentos ou um múltiplo dele, tornando-os completos quanto aos tratamentos em consideração. O que se procura é obter blocos com unidades semelhantes, às quais seriam atribuídos todos os tratamentos. Dessa maneira, a variabilidade entre blocos não afeta as diferenças entre os tratamentos, uma vez que cada tratamento aparece em todos os blocos.

Em experimentos com animais, estes devem ser agrupados por características que podem traduzir diferenças na resposta aos tratamentos, tais como peso inicial, sexo, idade, ninhada e outros. Em experimento de campo, os blocos são constituídos de parcelas próximas, com fertilidade semelhante, por exemplo. Não é necessário que os blocos fiquem próximos uns dos outros. No entanto, todas as unidades de um bloco devem ser tratadas de maneira uniforme, sendo a única diferença entre elas o tratamento recebido, objeto do estudo.

Existem outros delineamentos que utilizam a idéia de agrupar unidades similares, constituindo blocos. É o caso dos experimentos em Quadrado Latino que controlam duas causas de variação, isto é, têm dois tipos de blocos. Por exemplo, imagine que se pretende comparar o efeito de quatro rações no ganho de peso de suínos. Os animais disponíveis são, no entanto, diferentes em raça e em peso inicial.

O objetivo a ser alcançado com esta breve introdução aos princípios básicos da experimentação é chamar a atenção ao princípio da casualização (ou alaeatorização) que tem por finalidade propiciar a todos os tratamentos a mesma probabilidade de serem designados a qualquer das unidades experimentais.

Uma preocupação constante entre pesquisadores é com a independência entre as amostras. Segundo Guerra (1988), considerar que variáveis como teor do minério, espessura de um filão, densidade de vegetação, pluviometria, atributos físicos, químicos e granulométricos do solo e temperatura do ar são variáveis aleatórias, não é propriamente verdadeiro, visto que tais parâmetros variam de forma contínua ou descontínua de um ponto a outro, de acordo com a presença ou ausência de algum fenômeno natural.

Entende-se por variável aleatória aquela que pode tomar valores diferentes em diferentes lugares de observação, mostrando desta forma uma determinada independência de um lugar a outro (Guerra, 1988).

Com o objetivo de ajudar a decidir se um campo pode receber tratamento agronômico uniforme, tal como irrigação, Vieira *et al.* (1981) estudaram a magnitude da variação espacial da proporção de infiltração sobre uma área de aproximadamente 160 m por 55 m onde 1280 medidas de infiltração foram obtidas e detectaram uma forte dependência espacial nos dados. Assim, não é correto considerar este campo como excepcionalmente uniforme para propósitos da agricultura, como geralmente é feito. Vieira (1997) encontrou dependência espacial para todos os atributos químicos e granulométricos do solo analisados, concluindo que uma amostragem ao acaso falharia em detectar a variabilidade existente nos atributos químicos do solo.

Pereira *et al.* (1999) estudaram a variabilidade espacial da emissão de CO_2 e de propriedades do solo de uma área de solo preparado (sem vegetação) e constataram a existência de dependência espacial para a emissão de CO_2 e para todas as propriedades de solo analisadas, exceto para o teor de ferro, o qual apresentou distribuição aleatória.

Souza (1992) cita, em sua tese, vários trabalhos da década de 80 que mostram como a variabilidade espacial existente no solo pode conduzir a uma má interpretação dos efeitos de tratamentos em experimentos de campo.

Pode ser constatado que alguns pesquisadores da área agronômica já estão fazendo uso da Teoria das Variáveis Regionalizadas (Geoestatística), preocupados com os problemas relativos aos efeitos da variabilidade espacial (e/ou temporal) nos resultados obtidos nas análises.

Se a existência de tal dependência não for constatada, o uso do princípio da aleatorização nos experimentos atingirá seu objetivo que é o de controlar a variabilidade associada à heterogeneidade das unidades experimentais.

A análise de variância convencional de um clássico desenho aleatorizado é não espacial e é baseada em mínimos quadrados ordinários. Na presença de substancial dependência espacial, os estimadores de mínimos quadrados ordinários são não-tendenciosos, mas não são eficientes em estimar os efeitos de variedades (tratamentos), ou seja, minimizar a variância do erro (Watson, 2000).

Cressie (1993, Seção 5.6) relata sobre *desenho amostral espacial* que objetiva determinar o tamanho n da amostra e suas localizações na região (ou área) de interesse e *desenho experimental espacial* que objetiva estimar os efeitos de tratamentos de respostas que são espacialmente correlacionadas. Neste caso, desenho se refere à especificação de quais tratamentos são aplicados a quais unidades experimentais.

Na presença de dependência espacial, a escolha do desenho experimental também afeta a eficiência da estimação dos efeitos de variedades (tratamentos), devido à posição das variedades dentro do campo (Van Es & Van Es, 1993).

Timm *et al*. (2000) estudaram a relação entre comprimento da raiz e os atributos físicos do solo através de modelos de regressão, porém não de uma forma estática com coeficientes fixos, mas sim de uma forma dinâmica com coeficientes variáveis no espaço através de representação de espaço de estados, ao qual é atribuído o nome de Regressão Dinâmica. Os autores dizem, ainda, que pelo fato de serem modelos de ajuste local, torna-se possível estimar os coeficientes de regressão da(s) variável(is) em cada ponto de amostragem, obtendo-se, desta forma, uma estimativa mais precisa do comportamento da variável de interesse, o que ameniza o problema da variabilidade espacial, principalmente quando se tem interesse de aplicar uma agricultura de precisão.

Reichard *et al.* (1986) estudaram a variabilidade espacial de atributos do solo através de uma revisão bibliográfica e mostram que as técnicas estatísticas classicamente utilizadas e as técnicas geoestatísticas se complementam. Souza (1992) também aborda o assunto.

Quando existe a possibilidade de conhecer a distância entre as amostras, ou seja, quando as amostras são georreferenciáveis, existe a possibilidade de estudar a variabilidade espacial do fenômeno observado através da relação entre pontos vizinhos separados por uma determinada distância $\mathbf{h}$[1].

O fato é que, se a dependência espacial nos dados é constatada, o ideal é usar desenho espacial e um método espacial de análise. Porém, isto tem implicações em

[1] O vetor h é a distância entre dois pontos quaisquer $(\mathbf{u}_a)$ e $(\mathbf{u}_a + \mathbf{h})$, $a = 1,...,n$, pertencentes a área em estudo, logo, será impresso sempre em forma de vetor (em negrito) pois tem magnitude e direção, salvo no texto quando não causar confusão. Os pontos $(\mathbf{u}_a)$, $a = 1,...,n$, identificam uma posição na área em estudo e são vetores de coordenadas (x_a, y_a) por isso também serão impressos em negrito.

termos de análises e interpretações mais complexas e os benefícios dependem da força e do alcance da dependência espacial (Watson, 2000).

Com isto, depreende-se que a Teoria das Variáveis Regionalizadas (Geoestatística) complementa a Teoria da Estatística Clássica e fornece condições para estudar a variabilidade espacial entre os pontos amostrais. A seguir, na Seção 2.3, serão expostos alguns dos conceitos básicos da Geoestatística. Uma boa referência para o estudo desses conceitos é o trabalho de Vieira *et al.* (1983), cujo primeiro autor, é um dos precursores da Geoestatística no Brasil.

2.3 Princípios Básicos de Geoestatística

Definição 1- Um processo estocástico em um espaço de probabilidade $(\Omega, \Im, P)$ é uma coleção de variáveis aleatórias indexadas por um conjunto não vazio X, isto é, $\{Z(\mathbf{u}), \mathbf{u} \in X\}$.

O conjunto X acima, será aqui o espaço real p-dimensional, isto é, $X = \Re^p$.

Um processo estocástico é caracterizado pela distribuição conjunta de qualquer conjunto de variáveis aleatórias $Z(\mathbf{u}_1), Z(\mathbf{u}_2), ..., Z(\mathbf{u}_n)$ para todo n inteiro e $\mathbf{u}_a \in \Re^p$, onde $a = 1,2,..., n$. Para todo $\mathbf{u}_k$ fixo, $Z(\mathbf{u}_k)$ é uma variável aleatória. Assim:

?A média do processo estocástico, m($\mathbf{u}$), é definida como $m(\mathbf{u}) = E[Z(\mathbf{u})]$, quando a esperança existe.

?Dado $(\mathbf{u}_1, \mathbf{u}_2) \in \Re^p x \Re^p$, a função de covariância do processo estocástico, C($\mathbf{u}_1$,$\mathbf{u}_2$), é definida como

$C(\mathbf{u}_1,\mathbf{u}_2) = C[Z(\mathbf{u}_1), Z(\mathbf{u}_2)] = E\{[Z(\mathbf{u}_1) - m(\mathbf{u}_1)][Z(\mathbf{u}_2) - m(\mathbf{u}_2)]\}$, quando a covariância existe. Logo, $C(\mathbf{u},\mathbf{u}) = E\{[Z(\mathbf{u}) - m(\mathbf{u})]^2\} = V[Z(\mathbf{u})]$.

?A função *variograma* do processo estocástico, $2?(\mathbf{u}_1,\mathbf{u}_2)$, é definida como a variância do incremento $[Z(\mathbf{u}_1) - Z(\mathbf{u}_2)]$, ou seja, $2?(\mathbf{u}_1,\mathbf{u}_2) = V[Z(\mathbf{u}_1) - Z(\mathbf{u}_2)]$. A função $?(\mathbf{u}_1,\mathbf{u}_2)$ é denominada de *semivariograma*.

2.3.1 Variáveis Regionalizadas

Matheron (1963, citado por Vieira, 1996) define Variável Regionalizada (VR) como uma função espacial numérica, que varia de um local para outro, com uma continuidade (dependência espacial) aparente e cuja variação não pode ser representada por uma função matemática simples.

Considere um campo de área *W*, para o qual se tem um conjunto de valores medidos $\{z(\mathbf{u}_a), a = 1,..., n\}$, onde $\mathbf{u}_a$ identifica uma posição no espaço, ou no tempo, e é um vetor de coordenadas (x_a, y_a), isto é, os valores da VR são relacionados de algum modo com a posição espacial que ocupam.

Para uma dada posição $\mathbf{u}_k$, cada valor medido da variável em estudo, $z(\mathbf{u}_k)$, pode ser considerado como uma realização de uma certa variável aleatória, $Z(\mathbf{u}_k)$.

Uma função aleatória é definida como um conjunto de variáveis aleatórias dependentes $Z(\mathbf{u})$, uma para cada localização $\mathbf{u}$ na área em estudo *W*, $\{Z(\mathbf{u}), \forall \mathbf{u} \in W\}$. Para qualquer conjunto de n localizações $\mathbf{u}_a$, $a = 1,2,..., n$, corresponde um vetor de n variáveis aleatórias $Z(\mathbf{u}_1), Z(\mathbf{u}_2),...., Z(\mathbf{u}_n)$ que é caracterizado pela função de distribuição acumulada n-variada,

$$F(\mathbf{u}_1,...,\mathbf{u}_n; z_1,...,z_n) = Prob\{Z(\mathbf{u}_1) \le z_1,..., Z(\mathbf{u}_n) \le z_n\}.$$

A função de distribuição multivariada acumulada acima caracteriza a conjunta incertamente sobre os n reais valores $z(\mathbf{u}_1), z(\mathbf{u}_2),...., z(\mathbf{u}_n)$. O conjunto de tais n-variadas funções de distribuição acumuladas, para qualquer inteiro positivo n e para qualquer escolha das localizações $\mathbf{u}_a$, constituem a lei espacial da função aleatória Z($\mathbf{u}$). Na prática, a análise é limitada a funções de distribuições acumuladas envolvendo não mais que duas localizações de cada vez,

$F(\mathbf{u}_1, \mathbf{u}_2; z_1, z_2) = Prob\{Z(\mathbf{u}_1) \le z_1, Z(\mathbf{u}_2) \le z_2\}$, e seus primeiros momentos correspondentes. Desta forma, é necessário que haja repetidas medidas da variável regionalizada em cada localização para inferência da função aleatória e seus primeiros momentos (Goovaerts, 1997, p68; Vieira Braga, 1990, p3).

Ter várias medidas para a mesma localização $\mathbf{u}_a = (x_a, y_a)$ dificilmente é disponível, na prática. O que se tem é uma medida para cada localização.

A impossibilidade de repetir indefinidamente o experimento quando se trata de variáveis observadas, como, por exemplo, teores de um minério em uma jazida ou precipitação de chuva em um dia, é uma das características que limitam a modelagem de fenômenos naturais por variáveis aleatórias (Vieira Braga, 1990, p5).

Quando um conjunto de repetições consiste de medidas para a mesma localização, mas em tempos diferentes, temos a função aleatória Z($\mathbf{u}$; t) que é definida no espaço e no tempo. Um processo espaço-tempo é denotado como $\{Z(\mathbf{u}; t) : \mathbf{u} \in D(t), t \in T\}$.

2.3.2 Estacionariedade

Muitas vezes o principal interesse é estimar valores em locais não amostrados. Para tanto, é preciso introduzir a restrição de que a VR seja estacionária estatisticamente. Entende-se que uma variável regionalizada é estacionária se os momentos estatísticos da variável aleatória $Z(\mathbf{u}_a)$ forem invariantes para qualquer translação efetuada.

Usando todos os pares de medidas separados por um vetor $\mathbf{h}$ dentro da área $\Re^p$, $\{z(\mathbf{u}_\alpha), z(\mathbf{u}_\alpha + \mathbf{h}); \alpha = 1,..., n\}$, como um conjunto de repetições, a implícita suposição é que os correspondentes pares de variáveis aleatórias $\{Z(\mathbf{u}_\alpha), Z(\mathbf{u}_\alpha + \mathbf{h}); \alpha = 1,..., n\}$ venham de uma mesma distribuição. Assim, o modelo da função aleatória $Z(\mathbf{u})$ precisa ser escolhido para ser estacionário dentro de $\Re^p$, caso contrário, os dados representam uma incompleta amostragem de uma única realização, não sendo possível fazer inferência.

Definição 2- Um processo estocástico é dito estritamente estacionário se a distribuição conjunta de $\{Z(\mathbf{u}_1), Z(\mathbf{u}_2),...., Z(\mathbf{u}_n)\}$ é a mesma de $\{Z(\mathbf{u}_1 + \mathbf{h}), Z(\mathbf{u}_2 + \mathbf{h}),...., Z(\mathbf{u}_n + \mathbf{h})\}$ para qualquer conjunto finito de pontos $\{\mathbf{u}_1, \mathbf{u}_2,..., \mathbf{u}_n\}$ e qualquer vetor $\mathbf{h} \in \Re^p$, ou seja, o processo é invariante sob translação no conjunto de índices.

Definição 3- Um processo estocástico é dito estacionário de segunda ordem quando:

i) a esperança $E[Z(\mathbf{u})]$ existe e não depende do ponto $\mathbf{u}$, isto é,

$E[Z(\mathbf{u})] = m$, $\forall \mathbf{u} \in \Re^p$;

ii) $E[Z(\mathbf{u})]^2 < \infty$; e

iii) para cada par de variáveis aleatórias $\{Z(\mathbf{u} + \mathbf{h}), Z(\mathbf{u})\}$ a função de covariância existe e depende apenas do vetor $\mathbf{h} \in \Re^p$:

$$C(\mathbf{u}+\mathbf{h},\mathbf{u}) = E[Z(\mathbf{u}+\mathbf{h})Z(\mathbf{u})] - m^2 = C(\mathbf{h}),\ \forall \mathbf{u} \in \Re^p .$$

Definição 4- Um processo estocástico é dito intrínseco quando:

i) a esperança E [Z(**u**)] existe e não depende do ponto **u**, isto é, $E[Z(\mathbf{u})] = m$, $\forall \mathbf{u} \in \Re^p$;

ii) para todo $\mathbf{h} \in \Re^p$, a variância da diferença $[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})]$ existe e não depende do ponto $\mathbf{u}$, isto é, $V[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})] = 2?(\mathbf{u}+\mathbf{h},\mathbf{u}) = 2?(\mathbf{h})$, $\forall \mathbf{u} \in \Re^p$ (incremento tem variância finita).

Esta hipótese intrínseca é menos restritiva que as outras hipóteses existentes, visto que a estacionariedade de segunda ordem exige variância finita dos valores medidos. Alguns fenômenos físicos têm capacidade infinita de dispersão, como por exemplo, concentração de ouro, movimento Browniano e algumas cadeias de Markov.

Quando a hipótese de estacionariedade de segunda ordem é satisfeita, isto implica que a hipótese intrínseca também o é; mas, se a hipótese intrínseca é confirmada, isto não implica na estacionariedade de segunda ordem (Lúcio, 1991).

Se a hipótese de estacionariedade de segunda ordem for verificada, pode-se usar o variograma $2?(\mathbf{h})$ ou a covariância $C(\mathbf{h})$ para representar a dependência espacial existente nos dados.

Estará se referindo, ao assumir hipótese intrínseca, que a função aleatória dos incrementos $[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})]$ é que é estacionária de segunda ordem (Journel & Huijbregts, 1978, p33, citado por Goovaerts, 1997, p71). É a respeito da variabilidade dos incrementos que estará se referindo ao se efetuar os cálculos fundamentais, como a estimação do variograma.

Verificar se existe estacionariedade nos dados e identificar em qual tipo de estacionariedade os dados se enquadram, em geral, é uma tarefa difícil. O exame do semivariograma ou do covariograma e um teste conhecido como Validação Cruzada (ver Seção 2.5), são as principais ferramentas utilizadas para se conhecer indiretamente a estacionariedade dos dados (Vieira, 1996).

Até o início dos anos 60, a análise de dados era feita sob a hipótese de independência estatística ou distribuição espacial aleatória (ver Seção 2.2), para permitir o uso de métodos estatísticos como análise de variância e parâmetros como o coeficiente de variação. Entretanto, este tipo de hipótese não pode simplesmente ser feito antes que se prove a não existência de correlação de amostras com distâncias entre elas, ou seja, a não existência de dependência espacial (Ribeiro Jr., 1995).

Um dos métodos de se estimar a dependência de amostras vizinhas no espaço ou no tempo, é através da autocorrelação que tem sua origem em análise de séries temporais. Porém, quando as amostras forem coletadas nas duas dimensões do campo, ou seja, quando são georreferenciáveis e interpolações entre locais medidos forem necessárias para a construção de mapas de isolinhas, será preciso uma ferramenta mais adequada para medir a dependência espacial: o semivariograma ou o covariograma.

2.3.3 Semivariograma

Considerando então todos os pares de medidas, da variável aleatória de interesse, separados por um vetor **h** dentro da área W estudada $\{(z(\mathbf{u}_\alpha), z(\mathbf{u}_\alpha + \mathbf{h})); \alpha = 1,..., n\}$, como um conjunto de repetições, temos uma função que depende da distância **h**. As diferenças quadráticas entre valores que levam em conta a distância **h** que os separa (chamada de medida

de dissimilaridade) permitem a construção do semivariograma, que é a ferramenta base da Geoestatística.

A dissimilaridade entre os pares de valores medidos, $z(\mathbf{u}_\alpha + \mathbf{h})$ e $z(\mathbf{u}_\alpha)$, localizados nos pontos $(\mathbf{u}_\alpha + \mathbf{h})$ e $(\mathbf{u}_\alpha)$, $\alpha = 1,2,...,n$, em um domínio espacial W é dado por

$$\gamma^*(\mathbf{h}) = \frac{[z(\mathbf{u}_\alpha + \mathbf{h}) - z(\mathbf{u}_\alpha)]^2}{2},$$ ou seja, metade do quadrado da diferença entre os dois valores.

A dissimilaridade depende, então, do comprimento e da direção do vetor de separação entre os pontos, $\mathbf{h} \in W$. Ao construir um gráfico com as dissimilaridades de todos os pontos dois a dois, $\alpha = 1,2,...,n$, contra o vetor de separação **h**, tem-se o chamado (semi)variograma de nuvem ("(semi)variogram Cloud").

Como a dissimilaridade é uma diferença elevada ao quadrado, $[z(\mathbf{u}_\alpha + \mathbf{h}) - z(\mathbf{u}_\alpha)]^2$, a ordem em que as localizações são consideradas não altera o resultado, sendo assim a dissimilaridade simétrica com respeito ao vetor **h**:

$$[z(\mathbf{u}_\alpha + \mathbf{h}) - z(\mathbf{u}_\alpha)]^2 = [z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h})]^2$$

$$\gamma^*(\mathbf{h}) = \gamma^*(-\mathbf{h}). \qquad (1)$$

Diggle & Ribeiro Jr. (2000), apresentam um típico semivariograma de nuvem. Em geral, o variograma de nuvem é dominado por muitos pares com baixa dissimilaridade para todas as escalas de **h** (Figura 1).

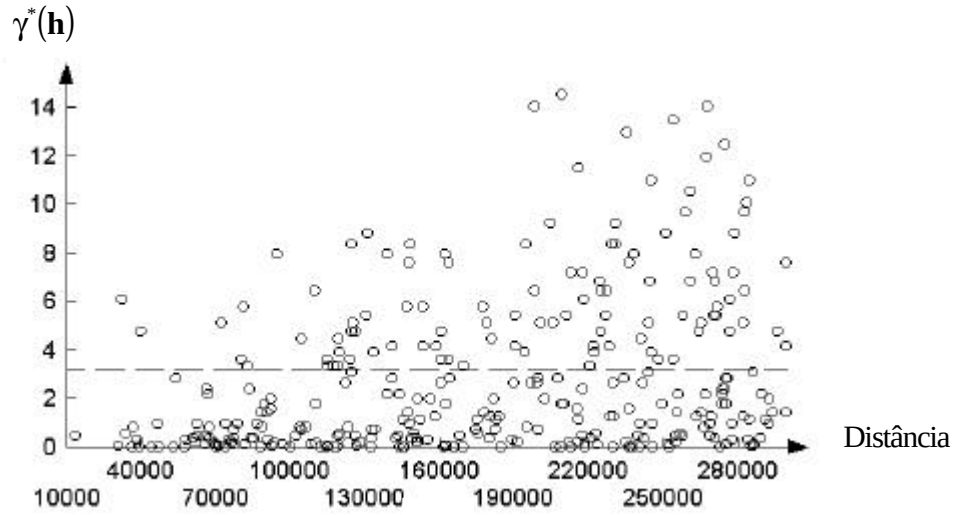

Figura 1 - Exemplo de semivariograma de nuvem.

Para um dado valor de **h**, temos uma nuvem de pontos formada por todos os pares amostrais $[Z(\mathbf{u}_a + \mathbf{h}), Z(\mathbf{u}_a)]$ separados por essa distância. Uma estimativa pontual do semivariograma, chamada de semivariância, pode ser interpretada como o momento de inércia sobre o 1º bissetor (veja Goovaerts, 1997, p29 e Pannatier, 1996, p38). Considerando alguns valores de **h**, ou seja, subdividindo o semivariograma de nuvem em intervalos (ou classes) e encontrando a dissimilaridade média em cada intervalo, forma-se uma seqüência de estimadores pontuais $\hat{\gamma}(\mathbf{h})$, o chamado semivariograma experimental:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2n(\mathbf{h})} \sum_{(\mathbf{u}_\alpha + \mathbf{h}, \mathbf{u}_\alpha)} [Z(\mathbf{u}_\alpha + \mathbf{h}) - Z(\mathbf{u}_\alpha)]^2 , \ \forall \mathbf{u}_\alpha, \mathbf{h} \in \Re^p , \qquad (2)$$

onde n(**h**) é o número de pares amostrais $[Z(\mathbf{u}_a + \mathbf{h}), Z(\mathbf{u}_a)]$ separados pelo vetor **h**. Este é o estimador baseado no método dos momentos proposto por Matheron em 1963, também chamado de clássico estimador do semivariograma.

O semivariograma mede a variação entre pares de valores separados por uma distância determinada, **h**.

Se Z(**u**) for uma função aleatória estacionária, então (2) é um estimador não tendencioso para o semivariograma.

Entretanto, "a priori" não há um teste que permita dizer se o conjunto particular de realizações de Z(**u**) provém de uma função aleatória estacionária (Vieira Braga, 1990, p11).

Isto só se torna uma hipótese aceitável após a obtenção dos resultados dos semivariogramas experimentais e sua comparação com as propriedades características de um semivariograma teórico.

O semivariograma dado em (2) é sensível a valores extremos. Isto se deve ao quadrado dos incrementos $[z(\mathbf{u}_\alpha + \mathbf{h}) - z(\mathbf{u}_\alpha)]^2$. Assim o semivariograma de nuvem pode levar a confundir assimetria com comportamento atípico. Cressie (1993, p40) ilustra este fato com um gráfico de caixa ("Box plot") do semivariograma de nuvem para vários valores de **h**.

O semivariograma de ordem w (3), definido como o desvio médio absoluto de expoente w pode ser usado para inferir características tais como alcance e anisotropia, mas não substitui o semivariograma tradicional, para w=2.

$$\hat{\gamma}_w(\mathbf{h}) = \frac{1}{2n(\mathbf{h})} \sum_{(\mathbf{u}_\alpha + \mathbf{h}, \mathbf{u}_\alpha)} \left| Z(\mathbf{u}_\alpha + \mathbf{h}) - Z(\mathbf{u}_\alpha) \right|^w, \text{ com } w \in [0,2]. \qquad (3)$$

Para w=1, (3) é chamado de "Madogram" e, para w=1/2 de "Rodogram" (Deutsch & Journel, 1992a, p41, citado por Goovaerts, 1997, p31).

Cressie (1993, p75) apresenta estimadores robustos (a contaminação por "outliers") para o variograma em que um deles é baseado na mediana.

Diblasi & Bowmam (2001) usam w = ½ em (3) como uma maneira de exibir informação sobre uma escala mais amena e robusta e propõem um teste estatístico para aferir independência espacial nos dados. Eles consideram, sob suposição de estacionariedade e isotropia, o modelo $Z(\mathbf{u}) = \mu + \varepsilon(\mathbf{u})$, onde $\varepsilon(\mathbf{u})$ é normalmente distribuído com média zero e variograma $\gamma(\mathbf{h})$ e testam sob a hipótese nula de independência, $\gamma(\mathbf{h}) = \sigma^2$, um valor constante.

Na prática, as distâncias (**h**) entre as localizações (chamadas de "lags"), são consideradas em até metade do diâmetro da região (Wackernagel, 1995).

Quando a distância **h** entre os pontos aumenta, o número de pontos considerados no cálculo do semivariograma diminui e a precisão nos resultados obtidos também. Por isso considera-se que um bom modelo deve se ajustar com maior precisão aos primeiros pontos experimentais. A dependência espacial tende a diminuir com o aumento na distância **h**, por isso na prática os semivariogramas experimentais são calculados apenas para distâncias que não excedam ½ das dimensões da região amostrada (em alguns casos em ¼, como descrito em Uzumaki, 1994, p19).

Guerra (1988, p62) diz que, na prática, o variograma experimental se calculará até no máximo ½ do campo geométrico. Depois desse valor, o semivariograma perde o significado experimental.

O semivariograma experimental é ajustado somente para semivariâncias de distâncias ("lags") de até metade da distância máxima porque a variância amostral das estimativas tende a aumentar excessivamente para grandes distâncias, devido ao decréscimo do número de observações envolvidas (Watson, 2000).

A falta de estrutura espacial nos dados é detectada se a dissimilaridade média é constante para todo valor de **h**.

Um conjunto de dados pode, freqüentemente, não apresentar qualquer estrutura de dependência espacial. A falta de evidente estrutura nas amostras disponíveis não justifica usar um modelo de função aleatória não correlacionada espacialmente. Segundo Isaaks & Srivastava (1989, p297), em Ciências da Terra, quase sempre existe alguma estrutura de dependência espacial, não sendo evidente nas amostras disponíveis por insuficiente número de repetições, erro amostral, valores erráticos ou possíveis valores discrepantes.

Os ajustes dos modelos de semivariogramas teóricos (Seção 2.3.9) ao semivariograma experimental (2) devem levar em conta n(**h**), ou seja, a quantidade de pares amostrais com que é avaliado cada ponto do semivariograma experimental (cada semivariância).

Em geral, observa-se que a dissimilaridade média entre os valores amostrais aumenta quando o espaçamento entre pares de pontos amostrais é aumentado. Para grandes espaçamentos o variograma experimental, algumas vezes, alcança um patamar que pode ser igual à variância dos dados (Wackernagel, 1995, p33).

Como a função semivariograma depende da distância **h**, e esta distância é considerada um vetor, tendo magnitude $\|\mathbf{h}\|$ e direção θ, o semivariograma deve ser construído em distintas direções no espaço (por exemplo: na direção X $\Rightarrow \theta = 0°$; na direção Y $\Rightarrow \theta = 90°$ e, nas direções das duas diagonais $\Rightarrow \theta = 45°$ e $\theta = -45°$) visando estudar como se deforma ao mudar a direção do vetor **h**.

Ao gráfico pontual de $[\hat{\gamma}(\mathbf{h}) \quad vs \quad \mathbf{h}]$ para cada direção θ, ajusta-se um modelo para a função semivariograma. Os modelos básicos que podem ser usados serão descritos na Seção 2.3.9.

Quando o gráfico do semivariograma experimental exibe o mesmo comportamento em todas as direções consideradas, ele é chamado isotrópico, isto é, depende apenas do comprimento[2] do vetor **h**, não de sua direção. Caso contrario, é chamado anisotrópico e deve sofrer transformações antes de ser usado.

Cressie (1993), Uzumaki (1994) e Goovaerts (1997), abordam os tipos de anisotropias (geométricas e zonais) e sugerem transformações (quando é possível) para se obter isotropia. Guerra (1988, p55) é a principal referência por conter figuras ilustrativas que facilitam a compreensão do fenômeno.

No caso isotrópico, os semivariogramas experimentais dependem somente da distância de separação entre os pontos e não de sua direção. Logo, todos os semivariogramas experimentais direcionais serão o mesmo. Assim, calculando um semivariograma experimental com um ângulo de tolerância de 90°, é equivalente a calcular o semivariograma médio considerando todas as direções. A este semivariograma experimental atribui-se o nome de "omnidirectional" e a modelagem é feita sobre ele. Uma outra opção é modelar os semivariogramas experimentais direcionais usando os modelos básicos e combiná-los linearmente para obter um ajuste satisfatório. Esta combinação linear de modelos de

[2] No caso isotrópico, onde o vetor **h** depende somente do comprimento e não da direção do vetor, o "lag" h pode não ser escrito, necessariamente, em negrito. Neste trabalho: h, $|\mathbf{h}|$ e $\|\mathbf{h}\|$ representam o comprimento do vetor **h**, desconsiderando-se sua direção.

semivariogramas positivos definidos, com coeficientes positivos, é também um modelo positivo definido: $\gamma(\mathbf{h}) = \sum_{i=1}^{N} |w_i| \gamma_i(\mathbf{h})$.

Em geoestatística esta combinação linear é chamada de modelo de estruturas próximas ("nested structures") (Isaaks & Srivastava, 1989, p376).

Os coeficientes de w_i, i=1,...,N precisam ser escolhidos de tal forma que a soma deles seja igual ao patamar do variograma experimental.

A condição de que os coeficientes precisam ser maiores que zero é suficiente para garantir a condição de positivo definido, mas não é necessária. Uma combinação linear de modelos positivos definidos com coeficientes negativos pode ou não ser positivo definido. Por exemplo, coeficientes negativos são requeridos para modelar o semivariograma cruzado quando duas variáveis são negativamente correlacionadas.

Na prática, em geral, tem-se dados irregularmente espaçados sobre a área W considerada. Então as distâncias consideradas para o cálculo do semivariograma deverão ser estimadas através de uma regularização angular e por classes de distância: considerando um ponto qualquer $\mathbf{u}_\alpha \in W$, segundo uma direção θ (pré-fixada), gera-se um cone de abertura $2\delta\theta$, onde, $2\delta\theta$ é dito ângulo de regularização[3], e são utilizados os pontos que "caiam dentro" do cone na classe de distância $\mathbf{h} \pm ? \mathbf{h}$. Isaaks & Srivastava (1989) e Cressie (1993) seguem recomendação de Journel & Huijbregts (1978, p194) de que o número de pares distintos que "caiam dentro" do cone (definido acima) seja no mínimo 30. Logo, as regiões de tolerância devem ser escolhidas de forma que esta condição seja satisfeita.

[3] Ou ângulo de tolerância como é chamado no manual do programa Variowin (Pannatier, 1996).

O semivariograma teórico $\gamma(\mathbf{h})$ é definido pela hipótese intrínseca (ver Definição 4 Seção 2.3.2), isto é, para quaisquer pares de pontos $(\mathbf{u}+\mathbf{h}, \mathbf{u}) \in W$ a esperança e variância dos incrementos são:

$$E[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})] = 0 \text{ , e} \tag{4}$$

$$V[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})] = E[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})]^2 - \{E[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})]\}^2$$

$$= E[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})]^2 = 2\gamma(\mathbf{h}), \text{ logo}$$

$$\gamma(\mathbf{h}) = \frac{1}{2} E[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})]^2 . \tag{5}$$

Assim, o semivariograma é definido pela variância dos incrementos e reflete a estrutura espacial do fenômeno estudado desde que pelo menos a hipótese intrínseca possa ser aceita.

Destacam-se as seguintes propriedades da função semivariograma (e do variograma):

i) $\gamma(0) = 0$; e como $\gamma^*(\mathbf{h}) = \gamma^*(-\mathbf{h})$

ii) $\gamma(\mathbf{h}) = \gamma(-\mathbf{h}) \geq 0$. (6)

O objetivo fundamental de um estudo semivariográfico (ou variográfico) é verificar qual o semivariograma teórico (de referência) que melhor se ajusta ao semivariograma observado (experimental), de tal modo que a partir deste modelo teórico possam ser feitas inferências em relação ao semivariograma verdadeiro (desconhecido).

A maneira de ajustar os modelos de semivariogramas teóricos (Seção 2.3.9) ao semivariograma experimental, que é chamado por Guerra (1988, p62) de “método das aproximações sucessivas”, consiste em alternar modelos e parâmetros até que se encontre

um modelo que se ajuste com maior precisão aos primeiros pontos experimentais. Ou seja, "quando as discrepâncias entre os valores experimentais e valores teóricos forem mínimas", como pode ser constatado, o ajuste, na realidade, é feito "a olho" (ou "a sentimento"), "by eye" como ressalta Mandallaz (2000).

Uma outra maneira de ajustar uma função às semivariâncias é pelo método dos mínimos quadrados ponderados (MQP). Este método, segundo Watson (2000), foi escolhido como um prático trabalho de conciliação entre mínimos quadrados generalizados (MQG), que leva em conta as diferentes variâncias e covariâncias das semivariâncias estimadas em diferentes distâncias ("lags"); e, mínimos quadrados ordinários (MQO) que os ignora. Mínimos quadrados ponderados considera o peso das observações inversamente proporcional a suas variâncias[4], mas não leva em consideração covariâncias entre as observações. Conseqüentemente, MQP provavelmente será mais eficiente que MQO, mas poderá ser menos eficiente que MQG. Em muitos trabalhos, MQP é escolhido por ser consideravelmente mais simples de ser aplicado que MQG. Mais sobre ajustes de mínimos quadrados pode ser visto em Cressie (1993, p94).

2.3.4 Covariograma

A função covariância $C(\mathbf{h})$ (covariograma) é definida com base nas hipóteses de estacionariedade dos dois primeiros momentos (média e covariância) da função aleatória $Z(u)$ (ver Definição 3 Seção 2.3.2).

[4] Variância de uma semivariância estimada $s(\mathbf{h})$ depende da distância ("lag"), do número de pares no qual isto é calculado, e da dependência espacial existente nos dados.

$$E[Z(\mathbf{u})] = m, \ \forall \mathbf{u} \in W, \text{ e} \tag{7}$$

$$C[Z(\mathbf{u}+\mathbf{h}), Z(\mathbf{u})] = E[Z(\mathbf{u}+\mathbf{h})Z(\mathbf{u})] - E[Z(\mathbf{u}+\mathbf{h})]E[Z(\mathbf{u})]$$

$$C(\mathbf{h}) = E[Z(\mathbf{u}+\mathbf{h})Z(\mathbf{u})] - m^2 . \tag{8}$$

Sob a hipótese de estacionariedade de segunda ordem (ou covariância estacionária), o covariograma e o semivariograma são duas ferramentas equivalentes para caracterizar as autocorrelações entre duas variáveis separadas por uma distância **h**.

Assim, as funções variância e semivariograma podem ser deduzidas da função covariância:

$$V[Z(\mathbf{u})] = E\{[Z(\mathbf{u}) - m]^2\} = C(0), \ \forall \mathbf{u} \in W, \text{ e} \tag{9}$$

multiplicando (8) por dois,

$$2C(\mathbf{h}) = 2E[Z(\mathbf{u}+\mathbf{h})Z(\mathbf{u})] - 2m^2 .$$

Somando-se e subtraindo-se $E[Z(\mathbf{u}+\mathbf{h})]^2$ e $E[Z(\mathbf{u})]^2$ no lado direito da igualdade, tem-se

$$2C(\mathbf{h}) = \{E[Z(\mathbf{u}+\mathbf{h})]^2 - m^2\} + \{E[Z(\mathbf{u})]^2 - m^2\}$$

$$- \{E[Z(\mathbf{u}+\mathbf{h})]^2 - 2E[Z(\mathbf{u}+\mathbf{h}).Z(\mathbf{u})] + E[Z(\mathbf{u})]^2\};$$

$C(\mathbf{h}) = C(0) - \gamma(\mathbf{h})$ e então, (10)

$$\gamma(\mathbf{h}) = \frac{1}{2} E\{[Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})]^2\} = C(0) - C(\mathbf{h}), \ \forall \mathbf{u} \in W, \text{ e } \forall \mathbf{h} \in W.$$

Seja $Z(\mathbf{u})$ uma função aleatória estacionária de esperança m, covariograma $C(\mathbf{h})$ e semivariograma $\gamma(\mathbf{h})$. O covariograma é uma função do vetor **h** (isto é, de seu módulo e direção) e suas propriedades são:

i) $C(\mathbf{0}) = Var[Z(\mathbf{u})] \geq 0$;

ii) $C(\mathbf{h}) = C(-\mathbf{h})$;

iii) a função covariância é limitada e seu valor absoluto não excede a variância

$$|C(\mathbf{h})| \leq C(\mathbf{0}) = \mathrm{Var}[Z(\mathbf{u})]. \quad (11)$$

Uma condição necessária e suficiente para $C(\mathbf{h})$ expressar a função covariância de $Z(\mathbf{u})$ é que $C(\mathbf{h})$ seja positiva semidefinida.

2.3.5 Correlograma

De (9) e (10) tem-se que, sob a hipótese de estacionariedade da covariância, a covariância e o variograma são duas ferramentas equivalentes para caracterizar as autocorrelações entre duas variáveis separadas por um vetor $\mathbf{h}$, $Z(\mathbf{u}+\mathbf{h})$ e $Z(\mathbf{u})$. Assim, o correlograma, $\rho(\mathbf{h})$, é definido como

$$\rho(\mathbf{u}+\mathbf{h},\mathbf{u}) = \frac{C(\mathbf{u}+\mathbf{h},\mathbf{u})}{\sqrt{C(\mathbf{u},\mathbf{u})C(\mathbf{u}+\mathbf{h},\mathbf{u}+\mathbf{h})}} = \frac{C(\mathbf{h})}{\sqrt{C(\mathbf{0})C(\mathbf{0})}} = \frac{C(\mathbf{h})}{C(\mathbf{0})} = \rho(\mathbf{h}).$$

De (10) tem-se que $C(\mathbf{h}) = C(\mathbf{0}) - \gamma(\mathbf{h})$, assim,

$$\rho(\mathbf{h}) = \frac{C(\mathbf{h})}{C(\mathbf{0})} = 1 - \frac{\gamma(\mathbf{h})}{C(\mathbf{0})}, \ \forall \mathbf{h} \in W. \quad (12)$$

Vieira *et al*. (1981) ilustram o uso de variogramas e autocorrelogramas como ferramentas para identificar o grau de dependência da proporção de infiltração com a distância entre os pares de medidas e como tirar proveito dessa dependência espacial.

2.3.6 Efeito Pepita

O comportamento para escalas muito pequenas, próximo da origem do semivariograma, é crítico quando isso indica o tipo de continuidade da variável regionalizada:

diferenciável, contínua, mas não diferenciável ou descontínua. Quando o semivariograma é descontínuo na origem, este é um sintoma do efeito pepita ("nugget effect"), o que significa que, para escalas muito pequenas, os valores da variável mudam abruptamente (Hudson & Wackernagel, 1994).

O nome efeito pepita é pelo fato de ser o caso que apresentam os minerais de distribuição geralmente errática, como depósitos de ouro (Guerra, 1988, p54).

Uma das propriedades da função semivariograma dada em (6) diz que $\gamma(\mathbf{0})=0$, mas o valor do semivariograma pode não tender a zero quando **h** tende a zero. Esta descontinuidade na origem do semivariograma é chamada de efeito pepita e representa as variações locais ou a pequena escala (menores do que a menor distância entre as amostras), como erros de análise, amostragem e outros.

É impossível de se quantificar qual contribui mais, se os erros de medição ou a variabilidade a uma escala menor do que aquela amostrada.

O efeito pepita é utilizado para representar uma descontinuidade na origem de um semivariograma e pode ser explicado fisicamente (por medida de erro ou por micro-estruturas). As micro-estruturas são estruturas espaciais abaixo da escala definida pelo espaço amostral (Guerra, 1988).

O chamado efeito pepita reflete o desconhecimento do comportamento da correlação para distâncias menores que a menor distância observada entre os pontos $\mathbf{u}_\alpha$ (Mandallaz, 2000).

Deve-se ressaltar que é praticamente impossível de diferenciar qual parte do efeito pepita é devida a erros de análise e qual a erros de pouca reprodutibilidade

(erros de análise ≅ erros de medição e erros de pouca reprodutibilidade ≅ variabilidade a uma escala menor do que aquela amostrada).

O semivariograma na Figura 2 é considerado ideal. Nele pode-se localizar o efeito pepita representado por c_0.

Em Cressie (1993, p127) encontra-se uma discussão sobre os tipos de erros que podem ser encontrados e como eles afetam o método de predição de Krigagem a ser visto na Seção 2.4.

2.3.7 Alcance

É de se esperar que em fenômenos naturais, a correlação espacial entre duas variáveis $Z(\mathbf{u}+\mathbf{h})$ e $Z(\mathbf{u})$ desapareça quando a distância **h** aumenta muito ($C(\mathbf{h})\rightarrow 0$ quando $\|\mathbf{h}\|\rightarrow\infty$). Ou seja, a distância a partir da qual um ponto da variável em estudo não tem mais influência sobre o ponto vizinho ($C(\mathbf{h})=0$ quando $\|\mathbf{h}\|>a$) marca o início da zona de pura aleatoriedade.

Mesmo que exista, teoricamente, tendência de $\gamma(\mathbf{h})$ crescer indefinidamente com **h**, isto é, que a partir de uma certa distância **h**, as amostras não tem influência umas sobre as outras. O valor *a*, denominado alcance (amplitude ou "range"), corresponde à idéia de zona de influência de uma amostra e é interpretado como (ver Figura 2):

$$\text{Se} \quad \begin{cases} \|\mathbf{h}\| < a, & \text{existe correlação espacial} \\ \|\mathbf{h}\| > a, & \text{ausência de correlação espacial.} \end{cases}$$

2.3.8 Patamar

A medida que **h** aumenta, $\gamma(\mathbf{h})$ também aumenta até um valor máximo no qual se estabiliza. Este valor, no qual $\gamma(\mathbf{h})$ se estabiliza, chama-se patamar ("sill") e é designado por c_0+c (ver Figura 2).

O patamar pode ser interpretado como a variância "a priori" da função aleatória. O patamar c_0+c marca o valor da variação total da variável em estudo.

$$\gamma(\infty) = \mathrm{Var}[Z(\mathbf{u})] = C(\mathbf{0}).$$

O valor c é denominado variância espacial e representa as diferenças espaciais entre os valores de uma variável tomada em dois pontos separados por distâncias cada vez maiores. Ocorre, com freqüência, em Ciências do Solo onde variáveis espacialmente distribuídas parecem não ter variância nem covariância "a priori" definida. Nestas circunstâncias, o semivariograma pode existir sem a covariância correspondente e isto justifica seu uso predominante nas análises (Ribeiro Jr., 1995).

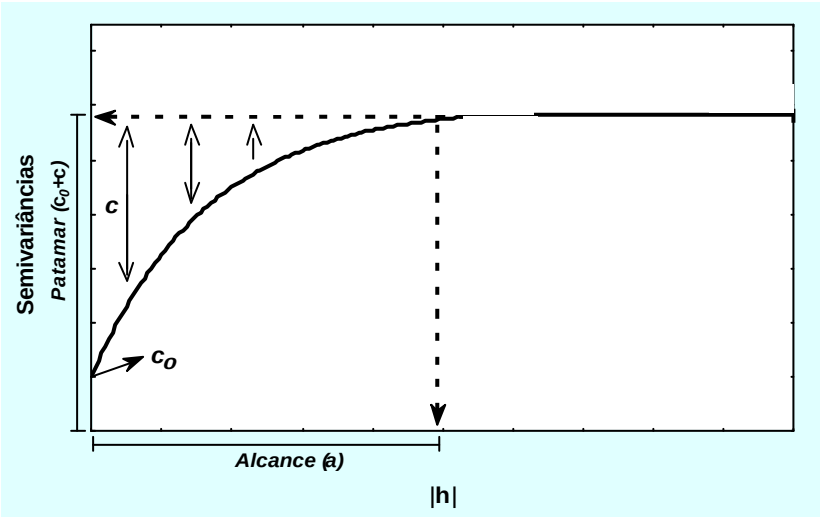

Figura 2 - Ilustração de um semivariograma ideal.

A Figura 2 ilustra um semivariograma ideal onde são localizados o efeito pepita (c_0), a variância espacial (c), o alcance (a) e o patamar (c_0+c).

Hamlett *et al.* (1986) apresentam técnicas de análise exploratória para ajudar a modelagem da estrutura espacial dos dados, "assegurando" a suposição de estacionariedade, e reforçam a importância da análise exploratória antes de se fazer uso de sofisticadas ferramentas de inferência.

2.3.9 Modelos Básicos

Os modelos teóricos de semivariograma que existem na literatura não surgiram espontaneamente e nem foram estabelecidos de forma arbitrária, eles correspondem a diferentes regionalizações que se apresentam na natureza (Guerra, 1988).

Estes modelos são classificados de acordo com seu comportamento no infinito, $?(\infty)$: limitados ou não limitados. O modelo potência, por exemplo, é não limitado.

Modelos limitados são chamados também de "*modelos de transição*" e a diferença entre eles é expressa por seu comportamento na origem, posto que a modelagem é feita tendo em vista pequenas distâncias, como: linear, parabólico ou descontínuo.

Para que um modelo de semivariograma ou covariograma seja permissível, é preciso garantir que a variância de qualquer combinação linear finita de variáveis aleatórias $Z(\mathbf{u}_\alpha), \mathbf{u}_\alpha \in W$ seja não negativa:

$$\mathrm{Var}\left\{\sum_{\alpha=1}^{n}\lambda_\alpha Z(\mathbf{u}_\alpha)\right\} \geq 0$$

$$\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_\alpha\lambda_\beta C(\mathbf{u}_\alpha - \mathbf{u}_\beta) \geq 0\,, \qquad (13)$$

para qualquer número finito de localizações espaciais $\{\mathbf{u}_\alpha : \alpha = 1,...,n\}$ e quaisquer pesos $\{\lambda_\alpha : \alpha = 1,...,n\}$.

Logo, o modelo de covariância $C(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C(\mathbf{h})$ precisa ser positivo definido.

De (10) $C(\mathbf{h}) = C(\mathbf{0}) - \gamma(\mathbf{h})$ e de (13) tem-se,

$$\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_\alpha\lambda_\beta C(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C(\mathbf{0})\sum_{\alpha=1}^{n}\lambda_\alpha\sum_{\beta=1}^{n}\lambda_\beta - \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_\alpha\lambda_\beta\gamma(\mathbf{u}_\alpha - \mathbf{u}_\beta) \geq 0\,. \qquad (14)$$

Logo, o modelo de semivariância $\gamma(\mathbf{u}_\alpha - \mathbf{u}_\beta) = \gamma(\mathbf{h})$ precisa ser condicionalmente negativo definido, isto é,

$$\sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_\alpha\lambda_\beta\gamma(\mathbf{u}_\alpha - \mathbf{u}_\beta) \leq 0\,, \qquad (15)$$

para qualquer número finito de localizações espaciais $\{\mathbf{u}_\alpha : \alpha = 1,...,n\}$ e quaisquer pesos $\{\lambda_\alpha : \alpha = 1,...,n\}$ satisfazendo $\sum_{\alpha=1}^{n}\lambda_\alpha = 0$. Esta condição sobre os pesos filtra $C(\mathbf{0})$ de (14) (Cressie, 1993).

Dentre os distintos modelos teóricos, que são considerados permissíveis na literatura Geoestatística[5], citam-se os mais importantes, outros podem ser obtidos por combinações lineares destes.

Os modelos apresentados, aqui, são isotrópicos (rotação invariante), ou seja, são definidos como função apenas do comprimento do vetor **h**, $h = |\mathbf{h}|$.

[5] Valente (1982); Guerra (1988); Isaaks & Srivastava (1989); Vieira Braga (1990); Cressie (1993); Uzumaki (1994); Queiroz (1995); Ribeiro Jr. (1995); Wackernagel (1995); Pannatier (1996); Vieira (1996); Goovaerts (1997) e outros.

2.3.9.1 Efeito Pepita ("Nugget Effect").

O modelo do efeito pepita (EP) é dado por:

$$?_{\mathrm{EP}}(\mathbf{h}) = \begin{cases} c_0 & \text{se} \quad |\mathbf{h}| \neq 0 \\ 0 & \text{se} \quad |\mathbf{h}| = 0 \end{cases} \tag{16}$$

onde c_0 é a descontinuidade na origem.

Se estacionariedade de segunda ordem (Definição 3, Seção 2.3.2) pode ser assumida, o modelo efeito pepita pode ser expresso pela função covariância através da relação (10):

$$\mathrm{C}_{\mathrm{EP}}(\mathbf{h}) = \begin{cases} c_0 & \text{se} \quad |\mathbf{h}| = 0 \\ 0 & \text{se} \quad |\mathbf{h}| \neq 0 \end{cases} \tag{17}$$

O efeito pepita puro corresponde a uma total ausência de correlação espacial entre duas variáveis Z(**u**) e Z(**u**+**h**) para qualquer distância h ($|\mathbf{h}| > ?$, onde $?>0$ é tão pequeno quanto se queira), é o caso limite quando $?(\mathbf{h})$ aparece exclusivamente como uma descontinuidade na origem (veja Figura 3) e é equivalentemente conhecido como o fenômeno do "ruído branco" na área da Física.

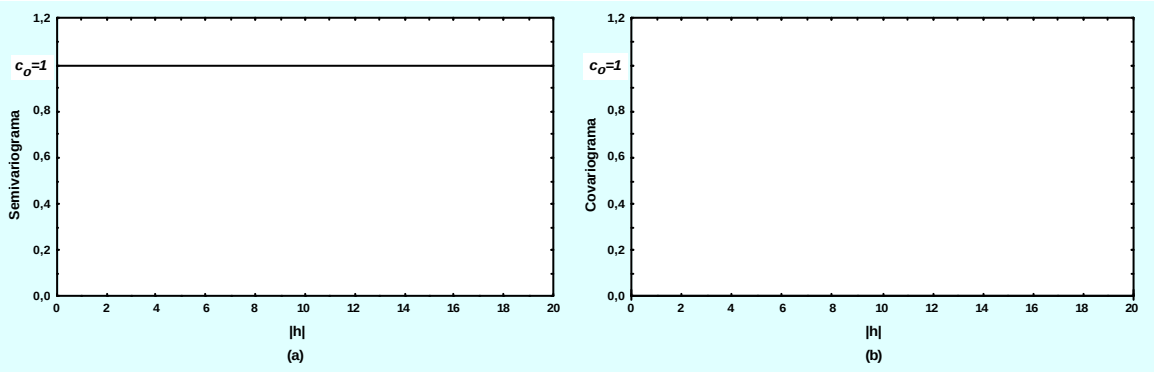

Figura 3 - (a) Semivariograma - Modelo efeito pepita puro com c_0=1; (b) Covariograma - Modelo efeito pepita puro com c_0=1.

Das Figuras 4 a 10, o efeito pepita foi considerado igual a zero (c_0=0), variância espacial igual a um (c=1) e alcance igual a dez (a=10). Estes valores foram escolhidos para facilitar a ilustração dos modelos.

2.3.9.2 Esférico

O modelo esférico (Esf) é dado por:

$$?_{\text{Esf}}(\mathbf{h}) = \begin{cases} c\left[\frac{3}{2}\left(\frac{|\mathbf{h}|}{a}\right) - \frac{1}{2}\left(\frac{|\mathbf{h}|}{a}\right)^3\right] & \text{se} \quad |\mathbf{h}| \le a \\ c & \text{se} \quad |\mathbf{h}| > a \end{cases} \qquad (18)$$

onde a é o alcance e c a variância espacial (veja Figura 4).

Este modelo tem um comportamento linear na origem e é um dos "modelos de transição" mais utilizado como pode ser visto em: Souza (1992); Mandallaz (2000) e Watson (2000). Qian (1997) usa, além do esférico, o exponencial e o gaussiano.

Se a estacionariedade de segunda ordem (Definição 3, Seção 2.3.2) pode ser assumida, o modelo esférico fica expresso pela função covariância através da relação (10) (veja Figura 5):

$$\mathrm{C}_{\text{Esf}}(\mathbf{h}) = \begin{cases} c\left[1 - \frac{3}{2}\left(\frac{|\mathbf{h}|}{a}\right) + \frac{1}{2}\left(\frac{|\mathbf{h}|}{a}\right)^3\right] & \text{se} \quad |\mathbf{h}| \le a \\ 0 & \text{se} \quad |\mathbf{h}| > a \end{cases} \qquad (19)$$

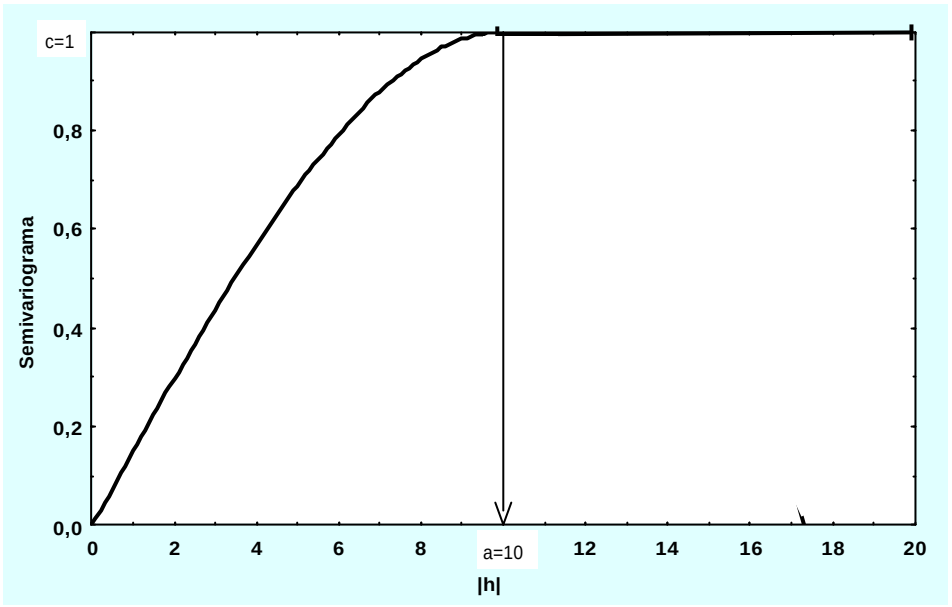

Figura 4 - Semivariograma, modelo esférico com a=10 e c=1.

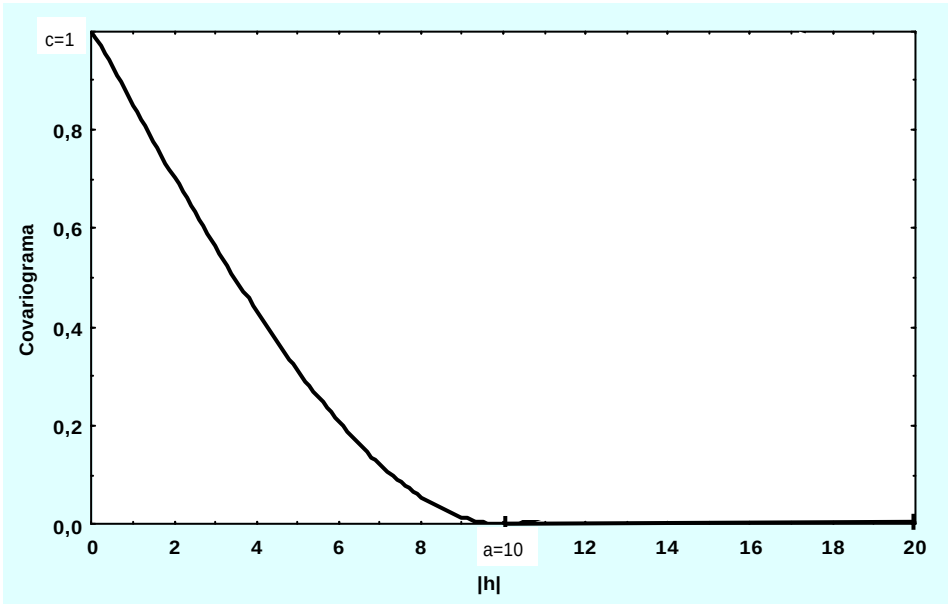

Figura 5 - Covariograma, modelo esférico com a=10 e c=1.

2.3.9.3 Exponencial

O modelo exponencial (Exp) é dado por:

$$?_{\text{Exp}}(\mathbf{h}) = c\left[1 - \exp\left(-\frac{3|\mathbf{h}|}{a}\right)\right], \quad a > 0 \tag{20}$$

onde c é a variância espacial e a é o "alcance prático", isto é, à distância na qual o valor do variograma é 95% do patamar (c_0+c) (veja Figura 6).

Este modelo alcança o patamar assintoticamente e tem um comportamento linear na origem.

Observa-se que a diferença entre o modelo exponencial e o esférico é que o exponencial atinge o patamar apenas assintoticamente, enquanto que o modelo esférico o atinge no valor do alcance.

Se a estacionariedade de segunda ordem (Definição 3, Seção 2.3.2) pode ser assumida, o modelo exponencial fica expresso pela função covariância através da relação (10) (veja Figura 7):

$$\mathrm{C}_{\text{Exp}}(\mathbf{h}) = c\exp\left(-\frac{3|\mathbf{h}|}{a}\right), \quad a > 0\,. \tag{21}$$

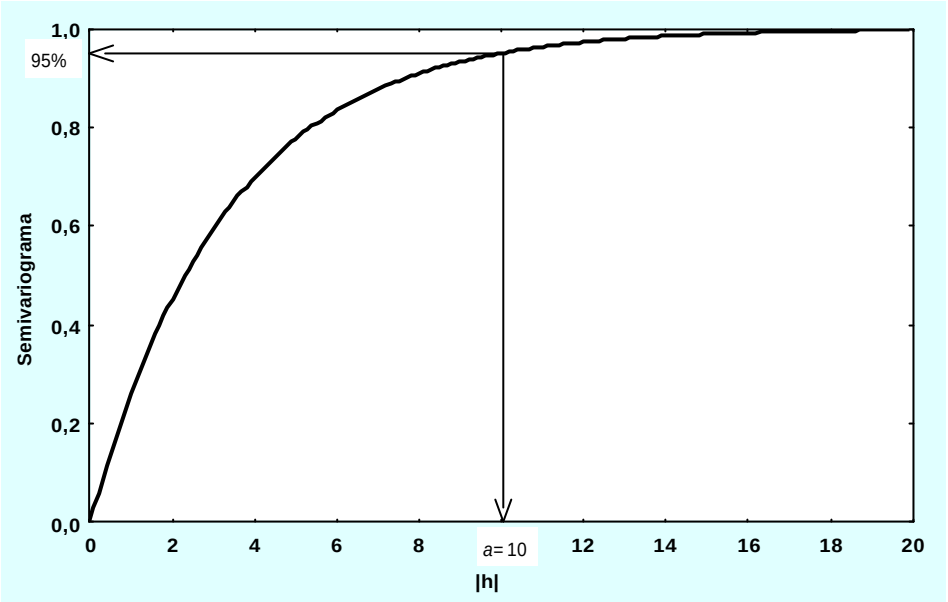

Figura 6 - Semivariograma, modelo exponencial com *a*=10 (“alcance prático”) e *c*=1.

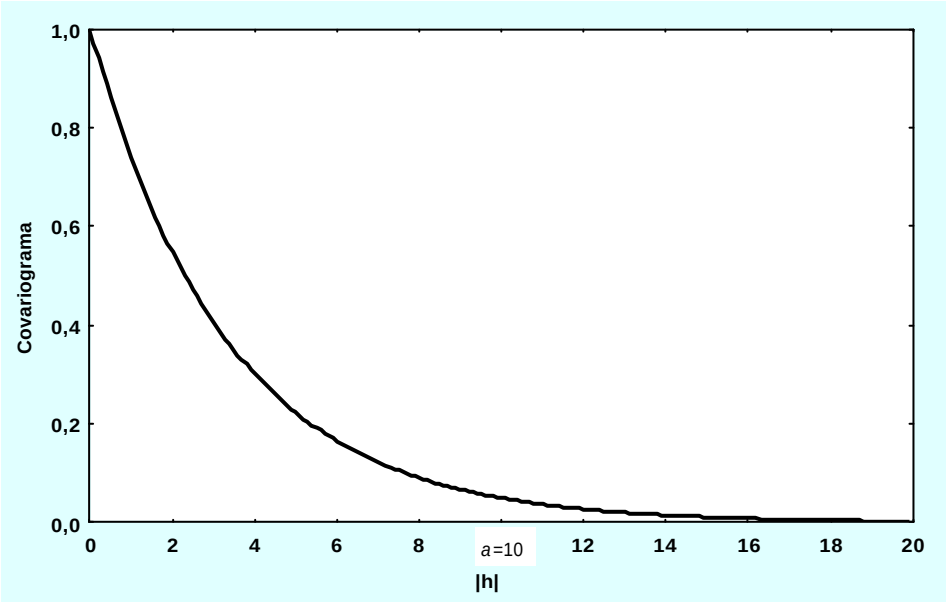

Figura 7 - Covariograma, modelo exponencial com *a*=10 (“alcance prático”) e *c*=1.

2.3.9.4 Gaussiano

O modelo gaussiano (Gau) é dado por:

$$?_{\text{Gau}}(\mathbf{h}) = c\left[1 - \exp\left(-\frac{3|\mathbf{h}|^2}{a^2}\right)\right], \quad a > 0 \tag{22}$$

onde c é a variância espacial e a é o "alcance prático", isto é, à distância na qual o valor do variograma é 95% do patamar (c_0+c) (veja Figura 8).

Este modelo alcança o patamar assintoticamente e tem um comportamento parabólico na origem.

Se a estacionariedade de segunda ordem (Definição 3, Seção 2.3.2) puder ser assumida, o modelo gaussiano fica expresso pela função covariância através da relação (10) (veja Figura 9):

$$\mathrm{C}_{\text{Gau}}(\mathbf{h}) = c\exp\left(-\frac{3|\mathbf{h}|^2}{a^2}\right), \quad a > 0\,. \tag{23}$$

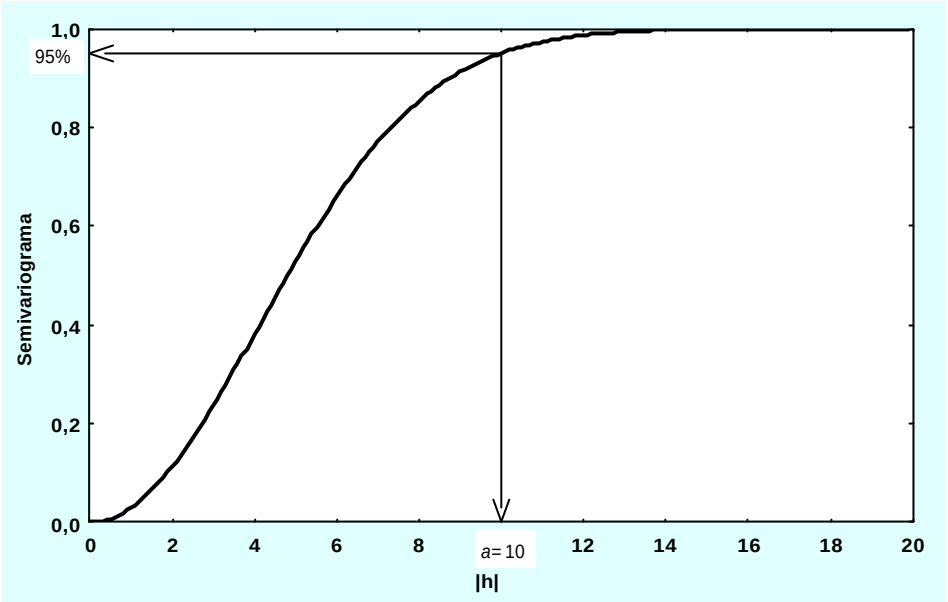

Figura 8 - Semivariograma, modelo gaussiano com a=10 (“alcance prático”) e c=1.

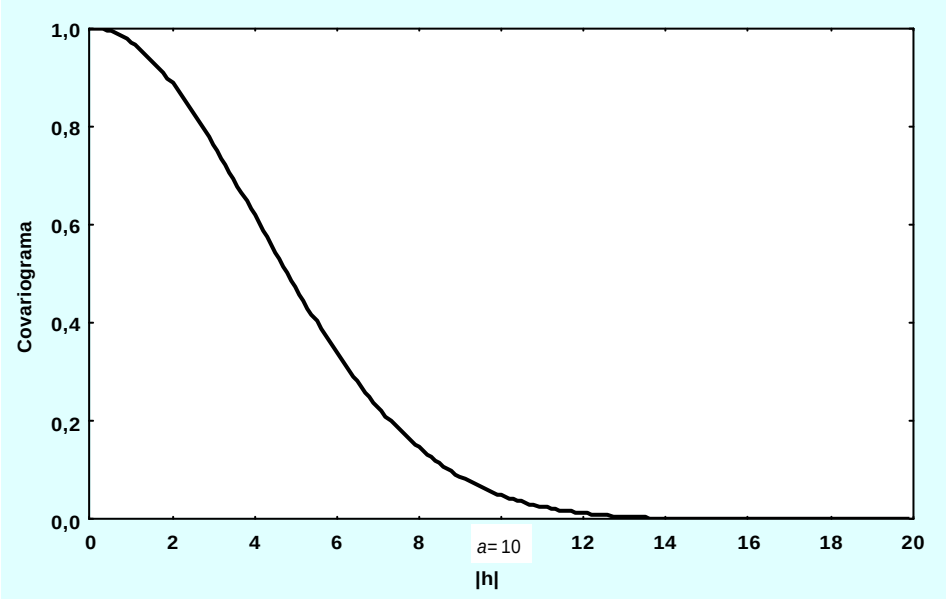

Figura 9 - Covariograma, modelo gaussiano com a=10 (“alcance prático”) e c=1.

2.3.9.5 Potência

Este modelo não é um "modelo de transição" (modelos caracterizados pelo alcance), pois ele não é limitado, seu valor aumenta com o comprimento de h.

O modelo potência (Pot) é dado por:

$$?_{\mathrm{Pot}}(\mathbf{h}) = b|\mathbf{h}|^{w}, \quad 0 < w < 2 \text{ e } b > 0, \tag{24}$$

onde b é a inclinação na origem (veja Figura 10).

Não existe covariância nesse caso, assim, não há possibilidade de assumir estacionariedade de segunda ordem.

Seu comportamento próximo da origem muda com o valor do parâmetro w: linear se w=1 e aproximadamente parabólico se w aumenta para 2.

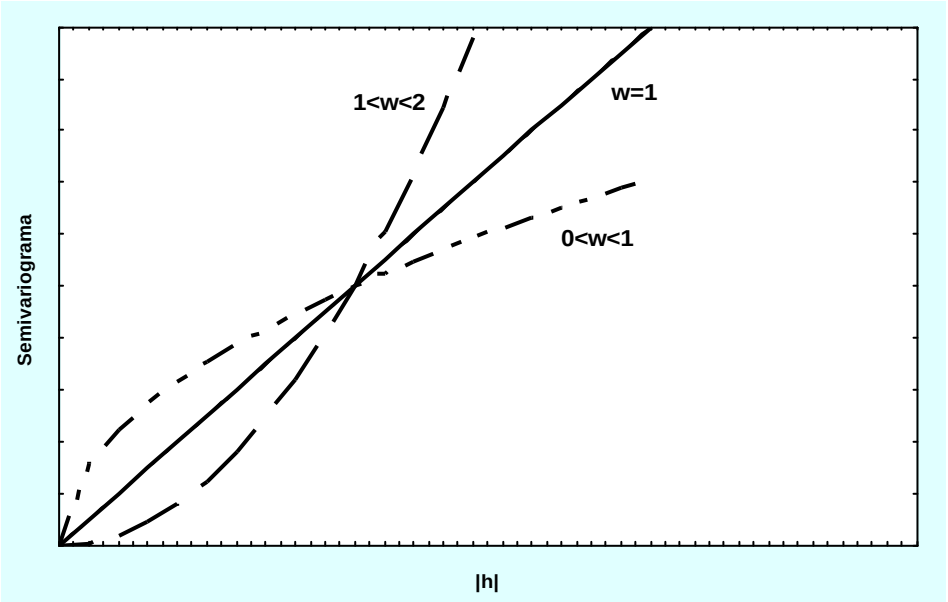

Figura 10 - Semivariograma, modelo potência, 0<w<2.

2.4 Krigagem

O nome Krigagem[6] ("Kriging") é uma referência a Danie G. Krige, pioneiro no uso de métodos estatísticos na indústria de mineração no Sul da África (Krige, 1951; citado por Diggle e Ribeiro Jr., 2000, p17).

Na modelagem do semivariograma (ou covariograma) está se ajustando a dependência espacial demonstrada nos dados e o preditor espacial (Krigagem) usa esta dependência através da relação entre os pontos vizinhos (entre os pontos observados e entre os pontos observados e a localização **u** de onde será predito o valor de z(**u**)).

Pode-se constatar, na literatura, que existem vários métodos de predição espacial que se dividem em estocásticos e não estocásticos. Estes últimos, que não serão abordados aqui, não necessitam do conhecimento dos parâmetros da modelagem espacial. Cressie (1993) trás alguns desses métodos, como, por exemplo: medida de tendência central global, média ponderada pelo inverso do quadrado das distâncias, médias móveis simples e outros.

Depois de verificar a existência de dependência espacial nos dados (para certas distâncias), pode-se usar o modelo do semivariograma (ou o covariograma) ajustado ao semivariograma experimental para predizer valores em pontos não amostrados. Este é o interesse na maioria das vezes.

---

[6] Embora em alguns trabalhos nacionais a palavra "Kriging" é traduzida como Krigeagem, no presente trabalho usa-se a tradução Krigagem.

Uma vez que as observações $(Z(\mathbf{u}_\alpha): \alpha = 1,...,n)$ são dependentes espacialmente, pode-se utilizar os vizinhos do ponto a ser predito $(Z(\mathbf{u}))$ para obter o valor predito para este ponto.

A idéia do método de Krigagem é atribuir pesos a estes vizinhos, de forma que o preditor seja, simplesmente, uma média ponderada dos valores observados na vizinhança. A maneira como estes pesos são atribuídos, relaciona-se com as distâncias euclidianas entre os pontos (distância dos vizinhos ao ponto a ser predito $\|\mathbf{u}_\alpha - \mathbf{u}\|$ com $\mathbf{u}_\alpha$ e $\mathbf{u} \in \Re^2$; e, distâncias entre os vizinhos $\|\mathbf{u}_\alpha - \mathbf{u}_\beta\|$, com $\mathbf{u}_\alpha$ e $\mathbf{u}_\beta \in \Re^2$), ou seja, segundo a posição geométrica dos pontos na área considerada.

O preditor de Krigagem considera que o modelo do semivariograma (ou covariograma) é correto[7] e que não há erros de medida. Neste caso, o preditor de Krigagem é obtido de forma que as condições de não tendência e de variância mínima do preditor sejam satisfeitas (ver equações 26 e 27).

Considere, então, o problema de predizer o valor de um atributo contínuo z^* para uma localização não amostrada, $\mathbf{u} = (x, y) \in \Re^2$, usando informações dos z_α dados disponíveis sobre a área em estudo *W*, ou seja, os n dados $\{z(\mathbf{u}_\alpha),\ \alpha = 1,2,...,n\}$. O modelo básico para predizer $Z^*(\mathbf{u})$ é definido como:

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})[Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)], \qquad (25)$$

[7] Isto significa que a modelagem do semivariograma (ou covariograma) foi a melhor no sentido de expressar a dependência espacial existente nos dados, até uma certa distância.

onde n(**u**) é a quantidade de localizações envolvidas na predição de z(**u**), e $\lambda_\alpha(\mathbf{u})$ são os pesos associados aos dados $z(\mathbf{u}_\alpha)$ interpretados como uma realização da variável aleatória $Z(\mathbf{u}_\alpha)$. As quantidades m(**u**) e $m(\mathbf{u}_\alpha)$ são os valores esperados das variáveis aleatórias Z(**u**) e $Z(\mathbf{u}_\alpha)$. O número de dados envolvidos na predição, bem como seus pesos, podem mudar de uma localização para outra.

A interpretação do valor desconhecido z(**u**) e dos valores $z(\mathbf{u}_\alpha)$ como realizações das variáveis aleatórias Z(**u**) e $Z(\mathbf{u}_\alpha)$, permite definir o erro de predição $Z^*(\mathbf{u})$- Z(**u**) como uma variável aleatória. As condições de não tendência e de variância mínima do preditor devem ser satisfeitas. Ou seja,

$$E\left[Z^*(\mathbf{u}) - Z(\mathbf{u})\right] = 0 \text{ e} \tag{26}$$

$$\sigma_E^2(\mathbf{u}) = \mathrm{Var}\left[Z^*(\mathbf{u}) - Z(\mathbf{u})\right] = E\left\{\left[Z^*(\mathbf{u}) - Z(\mathbf{u})\right]^2\right\} = \text{mínima.} \tag{27}$$

O preditor de Krigagem varia dependendo do modelo adotado para o covariograma (ou simivariograma) experimental, ou seja, para a função aleatória Z(**u**). A função aleatória pode ser decomposta em um componente de tendência m(**u**) e um componente residual R(**u**):

$$Z(\mathbf{u}) = m(\mathbf{u}) + R(\mathbf{u}). \tag{28}$$

O componente residual é modelado como uma função aleatória estacionária com média zero e covariância $C_R(\mathbf{h})$ (Goovaerts, 1997, p126):

$$E[R(\mathbf{u})] = 0 \text{ e} \tag{29}$$

$$Cov[R(\mathbf{u}), R(\mathbf{u}+\mathbf{h})] = E[R(\mathbf{u})R(\mathbf{u}+\mathbf{h})] = C_R(\mathbf{h})^{8}. \tag{30}$$

Desta forma, o valor esperado da variável aleatória Z para a localização **u** é o valor da componente de tendência para tal localização:

$$E[Z(\mathbf{u})] = m(\mathbf{u}). \tag{31}$$

O tipo de Krigagem a ser considerado dependerá da modelagem de m(**u**). Distinguem-se três tipos:

Krigagem Simples (KS);

Krigagem Ordinária (KO) e

Krigagem Universal (KU).

2.4.1 Krigagem Simples (KS)

A KS considera a média m(**u**) conhecida e constante em toda área estudada *W*:

m(**u**) = m, conhecida $\forall\ \mathbf{u} \in W$.

De (25) tem-se uma combinação linear de $(n(\mathbf{u})+1)$ informações, n(**u**) variáveis aleatórias $Z(\mathbf{u}_\alpha)$ e o valor médio m:

---

[8] De (29) e (30) a covariância espacial depende somente do vetor de separação (diferença) entre as localizações pertencentes à área estudada *W*. Sendo **h** o vetor de separação entre as localizações $\mathbf{u}_a$ e $\mathbf{u}_ß$, as covariâncias podem ser escritas de diferentes formas simplificadas:

$$Cov\ [R(\mathbf{u}_\alpha), R(\mathbf{u}_\beta)] = E[R(\mathbf{u}_\alpha)R(\mathbf{u}_\beta)] = C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C_R(\mathbf{h})$$

$$Cov[R(\mathbf{u}), R(\mathbf{u})] = E[R(\mathbf{u})R(\mathbf{u})] = C_R(\mathbf{u}-\mathbf{u}) = C_R(\mathbf{0})$$

$$Cov[R(\mathbf{u}_a), R(\mathbf{u})] = E[R(\mathbf{u}_a)R(\mathbf{u})] = C_R(\mathbf{u}_a - \mathbf{u})$$

$$Z^*_{KS}(\mathbf{u}) - m = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{KS}_{\alpha}(\mathbf{u})[Z(\mathbf{u}_\alpha) - m];$$

$$Z^*_{KS}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{KS}_{\alpha}(\mathbf{u}) Z(\mathbf{u}_\alpha) - m \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{KS}_{\alpha}(\mathbf{u}) + m;$$

$$Z^*_{KS}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{KS}_{\alpha}(\mathbf{u}) Z(\mathbf{u}_\alpha) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{KS}_{\alpha}(\mathbf{u})\right] m. \qquad (32)$$

Os $n(\mathbf{u})$ pesos $\lambda^{KS}_{\alpha}(\mathbf{u})$ são determinados para minimizar a variância do erro (27), sob a condição de não tendência (26),

$$E\left[Z^*_{KS}(\mathbf{u}) - Z(\mathbf{u})\right] = m - m = 0 .$$

Sob a hipótese de estacionariedade de segunda ordem (ver Definição 3, Seção 2.3.2):

$$E[Z(\mathbf{u})] = m, \ \forall \ \mathbf{u} \in W \text{ e}$$

$$Cov[Z(\mathbf{u}), Z(\mathbf{u}+\mathbf{h})] = C(\mathbf{h})^{9}, \ \forall \ \mathbf{u} \text{ e } \mathbf{h} \in W.$$

A média não depende da localização **u**, mas representa uma informação global comum para todas as localizações não amostradas dentro da área considerada.

O erro de predição $Z^*_{KS}(\mathbf{u}) - Z(\mathbf{u})$ pode ser visto como uma combinação linear de $(n(\mathbf{u})+1)$ variáveis aleatórias residuais $R(\mathbf{u}_\alpha)$ e $R(\mathbf{u})$:

$$Z^*_{KS}(\mathbf{u}) - Z(\mathbf{u}) = \left[Z^*_{KS}(\mathbf{u}) - m\right] - \left[Z(\mathbf{u}) - m\right], \text{ de (32),}$$

[9] Também se escreve indistintamente,

$$Cov\left[Z(\mathbf{u}), Z(\mathbf{u}+\mathbf{h})\right] = Cov\left[Z(\mathbf{u}_\alpha), Z(\mathbf{u}_\beta)\right] = C(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C(\mathbf{h}).$$

$$= \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{KS}(\mathbf{u})[Z(\mathbf{u}_\alpha) - m] - [Z(\mathbf{u}) - m]$$

$$= \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{KS}(\mathbf{u}) R(\mathbf{u}_\alpha) - R(\mathbf{u})$$

$$= R_{KS}^{*}(\mathbf{u}) - R(\mathbf{u}). \tag{33}$$

De (27),

$$\sigma_E^2(\mathbf{u}) = E\left\{\left[Z_{KS}^{*}(\mathbf{u}) - Z(\mathbf{u})\right]^2\right\}, \text{ por (33),}$$

$$= E\left\{\left[R_{KS}^{*}(\mathbf{u}) - R(\mathbf{u})\right]^2\right\}$$

$$= E\left\{\left[R_{KS}^{*}(\mathbf{u})\right]^2 + \left[R(\mathbf{u})\right]^2 - 2\left[R_{KS}^{*}(\mathbf{u}) R(\mathbf{u})\right]\right\}$$

$$= \underset{\text{(I)}}{E\left[R_{KS}^{*}(\mathbf{u})\right]^2} + \underset{\text{(II)}}{E\left[R(\mathbf{u})\right]^2} - \underset{\text{(III)}}{2E\left[R_{KS}^{*}(\mathbf{u}) R(\mathbf{u})\right]}. \tag{34}$$

(I) De (32)

$$E\left[R_{KS}^{*}(\mathbf{u})\right]^2 = E\left[\sum_{a=1}^{n(\mathbf{u})} ?_a^{KS}(\mathbf{u}) R(\mathbf{u}_a)\right]^2$$

$$= \sum_{a=1}^{n(\mathbf{u})} ?_a(\mathbf{u}) \sum_{ß=1}^{n(\mathbf{u})} ?_ß(\mathbf{u}) E\left[R(\mathbf{u}_a) R(\mathbf{u}_ß)\right], \text{ de (30),}$$

$$= \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta(\mathbf{u}) C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta). \tag{35}$$

(II) $E\left[R(\mathbf{u})\right]^2 = E\left[R(\mathbf{u}) R(\mathbf{u} + \mathbf{0})\right]$**,** de (30),

$$= C_R(\mathbf{0}). \tag{36}$$

(III) De (32)

$$-2E\left[R_{KS}^{*}(\mathbf{u}) R(\mathbf{u})\right] = -2E\left\{\left[\sum_{a=1}^{n(\mathbf{u})} ?_a^{KS}(\mathbf{u}) R(\mathbf{u}_a)\right] R(\mathbf{u})\right\}$$

$$= -2\sum_{a=1}^{n(\mathbf{u})} ?_a^{KS}(\mathbf{u})E[R(\mathbf{u}_a)R(\mathbf{u})], \text{ de (30),}$$

$$= -2\sum_{a=1}^{n(\mathbf{u})} ?_a^{KS}(\mathbf{u})C_R(\mathbf{u}_a - \mathbf{u}). \quad (37)$$

Substituindo (35), (36) e (37) em (34) tem-se,

$$\sigma_E^2(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})}\sum_{\beta=1}^{n(\mathbf{u})}\lambda_\alpha^{KS}(\mathbf{u})\lambda_\beta^{KS}(\mathbf{u})C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) + C_R(\mathbf{0}) - 2\sum_{\alpha=1}^{n(\mathbf{u})}\lambda_\alpha^{KS}(\mathbf{u})C_R(\mathbf{u}_\alpha - \mathbf{u}). \quad (38)$$

Os pesos $?_a^{KS}(\mathbf{u})$, $a = 1,2,...,n(\mathbf{u})$ são encontrados quando minimiza-se a variância do erro (38), ou seja, fazendo igual a zero todas as $n(\mathbf{u})$ primeiras derivadas parciais

$$\frac{\partial s_E^2(\mathbf{u})}{\partial ?_a^{KS}(\mathbf{u})} = 0 \text{ para } a = 1,2,...,n(\mathbf{u}).$$

Portanto,

$$\frac{1}{2}\frac{\partial s_E^2(\mathbf{u})}{\partial ?_a^{KS}(\mathbf{u})} = \sum_{ß=1}^{n(\mathbf{u})} ?_ß^{KS}(\mathbf{u})C_R(\mathbf{u}_a - \mathbf{u}_ß) - C_R(\mathbf{u}_a - \mathbf{u}) = 0 \quad (39)$$

para $a = 1,2,...,n(\mathbf{u})$.

O sistema (39) é conhecido como sistema de equações normais (Luenberger, 1969), aqui chamado de Sistema de Krigagem Simples (SKS).

Sob hipótese de estacionariedade de segunda ordem (Definição 3, Seção 2.3.2), a função covariância residual $C_R(\mathbf{h})$ é igual a função covariância $C(\mathbf{h})$ da função aleatória $Z(\mathbf{u})$. De (30)

$$C_R(\mathbf{h}) = E[R(\mathbf{u})R(\mathbf{u}+\mathbf{h})], \text{ de (28),}$$

$$= E\{[Z(\mathbf{u}) - m][Z(\mathbf{u}+\mathbf{h}) - m]\}$$

$$= E\{Z(\mathbf{u})Z(\mathbf{u}+\mathbf{h}) - mZ(\mathbf{u}) - mZ(\mathbf{u}+\mathbf{h}) + m^2\}$$

$$= E[Z(\mathbf{u})Z(\mathbf{u}+\mathbf{h})] - mE[Z(\mathbf{u})] - mE[Z(\mathbf{u}+\mathbf{h})] + m^2$$

$$= E[Z(\mathbf{u})Z(\mathbf{u}+\mathbf{h})] - m^2 - m^2 + m^2$$

$$= E[Z(\mathbf{u})Z(\mathbf{u}+\mathbf{h})] - m^2$$

$$= Cov[Z(\mathbf{u}), Z(\mathbf{u}+\mathbf{h})]$$

$$= C(\mathbf{h}). \qquad (40)$$

Com o resultado (40), o sistema (39) fica,

$$\sum_{ß=1}^{n(\mathbf{u})} ?_ß^{KS}(\mathbf{u})C(\mathbf{u}_a - \mathbf{u}_ß) = C(\mathbf{u}_a - \mathbf{u}), \text{ para } a = 1,2,...,n(\mathbf{u}). \qquad (41)$$

O lado esquerdo da equação (41) descreve as covariâncias entre localizações. O lado direito descreve as covariâncias entre cada localização e a localização de onde uma estimativa é procurada.

Quando a localização $\mathbf{u}$ do valor a ser predito coincide com uma localização nos dados, diga-se $\mathbf{u}_{a'}$, o SKS fica,

$$\sum_{\substack{ß=1 \\ ß\neq a'}}^{n(\mathbf{u})} ?_ß^{KS}(\mathbf{u})C(\mathbf{u}_a - \mathbf{u}_ß) + ?_{a'}^{KS}(\mathbf{u})C(\mathbf{u}_a - \mathbf{u}_{a'}) = C(\mathbf{u}_a - \mathbf{u}_{a'}),$$

$$\text{para } a = 1,2,...,n(\mathbf{u}). \qquad (42)$$

A única solução do sistema é obtida quando $?_{a'}^{KS}(\mathbf{u}) = 1$ e $?_ß^{KS}(\mathbf{u}) = 0$, $\forall$ $\mathbf{u}_ß \neq \mathbf{u}_{a'}$, isto é, o preditor dado pelo SKS é um interpolador exato (Goovaerts, 1997, p130; Cressie, 1993, p359).

A variância do erro $\left(s_E^2(\mathbf{u})\right)$, também chamada de variância de predição (ou "estimação"), pode ser obtida para cada localização $\mathbf{u}$ substituindo (41) em (38):

$$\sigma_{KS}^2(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{KS}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}) + C(\mathbf{0}) - 2\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{KS}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u});$$

$$\sigma_{KS}^2(\mathbf{u}) = C(\mathbf{0}) - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{KS}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}). \qquad (43)$$

O SKS (41) tem uma única solução e o resultado da variância de predição (43) é positivo se o modelo de covariância C(**h**) é permissível (ver Seção 2.3.9).

O SKS (41) pode ser escrito matricialmente como:

$\mathbf{K}_{KS}\ ?_{KS}(\mathbf{u}) = \mathbf{k}_{KS}$, onde,

$\mathbf{K}_{KS}$ é a matriz [n($\mathbf{u}$)xn($\mathbf{u}$)] de variâncias e covariâncias entre as localizações:

$$\mathbf{K}_{KS} = \begin{bmatrix} C(\mathbf{u}_1 - \mathbf{u}_1) & \cdots & C(\mathbf{u}_1 - \mathbf{u}_{n(\mathbf{u})}) \\ \vdots & \ddots & \vdots \\ C(\mathbf{u}_{n(\mathbf{u})} - \mathbf{u}_1) & \cdots & C(\mathbf{u}_{n(\mathbf{u})} - \mathbf{u}_{n(\mathbf{u})}) \end{bmatrix};$$

$?_{KS}(\mathbf{u})$ é o vetor [n($\mathbf{u}$)x1] de pesos da KS:

$$?_{KS}(\mathbf{u}) = \begin{bmatrix} ?_1^{KS}(\mathbf{u}) \\ \vdots \\ ?_{n(\mathbf{u})}^{KS}(\mathbf{u}) \end{bmatrix};$$

$\mathbf{k}_{KS}$ é o vetor [n($\mathbf{u}$)x1] de covariâncias entre as localizações e a localização a ser predita:

$$\mathbf{k}_{KS} = \begin{bmatrix} C(\mathbf{u}_1 - \mathbf{u}) \\ \vdots \\ C(\mathbf{u}_{n(\mathbf{u})} - \mathbf{u}) \end{bmatrix}.$$

Portanto, os pesos são obtidos por $?_{KS}(\mathbf{u}) = \mathbf{K}_{KS}^{-1}\mathbf{k}_{KS}$.

A formulação matricial da variância da Krigagem (43) é dada por,

$$s^2_{KS}(\mathbf{u}) = C(\mathbf{0}) - \boldsymbol{\lambda}^T_{KS}(\mathbf{u})\mathbf{k}_{KS} = C(\mathbf{0}) - \mathbf{k}^T_{KS}\mathbf{K}^{-1}_{KS}\mathbf{k}_{KS}.$$

Em geral, assumir m(**u**)=m, $\forall \mathbf{u} \in W$, e que a estrutura da covariância é conhecida nem sempre é possível na prática. Assim, o preditor de Krigagem simples, dificilmente é usado na prática.

2.4.2 Krigagem Ordinária (KO)

A Krigagem Ordinária (KO) considera flutuações locais da média, limitando o domínio de estacionariedade da média para uma vizinhança local *W*(**u**) centrada na localização **u** a ser predita:

$m(\mathbf{u}')$ = constante mas desconhecida $\forall\ \mathbf{u}' \in W(\mathbf{u})$.

O preditor linear $Z^*(\mathbf{u})$ é obtido como uma combinação linear de n(**u**) variáveis aleatórias $Z(\mathbf{u}_a)$ mais a média local constante m(**u**), de (25):

$$Z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})[Z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha)],$$ onde $m(\mathbf{u})$ e $m(\mathbf{u}_\alpha)$ são desconhecidas, mas constantes iguais a $m(\mathbf{u})$, pois estão num domínio de estacionariedade da média pertencente a *W*(**u**).

Logo,

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})Z(\mathbf{u}_\alpha) - m(\mathbf{u})\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) + m(\mathbf{u});$$

$$Z^*(\mathbf{u}) = \sum_{a=1}^{n(\mathbf{u})} \lambda_a(\mathbf{u})Z(\mathbf{u}_a) + \left[1 - \sum_{a=1}^{n(\mathbf{u})} \lambda_a(\mathbf{u})\right]m(\mathbf{u}). \qquad (44)$$

A média local desconhecida m(**u**) é filtrada do preditor linear (44), forçando a soma dos pesos de Krigagem a ser igual a 1 (um) (Goovaerts, 1997, p133):

$$Z^*_{KO}(\mathbf{u}) = \sum_{a=1}^{n(\mathbf{u})} \lambda^{KO}_a(\mathbf{u}) Z(\mathbf{u}_a), \text{ com } \sum_{a=1}^{n(\mathbf{u})} \lambda^{KO}_a(\mathbf{u}) = 1. \quad (45)$$

Isso também implica que o uso do semivariograma é autorizado na computação da variância do erro de predição.

O semivariograma é autorizado para a Krigagem Ordinária (KO), mas não para a Krigagem Simples (KS), porque esta última não inclui uma condição sobre os pesos (Wackernagel, 1995, p75).

O preditor (45) nada mais é que uma média móvel ponderada (Vieira, 1996, p16), que depende dos parâmetros do modelo do semivariograma ajustado para representar a dependência espacial nos dados.

Para que o preditor (45) seja ótimo, ele deve obedecer às condições de não tendência e variância mínima:

$$E\left[Z^*_{KO}(\mathbf{u}) - Z(\mathbf{u})\right] = 0 \text{ e} \quad (46)$$

$$s^2_E(\mathbf{u}) = Var\left[Z^*_{KO}(\mathbf{u}) - Z(\mathbf{u})\right] = E\left\{\left[Z^*_{KO}(\mathbf{u}) - Z(\mathbf{u})\right]^2\right\} = \text{mínima.} \quad (47)$$

A condição (46) é justificada com a unidade da soma dos pesos, ou seja, $\sum_{a=1}^{n(\mathbf{u})} \lambda^{KO}_a(\mathbf{u}) = 1$. De (45),

$$E\left[Z^*_{KO}(\mathbf{u}) - Z(\mathbf{u})\right] = E\left[\sum_{a=1}^{n(\mathbf{u})} \lambda^{KO}_a(\mathbf{u}) Z(\mathbf{u}_a) - Z(\mathbf{u})\right]$$

$$= \sum_{a=1}^{n(\mathbf{u})} \lambda^{KO}_a(\mathbf{u}) E\left[Z(\mathbf{u}_a)\right] - E\left[Z(\mathbf{u})\right], \text{ sob algum tipo de}$$

estacionariedade (ver Seção 2.3.2),

$$= \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{KO}(\mathbf{u})[m(\mathbf{u})] - m(\mathbf{u}) = m(\mathbf{u}) - m(\mathbf{u}) = 0 .$$

A condição (47) deve então ser minimizada sob a condição de não tendência (46), $\sum_{a=1}^{n(\mathbf{u})} ?_a^{KO}(\mathbf{u}) = 1$. De (47),

$$s_E^2(\mathbf{u}) = E\left\{\left[Z_{KO}^*(\mathbf{u}) - Z(\mathbf{u})\right]^2\right\}, \text{ por (33)},$$

$$= E\left\{\left[R_{KO}^*(\mathbf{u}) - R(\mathbf{u})\right]^2\right\}$$

$$= E\left\{\left[R_{KO}^*(\mathbf{u})\right]^2 + \left[R(\mathbf{u})\right]^2 - 2\left[R_{KO}^*(\mathbf{u})R(\mathbf{u})\right]\right\}$$

$$= \underset{(I)}{E\left[R_{KO}^*(\mathbf{u})\right]^2} + \underset{(II)}{E\left[R(\mathbf{u})\right]^2} - \underset{(III)}{2E\left[R_{KO}^*(\mathbf{u})R(\mathbf{u})\right]}. \qquad (48)$$

Trocando os índices (KS) por (KO) nas equações (35), (36) e (37) e as substituindo em (48) temos,

$$\sigma_E^2(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})}\sum_{\beta=1}^{n(\mathbf{u})} \lambda_\alpha^{KO}(\mathbf{u})\lambda_\beta^{KO}(\mathbf{u})C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) + C_R(\mathbf{0}) - 2\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{KO}(\mathbf{u})C_R(\mathbf{u}_\alpha - \mathbf{u}) . \qquad (49)$$

Sérias dificuldades serão encontradas ao minimizar diretamente (49): fazendo igual a zero, as n primeiras derivadas parciais de (49), produzem-se n equações com n parâmetros desconhecidos. A condição de não tendência (46) adicionará outra equação sem adicionar qualquer parâmetro desconhecido. Isto leva a um sistema com (n+1) equações e, somente n parâmetros desconhecidos, onde a solução não é simples. Para que este problema seja evitado, um parâmetro de Lagrange (μ) será introduzido na equação (49), da seguinte forma.

Aplicando o teorema dos multiplicadores de Lagrange[10] temos L(**u**), que é uma função dos pesos $\lambda_{\alpha}^{KO}(\mathbf{u})$ e, do parâmetro de Lagrange $\mu_{KO}(\mathbf{u})$:

$$L\left[?_{a}^{KO}(\mathbf{u}), a = 1,2,...,n(\mathbf{u}); 2\mu_{KO}(\mathbf{u})\right] = s_{E}^{2}(\mathbf{u}) + 2\mu_{KO}(\mathbf{u})\left[\sum_{a=1}^{n(\mathbf{u})} ?_{a}^{KO}(\mathbf{u}) - 1\right]. \quad (50)$$

Os pesos $?_{a}^{KO}(\mathbf{u})$, $a = 1,2,...,n(\mathbf{u})$ são encontrados quando a função L(**u**) (50) é minimizada, ou seja, fazendo-se igual a zero todas as (n(**u**)+1) primeiras derivadas parciais com respeito a cada uma dessas variáveis:

$$\frac{\partial L(\mathbf{u})}{\partial ?_{a}^{KO}(\mathbf{u})} = 0 \text{ para } a = 1,2,...,n(\mathbf{u}) \text{ e}$$

$$\frac{\partial L(\mathbf{u})}{\partial \mu_{KO}(\mathbf{u})} = 0.$$

Ao derivar parcialmente com relação a μ, tem-se a condição de não tendência. Portanto,

$$\frac{1}{2}\frac{\partial L(\mathbf{u})}{\partial ?_{a}^{KO}(\mathbf{u})} = \sum_{ß=1}^{n(\mathbf{u})} ?_{ß}^{KO}(\mathbf{u}) C_{R}(\mathbf{u}_{a} - \mathbf{u}_{ß}) - C_{R}(\mathbf{u}_{a} - \mathbf{u}) + \mu_{KO}(\mathbf{u}) = 0,$$

$$a = 1,2,...,n(\mathbf{u}) \text{ e}$$

$$\frac{1}{2}\frac{\partial L(\mathbf{u})}{\partial \mu_{KO}(\mathbf{u})} = \sum_{a=1}^{n(\mathbf{u})} ?_{a}^{KO}(\mathbf{u}) - 1 = 0.$$

O Sistema de Krigagem Ordinária (SKO) tem (n(**u**)+1) equações lineares: n(**u**) pesos $?_{a}^{KO}(\mathbf{u})$ e o parâmetro de Lagrange $\mu_{KO}(\mathbf{u})$ que é obtido pela restrição nos pesos:

[10] Encontrado em livros de cálculo avançado como: Fritzsche (1978); Edwards & Penney (1982).

$$\sum_{ß=1}^{n(\mathbf{u})} ?_ß^{KO}(\mathbf{u}) C_R(\mathbf{u}_a - \mathbf{u}_ß) + \mu_{KO}(\mathbf{u}) = C_R(\mathbf{u}_a - \mathbf{u}), \; a = 1,2,..., n(\mathbf{u}) \text{ e}$$

$$\sum_{ß=1}^{n(\mathbf{u})} ?_ß^{KO}(\mathbf{u}) = 1.$$

Embora a média m(**u**) seja assumida estacionária somente dentro de uma vizinhança local *W*(**u**), na prática de KO, a covariância residual é assimilada como um z-covariância global inferida de todos os dados disponíveis (Journel & Huijbregts, 1978, p33 e 34, citado por Goovaerts, 1997), levando ao seguinte SKO:

$$\sum_{ß=1}^{n(\mathbf{u})} ?_ß^{KO}(\mathbf{u}) C(\mathbf{u}_a - \mathbf{u}_ß) + \mu_{KO}(\mathbf{u}) = C(\mathbf{u}_a - \mathbf{u}), \; a = 1,2,..., n(\mathbf{u}) \text{ e} \qquad (51a)$$

$$\sum_{ß=1}^{n(\mathbf{u})} ?_ß^{KO}(\mathbf{u}) = 1. \qquad (51b)$$

A condição (51b) já está incluída nas (n+1) equações que resulta da diferenciação de (50) e a solução do sistema (51) produzirá um conjunto de pesos que minimizam (50) sob a condição de que a soma desses pesos seja igual a 1 (*um*). Esta solução também produzirá um valor para μ que será usado para o cálculo da variância do erro minimizada.

Para que o sistema (51) tenha uma e somente uma solução, é preciso assegurar que o sistema tenha a propriedade conhecida como positiva definida.

Existem muitas maneiras de conferir esta propriedade. No entanto, na prática, a existência e a unicidade da solução é garantida por ajustar (modelar) o semivariograma experimental com funções que são conhecidas para ser positiva definida (Isaaks & Srivastava, 1989, p297). Esta condição de positiva definida pode ser vista como

uma garantia de que a variância de qualquer variável aleatória formada por uma combinação linear de outras variáveis aleatórias é positiva, veja (13 na Seção 2.3.9).

A variância do erro $\left(s_E^2(\mathbf{u})\right)$ pode ser obtida para cada localização $\mathbf{u}$ sendo predita, substituindo-se (51a) em (49):

$$\sigma_{KO}^2(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{KO}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}) - \mu_{KO}(\mathbf{u}) + C(\mathbf{0}) - 2\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{KO}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u});$$

$$\sigma_{KO}^2(\mathbf{u}) = C(\mathbf{0}) - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{KO}(\mathbf{u}) C(\mathbf{u}_\alpha - \mathbf{u}) - \mu_{KO}(\mathbf{u}). \qquad (52)$$

O SKO (51) pode ser escrito em função do semivariograma usando a relação (10):

$C(\mathbf{h}) = C(\mathbf{0}) - \gamma(\mathbf{h})$, ou seja,

$C(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C(\mathbf{0}) - \gamma(\mathbf{u}_\alpha - \mathbf{u}_\beta)$, onde $\mathbf{h}$ é o vetor de separação entre as localizações $\mathbf{u}_a$ e $\mathbf{u}_ß$.

Então:

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{KO}(\mathbf{u}) \left[C(0) - \gamma(\mathbf{u}_\alpha - \mathbf{u}_\beta)\right] + \mu_{KO}(\mathbf{u}) = C(\mathbf{0}) - \gamma(\mathbf{u}_\alpha - \mathbf{u}),$$

$$a = 1,2,\ldots,n(\mathbf{u}) \text{ e}$$

$$\sum_{ß=1}^{n(\mathbf{u})} ?_ß^{KO}(\mathbf{u}) = 1.$$

$$C(\mathbf{0})\sum_{\beta=1}^{n(\mathbf{u})}\lambda_{\beta}^{KO}(\mathbf{u})-\sum_{\beta=1}^{n(\mathbf{u})}\lambda_{\beta}^{KO}(\mathbf{u})\gamma(\mathbf{u}_{\alpha}-\mathbf{u}_{\beta})+\mu_{KO}(\mathbf{u})=C(\mathbf{0})-\gamma(\mathbf{u}_{\alpha}-\mathbf{u}),$$

$$a=1,2,...,n(\mathbf{u}) \text{ e} \qquad (53a)$$

$$\sum_{ß=1}^{n(\mathbf{u})}?_{ß}^{KO}(\mathbf{u})=1. \qquad (53b)$$

Pela condição de não tendência $\sum_{ß=1}^{n(\mathbf{u})}?_{ß}^{KO}(\mathbf{u})=1$, cancelando C(**0**) e multiplicando ambos os lados da expressão (53a) por (-1) tem-se:

$$\sum_{ß=1}^{n(\mathbf{u})}?_{ß}^{KO}(\mathbf{u})?(\mathbf{u}_{a}-\mathbf{u}_{ß})-\mu_{KO}(\mathbf{u})=?(\mathbf{u}_{a}-\mathbf{u}),\ a=1,2,...,n(\mathbf{u}) \text{ e} \qquad (54a)$$

$$\sum_{ß=1}^{n(\mathbf{u})}?_{ß}^{KO}(\mathbf{u})=1. \qquad (54b)$$

A variância do erro $\left(s_{KO}^{2}(\mathbf{u})\right)$ (52) também pode ser escrita em função do semivariograma:

$$\sigma_{KO}^{2}(\mathbf{u})=C(\mathbf{0})-\sum_{\alpha=1}^{n(\mathbf{u})}\lambda_{\alpha}^{KO}(\mathbf{u})[C(\mathbf{0})-\gamma(\mathbf{u}_{\alpha}-\mathbf{u})]-\mu_{KO}(\mathbf{u});$$

$$s_{KO}^{2}(\mathbf{u})=\sum_{a=1}^{n(\mathbf{u})}?_{a}^{KO}(\mathbf{u})?(\mathbf{u}_{a}-\mathbf{u})-\mu_{KO}(\mathbf{u}). \qquad (55)$$

O SKO (54) tem solução única se o modelo de semivariograma $\gamma(\mathbf{h})$ for válido (permissível, ver Seção 2.3.9) e também é um interpolador exato, isto é, se o ponto a ser predito coincidir com um dos pontos amostrados, o valor predito será igual ao valor amostrado.

A KO pode ser usada para predizer um "bloco" em lugar de um valor pontual, com algumas modificações nas equações mostradas aqui (Guerra, 1988, CAP4; Isaaks & Srivastava, 1989, CAP13; Wackernagel, 1995, CAP11; Goovaerts, 1997, CAP5).

O SKO (51) pode ser escrito matricialmente como:

$\mathbf{K}_{KO}\ ?_{KO}(\mathbf{u}) = \mathbf{k}_{KO}$, onde,

$$\begin{bmatrix} C(\mathbf{u}_1-\mathbf{u}_1) & \cdots & C(\mathbf{u}_1-\mathbf{u}_{n(\mathbf{u})}) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C(\mathbf{u}_{n(\mathbf{u})}-\mathbf{u}_1) & \cdots & C(\mathbf{u}_{n(\mathbf{u})}-\mathbf{u}_{n(\mathbf{u})}) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} ?_1^{KO}(\mathbf{u}) \\ \vdots \\ ?_{n(\mathbf{u})}^{KO}(\mathbf{u}) \\ \mu_{KO}(\mathbf{u}) \end{bmatrix} = \begin{bmatrix} C(\mathbf{u}_1-\mathbf{u}) \\ \vdots \\ C(\mathbf{u}_{n(\mathbf{u})}-\mathbf{u}) \\ 1 \end{bmatrix}.$$

Portanto, os pesos são obtidos por $?_{KO}(\mathbf{u}) = \mathbf{K}_{KO}^{-1}\mathbf{k}_{KO}$.

Os pesos obtidos no sistema de Krigagem levam em consideração a proximidade dos dados à localização **u** sendo estimada através do vetor de covariâncias $[C(\mathbf{u}_a-\mathbf{u})]$; e a "superabundância" de dados através dos termos fora da diagonal da matriz de variâncias e covariâncias $[C(\mathbf{u}_\alpha-\mathbf{u}_\beta)]$.

Goovaerts (1997, p177) mostra que a medida que o efeito pepita aumenta para uma situação de efeito pepita puro, os pesos tendem a ser iguais a $\frac{1}{n(\mathbf{u})}$ (ver também Isaaks & Srivastava, 1989, p306).

Cressie (1993, p144) apresenta uma versão robusta para a Krigagem Ordinária, onde existe a possibilidade de editar os valores perdidos ("missing") e os discrepantes ("outliers"), com uma aplicação apresentada na página 160.

Note-se que, a média constante m(**u**) não precisa ser estimada. No entanto, algumas vezes é preferível assumir que a média é uma combinação linear desconhecida de funções conhecidas, como é o caso quando se usa a Krigagem Universal.

2.4.3 Krigagem Universal (KU)

Proposta por Matheron em 1969, a Krigagem Universal (KU) é também chamada de Krigagem com Modelo de Tendência (KMT), pois a componente de tendência m(**u**) (de 28) é modelada como uma combinação linear das funções $f_i(\mathbf{u})$, i=0,1,...,p, das coordenadas de **u**.

Se **u** é um ponto no plano, define-se x e y como suas coordenadas: **u**=(x,y). Se a componente de tendência m(**u**), chamada "drift", é um polinômio de grau k nas coordenadas de **u**, pode-se escrever (Dubrule, 1983; Hudson & Wackernagel, 1994),

$$m(\mathbf{u}) = \sum_{i=0}^{p} a_i(\mathbf{u}) f_i(\mathbf{u}). \qquad (56)$$

Considerando a decomposição de Z(**u**) com $\mathbf{u} \in \Re^2$ (aqui $\Re^2 = W$) dentro de uma tendência determinística m(**u**) e uma flutuação aleatória R(**u**), de (28):

$$Z(\mathbf{u}) = m(\mathbf{u}) + R(\mathbf{u}), \text{ de (56)},$$

$$Z(\mathbf{u}) = \sum_{i=0}^{p} a_i(\mathbf{u}) f_i(\mathbf{u}) + R(\mathbf{u}), \ \mathbf{u} \in W(\mathbf{u}), \qquad (57)$$

onde (a_i(**u**): i=0,...,p), são constantes desconhecidas, $\forall$ $\mathbf{u} \in W(\mathbf{u})$; $f_i(\mathbf{u})$ monômios de grau menor ou igual a k e, R(**u**) uma função aleatória, de média zero, intrinsecamente estacionária com variograma $2\gamma(\mathbf{h})$.

Embora (57) possa dar uma idéia de estar trabalhando com "superfície de resposta", em geoestatística as flutuações (nome mais comum entre os geoestatísticos ao se referir aos erros ou resíduos) não são assumidas independentes, como são em "superfície de resposta".

O tipo de função $f_i(\mathbf{u})$, pode ser diretamente sugerido por problemas físicos.

Hudson & Wackernagel (1994) assumem que a temperatura é uma função linear das coordenadas de $\mathbf{u}$=(x,y): $m(\mathbf{u}) = m(x,y) = a_0 + a_1 x + a_2 y$.

Goovaerts (1997, p142) relata que a tendência pode ser limitada a uma particular direção: $m(\mathbf{u}) = m(x,y) = a_0 + a_1(x,y)$, linear na direção 45°.

Pode-se, ainda, ter uma tendência quadrática:

$m(\mathbf{u}) = m(x,y) = a_0 + a_1 x + a_2 y + a_3 x^2 + a_4 xy + a_5 y^2$, Cressie (1993, p155).

Do modelo básico (25) tem-se,

$$Z^*(\mathbf{u}) = m(\mathbf{u}) + \sum_{a=1}^{n(\mathbf{u})} ?_a(\mathbf{u})[Z(\mathbf{u}_a) - m(\mathbf{u}_a)], \text{ e de (56),}$$

$$Z^*(\mathbf{u}) = \sum_{i=0}^{p} a_i(\mathbf{u}) f_i(\mathbf{u}) + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) \left[Z(\mathbf{u}_\alpha) - \sum_{i=0}^{p} a_i(\mathbf{u}) f_i(\mathbf{u}_\alpha) \right];$$

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) Z(\mathbf{u}_\alpha) + \sum_{i=0}^{p} a_1(\mathbf{u}) \left[f_i(\mathbf{u}) - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) f_i(\mathbf{u}_\alpha) \right]. \qquad (58)$$

Os (p+1) coeficientes $a_i(\mathbf{u})$ desconhecidos são filtrados do preditor (58) pelas seguintes condições:

$$\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) f_i(\mathbf{u}_\alpha) = f_i(\mathbf{u}), \; i=0, 1, 2,...,p. \qquad (59)$$

Estas condições sobre os pesos são necessárias e suficientes para um preditor uniformemente não tendencioso:

$$Z^*_{KU}(\mathbf{u}) = \sum_{a=1}^{n(\mathbf{u})} ?^{KU}_a(\mathbf{u}) Z(\mathbf{u}_a), \tag{60}$$

onde, para i=0, $f_0(\mathbf{u}) = 1$ por convenção, Wackernagel (1995, p177). Cressie (1993, p152) assume-se que $f_0(\mathbf{u}) \cong 1$ garantindo que $\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha = 1$.

Para que o preditor (60) seja ótimo, ele deve obedecer às condições de não tendência e de variância mínima (26) e (27), ou seja:

$$E\left[Z^*_{KU}(\mathbf{u}) - Z(\mathbf{u})\right] = E\left[\sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{KU}_\alpha(\mathbf{u}) Z(\mathbf{u}_\alpha) - Z(\mathbf{u})\right]$$

$$= \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{KU}_\alpha(\mathbf{u}) E\left[Z(\mathbf{u}_\alpha)\right] - E\left[Z(\mathbf{u})\right], \text{ de (31),}$$

$$= \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{KU}_\alpha(\mathbf{u}) m(\mathbf{u}_\alpha) - m(\mathbf{u}), \text{ por (56),}$$

$$= \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{KU}_\alpha(\mathbf{u}) \sum_{i=0}^{p} a_i(\mathbf{u}) f_i(\mathbf{u}_\alpha) - \sum_{i=0}^{p} a_i(\mathbf{u}) f_i(\mathbf{u})$$

$$= \sum_{i=0}^{p} a_i(\mathbf{u}) \left[\sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{KU}_\alpha(\mathbf{u}) f_i(\mathbf{u}_\alpha) - f_i(\mathbf{u})\right] = 0.$$ Isto só será igual a zero sob as condições de (59).

A condição de variância mínima deve então ser minimizada sob as condições de não tendência (59):

$$\sigma^2_E(\mathbf{u}) = \mathrm{Var}\left[Z^*_{KU}(\mathbf{u}) - Z(\mathbf{u})\right] = E\left\{\left[Z^*_{KU}(\mathbf{u}) - Z(\mathbf{u})\right]^2\right\} = \text{mínima.}$$

O procedimento é idêntico a KO, exceto pelos (p+1) multiplicadores de Lagrange.

O Sistema de Krigagem Universal (SKU) tem [n(**u**)+(p+1)] equações lineares: n(**u**) pesos $\lambda_\alpha^{KU}(\mathbf{u})$ e os (p+1) parâmetros de Lagrange $\mu_i^{KU}(\mathbf{u})$, i=0,1,2,...,p.

$$\sum_{\beta=1}^{n(\mathbf{u})}\lambda_\beta^{KU}(\mathbf{u})C_R(\mathbf{u}_\alpha-\mathbf{u}_\beta)+\sum_{i=0}^{p}\mu_i^{KU}(\mathbf{u})f_i(\mathbf{u}_\alpha)=C_R(\mathbf{u}_\alpha-\mathbf{u}),$$

$$a=1,2,...,n(\mathbf{u}) \text{ e}$$

$$\sum_{\beta=1}^{n(\mathbf{u})}\lambda_\beta^{KU}(\mathbf{u})f_i(\mathbf{u}_\beta)=f_i(\mathbf{u}),\ i=0,1,...,p. \qquad (61)$$

A variância do erro $\sigma_E^2(\mathbf{u})$ pode ser obtida para cada localização **u** através das primeiras n(**u**) equações do SKU (61):

$$\sigma_{KU}^2(\mathbf{u})=C_R(0)-\sum_{\alpha=1}^{n(\mathbf{u})}\lambda_\alpha^{KU}(\mathbf{u})C_R(\mathbf{u}_\alpha-\mathbf{u})-\sum_{i=0}^{p}\mu_i^{KU}(\mathbf{u})f_i(\mathbf{u}). \qquad (62)$$

Para i=0, o SKU (61) e a variância do erro (62) ficam idênticos ao SKO (54) e a variância do erro (55), respectivamente.

Qualquer diferença entre as estimativas de KO e KU, origina da diferença entre as duas estimativas de tendência, portanto, da decisão arbitrária de modelar a tendência local m(**u**) dentro da vizinhança *W*(**u**) como uma constante desconhecida (KO) ou como uma combinação linear desconhecida de funções conhecidas $\{f_0(\mathbf{u}),...,f_p(\mathbf{u})\}$, $\mathbf{u}\in W(\mathbf{u})$ (KU).

Para i=1, o SKU (61) se transforma em um caso particular, o da Krigagem com Tendência Externa (KTE), onde a tendência m(**u**) é modelada como uma função linear de uma variável secundária (externa) S(**u**),

$$E[Z(\mathbf{u})] = b_0(\mathbf{u}) + b_1(\mathbf{u})S(\mathbf{u}).$$

Para que a variável secundária S(**u**) possa ser usada dessa forma, o seu valor precisa ser conhecido para todas as localizações $\mathbf{u}_\alpha$, $\alpha = 1,..., n(\mathbf{u})$ onde os valores da variável primária são conhecidos e para toda localização **u** sendo predita. Há ainda a necessidade de existir uma forte correlação linear entre as variáveis secundária e primária. Assim, pode ser assumido que em média Z(**u**) é igual a S(**u**) sob uma constante $b_0(\mathbf{u})$ e um coeficiente $b_1(\mathbf{u})$, ambos constantes dentro da vizinhança *W*(**u**) e implicitamente estimados através do sistema de KTE:

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{KTE}(\mathbf{u}) C_R(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu_0^{KTE}(\mathbf{u}) + \mu_1^{KTE}(\mathbf{u}) y(\mathbf{u}) = C_R(\mathbf{u}_\alpha - \mathbf{u}),$$

$$\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha^{KTE}(\mathbf{u}) = 1 \qquad a = 1,2,..., n(\mathbf{u}) \text{ e}$$

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{KTE}(\mathbf{u}) s(\mathbf{u}_\beta) = s(\mathbf{u}). \qquad (63)$$

Hudson & Wackernagel (1994) usaram dados de altitude como variável secundária na KTE para modelar a tendência na temperatura média do mês de janeiro na Escócia.

O nome *tendência externa* é devido ao fato de que o modelo linear entre as variáveis primária (Z(**u**)) e secundária (S(**u**)) é incorporado no sistema de KU sob considerações que não interferem na modelagem do semivariograma (ou covariância generalizada) de Z(**u**). As variáveis de localização, coordenadas X e Y são internas, pois são usadas na modelagem do semivariograma. Libertá & Girolamo (1991) citados por Hudson &

Wackernagel (1994), ajustaram um variograma no espaço tridimensional (latitude, longitude e altitude) e assim usaram a elevação como uma tendência interna.

Mandallaz (2000) justificou o uso de KTE, em um caso onde a suposição de estacionariedade da média pode ser aceita (quando a KO é suficiente), pelo fato de querer usar a informação auxiliar disponível para todos os locais $\mathbf{u}_\alpha$, $\alpha = 1,2,...,n(\mathbf{u})$, na tentativa de diminuir a variância do erro da estimativa da variável resposta de interesse. Em artigo de 1991, o autor utilizou uma informação auxiliar através de mapas temáticos (baseado em interpretações de fotografias aéreas) os quais descrevem características qualitativas para todos os locais $\mathbf{u}_\alpha$, $\alpha = 1,2,...,n(\mathbf{u})$, com a finalidade de reduzir o erro de estimação do volume médio de floresta (número de pés por hectare) em m^3 por hectare obtido de medidas terrestres $Z(\mathbf{u}_a)$, $\alpha = 1,2,...,n(\mathbf{u})$, em Zurique - Suíça.

Carroll (1998) em seu trabalho de predição espacial de riqueza de espécies (área biológica), na Índia, considerou que a altitude do relevo[11] está relacionada com a diversidade do "habitat". Quando a altitude do relevo aumenta, aumenta também o alcance de temperatura que é um fator crítico para a diversidade do "habitat". Como a altitude do relevo é facilmente estimada de mapas topográficos, ela pode ser usada para auxiliar na predição de número de espécies. Técnicas de validação cruzada (ver Seção 2.5) foram usadas para ilustrar como a inclusão da altitude do relevo, como informação auxiliar, aumentou substancialmente a precisão na predição de riqueza de espécies. Carroll comenta, ainda, que em futuros trabalhos pretende investigar outros tipos de informação auxiliar, tais como

[11] Altitude do relevo é a diferença absoluta entre a maior e a menor elevação na região.

intensidade de radiação e temperatura, que podem ajudar na identificação de riqueza de espécies em outras regiões.

Mais sobre a metodologia da KU e suas variações é encontrado em: Myers (1982); Dubrule (1983); Cressie (1993, p151); Wackernagel (1995, p177); Qian (1997); Carroll (1998); Goudard *et al.* (1999) e Mandallaz (2000).

Observando o SKU (61), constata-se que para fazer a KU é necessário o conhecimento prévio do covariograma (ou do semivariograma) dos resíduos que, na prática, nem sempre é fácil ou possível obtê-lo. Por exemplo, usando o estimador de Matheron (2), para quaisquer $\mathbf{u}_\alpha$ e $\mathbf{u}_\beta \in W(\mathbf{u})$, tem-se um estimador do variograma com viés:

$$\mathrm{Var}\left[Z(\mathbf{u}_\alpha) - Z(\mathbf{u}_\beta)\right] = \mathrm{E}\left[Z(\mathbf{u}_\alpha) - Z(\mathbf{u}_\beta)\right]^2 - \left\{\mathrm{E}\left[Z(\mathbf{u}_\alpha) - Z(\mathbf{u}_\beta)\right]\right\}^2, \text{ de (31),}$$

$$\mathrm{E}\left[Z(\mathbf{u}_\alpha) - Z(\mathbf{u}_\beta)\right]^2 = \mathrm{Var}\left[Z(\mathbf{u}_\alpha) - Z(\mathbf{u}_\beta)\right] + \left\{m(\mathbf{u}_\alpha) - m(\mathbf{u}_\beta)\right\}^2$$

$$= \mathrm{Var}\left[Z(\mathbf{u}_\alpha) - Z(\mathbf{u}_\beta)\right] + \left\{m(\mathbf{u}_\alpha) - m(\mathbf{u}_\beta)\right\}^2$$

$$= 2\gamma(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \left\{m(\mathbf{u}_\alpha) - m(\mathbf{u}_\beta)\right\}^2, \text{ de (56),}$$

$$= 2\gamma(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \left\{\sum_{i=1}^{p} a_i \left[f_i(\mathbf{u}_\alpha) - f_i(\mathbf{u}_\beta)\right]\right\}^2.$$

Se os $(a_i(\mathbf{u})$: i=0,...,p) são conhecidos, um estimador do semivariograma pode ser baseado nos resíduos, $R(\mathbf{u}) \cong Z(\mathbf{u}) - \sum_{i=0}^{p} a_i f_i(\mathbf{u})$, porque $\mathrm{E}\left[R(\mathbf{u}_\alpha) - R(\mathbf{u}_\beta)\right]^2 = 2\gamma(\mathbf{u}_\alpha - \mathbf{u}_\beta)$. Mas os $(a_i(\mathbf{u})$: i=0,...,p) são desconhecidos (ver 56), pois a tendência $m(\mathbf{u})$ é uma combinação linear desconhecida de funções conhecidas.

Para que os $(a_i(\mathbf{u}): i=0,...,p)$ sejam estimados eficientemente, o conhecimento do semivariograma $\gamma(\mathbf{h})$ é necessário, precisamente a matriz de variâncias e covariâncias Var(Z). Segundo Cressie (1993, p165), esta circularidade tem levado a uma insatisfação prematura, de alguns autores, com a KU.

Armstrong (1984) comparou este problema (quando o variograma não é conhecido) com a velha questão da "*galinha-ovo*".

Cressie (1993, Seção 3.4.3) discutiu a questão da estimação do variograma baseado nos resíduos e mostra que, mesmo usando o estimador de mínimos quadrados generalizados, os resíduos $W \cong Z - \hat{a}_{mqg}$, onde $\hat{a}_{mqg} = \left(X'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}Z$, produzem um viés no variograma estimado.

Vale lembrar que se $\Sigma = \sigma^2 I$, onde I é a matriz identidade e Σ a matriz de variâncias e covariâncias, os resíduos são independentes e identicamente distribuídos (iid), não existindo correlação espacial. Nestes termos, $(Z - Xa)'(Z - Xa)/\sigma^2$ e $\hat{a}_{mqo} = \left(X'X\right)^{-1}X'Z$ torna-se o estimador de mínimos quadrados ordinários do vetor de coeficientes (a_i) de (56).

Se $\Sigma \neq \sigma^2 I$, o estimador de MQO pode ser ineficiente, ou seja os resíduos são espacialmente correlacionados e o melhor estimador não viciado do vetor de coeficientes (a_i) é o de mínimos quadrados generalizados.

Quase sempre a função covariância espacial, na KU, não é conhecida. O seguinte procedimento é usado com freqüência na literatura: obter a estimativa dos $(a_i(\mathbf{u}): i=0,...,p)$ por mínimos quadrados ordinários e ajustar (usualmente "a olho"), um variograma a estes empíricos resíduos para obter uma estimativa da matriz de variâncias e covariâncias dos

resíduos $\hat{\Sigma}$; atualizar então a estimativa dos ($a_i(\mathbf{u})$: i=0,...,p) usando mínimos quadrados generalizados com $\hat{\Sigma}$ e, iterar o procedimento algumas vezes (Mandallaz, 2000). Segundo Cressie (1993, p167), esta aproximação iterativa faz com que os resíduos de MQO se pareça mais e mais com os resíduos de MQG e diz, ainda, que o problema do viés não é resolvido, pois sempre que Σ é conhecido, o viés é diferente de zero.

Geralmente é verdadeiro que o viés do estimador do variograma, baseado nos resíduos, seja pequeno para distâncias ("lags") próximas a origem, mas mais substancial para grandes distâncias (Grondona & Cressie, 1995).

O preditor de Krigagem, tipicamente usa o modelo de variograma para pequenas distâncias. Assim, um apropriado ajuste ponderado do variograma estimado, pode minimizar o efeito do viés por atribuir mais peso a distâncias menores (Cressie, 1993, p170).

Observa-se, então, que computar o variograma dos resíduos $\gamma_R(\mathbf{h})$ não é muito simples, pois o que se tem são valores de Z e, portanto, $\gamma(\mathbf{h})$ e não valores residuais.

Algumas soluções são encontradas na literatura para casos específicos e algumas delas serão comentadas. Uma observação deve ser colocada, aqui, com respeito a estes casos específicos: em geral, no entanto, não é provável que a tendência ("drift") seja constante em qualquer direção do espaço. Portanto, uma abordagem mais sofisticada é necessária e esta deve estar baseada em funções aleatórias intrínsecas de ordem k e envolver covariância generalizada em vez de variogramas (Matheron, 1973; Chauvet & Galli, 1982; citados por Hudson & Wackernagel, 1994). Mais sobre funções aleatórias intrínsecas de ordem k envolvendo covariância generalizada pode ser visto em Wackernagel (1995, p185); Cressie (1993, p299) e Grondona & Cressie, (1995) e não serão abordados aqui.

Em alguns casos, o semivariograma (ou o covariograma) é conhecido "a priori" como no estudo de caso sobre "sea floor" feito por Chilès (1977), citado por Armstrong (1984), onde não existia tendência em uma direção. Assim, o variograma para esta direção foi usado como um variograma aproximado para todas as outras direções.

São "afortunados" os casos onde existe alguma direção livre de tendência no qual o variograma pode ser estimado (Cressie, 1993, p158). Neste caso, o inacessível variograma dos resíduos na direção da tendência é julgado similar àquele computado numa direção perpendicular. Tal decisão equivale a modelar o padrão de variação dos resíduos como isotrópico. Qualquer anisotropia nos dados é explicada no modelo de tendência, não na covariância residual (Goovaerts, 1997, p142).

Wackernagel (1995, p183) disse que a existência de tendência em certa direção pode mascarar o variograma em tal direção.

Construído sob a suposição de isotropia do variograma dos resíduos $\gamma_R(\mathbf{h})$, será possível inferir o variograma teórico do resíduo R(**u**) do variograma experimental de Z(**u**), construído na direção onde a tendência é constante (Hudson & Wackernagel, 1994).

Wackernagel (1995, p182) citou duas situações excepcionais onde o variograma é diretamente acessível:

i) quando a tendência ocorre somente em uma parte do domínio a qual é assumido homogêneo, então o modelo de variograma é inferido na outra parte estacionária do domínio e pode ser transferida para a parte não estacionária;

ii) quando a tendência não está ativa em uma particular direção do espaço, o variograma inferido nesta direção pode ser estendido para outras direções, sob uma suposição de comportamento isotrópico do variograma.

Considerando, como exemplo, o modelo de tendência linear em duas dimensões: $m(\mathbf{u}) = m(x, y) = a_0 + a_1 x + a_2 y$, onde (x,y) são as coordenadas da localização **u** sendo predita, o SKU (61) inclui $(n(\mathbf{u}) + 3)$ equações lineares:

$$\sum_{\beta=1}^{n(\mathbf{u})} ?_{\beta}^{KU}(\mathbf{u}) C_R(\mathbf{u}_a - \mathbf{u}_{\beta}) + \mu_0^{KU}(\mathbf{u}) + \mu_1^{KU}(\mathbf{u}) x_a + \mu_2^{KU}(\mathbf{u}) y_a = C_R(\mathbf{u}_a - \mathbf{u}),$$

$$\sum_{\beta=1}^{n(\mathbf{u})} ?_{\beta}^{KU}(\mathbf{u}) = 1 \qquad a = 1,2,\ldots, n(\mathbf{u})$$

$$\sum_{\beta=1}^{n(\mathbf{u})} ?_{\beta}^{KU}(\mathbf{u}) x_{\beta} = x$$

$$\sum_{\beta=1}^{n(\mathbf{u})} ?_{\beta}^{KU}(\mathbf{u}) y_{\beta} = y . \qquad (64)$$

Assim, o SKU (64) pode ser escrito matricialmente como:

$$\mathbf{K}_{KU} \; ?_{KU}(\mathbf{u}) = \mathbf{k}_{KU},$$

onde $\mathbf{K}_{KU}$ é uma matriz $(n(\mathbf{u}) + 3) \times (n(\mathbf{u}) + 3)$:

$$\begin{bmatrix} C_R(\mathbf{u}_1 - \mathbf{u}_{1)}) & \cdots & C_R(\mathbf{u}_1 - \mathbf{u}_{n(\mathbf{u})}) & 1 & x_1 & y_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ C_R(\mathbf{u}_{n(\mathbf{u})} - \mathbf{u}_1) & \cdots & C_R(\mathbf{u}_{n(\mathbf{u})} - \mathbf{u}_{n(\mathbf{u})}) & 1 & x_{n(\mathbf{u})} & y_{n(\mathbf{u})} \\ 1 & \cdots & 1 & 0 & 0 & 0 \\ x_1 & \cdots & x_{n(\mathbf{u})} & 0 & 0 & 0 \\ y_1 & \cdots & y_{n(\mathbf{u})} & 0 & 0 & 0 \end{bmatrix};$$

$?_{KU}(\mathbf{u})$, é o vetor $(n(\mathbf{u}) + 3)$x1 de pesos e, dos parâmetros de Lagrange:

$$\begin{bmatrix} ?_1^{KU}(\mathbf{u}) \\ \vdots \\ ?_{n(\mathbf{u})}^{KU}(\mathbf{u}) \\ \mu_0^{KU}(\mathbf{u}) \\ \mu_1^{KU}(\mathbf{u}) \\ \mu_2^{KU}(\mathbf{u}) \end{bmatrix} \text{ e}$$

$\mathbf{k}_{KU}$, é o vetor $(n(\mathbf{u})+3)$x1 de covariâncias entre as localizações e a localização a ser predita e, as funções de tendência:

$$\begin{bmatrix} C_R(\mathbf{u}_1 - \mathbf{u}) \\ \vdots \\ C_R(\mathbf{u}_{n(\mathbf{u})} - \mathbf{u}) \\ 1 \\ x \\ y \end{bmatrix}.$$

Portanto, os pesos são obtidos por $? _{KU}(\mathbf{u}) = \mathbf{K}_{KU}^{-1}\mathbf{k}_{KU}$.

A formulação matricial da variância da Krigagem (62) é dada por,

$$s_{KU}^2(\mathbf{u}) = C(\mathbf{0}) - ?_{KU}^T(\mathbf{u})\mathbf{k}_{KU} = C(\mathbf{0}) - \mathbf{k}_{KU}^T\mathbf{K}_{KU}^{-1}\mathbf{k}_{KU}. \qquad (65)$$

2.5 Validação Cruzada

Validação cruzada é uma ferramenta apropriada para testar métodos de interpolação (Krigagem). Com ela pode-se avaliar algumas suposições sobre o modelo de semivariograma (e seus parâmetros), ajustado ao semivariograma experimental.

Na prática, onde geralmente tem-se dados irregularmente espaçados, a determinação do número de vizinhos a ser usado na Krigagem para estimar determinado ponto ($\mathbf{u}_k$) não é tarefa simples. A validação cruzada pode ser usada para comparar qual o melhor número de vizinhos mais próximo a ser usado (Wackernagel, 1995).

Ribeiro Jr. (1995) observou que a validação cruzada não avalia simplesmente o modelo escolhido para o semivariograma. Toda a modelagem do processo estocástico em questão está sendo posta a prova. Isto inclui a decisão de estacionariedade, os estimadores utilizados, o tratamento dos dados atípicos, o modelo de semivariograma adotado e a decisão relativa à anisotropia.

No procedimento de validação cruzada, cada valor amostral $z(\mathbf{u}_a)$, $a = 1,..., n$, é removido (deletado) alternadamente do conjunto de dados e um valor $z^*(\mathbf{u}_a)$, $a = 1,..., n$, para cada localização é predito baseado no modelo de semivariograma ajustado e nos (n-1) valores amostrais. O objetivo é verificar como cada ponto retirado pode ser predito por seus vizinhos.

Quando os (n-1) valores amostrais $z(\mathbf{u}_a)$, $a \neq k$ (menos o que está sendo estimado $z^*(\mathbf{u}_k)$), são usados na validação cruzada, tem-se a vizinhança única; se alguns dos valores amostrais dos vizinhos mais próximos são usados, tem-se a vizinhança móvel.

Dubrule (1983) apresentou um método mais rápido para validação cruzada de uma única vizinhança onde apenas a inversão de uma matriz (nxn) é requerida. Em seu artigo diz, ainda, que mapas obtidos por Krigagem em uma vizinhança única são mais "suaves" e estéticos que aqueles obtidos em uma vizinhança móvel.

Para se usar vizinhança única é necessário que o semivariograma seja definido até a maior distância existente no campo. Uma grande vantagem desse método é que, uma vez invertida a matriz de coeficientes, então as estimativas podem ser feitas para qualquer espaçamento, com um pequeno consumo de tempo de processamento de computação.

Pode-se encontrar mais detalhes sobre a teoria de validação cruzada em: Dubrule (1983); Souza (1992); Cressie (1993); Uzumaki (1994); Ribeiro Jr. (1995); Wackernagel (1995) e Vieira (1996) e algumas aplicações como em Carroll (1998) e Host (1999).

Alguns critérios segundo Dubrule (1983), Wackernagel (1995, p80) e Cressie (1993, p102) podem ser utilizados para avaliar a qualidade das estimativas $z^*(\mathbf{u}_a)$, $a = 1,..., n$, através dos erros de predição:

$$\left[z(\mathbf{u}_a) - z^*(\mathbf{u}_a)\right], \; a = 1,..., n. \qquad (66)$$

Ou seja, a proximidade dos valores verdadeiros pode ser caracterizada de várias maneiras, por exemplo:

- Wackernagel (1995), disse que a média dos erros (66) deve ser próxima de zero,

$$\frac{1}{n}\sum_{a=1}^{n}\left[z(\mathbf{u}_a) - z^*(\mathbf{u}_a)\right] \cong 0;$$ e, a média do quadrado dos erros (66) padronizados é aproximadamente um,

$$\frac{1}{n}\sum_{a=1}^{n}\frac{\left[z(\mathbf{u}_a)-z^*(\mathbf{u}_a)\right]^2}{s_E^2(\mathbf{u}_a)} \cong 1,$$

onde o desvio padrão de Krigagem $s_E(\mathbf{u}_a)$ representa o erro do valor predito pelo modelo quando a Krigagem é feita para a localização $\mathbf{u}_a$ (omitindo o valor amostral da localização $\mathbf{u}_a$).

- O gráfico dos valores medidos $z(\mathbf{u}_a)$ *versus* os valores preditos $z^*(\mathbf{u}_a)$ deverá representar uma regressão linear $z^*(\mathbf{u}_a) = a + bz(\mathbf{u}_a)$, onde os coeficientes de correlação e de determinação sejam próximos de 1, o intercepto próximo de zero e o coeficiente angular próximo de 1. Se os valores do intercepto aumentam de zero para valores positivos, isto indica que o estimador $z^*(\mathbf{u}_a)$ está superestimando valores pequenos de $z(\mathbf{u}_a)$ e subestimando valores grandes. À medida que o intercepto decresce de zero para valores negativos, o contrário acontece (Vieira, 1996).

Uma aplicação deste gráfico pode ser vista em Host (1999), para dados de concentração de sulfato em países da Europa.

- O gráfico dos erros (66) *versus* os valores preditos $z^*(\mathbf{u}_a)$ deverá ser centrado ao redor da linha zero (uma propriedade denominada de "condição de não viciado"). Além disso, o gráfico deverá ter uma dispersão por igual, isto é, a variância do erro não deverá depender da grandeza do valor (uma propriedade conhecida como homocedasticidade do erro da variância) (Uzumaki, 1994).

3 MATERIAL E MÉTODOS

3.1 Material

Dados de temperatura do ar são diariamente medidos pelo Instituto Agronômico do Paraná – IAPAR em 33 estações agrometeorológicas distribuídas por todo o Estado do Paraná (Figura 11).

Algumas dessas estações existem há mais de 40 anos, como a do município de Ponta Grossa, na região central do Estado, desde 1954, e a do município de Cambará, ao norte do Estado, desde 1957.

As medidas diárias de temperatura do ar são feitas em termômetros de mercúrio em vidro, instalados a dois metros de altura, no interior do abrigo meteorológico. As leituras são realizadas três vezes ao dia, às 9 horas, 15 horas e 21 horas. São anotados também as temperaturas máximas e mínimas diárias dentro do abrigo e a temperatura média diária $\left(\overline{T}\right)$ é calculada pela seguinte expressão (Caviglione *et al.*, 2000):

$$\overline{T} = \frac{T_{9h} + \left(2T_{21h}\right) + T_{max} + T_{min}}{5}, \text{ onde:}$$

T_{9h} e T_{21h} são as temperaturas do ar às 9 horas e às 21 horas, respectivamente;

T_{max} e T_{min} são as temperaturas do ar máxima e mínima, respectivamente.

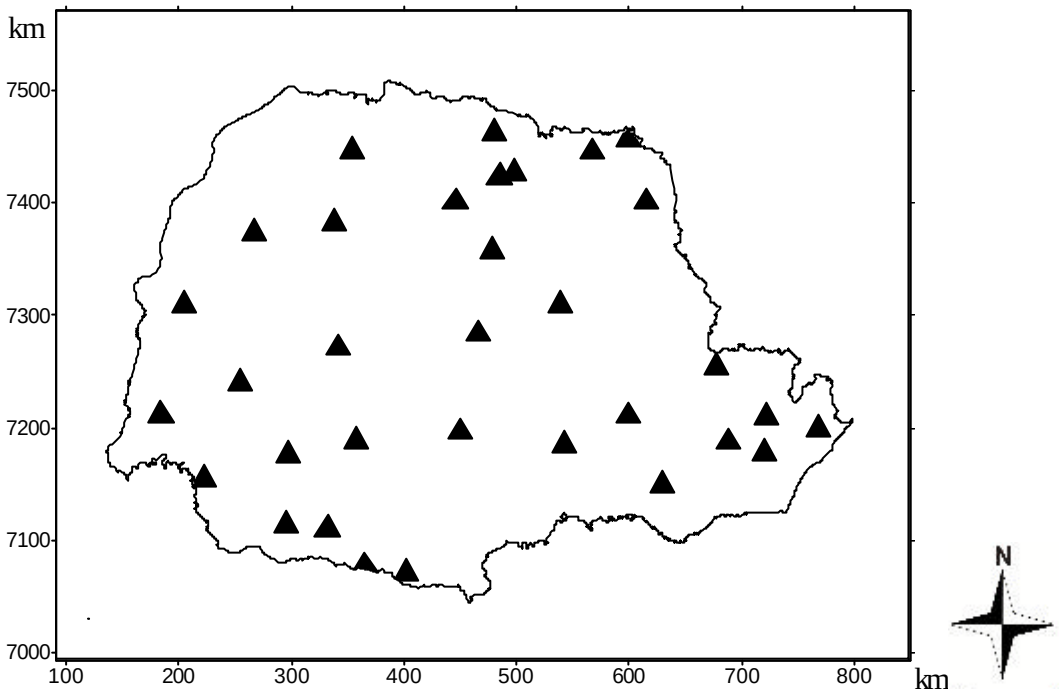

Figura 11 - Localização das estações agrometeorológicas do IAPAR.

Os dados de *temperatura média mensal*[12], fornecidos pelo Instituto Agronômico do Paraná - IAPAR situado na rodovia Celso Garcia Cid, km 375, CEP 86001-970, Londrina/Paraná, são obtidos em graus Celsius (°C), a *altitude* em metros (m) e as locações das estações agrometeorológicas em *latitude sul* e *longitude oeste* medidas em graus, minutos e segundos (xx°xx'xx'').

A temperatura média mensal para cada uma das 33 estações é calculada considerando um período de aproximadamente 20 anos.

[12] Para expressar a temperatura média mensal neste trabalho, será omitida a palavra média ou, em alguns casos, escrever-se-á apenas temperatura.

Os dados de latitude, longitude e altitude dos 399 municípios do Estado do Paraná foram obtidos no "site" http://www.paranacidade.org.br/, acessado entre os dias 22 e 28 de janeiro de 2001.

Como a maioria das estações agrometeorológicas do IAPAR tem suas localizações diferentes das localizações de referência de cada município (com exceção dos municípios de Cândido de Abreu e Lapa), os dados foram unidos formando um conjunto de 399 + (33-2) = 430 localizações. Estes valores de altitude foram utilizados na construção de gráficos de dispersão de altitude com valores preditos de temperatura média mensal pela Krigagem Universal.

3.2 Métodos

No conjunto de dados, o mês de julho é o que apresenta maior variação relativa na temperatura média através dos anos, por isso, foi um dos escolhidos para a aplicação das análises geoestatísticas. Por outro lado, em contraste ao mês de julho (inverno), as mesmas análises foram feitas para o mês de janeiro (verão) que é um dos meses que apresenta menor variação na temperatura média através dos anos. Os demais meses não foram analisados neste trabalho, pois o objetivo foi verificar a existência de dependência espacial em temperatura e obter mapas mais precisos por meio de alguns dos métodos de interpolação dados pela geoestatística, comparando-os com os existentes no IAPAR – Londrina, à disposição da população. A metodologia pode ser facilmente estendida aos outros meses, requerendo apenas familiarização com os métodos estatísticos e geoestatísticos básicos.

Para a aplicação das análises geoestatísticas, sabe-se que os dados precisam ter suas localizações conhecidas como num plano cartesiano, convencionalmente representado pelas coordenadas X e Y.

Para saber com exatidão onde se localiza qualquer ponto na superfície terrestre, é usado o sistema de coordenadas geográficas que se baseia em linhas imaginárias traçadas sobre o globo terrestre: os paralelos e os meridianos. Os paralelos são linhas circulares que dão volta na Terra (360°), sendo a linha do Equador o maior paralelo e o de referência para a medida da latitude (medida em graus de 0° a 90°). Os meridianos são linhas semicirculares, isto é, linhas de 180°, que vão do Pólo Norte ao Pólo Sul e cruzam com os paralelos. O meridiano considerado de referência para a medida de longitude (medida em graus de 0° a 180°) é o que passa pelos arredores de Londres (na antiga sede do observatório de Greenwich), na Inglaterra, por esta razão chamado de meridiano de Greenwich (meridiano tomado como origem do tempo universal).

Qualquer local ou objeto no globo terrestre, situado exatamente na linha do Equador, possui 0° de latitude. Para norte ou para sul desse paralelo, a latitude vai aumentando até chegar a 90°: o Pólo Norte possui 90° de latitude norte e o Pólo Sul, 90° de latitude sul. Da mesma forma, qualquer local ou objeto no globo terrestre situado exatamente sobre o meridiano de Greenwich possui 0° de longitude. Para oeste ou para leste desse meridiano, a longitude vai aumentando até chegar a 180°, no outro extremo do planeta. Por exemplo: em direção ao sul da linha do Equador a 23°23’30’’ e, em direção ao oeste do meridiano de Greenwich a 51°11’30’’, está a cidade de Londrina-Paraná (veja Figura 12 – a posição geográfica de alguns municípios extremos do Estado do Paraná).

O Paraná é um dos 26 estados brasileiros (com o Distrito Federal, totaliza 27 unidades político – administrativas) e situa-se na região sul do Brasil. Seu clima é subtropical, apresentando-se mais ameno na região norte e temperado no sul, onde os invernos podem ser rigorosos. Segundo a classificação de Wilhelm Köppen o tipo que corresponde à maior área é o CFA, que se caracteriza por ser subtropical úmido, mesotérmico, com verão quente, sem estação seca de inverno definida e geadas pouco freqüentes. O tipo CFB é subtropical úmido, mesotérmico, com verões frescos e geadas severas e freqüentes. Com médias térmicas anuais entre 14° e 19°C e pluviosidade entre 1250 e 2000 mm/ano.

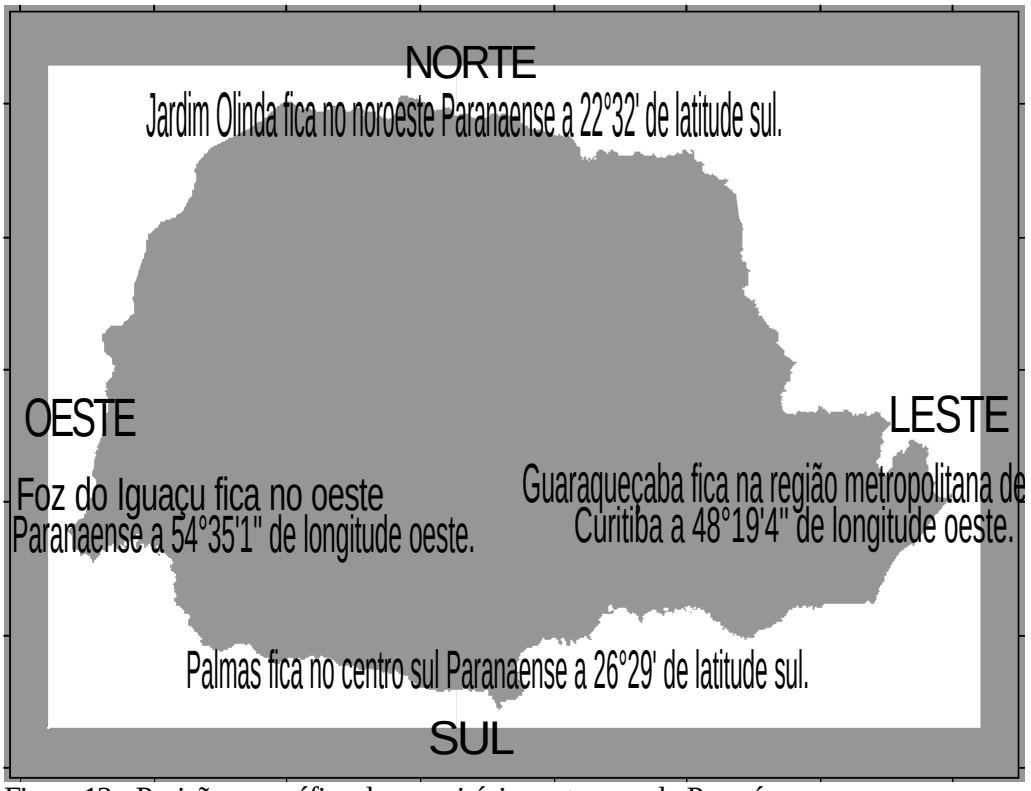

Figura 12 - Posição geográfica dos municípios extremos do Paraná.

As localizações das 33 estações são identificadas, a princípio, pelas coordenadas geográficas, latitude sul e longitude oeste, referidas ao equador e a um meridiano

- origem, fornecidas pelo IAPAR. Deve-se, então, transformá-las num sistema de unidades que permita o cálculo da distância (euclidiana) entre as localizações.

Para transformar as coordenadas geográficas: latitude e longitude, graus, minutos e segundos, em coordenadas X (leste) e Y (norte) apresentadas em metros (coordenadas UTM - Universal Transverse Mercator: coordenada leste-oeste, E-W, e coordenada norte-sul, N-S), foi utilizado o programa Arcview GIS 3.2a existente no IAPAR/Londrina.

Os limites efetivamente usados (compare-se com os da Figura 13) nos programas SURFER (Win 32) 6.01 e Geo-EAS (Engelud & Sparks, 1991) de modo a abranger toda a área do Estado do Paraná foram:

Quadro 1 - Limites usados na construção dos mapas.

	Coordenada leste (X)	**Coordenada norte (Y)**
Mínimo	130000 m	7040000 m
Máximo	820000 m	7520000 m

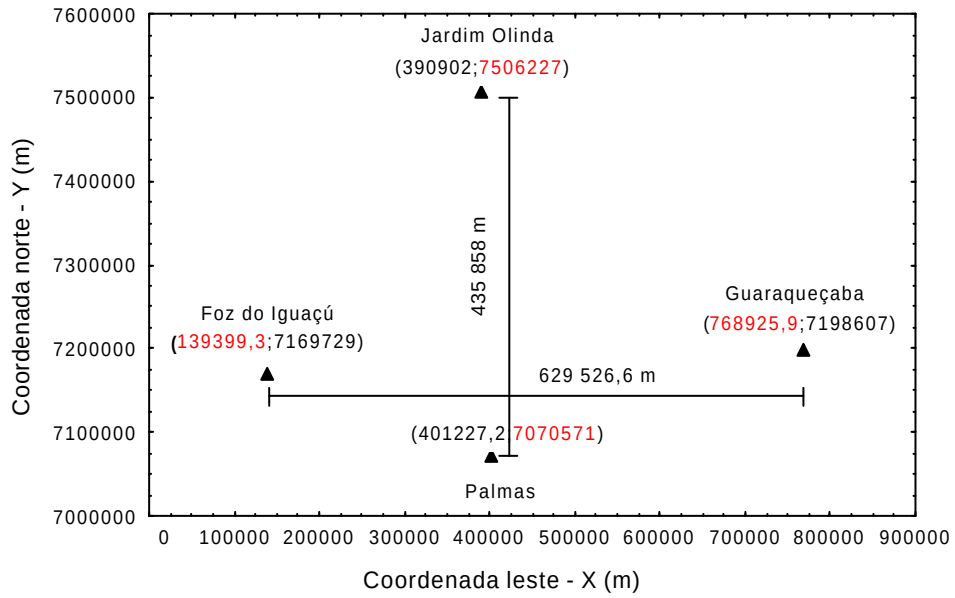

Figura 13 - Posição geográfica das estações que se situam nos extremos do Estado.

Os valores em vermelho, na Figura 13, indicam os valores mínimos e máximos das coordenadas (veja Quadro 1).

Para as análises geoestatísticas considerou-se os dados de *temperatura média dos meses de janeiro (TJA) e julho (TJU)* em graus Celsius (°C), *altitude (A)* em metros (m), e os locais de observação dados pelas *coordenadas leste (X) e norte (Y)* em metros (m), salvo em alguns mapas onde as coordenadas X e Y são dadas em km por questões de estética.

Logo, tem-se um conjunto de valores medidos $\{z(\mathbf{u}_a), a = 1,...,n\}$, $\alpha = 1,..., n = 33$, sendo que $z(\mathbf{u}_\alpha)$ representa a variável: temperatura média de janeiro (TJA) e temperatura média de julho (TJU), uma de cada vez e, $\mathbf{u}_a$ identifica uma posição no espaço, e é um vetor de coordenadas (x_a, y_a), x_α é o valor da coordenada leste (X) e y_α é o valor da coordenada norte (Y).

Inicialmente, como em todo trabalho estatístico, antes de se proceder com os métodos geoestatísticos, fez-se a análise exploratória dos dados (Seção 4.1). Isto foi realizado nos programas Excel (2000) e Statistica 6.0 e se faz necessário para o reconhecimento dos dados e para facilitar e orientar futuras interpretações e análises.

A análise exploratória dos dados pode induzir à escolha de um tipo de estacionariedade a ser assumida, mas não existe um teste objetivo para avaliar a significância da escolha.

Algum tipo de estacionariedade deve ser adotado, como mostrado na Seção 2.3.2. A estacionariedade permite obter dos dados diversas realizações de uma variável aleatória, de forma que pelo menos os dois primeiros momentos da distribuição desta variável

possam ser estimados. Conseqüentemente, isto permitirá a predição de valores em locais não amostrados.

Portanto, quando assumida alguma hipótese de estacionariedade, neste trabalho, será a hipótese intrínseca (Seção 2.3.2 Definição 4) que é um tipo de estacionariedade dos incrementos $[Z(\mathbf{u}+\mathbf{h})-Z(\mathbf{u})]$.

Uma análise exploratória espacial dos dados, considerando suas localizações, também é feita e o semivariograma (2) é a ferramenta mais adequada para detectar e expressar a existência de dependência espacial nos dados.

Os modelos básicos apresentados na Seção 2.3.9 foram ajustados aos semivariogramas experimentais das temperaturas médias dos meses de janeiro (TJA) e julho (TJU) e os que se ajustaram melhor foram analisados.

Os modelos considerados (Seção 2.3.9) são aqueles que melhor se ajustam aos primeiros pontos experimentais visto que, à medida que aumenta a distância entre os pontos, a dissimilaridade também aumenta, tornando-se pouco influentes uns pontos sobre os outros.

O semivariograma experimental é ajustado somente para semivariâncias de distâncias ("lags") de até metade da distância máxima, porque a variância amostral das estimativas tende a aumentar excessivamente para grandes distâncias, devido ao decréscimo do número de observações envolvidas (Watson, 2000).

Neste trabalho, a máxima distância considerada na construção dos semivariogramas experimentais foi de aproximadamente metade da maior distância existente entre as estações agrometeorológicas do IAPAR. Veja Quadro 1 e Figura 13.

O indicador de melhor ajuste dos modelos teóricos aos semivariogramas experimentais é o determinado pelo programa Variowin 2.2 (Pannatier, 1996, p56): o IGF ("Indicative Goodness of Fit"). O IGF é um número sem unidades (adimensional), sendo que valores próximos de zero indicam bom ajuste. O autor do programa adverte que um bom ajuste, considerando apenas o valor do IGF, pode não ser o mais apropriado, principalmente quando se considera o semivariograma construído até a maior distância existente no campo.

Nesta etapa do trabalho, uma integração entre as análises exploratórias clássica (estatísticas como: média, desvio padrão, coeficientes de assimetria e curtose, histogramas, e outras) e geoestatística (variogramas e outros) são muito importantes e a necessidade de "ir e vir" entre estes recursos é constante em qualquer trabalho que envolva dados georreferenciáveis.

Para avaliar a modelagem, também se utilizou a técnica de Validação Cruzada (Seção 2.4), por meio do programa Geo-EAS (Engelud & Sparks, 1991) e GS+ para Windows 5.0.3 Beta.

De acordo com Wackernagel (1995), aplicando-se a técnica de Validação Cruzada, a média dos erros, eq. (66), deve ser próxima de 0 (*zero*) e a média do quadrado dos erros padronizados próximo de 1 (*um*). Veja Seção 2.5.

Na hipótese da existência de dependência espacial nos dados considerados, valores de temperatura média mensal em locais não amostrados, foram preditos através da Krigagem Ordinária (Seção 2.4.2) e Krigagem Universal (Seção 2.4.3), para a obtenção de uma malha bem mais fina que a amostrada e, conseqüentemente, mapas de temperatura bem mais precisos. Os resultados das análises são apresentados no Capítulo 4.

Todos os semivariogramas experimentais e toda a modelagem da estrutura de variabilidade espacial, detectada nos semivariogramas, foram feitos através do programa Variowin 2.2 (Pannatier, 1996).

O número de vizinhos a ser considerado na predição de cada ponto é um fator crítico nesse trabalho, pois os dados são irregularmente espaçados e escassos. Veja Wackernagel (1995) que descreve sobre o número de vizinhos necessários nos procedimentos de interpolação.

O número de vizinhos foi então considerado da seguinte forma: a área em torno do ponto a ser predito, num raio igual ao valor do alcance (*a*) foi dividida em 4 partes, os chamados quadrantes, sendo permitido até 3 (*três*) pontos amostrais por quadrante. O número máximo permitido de quadrantes vazios foi de 2 (*dois*) e o número mínimo total de pontos amostrais na predição foi de 2 (*dois*). Estas restrições foram impostas pela existência de poucos valores em determinadas regiões e para que o número máximo de vizinhos em regiões onde existem mais amostras não ultrapassasse 12 (*doze*).

Um estudo descritivo das variáveis de temperatura média dos meses de janeiro e julho com as coordenadas X e Y foi feito para identificar a presença de tendências. Sabe-se que a temperatura decresce com o aumento da latitude (Coelho, 1992) assim, é importante que seja verificada a existência de tendência nessa direção.

A associação existente entre as variáveis de temperatura média de janeiro (TJA) e de julho (TJU) com coordenada Y, motivou o uso da Krigagem Universal como método de interpolação.

Se a tendência em alguma direção for confirmada, a KO não é a melhor alternativa como método de interpolação, pois requer estacionariedade da média.

Assim, técnicas geoestatísticas não estacionárias de predição são necessárias, como a Krigagem Universal que incorpora esta tendência (determinística) existente nos dados, no sistema de krigagem, sem interferir na modelagem do variograma (ver Seção 2.4.3).

Por assumir isotropia, pode-se usar os modelos ajustados ao semivariograma obtido para a direção X (longitude), também na direção Y (latitude). A interpretação dada por Wackernagel (1995) para o semivariograma na direção Y é que a tendência pode estar mascarando o semivariograma em tal direção.

As considerações discutidas na revisão de literatura (Seção 2.3) foram analisadas à medida do possível, pois a pouca reprodutibilidade dos dados amostrais limitam a realização de estudos mais elaborados exigidos pela teoria Geoestatística. A quantidade de pares amostrais com que foi avaliado cada ponto dos semivariogramas experimentais foi sempre superior a 30, seguindo recomendações encontradas na literatura, atribuídas a Journel & Huijbrets, 1978.

Depois de modelar a dependência espacial detectada nos semivariogramas experimentais, o interesse está em construir mapas de temperatura média dos meses de janeiro e julho. Para isso, uma maneira de representar os resultados da krigagem é através dos mapas de isolinhas.

Para a construção dos mapas, um “grid” regular foi construído sobre a área do Estado do Paraná (considerando os limites dados no Quadro 1) contendo 50 colunas e 35 linhas, no programa SURFER, com 1750 pontos a serem preditos pelos métodos de interpolação apresentados na Seção 2.4.

A Krigagem Ordinária e a Universal foram feitas por meio do programa SURFER (Win32) 6.01 e do GS+ para Windows 5.0.3 Beta. Os mapas foram construídos com o programa SURFER (Win32) 6.01 que é o de melhor qualidade.

4 RESULTADOS E DISCUSSÃO

4.1 Análise Exploratória dos Dados

Somente os meses de janeiro e julho foram considerados nas análises geoestatísticas, pois são os que apresentam o menor e o maior coeficientes de variação, respectivamente. Os Quadros 2 e 4 apresentam todos os meses, apenas para comparações. Estas estatísticas foram obtidas computando os valores de todos os anos disponíveis.

Quadro 2 - Estatísticas descritivas das temperaturas médias (°C) dos meses de janeiro a dezembro no Estado do Paraná.

Mês	n	Média	Mínimo	Máximo	Variância	DP	Assimetria	Curtose	CV
JAN	**33**	**23,361093**	**20,162963**	**25,720000**	**2,782321**	**1,668029**	**-0,600694**	**-0,969081**	**0,071402**
FEV	33	23,192366	19,966667	24,959091	2,707384	1,645413	-0,595994	-1,147764	0,071946
MAR	33	22,400698	18,923810	24,425926	3,095429	1,759383	-0,635590	-1,020682	0,078541
ABR	33	20,142349	16,395238	22,266667	3,381785	1,838963	-0,612506	-1,051753	0,091298
MAI	33	17,228188	13,280952	19,464000	3,485144	1,866854	-0,536538	-1,072927	0,108360
JUN	33	15,446630	11,676190	17,804000	3,136218	1,770937	-0,468340	-1,040731	0,114649
JUL	**33**	**15,533146**	**11,638095**	**18,148000**	**3,390577**	**1,841352**	**-0,378780**	**-0,975221**	**0,118543**
AGO	33	17,044842	13,338095	19,928000	3,722519	1,929383	-0,312099	-0,969457	0,113195
SET	33	18,178332	14,195238	20,736000	3,924159	1,980949	-0,465143	-0,907925	0,108973
OUT	33	20,293637	16,271429	22,888000	4,014489	2,003619	-0,564727	-0,798081	0,098731
NOV	33	21,808416	17,875000	24,214286	3,602100	1,897920	-0,636903	-0,765610	0,087027
DEZ	33	22,883611	19,364286	25,392857	3,007287	1,734153	-0,623076	-0,854638	0,075781

n → # de estações agrometeorológicas. DP → Desvio Padrão. CV → Coeficiente de Variação.

No Quadro 3 tem-se as principais estatísticas descritivas das variáveis: coordenada leste (X), coordenada norte (Y) e altitude (A) em metros (m) e, temperatura média dos meses de janeiro (TJA) e julho (TJU) em graus Celsius (°C) e nas Figuras 14, 15 e 16 os histogramas (com as respectivas curvas normais esperadas).

Quadro 3 - Estatísticas descritivas: da altitude (A), em m e; das temperaturas médias dos meses de janeiro (TJA) e julho (TJU), em °C.

	n	Média	Mediana	Mínimo	Máximo	Variância	DP	Assimetria	Curtose	CV
A	33	600,76	585,00	40,00	1100,00	80190,91	283,18	-0,180101	-0,442507	0,471367
TJA	33	23,37	23,90	20,20	25,70	2,76	1,66	-0,608226	-0,974465	0,070966
TJU	33	15,53	16,00	11,60	18,10	3,39	1,84	-0,387803	-0,957059	0,118210

n → # de estações agrometeorológicas. DP → Desvio Padrão. CV → Coeficiente de Variação.

Os coeficientes de assimetria indicam um comportamento aproximadamente simétrico (leve assimetria negativa) e os coeficientes de curtose negativos indicam uma distribuição platicúrtica. Para distribuição normal (DeMoivre-Laplace-Gauss) os coeficientes são: 0 (zero) e 0 (zero), respectivamente (Snedecor e Cochran, 1967).

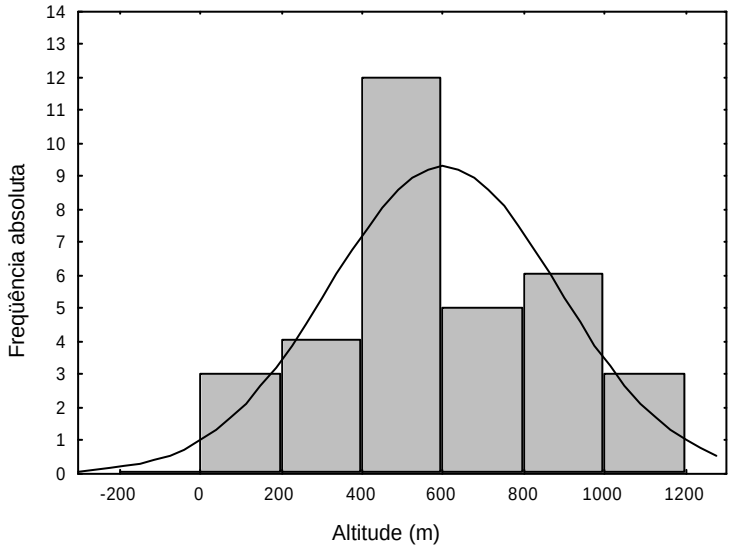

Figura 14 - Histograma e curva normal esperada da altitude.

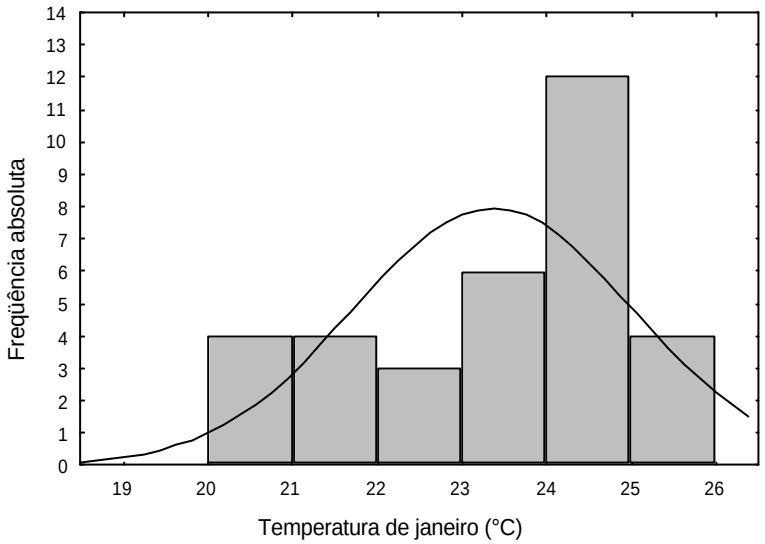

Figura 15 - Histograma e curva normal esperada da temperatura de janeiro.

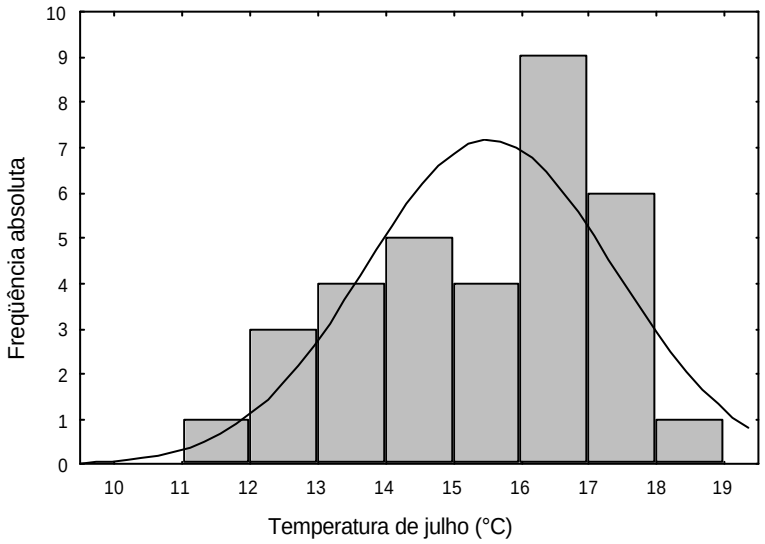

Figura 16 - Histograma e curva normal esperada da temperatura de julho.

No conjunto de dados não se encontrou valores atípicos e nem foram detectados sub-regiões que exigissem análises separadas (uma para cada sub-região). Pelo fato já comentado, a escassez de dados é um fator limitante a análises mais elaboradas.

Quadro 4 - Coeficientes de correlação de Pearson das temperaturas dos meses de janeiro a dezembro com as variáveis X, Y e A.

	Coordenada leste (X)	Coordenada norte (Y)	Altitude (A)
JAN	**-0,2193**	**0,4518**	**-0,8602**
	p=0,220	**p=0,008**	**p=0,000**
FEV	-0,0868	0,5235	-0,8873
	p=0,631	p=0,002	p=0,000
MAR	-0,1557	0,5798	-0,8376
	p=0,387	p=0,000	p=0,000
ABR	-0,0683	0,6478	-0,8231
	p=0,706	p=0,000	p=0,000
MAI	-0,0158	0,6696	-0,8045
	p=0,930	p=0,000	p=0,000
JUN	-0,0310	0,7191	-0,7445
	p=0,864	p=0,000	p=0,000
JUL	**-0,1028**	**0,7328**	**-0,6860**
	p=0,569	**p=0,000**	**p=0,000**
AGO	-0,2223	0,7718	-0,5787
	p=0,214	p=0,000	p=0,000
SET	-0,2345	0,7582	-0,6290
	p=0,189	p=0,000	p=0,000
OUT	-0,3047	0,7053	-0,6553
	p=0,085	p=0,000	p=0,000
NOV	-0,2836	0,6352	-0,7311
	p=0,110	p=0,000	p=0,000
DEZ	-0,2730	0,4852	-0,8177
	p=0,124	p=0,004	p=0,000

No Quadro 4 os valores de $p<0,01$ indicam que as temperaturas médias dos meses (janeiro a dezembro) são significativamente associadas com a coordenada Y (latitude) e com a altitude (A).

A associação da temperatura média do ar com a latitude motivou o uso da Krigagem Universal, Seção 4.2.2. E a associação da temperatura média do ar com a altitude foi utilizada no Modelo de Regressão Linear Múltipla (RLM) feito pelo IAPAR.

4.2 Aplicação dos Métodos de Predição

4.2.1 Krigagem Ordinária (KO)

Para que ao menos a hipótese intrínseca possa ser aceita com segurança, mais atenção deve ser dada ao comportamento dos dados na análise exploratória. Por ora, vamos supor que seja plausível assumir estacionariedade intrínseca.

Realizou-se a predição por KO para que comparações possam ser feitas, posteriormente, com os resultados obtidos pela Krigagem Universal.

4.2.1.1 Dados de temperatura de janeiro

O semivariograma para os dados de temperatura de janeiro (TJA), usando-se o modelo esférico foi ajustado (IGF = 0,011801) com os seguintes parâmetros: efeito pepita de 0,83691; variância espacial de 2,15991 e alcance de 310000 metros.

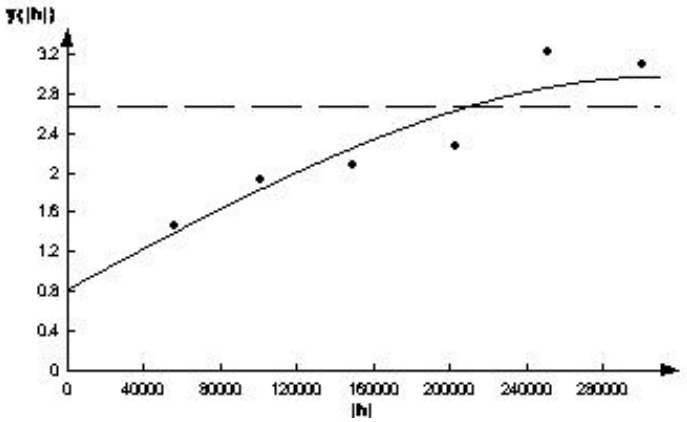

Figura 17 - Semivariograma dos dados de TJA, modelo esférico.

O semivariograma para os dados de temperatura de janeiro (TJA), usando-se o modelo gaussiano foi ajustado (IGF = 0,012885) com os seguintes parâmetros: efeito pepita de 1,16091; variância espacial de 2,15991 e alcance de 310000 metros.

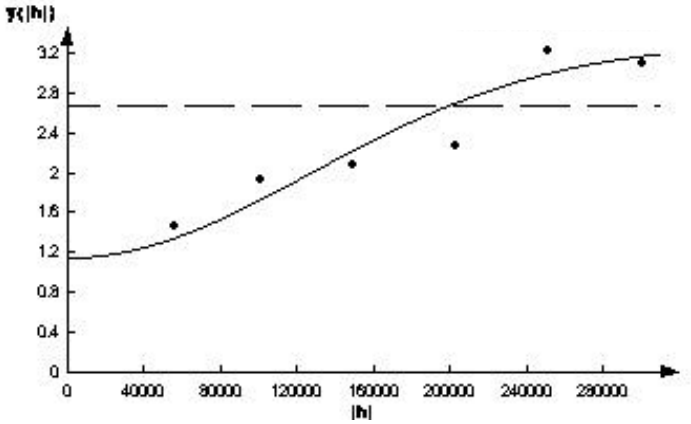

Figura 18 - Semivariograma dos dados de TJA, modelo gaussiano.

O alcance (a = 310000 m) dos semivariogramas para o mês de janeiro foi de aproximadamente metade da maior distância existente no campo, ou seja, a maior distância entre as estações agrometeorológicas do IAPAR.

As semivariâncias são calculadas a cada 50000 m (chamado de "lag").

Os valores do efeito pepita foram de $c_0 = 0{,}83691$ e $c_0 = 1{,}16091$ para os modelos esférico e gaussiano, respectivamente, indicando que:

$$\frac{c_0}{c_0 + c} = 0{,}2793 \cong 28\% \text{ (modelo esférico) e,}$$

$$\frac{c_0}{c_0 + c} = 0{,}3496 \cong 35\% \text{ (modelo gaussiano),}$$

da variância total é causada por variações aleatórias segundo Journel & Huijbregts, 1975, citado por Guerra (1989, p55).

Os autores indicam que quando esta proporção for superior a 0,8, a Estatística Clássica e a Geoestatística não se diferenciam, ou seja, a dependência espacial é desprezível.

O patamar $(c_0 + c)$ representa a variação total da variável TJA e é dado por:

$c_0 + c = 2{,}99682$ (modelo esférico) e,

$c_0 + c = 3{,}32082$ (modelo gaussiano).

Aplicando-se a técnica de Validação Cruzada para os dados de temperatura média do mês de janeiro, obteve-se para os erros, eq. (66), média (desvio padrão) de –0,037 (0,852) e –0,034 (0,864) para os modelos esférico e gaussiano, respectivamente. Considerando-se esses resultados, os dois modelos, esférico e gaussiano, representam bem a dependência espacial existente na temperatura média do mês de janeiro.

Os semivariogramas construídos para o mês de janeiro (Figuras 17 e 18) mostram-se limitados. Isto pode sugerir que os dados sejam estacionários e, assim, a hipótese de estacionariedade intrínseca, inicialmente assumida, é válida permitindo a realização da predição por Krigagem Ordinária.

Os mapas construídos com os modelos esférico e gaussiano através da KO, para a temperatura média do mês de janeiro são mostrados nas Figuras 19 e 20, respectivamente.

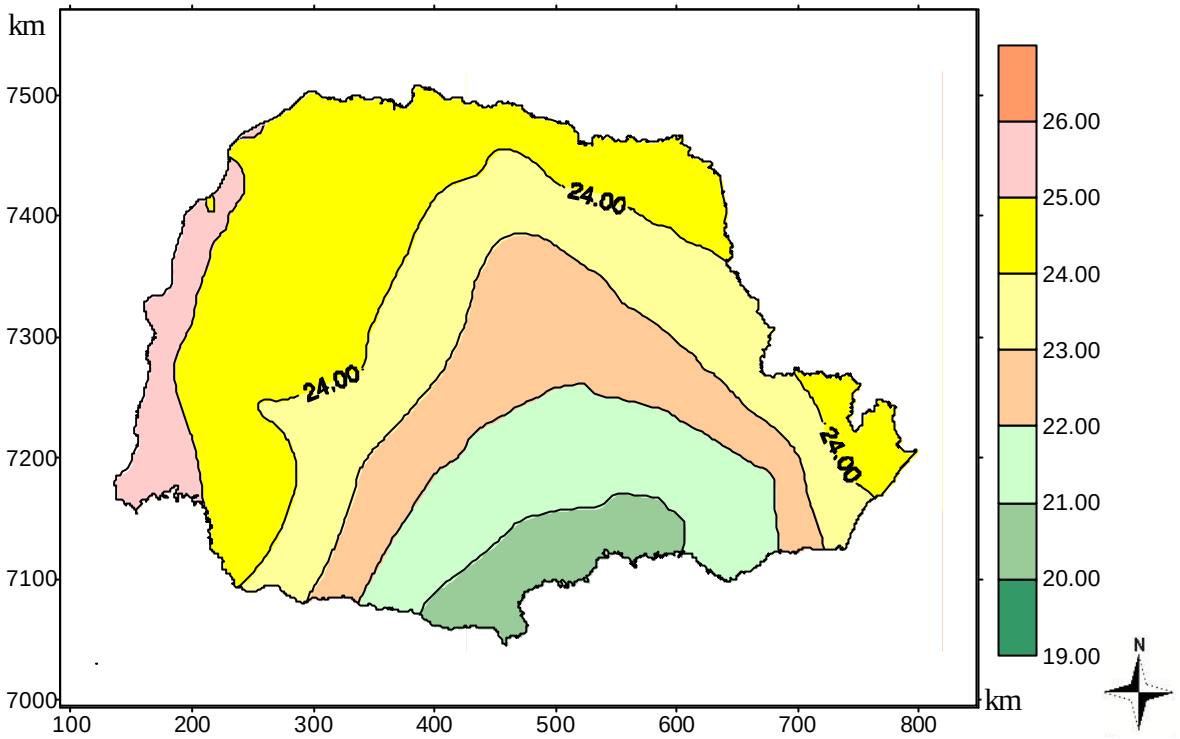

Figura 19 - Mapa de temperatura média do mês de janeiro (°C), modelo esférico, KO.

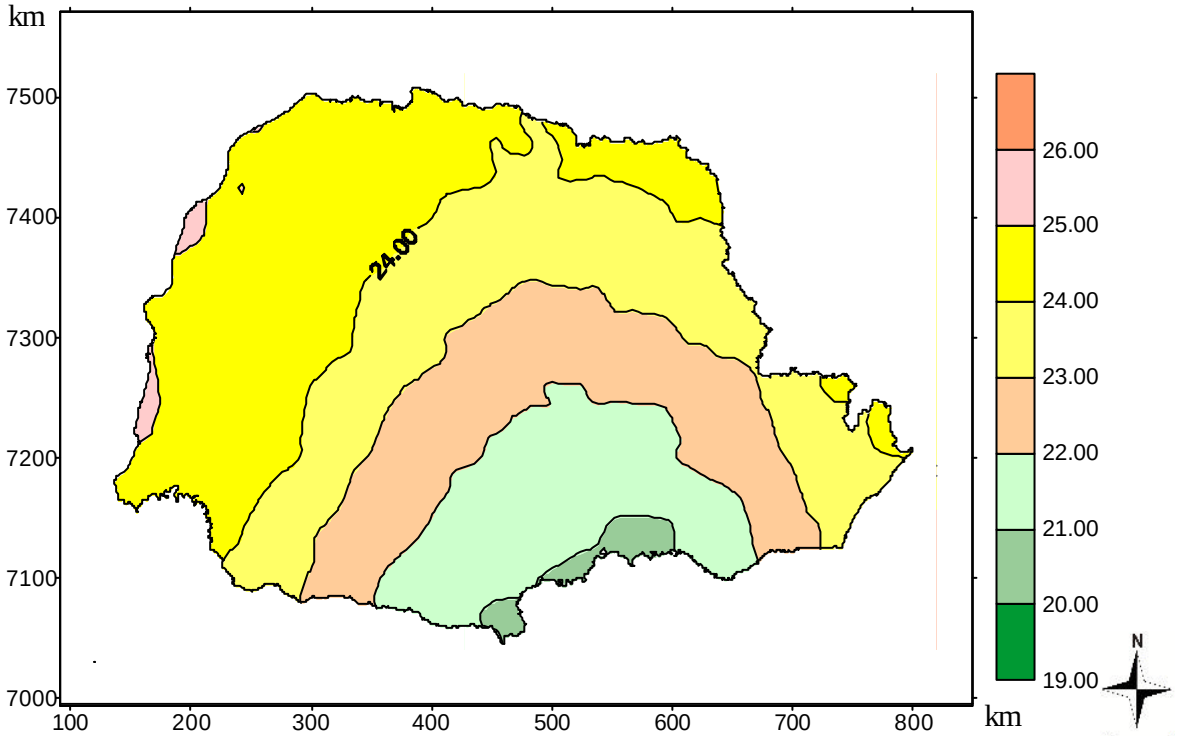

Figura 20 - Mapa de temperatura média do mês de janeiro (°C), modelo gaussiano, KO.

4.2.1.2 Dados de temperatura de julho

O semivariograma para os dados de temperatura de julho (TJU), usando-se o modelo esférico foi ajustado (IGF = 0,021012) com os seguintes parâmetros: efeito pepita de 0,165; variância espacial de 3,3 e alcance de 280000 metros.

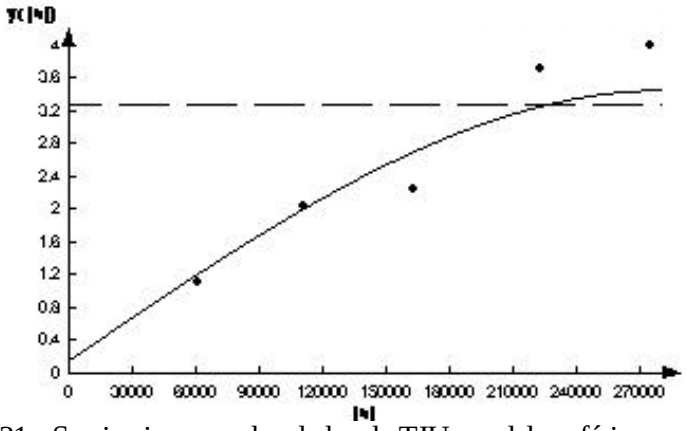

Figura 21 - Semivariograma dos dados de TJU, modelo esférico.

O semivariograma para os dados de temperatura de julho (TJU), usando-se o modelo gaussiano foi ajustado (IGF = 0,012844) com os seguintes parâmetros: efeito pepita de 0,65967; variância espacial de 3,3 e alcance de 280000 metros.

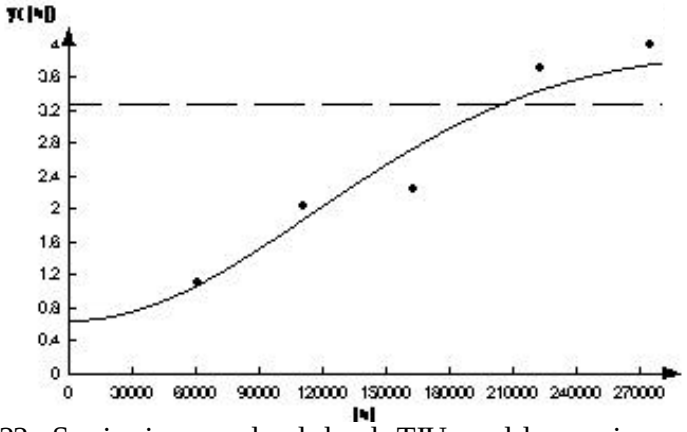

Figura 22 - Semivariograma dos dados de TJU, modelo gaussiano.

O alcance (a = 280000 m) dos semivariogramas para o mês de julho foi de aproximadamente metade da maior distância existente no campo, ou seja, a maior distância entre as estações agrometeorológicas do IAPAR. Este valor é um pouco menor que o obtido para o mês de janeiro.

As semivariâncias são calculadas a cada 55000 m (chamado de "lag").

Os valores do efeito pepita foram de $c_0 = 0{,}165$ e $c_0 = 0{,}65967$ para os modelos esférico e gaussiano, respectivamente, indicando que:

$$\frac{c_0}{c_0 + c} = 0{,}0476 \cong 5\% \text{ (modelo esférico) e,}$$

$$\frac{c_0}{c_0 + c} = 0{,}1666 \cong 17\% \text{ (modelo gaussiano),}$$

da variância total é causada por variações aleatórias segundo Journel & Huijbregts (1975), citado por Guerra (1989, p55).

O patamar $(c_0 + c)$ representa a variação total da variável TJU e, é dado por:

$$c_0 + c = 3{,}465 \text{ (modelo esférico) e,}$$

$$c_0 + c = 3{,}959 \text{ (modelo gaussiano).}$$

Aplicando-se a técnica de Validação Cruzada para os dados de temperatura média do mês de julho, obteve-se para os erros, eq. (66), média (desvio padrão) de –0,013 (0,865) e –0,012 (0,887) para os modelos esférico e gaussiano, respectivamente. Considerando-se esses resultados, os dois modelos, esférico e gaussiano, representam bem a dependência espacial existente na temperatura média do mês de julho.

Os semivariogramas construídos para o mês de julho (Figuras 21 e 22) mostram-se limitados. Isto pode sugerir que os dados sejam estacionários e, assim, a hipótese de estacionariedade intrínseca, inicialmente assumida, é válida permitindo a realização da predição por Krigagem Ordinária.

A tendência relacionada com a latitude (coordenada Y), não é aparente nos semivariogramas. Isto pode ter acontecido devido à escassez de dados de temperatura, no Estado do Paraná, que levou à necessidade de construir semivariogramas com ângulos de tolerância de 90°C e, com este ângulo, os semivariogramas direcionais (por exemplo, 0°C, 25°C, 45°C, 90°C e 135°C) foram todos iguais para os dois meses estudados.

Portanto, os semivariogramas foram considerados isotrópicos, além disso, com poucas amostras não seria possível modelar um comportamento anisotrópico, caso este fosse confirmado.

Os mapas construídos com os modelos esférico e gaussiano através da KO para a temperatura média do mês de julho são mostrados nas Figuras 23 e 24, respectivamente.

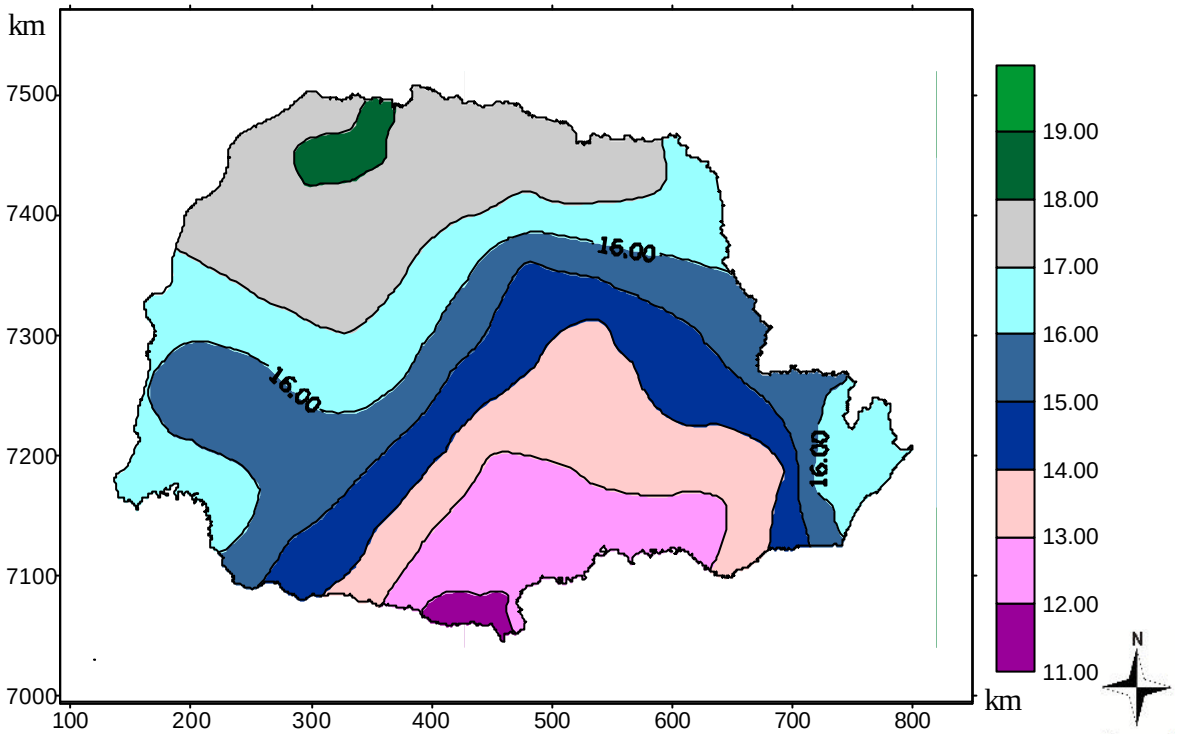

Figura 23 - Mapa de temperatura média do mês de julho (°C), modelo esférico, KO.

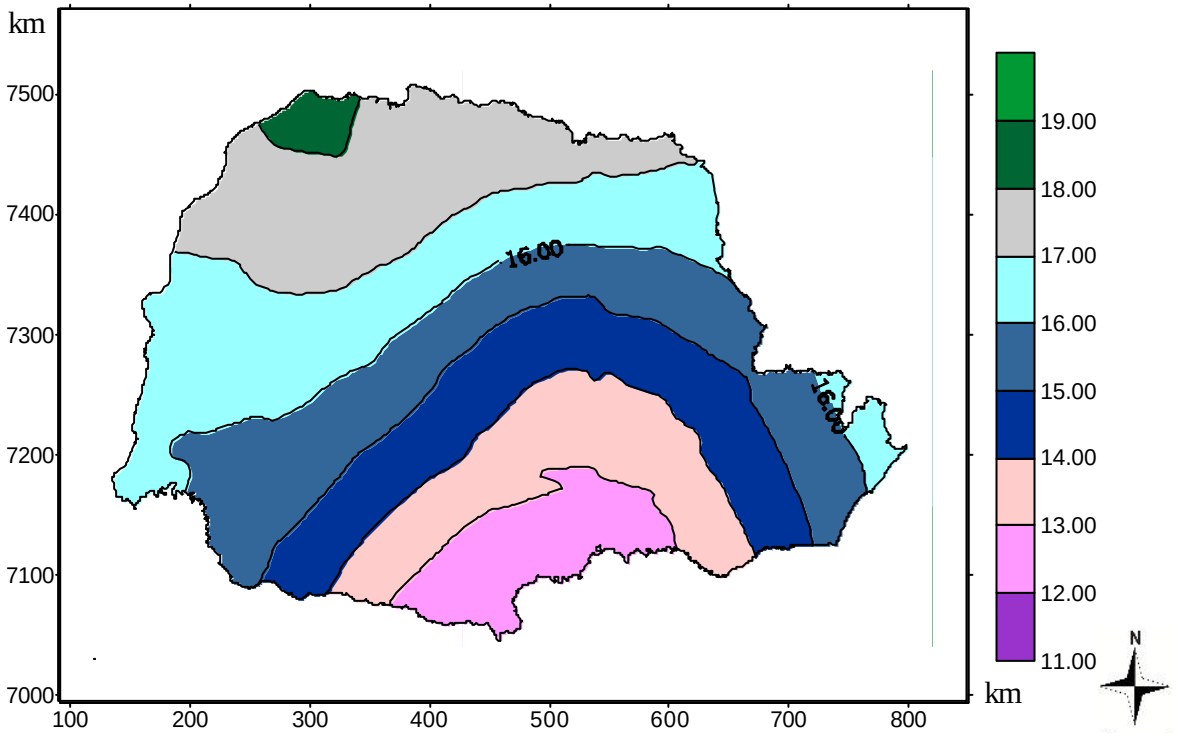

Figura 24 - Mapa de temperatura média do mês de julho (°C), modelo gaussiano, KO.

4.2.2 Krigagem Universal (KU)

Apresenta-se a seguir os resultados obtidos com a Krigagem Universal.

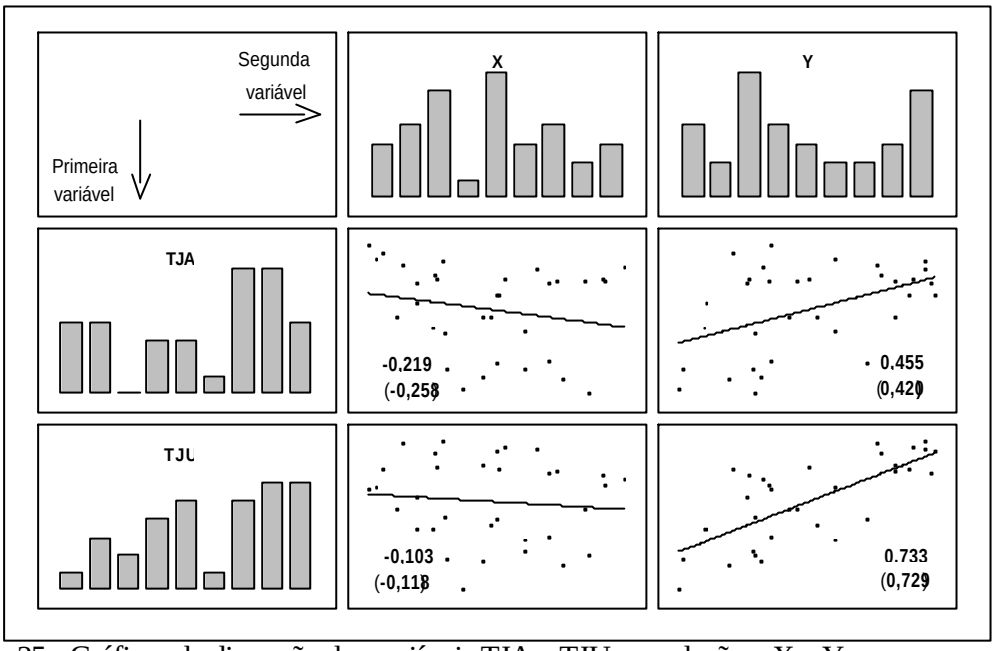

Figura 25 - Gráficos de dispersão das variáveis TJA e TJU em relação a X e Y.

Como mostra a Figura 25, as variáveis TJA e TJU são associadas principalmente com a coordenada Y, constatações feitas pelos valores do coeficiente de correlação de Pearson e o coeficiente de correlação de ordem de Spearman (entre parênteses).

As temperaturas de janeiro e julho aumentam com o aumento da coordenada Y, tendo um coeficiente de correlação de Pearson (Spearman) de 0,455 (0,420) para o mês de janeiro e 0,733 (0,729) para o mês de julho. Neste caso, diz-se que existe uma tendência nessa direção (também chamado de "drift" em geoestatística).

As correlações existentes na direção da coordenada X, para as variáveis TJA e TJU, não foram consideradas significativas (ver Quadro 4 e Figura 25). Logo, segundo descrito na Seção 2.4.3, esta foi a direção considerada livre de tendência, na qual os semivariogramas foram construídos.

Logo, tomando-se a direção X livre de tendência, o semivariograma encontrado em tal direção foi usado para representar o semivariograma dos resíduos na KU (Seção 2.4.3). Estes semivariogramas são os mesmos encontrados nas Figuras 17 e 18 para o mês de janeiro e Figuras 21 e 22 para o mês de julho. Isto porque todos os semivariogramas foram obtidos considerando-se direção 0°C e ângulo de tolerância de 90°C, devido à escassez de dados de temperatura, fato recomendado na literatura por autores como Cressie (1993), Goovaerts (1997) e outros.

Os mapas construídos com os modelos esférico e gaussiano para a temperatura do mês de janeiro, através da KU com modelo de tendência linear (KU-L), são mostrados nas Figuras 26 e 27, respectivamente. Os construídos para a temperatura do mês de julho, nas Figuras 28 e 29, respectivamente.

Nos Quadros 5 e 6 são apresentados os valores do quadrado das diferenças (Δ) entre os valores observados da variável (TJA e TJU, respectivamente) e seus respectivos valores preditos pela krigagem ordinária (KO), krigagem universal com modelo de tendência linear (KU-L) e pelo quadrado dos resíduos do modelo de regressão linear múltipla (RLM) dado pela equação (67).

4.2.2.1 Dados de temperatura de janeiro

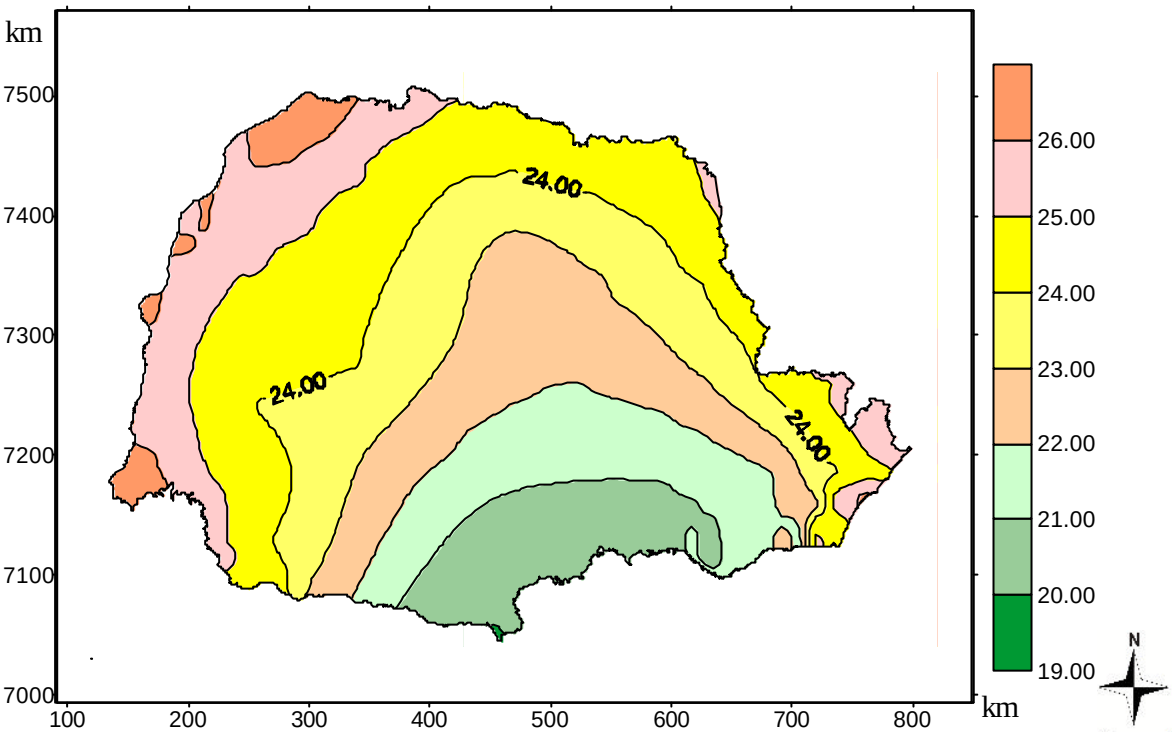

Figura 26 - Mapa de temperatura média do mês de janeiro (°C), modelo esférico, KU – L.

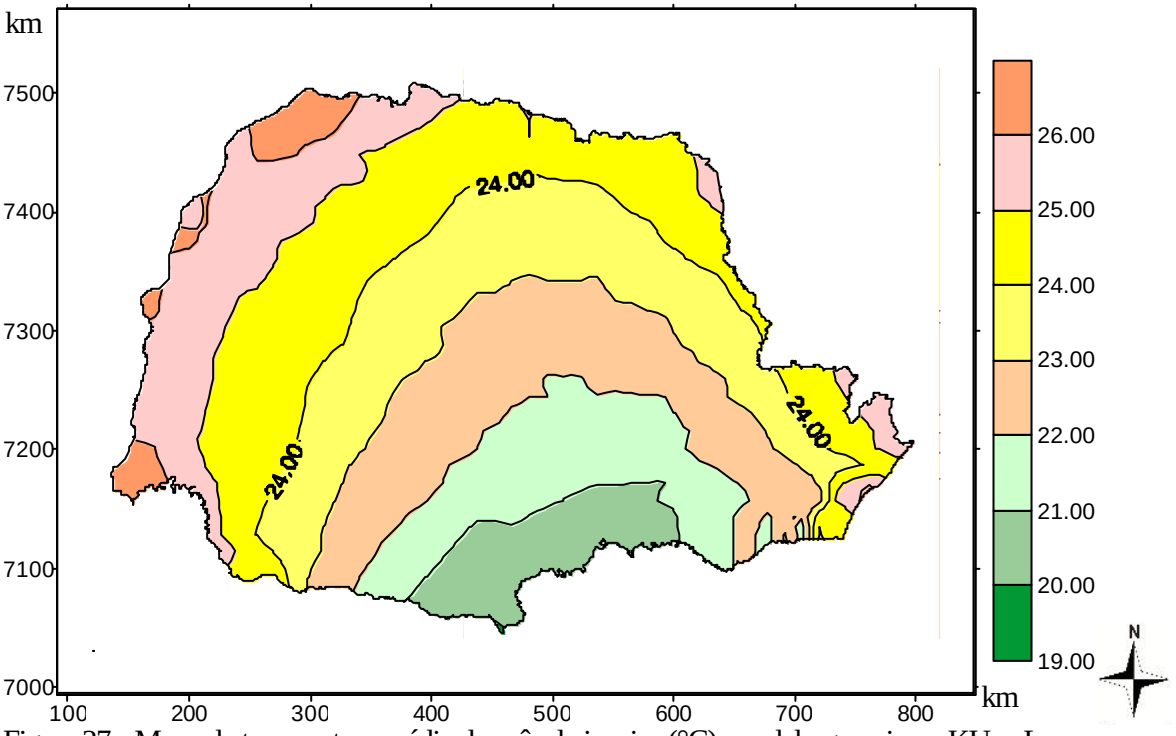

Figura 27 - Mapa de temperatura média do mês de janeiro (°C), modelo gaussiano, KU – L.

4.2.2.2 Dados de temperatura de julho

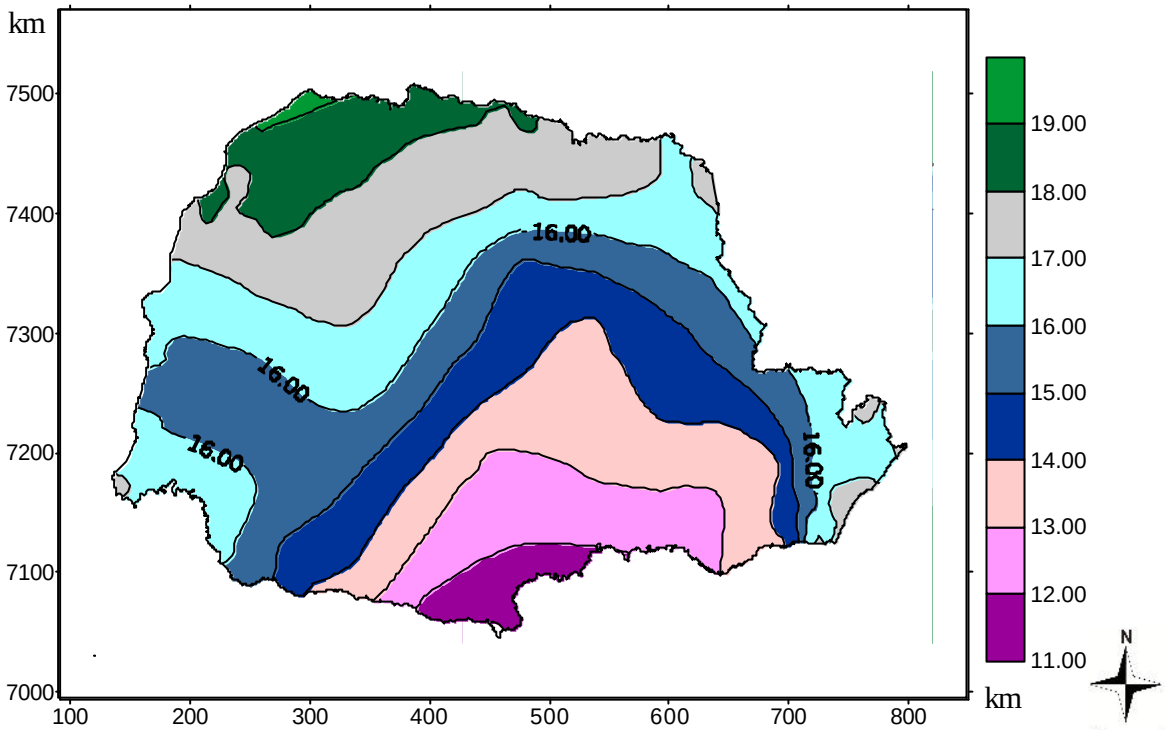

Figura 28 - Mapa de temperatura média do mês de julho (°C), modelo esférico, KU - L.

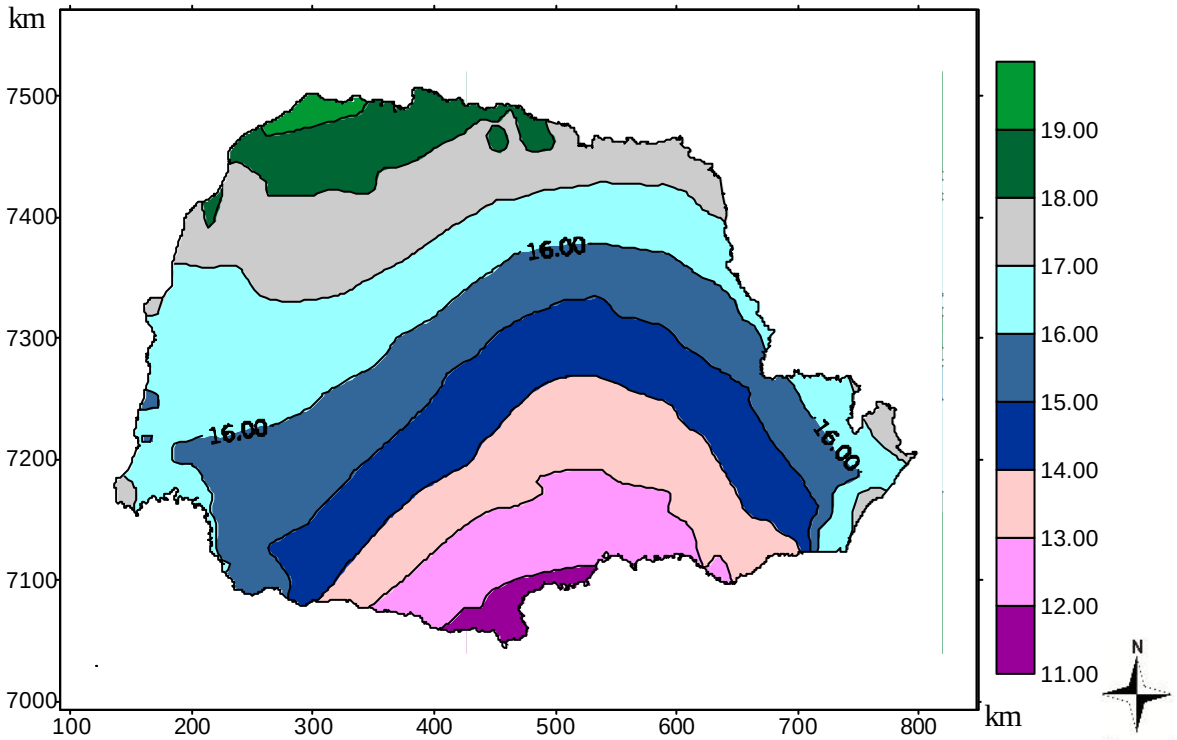

Figura 29 - Mapa de temperatura média do mês de julho (°C), modelo gaussiano, KU - L.

Quadro 5 - Quadrado das diferenças entre valores observados e preditos, TJA.

TEMPERATURA DE JANEIRO (TJA)				
MODELO ESFÉRICO		MODELO GAUSSIANO		
KO	KU-L	KO	KU-L	RLM(*)
0,303963	0,116965	0,899786	0,212604	0,276192
0,074099	0,086361	0,372368	0,377697	0,010883
0,136540	0,052988	0,408533	0,054661	0,098657
0,037516	0,450054	0,027571	1,060842	0,024351
0,012887	0,034484	0,004887	0,058083	0,003876
0,301487	0,308591	0,387731	0,436008	0,093431
0,885043	0,924765	1,908680	2,213897	0,349783
0,667775	0,267335	1,683597	0,678645	0,421885
0,006996	0,000658	0,098344	0,011071	0,000005
0,106439	0,019472	0,420033	0,053945	0,040738
0,001985	0,000547	0,100865	0,058830	0,032130
0,039541	0,011683	0,225868	0,057295	0,247591
0,389550	0,356430	1,140891	1,044654	0,000412
0,236914	0,033790	1,247358	0,003240	0,006406
0,301047	0,253430	0,474504	0,273775	0,004063
0.032639	0,018661	0,123585	0,056682	0,391264
0,218800	0,061464	0,970621	0,556798	0,003583
0,075789	0,074112	0,124716	0,146216	0,149624
0,009462	0,005854	0,013314	0,000578	0,003850
1,456962	1,377975	3,474291	3,423370	0,056359
1,101998	1,321284	1,680495	1,704704	0,099893
0,275132	0,283320	0,979299	0,938048	0,156028
0,317615	0,027296	1,115542	0,107998	0,000139
0,053922	0,000795	0,102162	0,000908	0,493256
0,122175	0,013301	0,361162	0,025971	0,002533
0,003811	0,007944	0,001047	0,019840	0,011569
4,217400	3,884084	7,014605	6,820110	0,146655
0,279153	0,028545	1,146872	0,103457	0,384517
0,045837	0,039913	0,199989	0,125366	0,025390
0,201289	0,209977	0,627098	0,510849	0,073673
0,234348	0,071103	0,719255	0,188114	0,019788
0,001120	0,000173	0,059606	0,034640	0,009721
0,013800	0,004862	0,089682	0,005122	0,005001
0,368577	**0,313582**	**0,854678**	**0,647394**	**0,110401**

Quadro 6 - Quadrado das diferenças entre valores observados e preditos, TJU.

TEMPERATURA DE JULHO				
MODELO ESFÉRICO		MODELO GAUSSIANO		
KO	KU-L	KO	KU-L	RLM(*)
0,024543	0,015505	0,429888	0,171422	0,020021
0,027729	0,025765	0,068091	0,055227	0,696841
0,026161	0,023217	0,355371	0,170437	0,076073
0,000044	0,015134	0,002947	0,272577	0,158569
0,035154	0,048555	0,176315	0,249061	0,648105
0,005370	0,004550	0,001122	0,000676	1,335523
0,059303	0,059086	0,791663	0,825930	0,138177
0,001736	0,001961	0,000293	0,095921	0,299179
0,000591	0,000342	0,026955	0,006395	0,180300
0,000285	0,005295	0,030355	0,017007	0,026563
0,002513	0,004414	0,034190	0,060820	0,054688
0,034013	0,028574	0,132376	0,137565	0,000186
0,037987	0,039474	0,624658	0,624982	0,124124
0,000026	0,023003	0,140671	0,025995	0,089080
0,164910	0,170823	0,747459	0,593606	0,289463
0,011538	0,002179	0,075154	0,001804	0,008681
0,049962	0,027709	0,922661	0,565837	0,048959
0,000285	0,000428	0,022050	0,016288	0,662543
0,072034	0,064786	0,016715	0,031765	0,006579
0,099204	0,083942	1,152108	1,034315	0,001037
0,365951	0,371231	1,796310	1,625753	0,295675
0,025267	0,029936	0,997023	1,027002	0,828676
0,075801	0,025417	1,105650	0,350452	0,112032
0,006262	0,007933	0,138114	0,144987	2,474029
0,003823	0,000250	0,056873	0,009413	0,027296
0,028219	0,024806	0,100701	0,157484	0,089681
0,567166	0,533197	2,921816	2,742058	0,025138
0,088910	0,074423	1,291582	0,948278	1,142690
0,017179	0,013327	0,078002	0,087571	0,105195
0,008051	0,010263	0,146143	0,147170	0,081225
0,002137	0,001322	0,000271	0,000003	0,976447
0,030366	0,032384	0,676149	0,655226	1,626419
0,022343	0,014773	0,409116	0,297318	0,281674
0,057420	**0,054061**	**0,468751**	**0,398495**	**0,391845**

(*) Valores do quadrado dos resíduos da regressão linear múltipla (RLM) obtidos pela equação: $T_i = \alpha + \beta_1 X_i + \beta_2 Y_i + \beta_3 A_i + \xi_i$, (67)

onde: i = 1,...,33; T_i é a temperatura média mensal; X_i é a coordenada leste (longitude); Y_i é a coordenada norte (latitude); A_i é a altitude e ξ_i é o erro aleatório.

Na última linha dos Quadros 5 e 6 (em negrito), tem-se a média do quadrado das diferenças $\Delta_{(\text{MÉDIA})}$ obtidos pela expressão:

$$\Delta_{(\text{MÉDIA})} = \frac{1}{33}\sum_{\alpha=1}^{33}\left[z(\mathbf{u}_\alpha) - z^*_G(\mathbf{u}_\alpha)\right]^2 , \qquad (68)$$

onde G representa as predições por KO e KU-L e RLM.

Nos Quadros 5 e 6 observa-se que o menor valor da média do quadrado das diferenças $\left(\Delta_{(\text{MÉDIA})}\right)$ é obtido quando utiliza-se o método de interpolação por Krigagem Universal com modelo de tendência linear (KU-L) e o semivariograma ajustado pelo modelo esférico. Observa-se, também, que para o mês de julho os métodos de interpolação dados pela geoestatística foram mais precisos que o modelo de Regressão Linear Múltipla (RLM) utilizado pelo IAPAR.

Mesmo com o modelo gaussiano apresentando um ajuste um pouco melhor aos semivariogramas experimentais, o modelo esférico foi preferido por representar melhor a realidade, pois tem um comportamento linear na origem.

O modelo de RLM, eq. (67), usado pelo IAPAR, considera a variável altitude na predição dos valores de temperatura em locais não medidos. Como a temperatura média do mês de janeiro tem uma associação mais forte com a altitude (-0,86) que a temperatura média do mês de julho com altitude (-0,69), o modelo de RLM para o mês de janeiro se ajustou melhor às temperaturas médias que o mês de julho (0,110401 e 0,391845, respectivamente).

Também é percebido, no ajuste dos semivariogramas, que a temperatura média para o mês de julho apresentou uma dependência espacial mais forte que a do mês de janeiro.

Assim, predizendo-se os valores de temperatura média para os 399-2 = 397 municípios do Estado do Paraná (duas localizações de municípios coincidem com as localizações das estações do IAPAR), pela Krigagem Universal com modelo de tendência linear (KO-L), utilizando-se o modelo esférico ajustado ao semivariograma, mais os valores observados de temperatura média das 33 estações agrometeorológicas, construiu-se os mapas das Figuras 32 e 33, para os meses de janeiro e julho, respectivamente.

No Quadro 7, tem-se as estatísticas dos 430 valores de altitude (A) (397 municípios do Estado do Paraná, mais 33 estações agrometeorológicas) e dos valores preditos de temperatura média de janeiro pela KU-L, utilizados na construção do mapa da Figura 32.

Quadro 7 - Estatísticas dos valores de altitude (A) e valores preditos de temperatura média do mês de janeiro ($TJA_{preditos}$).

	n	Média	Mínimo	Máximo	DP	Assimetria	Curtose
A	430	615,6163	3	1200	223,3454	0,036412	0,10946
$TJA_{preditos}$	430	23,80075	20,2	27,56536	1,429017	-0,56237	-0,10116

n→ (33 estações + 397 municípios do Estado do Paraná). DP→ Desvio Padrão.

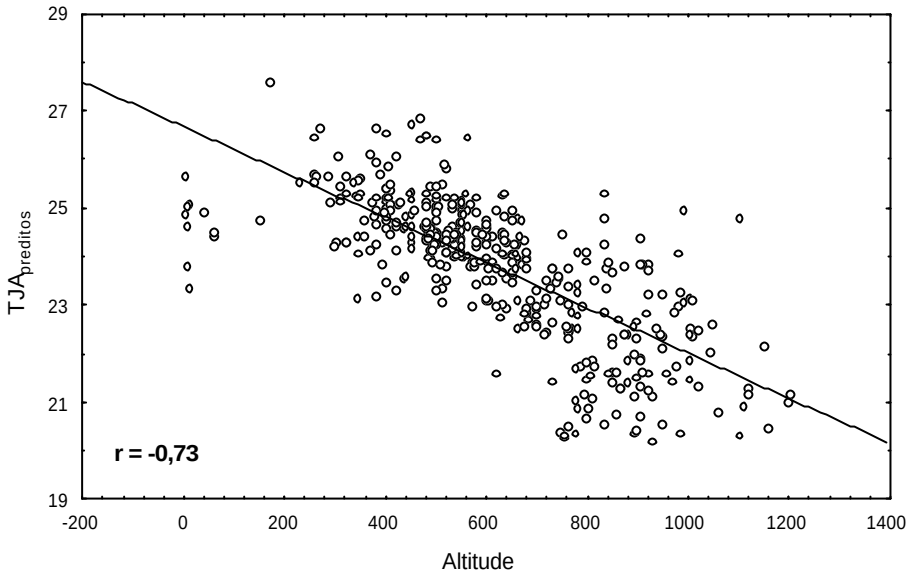

Figura 30 - Gráfico de dispersão entre $TJA_{preditos}$ (°C) e altitude (m).

Comparando os valores de mínimo e máximo de $TJA_{preditos}$ do Quadro 7 com os respectivos valores no Quadro 3, observa-se que estes limites são próximos, significando que o método não estimou valores muito discrepantes. Observa-se, também, na Figura 30 que os valores de $TJA_{preditos}$ têm uma associação negativa com altitude de –0,73.

Não necessariamente as localizações de referência dos 399 municípios do Paraná devem ser usadas para a obtenção de valores estimados de temperatura. Quaisquer localizações poderiam ter sido usadas.

No Quadro 8, tem-se as estatísticas dos valores de altitude (A) e dos valores preditos de temperatura média de julho pela KU-L, utilizados na construção do mapa da Figura 33.

Quadro 8 - Estatísticas dos valores de altitude (A) e valores preditos de temperatura média do mês de julho ($TJU_{preditos}$).

	n	Média	Mínimo	Máximo	DP	Assimetria	Curtose
A	430	615,6163	3	1200	223,3454	0,0364118	0,10946
$TJU_{preditos}$	430	16,02964	11,41384	19,08132	1,719664	-0,562785	-0,41397

n→ (33 estações + 397 municípios do Estado do Paraná). DP→ Desvio Padrão.

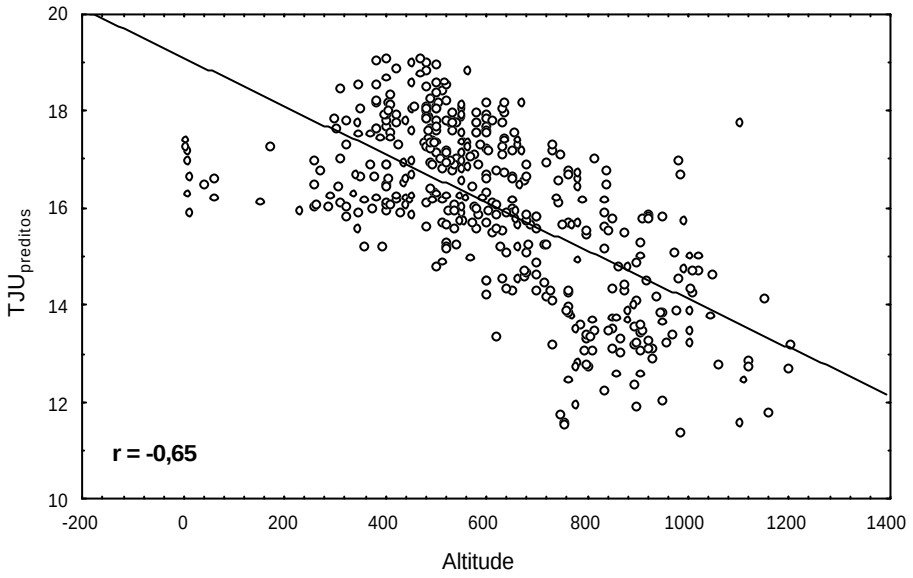

Figura 31 - Gráfico de dispersão entre $TJU_{preditos}$ (°C) e altitude (m).

Comparando os valores de mínimo e máximo de $TJU_{preditos}$ do Quadro 8 com os respectivos valores no Quadro 3, observa-se que estes limites são próximos, significando que o método não estimou valores muito discrepantes. Observa-se, também, na Figura 31, que os valores de $TJU_{preditos}$ têm uma associação negativa com altitude de –0,65.

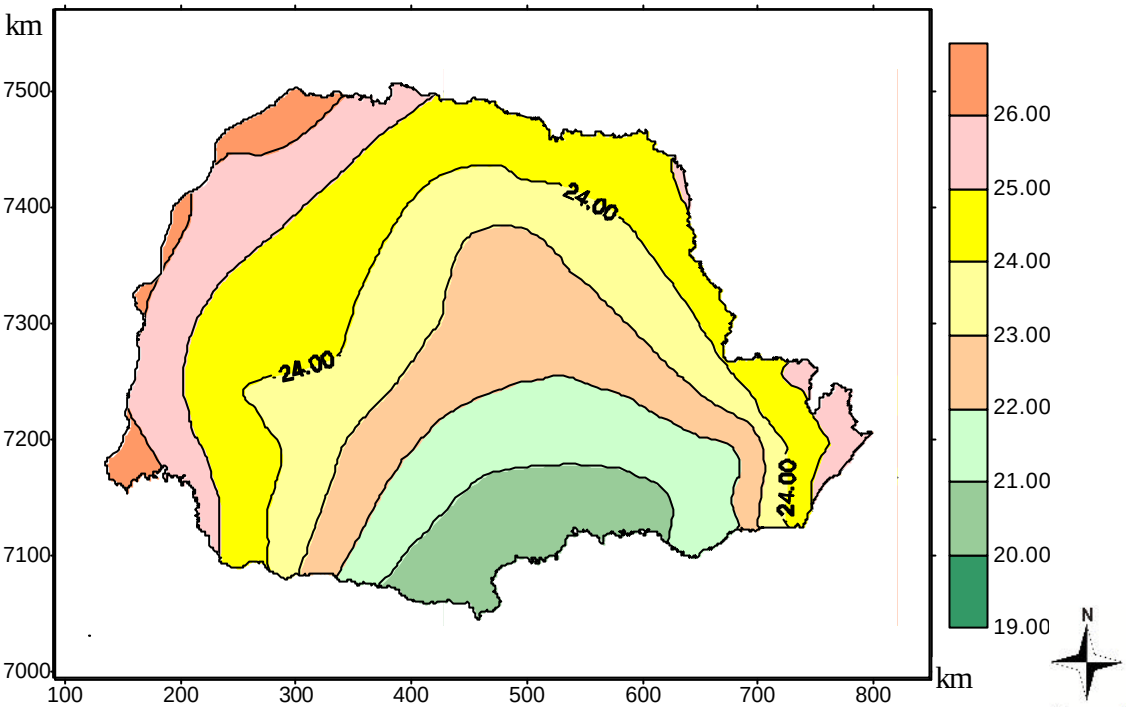

Figura 32 - Mapa de temperatura média do mês de janeiro (°C) – TESE/2002.

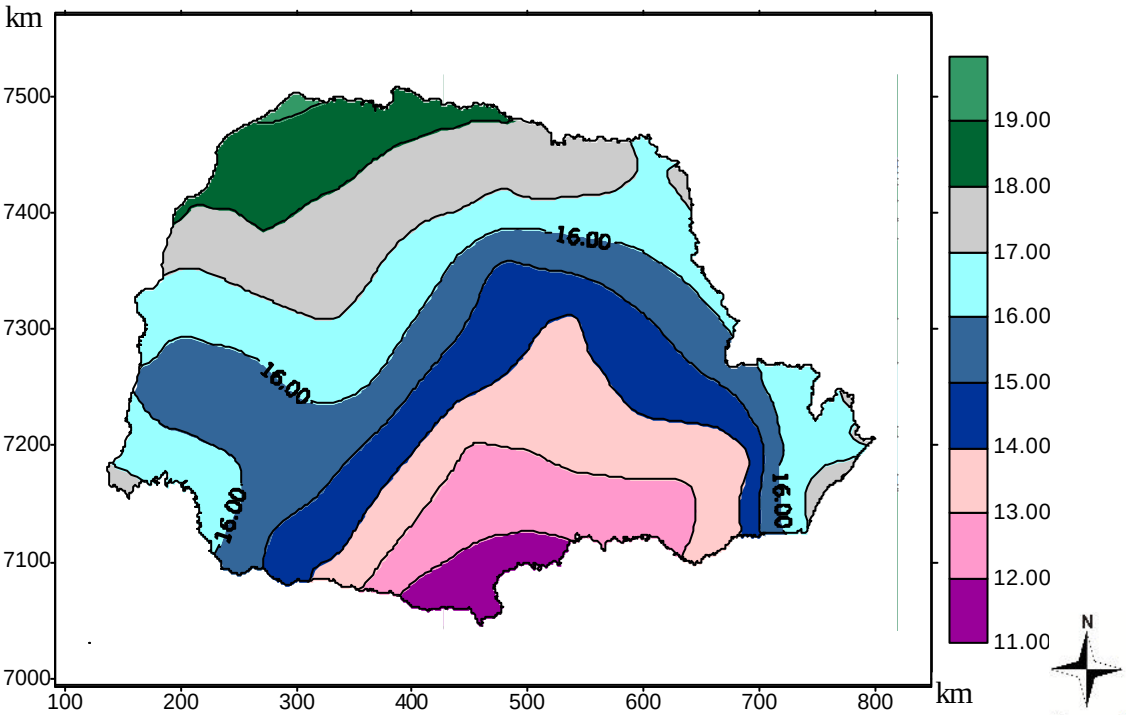

Figura 33 - Mapa de temperatura média do mês de julho (°C) – TESE/2002.

Embora considerando que os mapas construídos até aqui (pelas técnicas geoestatísticas) não utilizaram a variável altitude nas análises e foram confeccionados com apenas 33 dados de temperatura média mensal, houve semelhança (e em alguns casos foram mais precisos) com os obtidos pelo IAPAR (Figuras 34 e 35), que usou um modelo de regressão linear múltipla, eq. (67), envolvendo a variável altitude.

Os mapas das Figuras 34 e 35 foram obtidos pelo Instituto Agronômico do Paraná – IAPAR, através do modelo de regressão linear múltipla, eq. (67), para a temperatura média dos meses de janeiro e julho, respectivamente. Estes e outros mapas podem ser vistos no CD–ROM: Cartas Climáticas do Paraná – IAPAR – Londrina, Caviglione *et al.* (2000).

4.2.3 Cartas Climáticas do Paraná

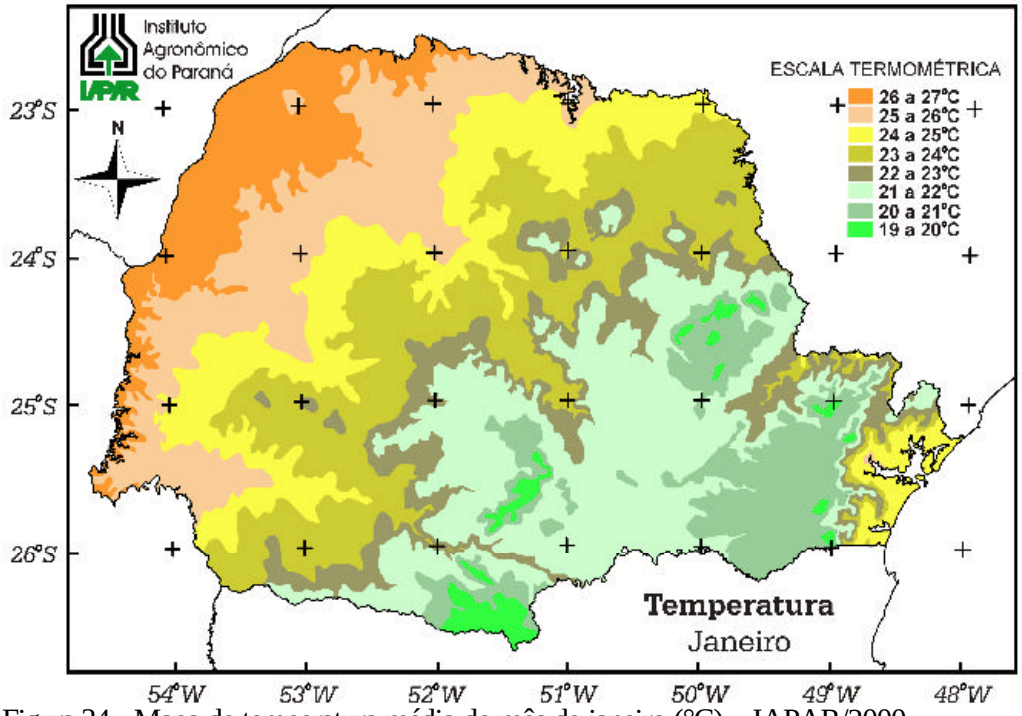

Figura 34 - Mapa de temperatura média do mês de janeiro (°C) – IAPAR/2000.

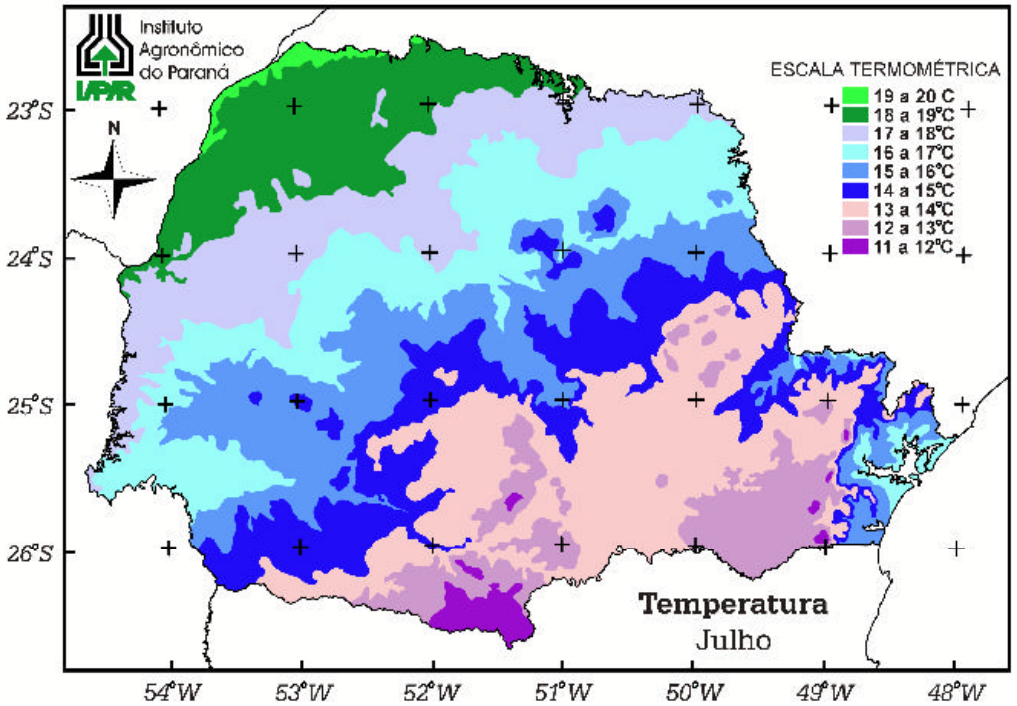

Figura 35 - Mapa de temperatura média do mês de julho (°C) – IAPAR/2000.

Sabe-se que a variável altitude interfere no comportamento da temperatura e esta variável foi utilizada no modelo de regressão linear múltipla, eq. (67), pelo IAPAR, para estimar valores de temperatura em locais não amostrados. Desta forma, muita importância foi dada à variável altitude, de maneira local, não levando em consideração a dependência espacial existente nos dados de temperatura, isto é, as associações entre pontos amostrais de acordo com a proximidade uns dos outros.

Observa-se na Figura 32 que, no litoral do Paraná, a KU atribuiu valores estimados de temperatura média do mês de janeiro entre 25°C e 26°C, o mesmo não ocorreu com a RLM, Figura 34. Observa-se, também, na Figura 33 que, no litoral do Paraná, a KU atribuiu valores estimados de temperatura média do mês de julho entre 17°C e 18°C, o mesmo não ocorreu com a RLM, Figura 35.

Os mapas de temperatura média mensal dos meses de janeiro (Figura 32) e julho (Figura 33) obtidos através das técnicas geoestatísticas, apresentaram contornos bem mais suaves que os obtidos pelo IAPAR através da RLM considerando a variável altitude (Figuras 34 e 35). Nas Figuras 34 e 35 observa-se, também, como a variável altitude influenciou os valores estimados de temperatura, pela semelhança entre os dois mapas e por não ter sido considerado a existência de dependência espacial.

4.2.4 Sugestão para Futuros Trabalhos

Até mesmo em livros básicos de Geografia do Brasil (Coelho (1992) e Vesentini & Vlach (1999)), pode-se constatar que a temperatura diminui com o aumento da altitude. Por existir associação entre temperatura e altitude (Figura 36), deve-se aproveitar deste fato para melhorar as predições de temperatura em localizações onde esta medida não é

disponível. Com a ajuda das medidas de altitude que são facilmente obtidas em qualquer localização por meio do GPS – Global Positioning Systen, ou mesmo utilizando um instrumento chamado *teodolito* (instrumento óptico para medir com precisão ângulos horizontais e ângulos verticais, muito usado em trabalhos topográficos e geodésicos) e as técnicas geoestatísticas que levam em conta a dependência espacial nas amostras, esta situação torna-se completamente resolvida.

Como um possível trabalho futuro, seria interessante realizar um estudo que levasse em conta a dependência espacial existente nos dados de temperatura, considerando a variável altitude como co-variável para melhorar, ainda mais, as estimativas de temperatura média mensal. A Krigagem com tendência externa - KTE (brevemente comentada na Seção 2.4.3) e a Cokrigagem são algumas das opções.

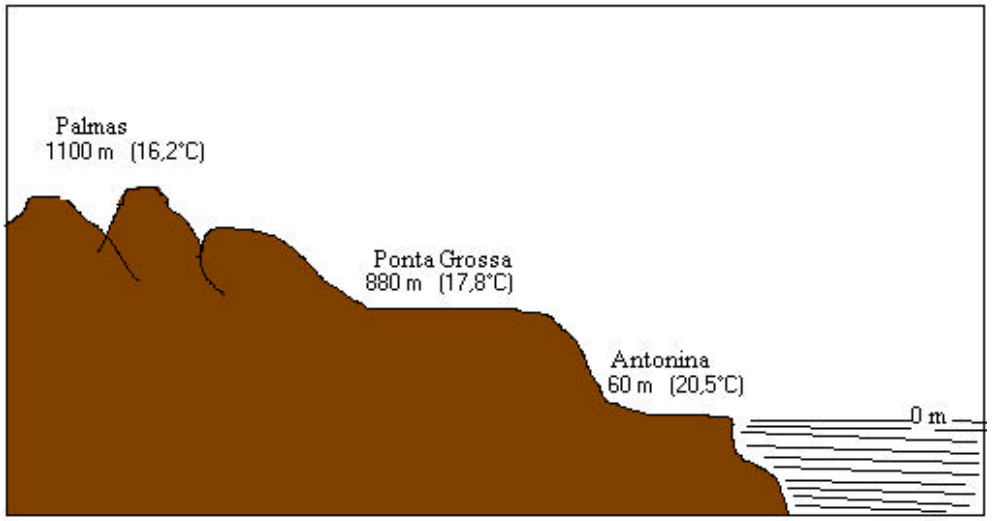

Figura 36 - Temperatura média anual diminuindo com o aumento da altitude.

Estas técnicas geoestatísticas não foram feitas, aqui, por este trabalho objetivar apresentar uma técnica eficiente e simples, que levasse em conta a estrutura espacial das amostras para se construir mapas de temperatura e por serem necessários programas computacionais específicos e análises mais elaboradas.

5 CONCLUSÕES

Os dados apresentados no presente trabalho remetem às seguintes considerações:

A existência de dependência espacial na temperatura média mensal foi confirmada pelos semivariogramas do mês de janeiro e, principalmente, pelos do mês de julho.

O modelo esférico se ajustou bem aos semivariogramas experimentais e foi o que melhor expressou a realidade da temperatura média dos meses de janeiro e julho, por ter um comportamento linear na origem.

O alcance de aproximadamente 310000 m, para o mês de janeiro, indicou que pontos amostrais separados por até esta distância são associados espacialmente. O mês de julho acusou um alcance um pouco menor, 280000 m.

As estimativas obtidas pela Krigagem Universal com modelo de tendência linear (KU-L) e o modelo esférico ajustado ao semivariograma experimental, apresentaram melhores resultados no sentido de proximidade com os respectivos valores observados. Para o mês de julho, estas estimativas foram melhores que as obtidas pela regressão linear múltipla feita pelo IAPAR usando a variável altitude.

Mesmo não utilizando a variável altitude nas técnicas geoestatísticas para auxiliar na estimativa de valores de temperatura média mensal, verificou-se melhoria na precisão das estimativas comparando-se com o modelo de regressão linear múltipla produzido pelo IAPAR e, conseqüentemente, a elaboração dos mapas de temperatura média mensal (cartas climáticas) mostrou-se mais suave e representativa.

De mesmo raciocínio, ainda com a escassez de dados de temperatura média mensal (n=33), conseguiu-se detectar e modelar dependência espacial nos dados, com resultados que asseguram a melhoria na precisão das estimativas de temperatura média dos meses de janeiro e julho para a construção das cartas climáticas do Estado do Paraná.

Os resultados alcançados, neste trabalho, foram altamente satisfatórios e motivam futuras pesquisas envolvendo, também, outras variáveis climáticas que certamente trarão um refinamento na acurácia das previsões.

REFERÊNCIAS BIBLIOGRÁFICAS[*]

ARMSTRONG, M. Problems with Universal Kriging. **Math. Geol.**, v.16, p.101-8, 1984.

BERLATO, M. A., MATZENAUER, R., SUTILI, V. R. Relação entre temperatura e o aparecimento de fases fenológicas do milho (Zea mays L.). **Agronomia Sulriograndense**, v.20, p.111-32, 1984.

BUSSAB, W. O., BOLFARINE, H. **Elementos de amostragem** In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 11, 1994, Belo Horizonte. 161p.

CÂMARA, G. M. S., SEDIYAMA, T., DOURADO-NETO, D., BERNARDES, M. S. Influência do fotoperíodo e da temperatura do ar no crescimento, floração e maturação da soja (Glycine max (L.) Merrill). **Sci. Agric.**, v.54, Número Especial, 1997.

CARROL, S. S. Modeling abiotic indicators when obtaining spatial predictions of species richness. **Environ. Ecol. Stat.**, v.5, p.257-76, 1998.

CAVIGLIONE, J. H., KIIHL, L. R. B., CARAMORI, P. H., OLIVEIRA, D., GALDINO, J., BORROZINO, E., GIACOMINI, C. C., SONOMURA, M. G. Y., PUGSLEY, L. **Cartas climáticas do Paraná**. Londrina: Instituto Agronômico do Paraná, CD-ROM.

CRESSIE, N. A. C. **Statistics for spatial data**. Rev. ed. New York: John Wiley, 1993. 900p.

COELHO, M. A. **Geografia do Brasil**. 3.ed. São Paulo: Moderna,1992. 281p.

DEUTSCH, C. V., JOURNEL, A. G. **GSLIB**: geoestatistical software library and user's guide. New York: Oxford University Press, 1992.

[*] UNIVERSIDADE ESTADUAL PAULISTA. Faculdade de Ciências Agronômicas. Normas para a elaboração de dissertações e teses. Botucatu, 1997. 35p.

DIBLASI, A., BOWMAN, A. W. On the Use of the Variogram in Checking for Independence in Spatial Data. **Biometrics**, v.57, p.211-8, 2001.

DIGGLE, P. J., RIBEIRO Jr., P. J. **Model based geostatistics**. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 14, 2000, Caxambu. 129p.

DUBRULE, O. Cross Validation of Kriging in a Unique Neighborhood. **Math. Geol.**, v.15, p.687-99, 1983.

EDWARDS, C., PENNEY, D. **Calculus and analytical geometry**. Englewood Cliffs: Prentice-Hall, 1982.

ENGELUD, E., SPARKS, A. **Geo-EAS**: geostatistical enviromental assessment software. Las Vegas: U.S. Enviromental Protency Agency, 1991.

FISHER, R.A. Edited by BENNETT, J.H. **With a Foreword by Yates, F. Statistical Methods, Experimental Design, and Scientific Inference**. A Re-issue of Statistical Methods for Research Workers, The Design of Experiments, and Statistical Methods and Scientific Inference. Oxford University Press – Oxford, 1993. 362p, 248p and 182p respectively.

FISHER, R.A. **Statistical methods and scientific inference**. Edinburgh: Oliver and Boyd, 1956. 175p.

FRITZSCHE, H. **Programação não-linear**: análise e métodos. São Paulo: Edgard Blücher, Ed. da Universidade de São Paulo, 1978. 954p.

GADIOLI, J. L., DOURADO-NETO, D., GARCIA Y GARCIA, A., BASANTA, M. V. Temperatura do ar, rendimento de grãos de milho e caracterização fenológica associada a somo calórica. **Sci. Agric.**, v.57, p.377-83, 2000.

GOUDARD, M., KARSON, M., LINDER, E., SINHA, D. Bayesian spatial prediction. **Environ. Ecol. Stat.**, v.6, p.147-71, 1999.

GOOVAERTS, P. **Geostatistics for natural resources evaluation**. New York: Oxford University Press, 1997. 483p.

GRONDONA, M. O., CRESSIE N. Residuals based estimators of the covariogram. **Statistics**, v.26, p.209-18, 1995.

GS+ for Windows 5.0.3 Beta: Geoestatistics for the environmental sciences. Michigan: Gamma Design Software, 2000. Disponível em www.gammadesign.com.

GUERRA, P.A.G. **Geoestatística operacional**. Brasília: Departamento Nacional da Produção Mineral, Ministério das Minas e Energia, 1988. 145p.

HAMLETT, J. M., HORTON, R., CRESSIE, N. A. C. Resistant and exploratory techniques for use in semivariogram analyses. **Soil Sci. Soc. Am. J.**, v.50, p.868-75, 1986.

HINKELMANN, K., KEMPTHORNE, O. **Design and analysis of experiments**: introduction to experimental design. New York: John Wiley, 1994. 495p.

HOFFMANN, R., VIEIRA, S. **Estatística experimental**. São Paulo: Atlas, 1989. 179p.

HOST, G. Bayesian estimation of European sulphur emissions using monitoring data and an acid deposition model. **Environ. Ecol. Stat.**, v.6, p.381-99, 1999.

HUDSON, G., WACKERNAGEL, H. Mapping temperature using kriging with external drift: theory and an example from Scotland. **Int. J. Climatol.**, v.14, p.77-91, 1994.

ISAAKS, E. H., SRIVASTAVA, R. M. **Applied Geostatistics**. New York: Oxfor University Press, 1989. 561p.

KEMPTHORNE, O. **The design and analysis of experiments**. New York: John Wiley, 1967. 631p.

KRONKA, S.N., BANZATTO, D.A. **Experimentação agrícola**. Jaboticabal: FUNEP, 1989. 247p.

LUCIO, P. S. **Kriging disjuntivo usando anamorfismo Gaussiano**: um método estocástico para predição de reservas recuperáveis. Campinas, 1991. Dissertação (Mestrado em Estatística) – Universidade Estadual de Campinas - UNICAMP.

LUENBERGER, D. G. **Optimization by vector space methods**. New York: John Wiley, 1969.

MANDALLAZ, D. Estimation of the spatial covariance in universal kriging: application to forest inventory. **Environ. Ecol. Stat.**, v.7, p.263-84, 2000.

MARCOTTE, D. **Cokriging with matlab**. *Comp. Geosci.*, v.17, p.1265-80, 1991.

MARTORANO, L. G., ANGELOCCI, L. R., VETTORAZZI, C. A., VALENTE, R. O. A. Zoneamento agroecológico para a região de Ribeirão Preto utilizando um sistema de informações geográficas. **Sci. Agric.**, v.56, n.3, 1999.

MISCHAN, M.M., PINHO, S.Z. **Experimentação agronômica:** dados não-balanceados. Botucatu: FUNDIBIO, 1996. 457p.

MYERS, D. E. Matrix formulation of co-kriging. **Math. Geol.**, v.14, p.249-57, 1982.

PANNATIER, I. **VARIOWIN 2.2:** software for spatial data analysis in 2D. Statistics and computing. New York: Springer-Verlag, 1996. 93p.

PARANÁ Cidade: Serviço Social Autônomo. Municípios do Paraná. Fundo de Desenvolvimento Urbano - FDU. Disponível em: http://www.paranacidade.org.br/base/municipios/shtml. Acessado em: 22 a 28 de janeiro de 2001.

PEREIRA, G.T., SCALA JUNIOR., N., MARQUES JR. J., CORÁ, J.E. **Variabilidade espacial da emissão de CO_2 de um Latossolo Vermelho-Escuro Argiloso na Região de Jaboticabal, SP**. In. REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 44, 1999, Botucatu.

PEZZI, L. P., CAVALCANTI, I. F. A. Temperature and precipitation anomalies over Brazil during the 1995 winter season – atmospheric and oceanic characteristics. **Rev. Bras. Geof.**, v.16, n.2-3, 1998.

PIMENTEL GOMES, F. **Curso de estatística experimental**. 13.ed. Piracicaba: Nobel, 1990. 468p.

QIAN, S. S. Estimating the area affected by phosphorus runoff in an Everglades wetland: a comparison of universal kriging and Bayesian kriging. **Environ. Ecol. Stat.**, v.4, p.1-29, 1997.

QUEIROZ, J. E. **Parâmetros hidrodinâmicos de um solo de várzea para fins de drenagem subterrânea**. Piracicaba, 1995. 167p. Tese (Doutorado em Agronomia/Irrigação e Drenagem) – Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo - ESALQ.

REICHARDT, K., VIERIA, S.R., LIBARDI, P.L. Revisão de Literatura: variabilidade espacial de solos e experimentação de campo. **Rev. Bras. Ciênc. Solo**, v.10, p.1-6, 1986.

RIBEIRO JUNIOR, P. J. **Métodos Geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. Piracicaba, 1995. 99p. Dissertação (Mestrado em Agronomia/ Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo - ESALQ.

SNEDECOR, G. W., COCHRAN, W. G. **Statistical methods**. 6.ed. Ames: Iowa State University, 1967. 593p.

SOUZA, L. S. **Variabilidade espacial do solo em sistemas de manejo**. Porto Alegre, 1992. 162p. Tese (Doutorado em Agronomia/Ciência do Solo) – Universidade Federal do Rio Grande do Sul - UFRGS.

STATISTICA for Windows 5.1. User's guide: General conventions and statistics I. StatSoft, Inc., 1984-1996. 2300 East 14th Street, Tulsa, OK 74104, U.S.A. Fax: (918) 749-2217. Web site: www.statsoft.com.

SURFER for Windows 6.01. User's guide: Contouring and 3D Surface Mapping. New York: Golden Software, Inc., 1995. 809 14th Street, Golden, CO 80401-1866, U.S.A. Fax: (303) 279-0909. Web site: www.goldensoftware.com.

TIMM, L.C., FONTE JR. L., BARBOSA, E.P., REICHARDT, BACCHI, O.O.S. Interação solo-planta avaliada por modelagem estatística de espaço de estados. **Sci. Agric.**, v.57, p.751-60, 2000.

UZUMAKI, E. T. **Geoestatística multivariada**: estudo de métodos de predição. Campinas, 1994. 104 p. Dissertação (Mestrado em Estatística) – Universidade Estadual de Campinas - UNICAMP.

VALENTE, J. M. G. P. **Geomatemática: lições de geoestatística, vol. I, II, III e IV**. Ouro Preto: Fundação Gorceix, 1982.

VAN ES, H. M., VAN ES, C. L. Spatial nature of randomization and its effect on the outcome of field experiments. **Agron. J.**, v.85, p.420-8, 1993.

VESENTINI, J. W., VLACH, V. **Geografia Crítica**: espaço natural e a ação humana. 25.ed. São Paulo: Ática, 1999. 192p.

VIEIRA BRAGA, L. P. **Geoestatística e aplicações**. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 9, 1990, São Paulo. 36p.

VIEIRA, S.R., NIELSEN, D. R., BIGGAR, J. W. Spatial variability of field-measured infiltration rate. **Soil Sci. Soc. Am. J.**, v.45, p.1040-47, 1981.

VIEIRA, S. R., HATFIELD, J. L., NIELSEN, D. R., BIGGAR, J. W. Geostatistical Theory and Application to Variability of Some Agronomical Properties. **Hilgardia**, v.51, p.1-75, 1983.

VIEIRA, S. R. **Geoestatística**: curso de extensão universitária. Botucatu, Faculdade de Ciências Agronômicas, UNESP, 1996. 72p.

VIEIRA, S.R. Variabilidade espacial de argila, silte e atributos químicos em uma parcela experimental de um latossolo roxo de Campinas (SP). **Bragantia**, v.56, p.1-16, 1997.

WACKERNAGEL, H. **Multivariate geostatistics**: an introduction with applications. New York: Springer-Verlag, 1995. 257p.

WATSON S. Spatial dependence and blok designs in spaced plant herbage trials. **J. Agric. Sci.**, v.134, p.245-58, 2000.