



GÉRSON RODRIGUES DOS SANTOS

**HIERARQUIZAÇÃO GEOMÉTRICA DOS PREDITORES
GEOESTATÍSTICOS**

LAVRAS - MG

2010

GÉRSON RODRIGUES DOS SANTOS

**HIERARQUIZAÇÃO GEOMÉTRICA DOS PREDITORES
GEOESTATÍSTICOS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Orientador

Dr. Marcelo Silva de Oliveira

LAVRAS - MG

2010

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da
Biblioteca Central da UFLA**

Santos, Gérson Rodrigues dos.

Hierarquização geométrica dos preditores geoestatísticos / Gérson Rodrigues dos Santos. – Lavras : UFLA, 2010.

200 p. : il.

Tese (Doutorado) - Universidade Federal de Lavras, 2010.

Orientador: Marcelo Silva de Oliveira.

Bibliografia.

1. Geoestatística. 2. Krigagem. 3. Esperança condicional. 4. Espaços de Hilbert. I. Universidade Federal de Lavras. II. Título.

CDD-519.54

GÉRSON RODRIGUES DOS SANTOS

**HIERARQUIZAÇÃO GEOMÉTRICA DOS PREDITORES
GEOESTATÍSTICOS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 16 de Novembro de 2010.

Dr. Denismar Alves Nogueira UNIFAL-MG

Dr. Eric Batista Ferreira UNIFAL-MG

Dr. João Domingos Scalon UFLA

Dr. Lucas Monteiro Chaves UFLA

Dr. Marcelo Silva de Oliveira
Orientador

LAVRAS - MG

2010

A Deus, Autor e Consumador da minha fé.

A Manoel, meu pai (in memorian), aquele que me ensinou a combater o bom combate, completar a carreira e guardar a fé.

A Joanina, minha mãe, a guerreira sublime que suportou tudo pela felicidade dos filhos.

A Marilane, Rosilane e Marcos, meus irmãos, que me mostram a cada dia como Deus tem sido misericordioso conosco.

A Belmer Negrillo e Marcelo Oliveira, meus pais científicos, que me ensinaram o valoroso ofício de um pesquisador, principalmente, quando está em sintonia com o criador do Universo, o grande “EU SOU”.

A Edcarlos Miranda, meu amigo, companheiro e irmão, a quem devo muito mais que gratidão.

A Família Rocha Aguiar (Dinha, Nilson, Rommel, Nina e tia Rita), que me acolheram em seu lar, como um filho.

Em especial, a razão de minha vida: DRICA e BABÍ, amada esposa e minha linda filha,

DEDICO

AGRADECIMENTOS

A Deus, por endireitar as minhas veredas e pelo imenso amor demonstrado até aqui.

A minha família, pela paciência, compreensão e carinho em todos os momentos, principalmente, minhas “filhas” (Adriana e Bárbara).

A família Rocha Aguiar (Dinha, Nilson, Rommel, Nina e tia Rita), pelo cuidado, paciência e auxílio sempre presente.

À Universidade Federal de Lavras (UFLA) e ao Departamento de Ciências Exatas (DEX), pela oportunidade concedida para a realização do doutorado.

À União Este-Brasileira da Igreja Adventista do Sétimo Dia (UEB-IASD) e Faculdades Integradas Adventistas de Minas Gerais (FADMINAS), especialmente, Pr. Waldomiro Passos, Prof. Edson Gasquez e Prof. Samuel De Benedicto, pela permissão e apoio financeiro nessa grande realização.

À Universidade Federal de Viçosa (UFV) e ao Departamento de Estatística (DET), pela compreensão e apoio na conclusão desse sonho, especialmente, Dr. Sebastião Martins Filho.

Ao professor, orientador, pastor, amigo e inspiração, Dr. Marcelo Silva de Oliveira, pela grande relevância neste trabalho e em minha vida.

Ao amigo Edcarlos Miranda, pela amizade, companheirismo, ensinamentos e pela imensa contribuição em minha nova carreira profissional.

Ao grande servo de Deus, João Marcos Louzada, pelos fundamentais ensinamentos em Geoestatística, sobretudo, na parte computacional, sem os quais este trabalho não seria uma realidade.

Ao amigo, professor e inspiração, Dr. Daniel Furtado Ferreira, pelos ensinamentos estatísticos e, também, pela imensa contribuição em minha nova carreira profissional.

Aos professores e servidores do DEX, pelos ensinamentos e grande estima.

Aos amigos de estudo, Ana Paula, Ademária, Augusto, Crysttian, Ricardo Olinda, Ana Lúcia, Paty Baiana, Tânia, Richardson, Altemir, Verônica e Eustáquio, em especial a Paulo Emiliano por transmitir tantos conhecimentos.

Aos professores e servidores do EDESSA, IAE-SP e UNINCOR, pela sólida e duradoura formação, essencial nesse nível acadêmico.

Aos amigos Pr. Zizion (*in memorian*), Josias Lacerda, Ervino, Zeferino, Selzi, João Cardoso, Elza, Noel, Iranildo, Marilene Miranda e Ivanildo Prado, pela lapidação inicial de minha vida acadêmica.

Aos membros da banca examinadora, Doutores: Eric, Denismar, Scalon, Lucas, Joel, Renato e Marcelo, pela disposição em julgar e contribuir com essa Tese.

“Once a map is drawn people tend to accept it as reality”.

Bert Friesen

RESUMO

A predição geoestatística faz parte da modelagem probabilística que usa a vizinhança amostrada com o intuito de fazer valer a percepção de que a estrutura de dependência espacial de um processo estocástico otimiza as predições, sem viés e com variância mínima. Assim, objetiva-se neste trabalho de tese apresentar a hierarquização geométrica dos preditores lineares geoestatísticos, considerando suas restrições e respectivos subespaços de projeção. Para tanto, adota-se a abordagem geométrica desses preditores, visando a projeção ortogonal de uma função-vetor $Y(\mathbf{x}_0)$ (que representa um valor a ser predito na localização \mathbf{x}_0) no subespaço vetorial resultante das restrições impostas. Como resultado, cita-se a hierarquia geométrica obtida: a esperança condicional, em primeiro lugar, a krigagem linear simples, em segundo lugar, e a krigagem linear universal, em terceiro lugar, devido a dimensão dos subespaços de projeção, com ilustração, também, através de simulação estocástica. Além disso, ficam estabelecidas as condições para a utilização prática de tais preditores.

Palavras-Chave: Geoestatística. Espaços de Hilbert. Esperança Condicional. Krigagem. Hierarquização Geométrica.

ABSTRACT

Geostatistics estimation is a probabilistic modeling strategy that uses sample points neighborhood information in order to consider the spatial dependence of the structure of a stochastic process to obtain, unbiased and minimum variance estimates. Hence, the main goal of this thesis is to present the geometric hierarchy of the geostatistical linear predictors, taking into account their restrictions and respective projection subspaces. We apply the orthogonal geometric projection of the predictors in order to obtain a vector function $Y(\mathbf{x}_0)$, (that represents a value to be predicted in the location (\mathbf{x}_0)) in the resulting vector subspace from the restrictions imposed to the predictors. As an important result obtained in our work, we show how to obtain geometric hierarchy: Conditional expectation, simple linear kriging and linear universal kriging, 1st, 2nd and 3rd, respectively. We establish practical conditions for using such predictors, and we illustrate our findings with stochastic simulations.

Keywords: Geostatistics. Hilbert Spaces. Conditional Expectation. Kriging. Geometric Hierarchy.

LISTA DE FIGURAS

Figura 1	Representação de 3 possíveis realizações conjuntas do processo estocástico, em \mathbb{D}	21
Figura 2	Representação dos componentes de uma variável regionalizada com (a) uma componente determinística constante e (b) uma componente determinística com tendência	30
Figura 3	Ilustração da amostragem em duas dimensões	33
Figura 4	Ilustração de um variograma experimental com a apresentação dos parâmetros desejados	34
Figura 5	Esboço de uma amostragem regularmente espaçada juntamente com a configuração direcional adotada na Geoestatística	35
Figura 6	Ilustração do cálculo do semivariograma experimental a partir de uma amostragem regularmente espaçada	36
Figura 7	Ilustração do cálculo do semivariograma experimental a partir de uma amostragem regularmente espaçada	37
Figura 8	Exemplo de um mapa de isolinhas (krigagem)	41
Figura 9	Mapa de isocores das classes de precipitação pluvial total média anual (mm) para o Estado de Minas Gerais - ZEE/MG	42
Figura 10	Ilustração da distância entre um ponto e um plano	48
Figura 11	Representação das 3 características de um vetor	49
Figura 12	Representação geométrica da extremidade de um vetor	50
Figura 13	Representação geométrica do produto de um vetor por um escalar	51
Figura 14	Representação geométrica da soma de 2 vetores	51
Figura 15	Representação geométrica de uma variedade linear para $\mathbf{N} = \mathbb{R}^2$	59
Figura 16	Representação geométrica de um conjunto aberto para $\mathbf{V} = \mathbb{R}^2$	66
Figura 17	Representação geométrica de um conjunto fechado para $\mathbf{V} = \mathbb{R}^2$	67
Figura 18	Representação do teorema da projeção quando $\mathbf{H} = \mathbb{R}^3$	73
Figura 19	Diagrama em \mathbb{R}^2 da ortonormalização de Gram-Schmidt	82
Figura 20	<i>grid's</i> regulares de dados simulados, considerando um Processo Gaussiano de média zero e com tendência	98
Figura 21	Representação do vetor-função $Y(\mathbf{x}_0) \in \mathbf{U}$ para $\mathbf{U} = \mathbb{R}^3$	105
Figura 22	Representação da projeção ortogonal do melhor preditor $\hat{Y}_{EC}(\mathbf{x}_0)$ para $\mathbf{U} = \mathbb{R}^3$	107
Figura 23	Diagrama em \mathbb{R}^3 de duas projeções de $Y(\mathbf{x}_0)$	115
Figura 24	Diagrama em \mathbb{R}^3 da projeção de $Y(\mathbf{x}_0)$ em $\mathbf{B}_{(n+1)}$	115
Figura 25	<i>Grid's</i> regulares de dados simulados, considerando um Processo Gaussiano de média zero e com tendência	160

Figura 26	<i>Grid's</i> regulares de dados simulados, considerando um Processo Gaussiano de média zero e com tendência	161
Figura 27	Análise geoestatística de um <i>grid</i> regular de 150 pontos simulados, considerando o modelo esférico de semivariograma	162
Figura 28	Análise geoestatística de um <i>grid</i> regular de 130 pontos simulados, considerando o modelo esférico de semivariograma	163
Figura 29	Análise geoestatística de um <i>grid</i> regular de 110 pontos simulados, considerando o modelo esférico de semivariograma	164
Figura 30	Análise geoestatística de um <i>grid</i> regular de 90 pontos simulados, considerando o modelo esférico de semivariograma	165
Figura 31	Análise geoestatística de um <i>grid</i> regular de 70 pontos simulados, considerando o modelo esférico de semivariograma	166
Figura 32	Análise geoestatística de um <i>grid</i> regular de 50 pontos simulados, considerando o modelo esférico de semivariograma	167
Figura 33	Análise geoestatística de um <i>grid</i> regular de 30 pontos simulados, considerando o modelo esférico de semivariograma	168
Figura 34	Análise geoestatística de um <i>grid</i> regular de 20 pontos simulados, considerando o modelo esférico de semivariograma	169
Figura 35	<i>Grid's</i> regulares de dados simulados, considerando um Processo Gaussiano com perturbação e tendência	191
Figura 36	<i>Grid's</i> regulares de dados simulados, considerando um Processo Gaussiano com perturbação e tendência	192
Figura 37	Análise geoestatística de um <i>grid</i> regular de 150 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano	193
Figura 38	Análise geoestatística de um <i>grid</i> regular de 130 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano	194
Figura 39	Análise geoestatística de um <i>grid</i> regular de 110 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano	195
Figura 40	Análise geoestatística de um <i>grid</i> regular de 90 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano	196
Figura 41	Análise geoestatística de um <i>grid</i> regular de 70 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano	197

Figura 42	Análise geoestatística de um <i>grid</i> regular de 50 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano	198
Figura 43	Análise geoestatística de um <i>grid</i> regular de 30 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano	199
Figura 44	Análise geoestatística de um <i>grid</i> regular de 20 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano	200

LISTA DE TABELAS

Tabela 1	Hierarquização dos principais preditores geoestatísticos considerando aspectos geométricos.	133
Tabela 2	Resultados das simulações que ratificam numericamente a hierarquização geométrica dos principais preditores geoestatísticos.	135

SUMÁRIO

1	INTRODUÇÃO	14
2	REFERENCIAL TEÓRICO	17
2.1	Processo estocástico	17
2.2	Visão geral da Geoestatística	27
2.2.1	Variáveis regionalizadas	28
2.2.2	Hipóteses	29
2.2.3	Semivariograma	32
2.2.4	A inferência geoestatística	39
2.2.5	Vizinhança de krigagem	45
2.2.6	Autovalidação	45
2.3	Espaços de Hilbert	47
2.3.1	Conceitos e ilustrações	49
2.3.2	O teorema da projeção	70
2.4	Espaços de Hilbert de variáveis aleatórias	91
3	METODOLOGIA	97
3.1	Materiais	97
3.2	Métodos	99
4	RESULTADOS E DISCUSSÃO	101
4.1	Função de predição	101
4.2	Predição ótima	102
4.3	A krigagem em termos de projeção	112
4.4	Preditores lineares	115
4.4.1	Krigagem linear simples	116
4.4.2	Krigagem linear ordinária	123
4.4.3	Krigagem linear universal	129
4.5	Hierarquização	132
4.6	Simulação	134
5	CONCLUSÃO	136
	REFERÊNCIAS	138
	ANEXOS	142

1 INTRODUÇÃO

A descrição quantitativa de fenômenos, assim como a predição de seus valores desconhecidos (por não observação) compõem os objetivos principais da atividade científica. Tal descrição e predição se fazem, geralmente, utilizando-se de uma prévia modelagem. Quanto à modelagem, existem, basicamente, duas abordagens, conhecidas como “determinística” (modelagem que não considera a incerteza) e “probabilística” (modelagem que quantifica a incerteza inerente ao fenômeno estudado).

Dentro da modelagem probabilística encontram-se os preditores geoestatísticos, aqueles que usam a vizinhança amostrada com o intuito de fazer valer a percepção de que a estrutura de dependência espacial de um processo estocástico otimiza as predições, a saber, sem viés e com variância mínima.

Os preditores geoestatísticos utilizam-se do método de predição chamado de *Krigagem*, nome dado em homenagem ao engenheiro de minas Sul-Africano Daniel Gerhardus Krige, feita pelo matemático e geólogo Francês Georges François Paul Marie Matheron. Muitos são os estudos, e em diversas áreas do conhecimento, que têm usado os procedimentos da Krigagem, recebendo ao longo do tempo a seguinte subdivisão:

- a) Krigagens lineares: krigagem simples, krigagem ordinária e krigagem universal;
- b) Krigagens não lineares: Esperança condicional e krigagem disjuntiva.

Apesar da vasta utilização desses preditores, encontra-se muita divergência em publicações sobre a escolha e utilização destes. Assim, sendo, percebe-se

uma necessidade da explicitação do que é comum (dito *backbone*) entre os preditores geoestatísticos e uma diferenciação dos mesmos, principalmente, quanto à utilização de cada um e quanto às suas variâncias. Esta necessidade traduzida como uma hierarquização constitui o problema de pesquisa dessa tese.

Portanto, adota-se como hipótese, neste trabalho, que a abordagem geométrica dos preditores lineares geoestatísticos viabiliza hierarquizá-los, ou seja, chegar a uma classificação sob os critérios de viés e variância (precisão).

De forma geral, objetiva-se apresentar a hierarquização geométrica dos preditores lineares geoestatísticos considerando suas restrições e respectivos subespaços de projeção, o que levará à classificação desejada.

Especificamente, pretende-se atingir este objetivo geral através dos específicos:

- a) Apresentar a fundamentação teórico-matemática de processos estocásticos e espaços vetoriais;
- b) Demonstrar que a Esperança Condicional (doravante EC) é o melhor de todos os preditores geoestatísticos (isto é, a melhor krigagem possível);
- c) Demonstrar que a Krigagem Linear Simples (doravante KLS), sob o aspecto geométrico é o melhor dos preditores lineares da Geoestatística;
- d) Demonstrar que a Krigagem Linear Ordinária (doravante KLO), sob o aspecto geométrico é o segundo melhor dos preditores lineares da Geoestatística;
- e) Demonstrar que a Krigagem Linear Universal (doravante KLU), sob o aspecto geométrico é o terceiro melhor dos preditores lineares da Geoestatística;

tica;

- f) Esclarecer a correta utilização das krigagens, considerando as restrições geométricas e de processos estocásticos.

Essa tese está organizada na forma tradicional de trabalhos científicos (elementos pré-textuais, textuais e pós-textuais), onde a seção de Referencial Teórico está subdividida em:

- a) Processo estocástico: apresentação geral do tema e sua relação com a Geoestatística;
- b) Visão geral da Geoestatística: apresentação do *backbone* da Geoestatística clássica, abordando tópicos como, variáveis regionalizadas, principais hipóteses da geoestatística, ajuste de semivariogramas, predição geoestatística, entre outros;
- c) Espaços de Hilbert: apresentação dos conceitos, teoremas, proposições, entre outros, necessários para a fundamentação da abordagem geométrica do assunto;
- d) Espaços de Hilbert de variáveis aleatórias: apresentação complementar de itens que possibilitam a abordagem escolhida.

Após o Referencial Teórico, apresenta-se a Metodologia, subdividida em Materiais e em Métodos utilizados em busca da confirmação da hipótese proposta; e a seção de Resultados e Discussão, que por sua vez está subdividida em resultados teóricos e validações aplicadas destes resultados.

Finalmente, apresenta-se a Conclusão do trabalho.

2 REFERENCIAL TEÓRICO

Na atividade científica percebe-se um esforço para a coleta de informações objetivando a construção de modelos que expliquem os fenômenos em estudo (OLIVEIRA et al., 2009).

Conforme citado, os modelos utilizados na ciência podem ser classificados como *determinísticos e não determinísticos*.

É certo que os modelos determinísticos contribuem para a compreensão do comportamento dos fenômenos, porém, em um nível básico inicial. Um dos motivos dessa limitação é não lidar com a incerteza envolvida na ocorrência dos mesmos.

Recorrer a processos estocásticos é uma forma de tratar quantitativamente a incerteza desses fenômenos, que, aproveitando certas características de regularidade que estes apresentam, pode-se descrevê-los por modelos probabilísticos.

Tomando o exemplo da variável indexadora *tempo*, compara-se as duas formas de modelagem: enquanto uma função $f(t)$, no caso determinístico, toma valores bem definidos ao longo do tempo, um processo estocástico toma valores aleatórios em torno de uma função média.

2.1 Processo estocástico

De forma simples, define-se um processo estocástico como um conjunto de variáveis aleatórias indexadas em um conjunto arbitrário de índices. Este índice pode ser unidimensional, como o exemplo do tempo, conforme abordagem de Clarke e Disney (1979), e pode ser multidimensional, constituindo vetores, conforme Kannan (1979).

Aos valores que um processo estocástico pode assumir denomina-se *estados*, e ao conjunto de todos os valores possíveis denomina-se *espaço de estados*.

Usando aqui uma notação própria, seguindo Alves e Delgado (1997), um processo estocástico é representado por $\{Y(\mathbf{x}) : \mathbf{x} \in \mathbb{D} \subset \mathbb{R}^p\}$. Em que, para o tempo, $\mathbf{x} = t$, para uma superfície, $\mathbf{x} = (x, y)$, e assim, sucessivamente. Além disso, os processos estocásticos são classificados segundo:

- a) O espaço dos estados;
- b) A natureza do conjunto \mathbb{D} ;
- c) As características estatísticas das variáveis aleatórias que definem o processo.

Em relação ao espaço de estados, se um conjunto de estados for finito ou enumerável, $Y(\mathbf{x})$ é um “processo de estados discretos”, ou comumente, uma “cadeia”. Para qualquer outro caso, tem-se um “processo de estados contínuos”. Alves e Delgado (1997) citam três exemplos de processos discretos (para a variável tempo): a condição de uma máquina (ligada/desligada) no momento t , o número de clientes em uma loja no instante t e o número de máquinas avariadas no fim do dia t .

Em relação à natureza do conjunto \mathbb{D} , se este conjunto for finito ou enumerável, $Y(\mathbf{x})$ é um “processo de índice discreto” e é normalmente composto por coordenadas que pertençam ao conjunto dos inteiros não negativos. Caso contrário, $Y(\mathbf{x})$ é denominado “processo de índice contínuo”. Como exemplos do caso discreto, cita-se a cotação do dólar ao final do dia $\mathbf{x} = t$ ou qualquer análise da estatística espacial para dados de área, com $\mathbf{x} = (x, y)$ sendo a coordenada do centróide arbitrário da área, ou qualquer outro ponto que a caracterize.

Por fim, em relação às características estatísticas das variáveis aleatórias que definem o processo, classifica-se tal processo através de propriedades envolvendo a distribuição conjunta do processo, em especial, a propriedade de *estacionariedade*, que é de grande importância. Um processo estocástico é considerado estacionário se o seu comportamento estocástico for independente de \mathbf{x} . Como exemplos, pode-se citar processos markovianos (processos estacionários que possuem a propriedade de Markov ou da “perda de memória”, isto é, o comportamento futuro do processo é condicionado apenas pelo estado presente) e processos geoestatísticos (foco desse trabalho).

Não faz parte da abordagem desse trabalho todos os tipos de processos estocásticos existentes, porém, aqueles de interesse da Geoestatística. Todos esses processos interessantes à Geoestatística derivam de um processo estocástico fundamental, conforme apresentação que segue.

De forma introdutória, já se definiu um processo estocástico. Contudo, uma formalização se faz necessária, conforme Definição 1.

Definição 1 (Processo Estocástico). *Um processo estocástico é um conjunto de variáveis aleatórias $\{Y(\mathbf{x}) : \mathbf{x} \in \mathbb{D} \subset \mathbb{R}^p\}$ definidas em um dado espaço de probabilidade (Ω, \mathbb{A}, P) e indexadas pelo índice \mathbf{x} , com \mathbf{x} variando em \mathbb{D} .*

Então, dado um espaço $\mathbb{D} \subset \mathbb{R}^p$, esse \mathbb{D} pode corresponder a um espaço físico, para o qual cita-se alguns como exemplos, com suas respectivas dimensões p :

- a) Uma região geográfica ($p = 2$);
- b) Uma área de solo ($p = 2$);

- c) Uma jazida de minério ($p = 3$);
- d) O volume de água lacustre (lagos) ou oceânica ($p = 3$);
- e) A massa de ar atmosférico ($p = 3$);
- f) Uma superfície de um disco de arado ($p = 2$ ou $p = 3$);
- g) Uma superfície de chapas de aço ($p = 2$);
- h) Perfurações verticais para prospecção de petróleo ($p = 1$);
- i) Uma reserva petrolífera ($p = 3$);
- j) A linha do tempo ($p = 1$);
- k) Eventos no espaço-tempo ($p = 4$);
- l) Espaços de dimensão $p > 4$, dificilmente com significado físico, mas tal modelagem adotada aqui não os exclui.

Em geral, pode-se associar a cada ponto \mathbf{x} de \mathbb{D} o valor $y(\mathbf{x})$ de uma variável de interesse. Essa variável “localizada” pode ser:

- a) A densidade do solo no ponto \mathbf{x} ;
- b) A temperatura do ar no ponto \mathbf{x} ;
- c) A tensão sofrida pelo disco de arado em um ponto \mathbf{x} de sua superfície;
- d) O valor do dólar no tempo $\mathbf{x} = t$;
- e) O valor da umidade do solo no índice $\mathbf{x} = (x_1, x_2, x_3, t)$;
- f) Entre outras.

Os valores $\{y(\mathbf{x}) : \mathbf{x} \in \mathbb{D} \subset \mathbb{R}^p\}$ podem ser vistos, sem inconsistência lógica (referência feita a processos com uma única realização do fenômeno), como uma realização do processo estocástico $PE = \{Y(\mathbf{x}) : \mathbf{x} \in \mathbb{D} \subset \mathbb{R}^p\}$, ou seja, podem ser vistos como a realização $\{Y(\mathbf{x}, \omega) : \mathbf{x} \in \mathbb{D} \subset \mathbb{R}^p, \omega \in \Omega\}$, onde Ω é o espaço amostral de todas as realizações conjuntas possíveis desse processo estocástico, e ω um dos eventos elementares de Ω , conforme representa-se na Figura 1. Nessa Figura, $y(\mathbf{x}) = Y(\mathbf{x}, \omega)$.

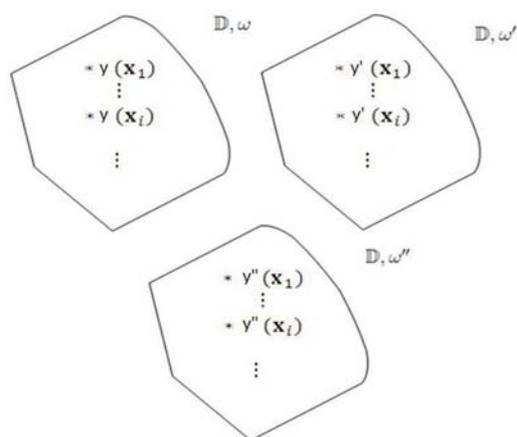


Figura 1 Representação de 3 possíveis realizações conjuntas do processo estocástico, em \mathbb{D}

Este Ω que identifica cada realização ω particular de todo o processo estocástico PE induz (ou subsidia) o espaço de probabilidade (Ω, \mathbb{A}, P) , que garante a existência de uma distribuição de probabilidade conjunta para todo o processo estocástico considerado. Um subconjunto qualquer deste processo estocástico, em especial uma amostra $A = \{Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)\}$, tem, por consequência, garan-

tida também a existência de uma distribuição de probabilidade conjunta

$$F_{Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)}(\cdot, \dots, \cdot), \forall n \in \mathbb{N}, \forall \mathbf{x}_i \in \mathbb{D}.$$

Em outras palavras, essa distribuição $F_{Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)}(\cdot, \dots, \cdot)$ tem existência garantida pelo mesmo espaço (Ω, \mathbb{A}, P) do processo estocástico todo, sendo desnecessário invocar outro espaço específico para a amostra A .

Nota: A variável $Y(\mathbf{x})$ pode ser de qualquer tipo numérico, discreta ou contínua, isto é, $Y(\mathbf{x})$ pode ser, inclusive, contagem de algum evento de interesse, além da medição de qualquer quantidade, conforme classificação já apresentada.

Todo o zelo apresentado em explicitar e comentar sobre o espaço de probabilidade (Ω, \mathbb{A}, P) reside apenas em garantir o cálculo de probabilidades sobre variáveis de interesse. Em outras palavras, ao explicitar o espaço (Ω, \mathbb{A}, P) , fenômenos “complexos” (ou “complicados”), para os quais não é possível calcular consistentemente probabilidades não estão sendo considerados neste trabalho. Verdadeiramente, pode-se considerar tal atitude um “excesso” de zelo, pois, tais fenômenos “complicados” talvez nem mesmo existam fisicamente ou ocorram empiricamente. Porém, a modelagem é, de fato, algo rigoroso, onde a precaução contra situações inválidas é constante. Para os propósitos deste trabalho, ao definir (Ω, \mathbb{A}, P) de acordo com Mood, Graybill e Boes (1974), todo e qualquer cálculo de probabilidade feito tem a garantia de ser válido (não existindo probabilidade infinita, negativa, maior que 1, entre outros).

Avançando na modelagem, pode-se detalhar a variável aleatória $Y(\mathbf{x})$ um pouco mais, segundo Webster e Oliver (2007) e Cressie (1993), entendendo-a

como constituída pelos componentes:

$$Y(\mathbf{x}) = \mu(\mathbf{x}) + \epsilon'(\mathbf{x}) + \epsilon''. \quad (1)$$

Essa componentização de $Y(\mathbf{x})$ serve para, pelo menos, duas funções:

- a) A primeira é apresentar os 3 tipos fundamentais de variação (ou variabilidade) espacial, a saber: variação *determinística*, representada por $\mu(\mathbf{x})$, a qual quase sempre representa uma variação de grande escala no fenômeno, como por exemplo, variações observadas na paisagem (*landscape*) região de estudo. Dessa forma, a dependência espacial de $\mu(\mathbf{x})$ com $\mu(\mathbf{x}')$ não é descrita (governada) probabilisticamente, mas, sim, matematicamente, por meio de funções matemáticas definidas sobre as coordenadas espaciais; variação *probabilística regionalizada*, representada por $\epsilon'(\mathbf{x})$, que abriga e dá conta da dependência espacial estocástica presente na vizinhança de \mathbf{x} (daí o termo “regionalizada”, criado por Georges Matheron). Nesta componente está toda a estrutura de dependência espacial de natureza probabilística (também conhecida como natureza estatística ou estocástica). Em outras palavras, a Geoestatística existe devido essa componente, pois, sem $\epsilon'(\mathbf{x})$, a variável aleatória pode ser analisada via Estatística *Clássica* (expressão comumente usada na literatura especializada para a abordagem em que a “localização” não é considerada). Sem perda de generalidade e por conveniência conceitual, faz-se $E[\epsilon'(\mathbf{x})] = 0, \forall \mathbf{x} \in \mathbb{D}$. Alguns autores chamam esta componente de “variação de média escala”, porém, o termo “variação regionalizada” é mais enraizado na Geoestatística; variação *probabilística completamente independente espacialmente*, ou seja, absolutamente

não regionalizada, representada por ϵ'' . Propositadamente, a componente ϵ'' não apresenta o argumento \mathbf{x} , pois não depende de tal localização, ou, ainda, não tem dependência espacial, nem determinística e nem probabilística. Também, sem perda de generalidade e por conveniência conceitual, faz-se $E[\epsilon''] = 0$. De fato, se a variável aleatória for contínua, espera-se mesmo que ϵ'' seja um “ruído normal”, isto é, $\epsilon'' \sim N(0, \tau^2)$, porém, tal exigência de normalidade não é uma condição necessária para a Geoestatística, pelo menos em sua forma convencional;

- b) A segunda é permitir um tratamento matemático e estatístico mais elucidativo do modelo para o fenômeno representado pelo processo estocástico fundamental. Observa-se que pela equação (1): $E[Y(\mathbf{x})] = \mu(\mathbf{x}), \forall \mathbf{x} \in \mathbb{D}$ e exige-se, ou pressupõe-se, implicitamente, a existência de $E[Y(\mathbf{x})]$, $\forall \mathbf{x} \in \mathbb{D}$, isto é, o processo estocástico é de 1ª ordem. Justifica-se essa suposição pela argumentação de que processos estocásticos que não possuem média (aqueles com distribuição de Cauchy, por exemplo) não ocorrem na prática, ou na realidade do interesse desse trabalho. Caso tais processos ocorram, a solução pode estar na análise que utiliza apenas as distribuições de probabilidade, e não envolve momentos estatísticos.

Por fim, aperfeiçoando mais ainda a modelagem, pode-se considerar vetores aleatórios em lugar de variáveis aleatórias no processo estocástico fundamental,

a saber:

$$\left\{ \begin{bmatrix} Y_1(\mathbf{x}) \\ Y_2(\mathbf{x}) \\ \vdots \\ Y_m(\mathbf{x}) \end{bmatrix} : \mathbf{x} \in \mathbb{D} \subset \mathbb{R}^p \right\}.$$

Uma realização desse processo estocástico multivariado para um particular ω de Ω pode ser representado por

$$\left\{ \begin{bmatrix} Y_1(\mathbf{x}, \omega) = y_1(\mathbf{x}) \\ Y_2(\mathbf{x}, \omega) = y_2(\mathbf{x}) \\ \vdots \\ Y_m(\mathbf{x}, \omega) = y_m(\mathbf{x}) \end{bmatrix} : \mathbf{x} \in \mathbb{D} \subset \mathbb{R}^p \right\}.$$

De igual forma, o espaço (Ω, \mathbb{A}, P) , conforme assumido, permite para os vetores aleatórios:

- a) Obter (calcular) a distribuição conjunta espacial (no espaço de índices \mathbb{D}), isto é, definir

$$F_{Y_i(\mathbf{x}_1), \dots, Y_i(\mathbf{x}_n)}(\cdot, \dots, \cdot), \forall i = 1, \dots, m; \forall n = 1, 2, \dots; \forall \mathbf{x}_n \in \mathbb{D};$$

- b) Obter (calcular) a distribuição conjunta multivariada (ou “cruzada”)

$$F_{Y_1(\mathbf{x}), \dots, Y_m(\mathbf{x})}(\cdot, \dots, \cdot), \forall m = 1, 2, \dots; \forall \mathbf{x} \in \mathbb{D}.$$

A dependência em (i) é espacial, mas em (ii) não é, sendo, em vez disso, uma dependência multivariada entre variáveis aleatórias medidas no mesmo ponto

x. Em (i), a dependência conjunta da mesma variável aleatória Y_i localizada em diferentes pontos \mathbf{x}_i do espaço (o que justifica o termo “dependência espacial”) é medida por $F_{Y_i(\mathbf{x}), \dots, Y_i(\mathbf{x})}(\cdot, \dots, \cdot)$, enquanto em (ii) mede-se a dependência conjunta de diferentes (ou múltiplas) variáveis aleatórias $Y_1(\mathbf{x}_1), \dots, Y_m(\mathbf{x}_m)$ localizadas no mesmo ponto \mathbf{x} do espaço (o que justifica o termo “dependência cruzada”).

Pode-se, ainda, admitir as dependências híbridas, do tipo $F_{Y_i(\mathbf{x}_k), Y_j(\mathbf{x}_l)}(\cdot, \cdot)$. Tais hibridizações são chamadas de “dependência espacial cruzada”, pois varia-se tanto a localização no espaço ($\mathbf{x}_k \neq \mathbf{x}_l$) quanto as variáveis envolvidos ($Y_i \neq Y_j$).

Assim, cita-se algumas possíveis situações na análise estatística espacial, usando como exemplo a área de Ciência de Solos:

- a) Análise da medida da densidade do solo em diferentes pontos, isto é, uma situação que envolve a autodependência espacial (ou autovariabilidade espacial) da densidade do solo (única variável medida naquele solo);
- b) Análise da medida da densidade do solo, da velocidade de infiltração básica e da tensão de pré-consolidação (três variáveis distintas) em um mesmo ponto \mathbf{x} do solo, isto é, uma situação de dependência (ou variabilidade) multivariada das 3 variáveis em uma única localização espacial;
- c) Análise da medida das mesmas 3 variáveis anteriores em diferentes pontos da região de estudo indica uma situação de dependência espacial cruzada.

Conforme todo o exposto nessa seção, pode-se afirmar que o processo estocástico fundamental é aquele em que suas realizações garantem a existência de um espaço de probabilidade (Ω, \mathbb{A}, P) , em outras palavras, a existência da distribuição conjunta para todo o processo estocástico considerado.

2.2 Visão geral da Geoestatística

Considerando a literatura geoestatística, em especial, Andriotti (2003), Chilès e Del'ner (1999), Cressie (1993), Isaaks e Srivastava (1989), Journel e Huijbregts (2003), Vieira (2000) e Webster e Oliver (2007), aborda-se todo o conteúdo dessa subseção.

Com considerável frequência encontra-se a errônea ideia de que a Geoestatística é a própria Estatística Espacial. De fato, a Geoestatística constitui uma das técnicas dessa área da Estatística, que é composta por também outras metodologias que visam as especificações de modelos inferenciais que consideram explicitamente a estrutura de correlação espacial das observações dos fenômenos em estudo.

A busca por metodologias, que descrevessem os efeitos de uma das preocupações mais antigas da ciência experimental, *a variabilidade espacial*, data publicações de 1910, 1913, 1915, 1918, 1919, e assim por diante, de pesquisadores como Smith, Montgomery, Robinson e Lloyd, Waynick e Sharp, entre outros.

Porém, os métodos propostos geralmente supunham que as realizações das variáveis aleatórias fossem independentes entre si, ou seja, observações não deviam exercer influências sobre suas vizinhas, contradizendo, durante décadas, a prática.

Com o intuito de modelar a variabilidade, pesquisadores como Fisher, Snedecor e Cochran, adotaram técnicas como *casualização*, *controle local* e *repetição* que levaram à utilização de amostragens ao acaso, desprezando, dessa forma, as referências dos pontos amostrados.

Os métodos geoestatísticos, ou simplesmente Geoestatística, conforme ci-

tado, foram desenvolvidos graças aos estudos do Matemático e Geólogo Georges Matheron na *Écoles des Mines de Fontainebleau* - França, no final da década de 50 e início da década de 60. Estes métodos estão fundamentados na “Teoria das Variáveis Regionalizadas”, que foi formalizada por Matheron, a partir dos estudos práticos desenvolvidos pelo Engenheiro de Minas Daniel Krige no cálculo de reservas nas minas de ouro na África do Sul. Este pesquisador não conseguia encontrar sentido nas variâncias estimadas, se não levasse em consideração a distância e a direção da amostragem.

Durante um período, o principal obstáculo encontrado pela Geoestatística foi a pequena capacidade de processamento dos computadores, pois, muitos são os cálculos envolvidos na aplicação desta metodologia.

2.2.1 Variáveis regionalizadas

A teoria das variáveis regionalizadas pressupõe que a variação de uma variável pode ser expressa pela soma de três componentes:

- a) Uma componente estrutural, relativa a um valor médio ou a uma tendência;
- b) Uma componente aleatória, espacialmente correlacionada;
- c) Um ruído aleatório.

Dado que \mathbf{x} representa uma posição de uma, duas ou mais dimensões da região \mathbb{D} , então a variável regionalizada $Y(\mathbf{x})$ pode ser denotada por

$$Y(\mathbf{x}) = \mu(\mathbf{x}) + \epsilon'(\mathbf{x}) + \epsilon'',$$

(uma componentização representada anteriormente e repetida por conveniência) em que:

- a) $\mu(\mathbf{x})$ é uma função determinística que representa a componente estrutural;
- b) $\epsilon'(\mathbf{x})$ é um termo estocástico que varia localmente e depende espacialmente de $\mu(\mathbf{x})$;
- c) ϵ'' um ruído aleatório, não correlacionado, com distribuição normal com média zero e variância τ^2 .

Representa-se, graficamente através da Figura 2, essa componentização de uma variável regionalizada, em que $Y(\mathbf{x}) = Z$ e $\mu(\mathbf{x}) = m(\mathbf{x})$.

2.2.2 Hipóteses

Conforme apresentado, Matheron fundamentou os conceitos teóricos da Geoestatística em processos estocásticos, o qual ele denominou como funções aleatórias. Uma função aleatória pode ser entendida como o conjunto de variáveis aleatórias existentes dentro de uma região de interesse, isto é, um processo estocástico. Assim definida e pelo fato de ser contínua, uma função aleatória pode ser submetida a uma grande quantidade de hipóteses, sem as quais a dedução de equações não seria viável. Em especial, a restrição de que normalmente os processos estocásticos espaciais são realizados uma única vez.

Portanto, para estimar valores em locais não amostrados, deve-se introduzir a restrição de que a variável regionalizada seja estacionária estatisticamente, isto é, os momentos estatísticos da variável aleatória $Y(\mathbf{x} + \mathbf{h})$ sejam os mesmos para qualquer vetor distância \mathbf{h} .

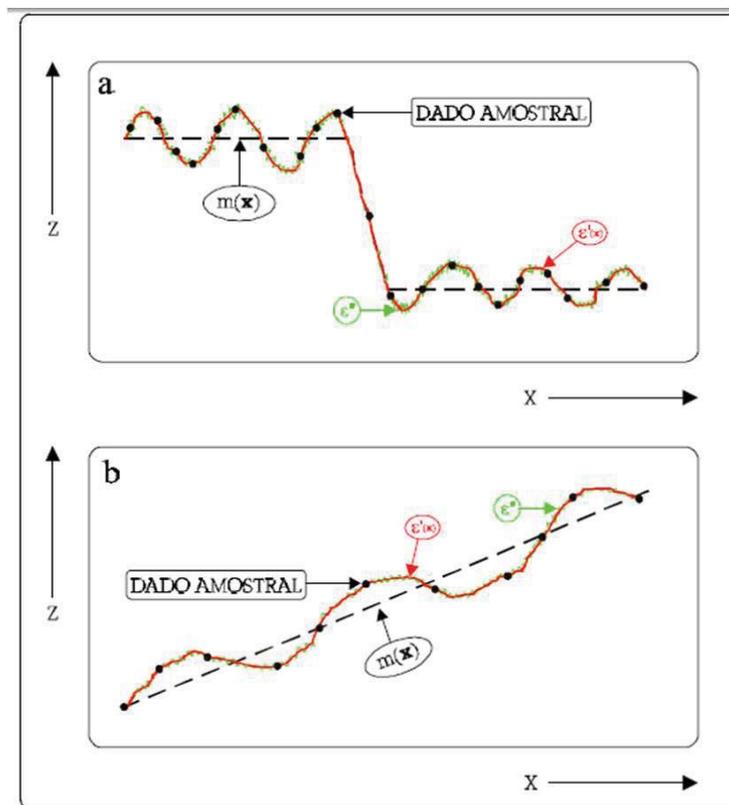


Figura 2 Representação dos componentes de uma variável regionalizada com (a) uma componente determinística constante e (b) uma componente determinística com tendência

Fonte: Modificada de Burrough (1987 *apud* CAMARGO, 1998)

De acordo com o número k de momentos estatísticos que são constantes, a variável aleatória é chamada de estacionária de ordem k , sendo que na Geoestatística a estacionariedade de ordem 2 ($k = 2$) é o máximo o que se requer.

Conforme apresentado, a função aleatória $Y(\mathbf{x})$ tem valores esperados $E[Y(\mathbf{x})] = \mu(\mathbf{x})$ e $E[Y(\mathbf{x} + \mathbf{h})] = \mu(\mathbf{x} + \mathbf{h})$ e variâncias $\text{var}[Y(\mathbf{x})]$ e $\text{var}[Y(\mathbf{x} + \mathbf{h})]$, respectivamente, para os locais \mathbf{x} e $\mathbf{x} + \mathbf{h}$, para qualquer vetor \mathbf{h} . Então, a

covariância $\text{cov}(\mathbf{x}, \mathbf{x} + \mathbf{h}) = \text{cov}(Y(\mathbf{x}), Y(\mathbf{x} + \mathbf{h}))$ é dada por

$$\text{cov}(\mathbf{x}, \mathbf{x} + \mathbf{h}) = E[Y(\mathbf{x}).Y(\mathbf{x} + \mathbf{h})] - \mu(\mathbf{x}).\mu(\mathbf{x} + \mathbf{h}) \quad (2)$$

e o parâmetro *semivariância* por

$$2\gamma(\mathbf{x}, \mathbf{x} + \mathbf{h}) = E\{[Y(\mathbf{x}) - Y(\mathbf{x} + \mathbf{h})]^2\} - \{[\mu(\mathbf{x} + \mathbf{h}) - \mu(\mathbf{x})]^2\} \quad (3)$$

Salienta-se que $\text{var}[Y(\mathbf{x})] = \text{cov}(\mathbf{x}, \mathbf{x})$, denotado simplesmente por $\text{cov}(0)$ e que $\text{var}[Y(\mathbf{x} + \mathbf{h})] = \text{cov}(\mathbf{x} + \mathbf{h}, \mathbf{x} + \mathbf{h})$, também denotado por $\text{cov}(0)$.

Diante dessas definições, uma função aleatória $Y(\mathbf{x})$ é estacionária de 1ª ordem se

$$E[Y(\mathbf{x})] = \mu \quad (4)$$

ou seja, o valor esperado existe e não depende da posição \mathbf{x} . Nesse caso, $\gamma(\mathbf{x}, \mathbf{x} + \mathbf{h}) = \frac{1}{2}E\{[Y(\mathbf{x}) - Y(\mathbf{x} + \mathbf{h})]^2\}$.

Analogamente, uma função aleatória $Y(\mathbf{x})$ é estacionária de 2ª ordem se for estacionária de 1ª ordem e, para cada par de variáveis aleatórias $Y(\mathbf{x})$ e $Y(\mathbf{x} + \mathbf{h})$, a covariância é dada por

$$\text{cov}(\mathbf{h}) = E[Y(\mathbf{x}).Y(\mathbf{x} + \mathbf{h})] - \mu^2 \quad (5)$$

ou seja, a covariância existe e está em função apenas de \mathbf{h} .

Através das equações (4) e (5) é possível demonstrar que a covariância e a semivariância são mecanismos equivalentes na caracterização da dependência espacial.

A hipótese de estacionariedade de 2ª ordem implica a existência de uma variância finita das realizações, porém, essa hipótese não pode ser satisfeita para alguns fenômenos de infinita dispersão. Para tais situações, uma hipótese menos restritiva, a *hipótese intrínseca*, se faz necessária.

Uma função aleatória é intrínseca quando satisfaz a estacionariedade de 1ª ordem e o incremento $\{Y(\mathbf{x}) - Y(\mathbf{x} + \mathbf{h})\}$ tem variância finita e não depende de \mathbf{x} , para qualquer vetor \mathbf{h} . Matematicamente, isso pode ser escrito por

$$\gamma(\mathbf{h}) = \frac{1}{2} E \left\{ [Y(\mathbf{x}) - Y(\mathbf{x} + \mathbf{h})]^2 \right\} \quad (6)$$

Assim, ficam descritas as principais hipóteses da Geoestatística, estacionariedade de 1ª ordem, de 2ª ordem e intrínseca.

2.2.3 Semivariograma

O semivariograma, ou simplesmente variograma, é uma ferramenta fundamental para descrever quantitativamente a variação no espaço de um fenômeno regionalizado. A natureza estrutural de um conjunto de dados é definido a partir da comparação de valores tomados em 2 pontos, obedecendo uma determinada direção, conforme ilustra-se na Figura 3.

O nível de dependência entre duas variáveis regionalizadas é representado por

$$2\gamma(\mathbf{h}) = E \left\{ [Y(\mathbf{x}) - Y(\mathbf{x} + \mathbf{h})]^2 \right\} = \text{var} [Y(\mathbf{x}) - Y(\mathbf{x} + \mathbf{h})] \quad (7)$$

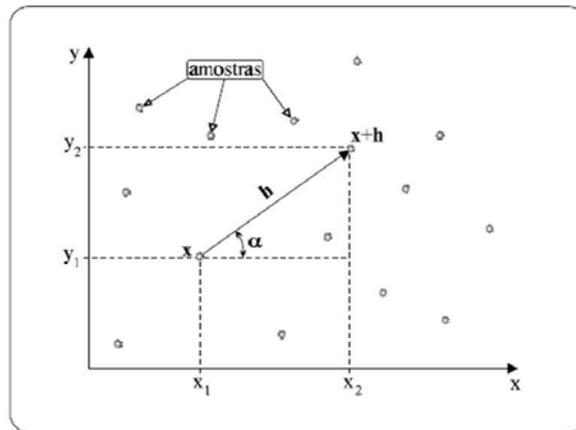


Figura 3 Ilustração da amostragem em duas dimensões

Fonte: Camargo (1998)

e pode ser estimada por

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [Y(\mathbf{x}_i) - Y(\mathbf{x}_i + \mathbf{h})]^2 \quad (8)$$

em que $N(\mathbf{h})$ é o número de pares de valores medidos $Y(\mathbf{x}_i)$ e $Y(\mathbf{x}_i + \mathbf{h})$, separados por uma vetor distância \mathbf{h} .

Graficamente, a representação das semivariâncias estimadas $\hat{\gamma}(\mathbf{h})$ em função dos valores correspondentes de \mathbf{h} (que visa esboçar o comportamento da dependência espacial do fenômeno) é denominada semivariograma *experimental* ou *empírico*.

Conforme apresentado, a estacionariedade é uma propriedade do modelo probabilístico, que pode variar com a mudança de escala do estudo e/ou à medida que novos dados são disponibilizados. Logo, é a decisão de estacionariedade que permite agrupar observações obtidas em diferentes posições da área de estudo,

gerando os diferentes semivariogramas existentes.

A Figura 4 ilustra um variograma experimental com características bem próximas do ideal. Intuitivamente, espera-se dos dados amostrais que as diferenças $Y(x_i) - Y(x_i + h)$ decresçam à medida que h também decresça. O que se espera é que observações mais próximas tenham um comportamento mais semelhante entre si do que aquelas mais distantes, até que essa dessemelhança se estabilize.

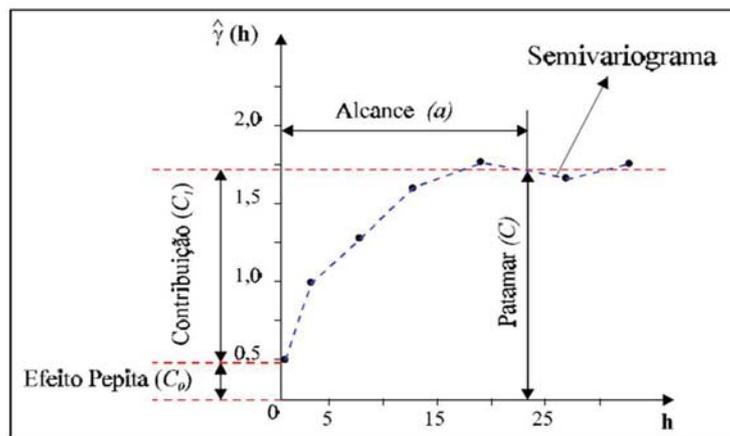


Figura 4 Ilustração de um variograma experimental com a apresentação dos parâmetros desejados

Fonte: Camargo (1998)

Através do variograma deve-se obter os parâmetros da estrutura de dependência espacial, úteis no processo de predição geoestatística. Conforme representado na Figura 4, os parâmetros são:

- Alcance (a): distância dentro da qual as amostras apresentam-se correlacionadas espacialmente, então, além do alcance considera-se que não exista dependência espacial;
- Patamar (C): é o valor da semivariância correspondente ao alcance;

- c) Efeito pepita (C_0): idealmente $\gamma(\mathbf{0}) = 0$, porém, há sempre uma descontinuidade para as menores distâncias entre os pontos, o que pode ocorrer devido a erros no processo de amostragem, por exemplo, entre outros motivos;
- d) Contribuição (C_1): é a diferença entre o patamar (C) e o efeito pepita (C_0).

A amostragem é outro aspecto muito importante para a predição geoestatística, pois, é através desse processo que os parâmetros do semivariograma podem ser melhor estimados.

A amostragem geoestatística pode ser realizada de duas formas, a princípio: com malhas regular ou irregularmente espaçadas, conforme representação feita nas Figuras 5 e 7.

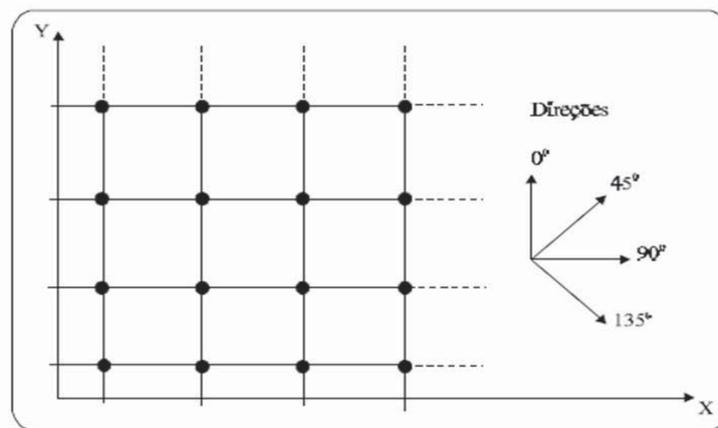


Figura 5 Esboço de uma amostragem regularmente espaçada juntamente com a configuração direcional adotada na Geoestatística

Fonte: Camargo (1998)

Considerando, inicialmente, um conjunto de amostras em 2 dimensões regularmente espaçadas (Figura 5), a construção do variograma experimental con-

siste em calcular as semivariâncias para a menor distância entre os pontos, em uma mesma direção e plotá-las no gráfico (pode-se plotar todas as parcelas das semivariâncias estimadas ou apenas a sua média). Em seguida, o mesmo deve ser feito para a segunda menor distância, na mesma direção (conforme Figura 6), e o procedimento finaliza quando todas as distâncias e todas as direções são consideradas. Salienta-se, porém, que para cada direção deve-se construir um semivariograma diferente.

Quando o variograma apresenta um comportamento idêntico em qualquer direção de h , denomina-se *variograma direcional isotrópico*, caso contrário, *variograma direcional anisotrópico*.

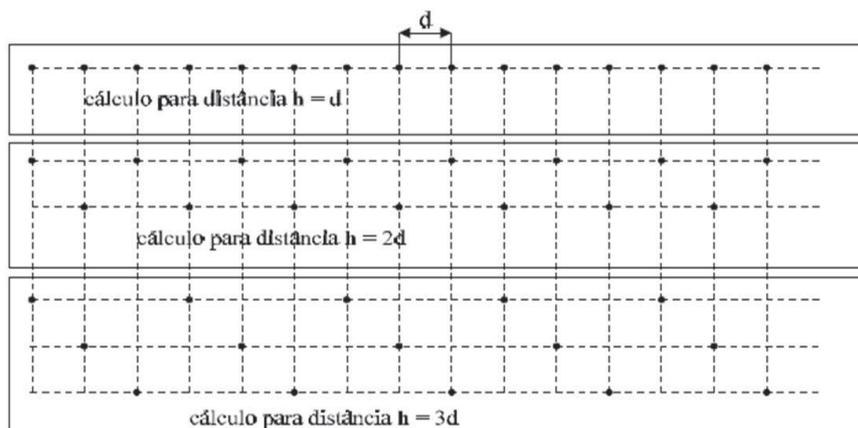


Figura 6 Ilustração do cálculo do semivariograma experimental a partir de uma amostragem regularmente espaçada

Fonte: Camargo (1998)

Considerando, então, uma amostragem irregularmente espaçada em duas dimensões (Figura 7), deve-se introduzir *limites de tolerância* para a escolha de pontos amostrais em função da distância e direção.

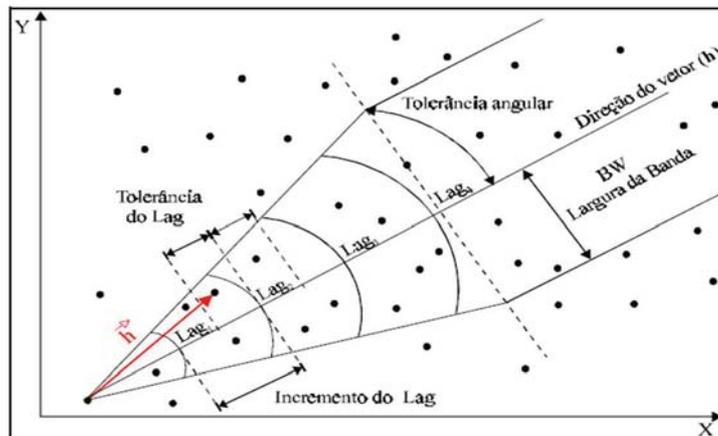


Figura 7 Ilustração do cálculo do semivariograma experimental a partir de uma amostragem regularmente espaçada

Fonte: Camargo (1998)

Tomando como referência o lag_1 (lag é o termo utilizado na Geoestatística como uma distância entre 2 pontos) e direção de exemplo 45° , utiliza-se a tolerância angular de $22,5^\circ$ para a obtenção dos pares de pontos que serão incluídos no cálculo. Este procedimento deve ser repetido para todos os lag 's, todos os pontos e todas as direções, sempre obedecendo a análise anisotrópica.

Uma vez que o semivariograma experimental é obtido, pode-se então ajustá-lo através de modelos teóricos próprios. Este processo de ajuste não é direto e automático, como em outras metodologias da Estatística, pois depende do julgamento de um especialista, do modelo escolhido e da natureza do problema estudado.

Os principais modelos isotrópicos disponíveis na literatura contemplam variogramas sem patamar e com patamar. Os modelos sem patamar são próprios para a estimação de parâmetros em fenômenos com dispersão infinita (aqueles que não atendem a hipótese intrínseca). Entre estes, cita-se o modelo potência (em que

b é o parâmetro de intensidade da dispersão, ou seja, se $b > 1$ a dispersão é mais acentuada que $b = 1$ e $b < 1$), dado por

$$\gamma(\mathbf{h}) = \begin{cases} 0, & \text{se } |\mathbf{h}| = 0 \\ C_0 + C_1 \cdot |\mathbf{h}|^b, & \text{se } |\mathbf{h}| \neq 0 \end{cases} \quad (9)$$

Os modelos com patamar são aqueles que atingem o patamar, pelo menos assintoticamente. Entre estes, cita-se os três principais e mais utilizados: *esférico* (também chamado de *Modelo de Matheron*), *exponencial* (modelo pertencente à importante família *Matérn*) e o *gaussiano* (um dos modelos utilizados no ajuste de fenômenos com semivariâncias que demoram alguns *lag's* para aumentar). Respectivamente, estes modelos são dados por

$$\gamma(\mathbf{h}) = \begin{cases} 0 & , \text{ se } |\mathbf{h}| = 0 \\ C_0 + C_1 \left[\frac{3}{2} \left(\frac{|\mathbf{h}|}{a} \right) - \frac{1}{2} \left(\frac{|\mathbf{h}|}{a} \right)^3 \right] & , \text{ se } 0 < |\mathbf{h}| \leq a \\ C_0 + C_1 & , \text{ se } |\mathbf{h}| > a \end{cases} \quad (10)$$

$$\gamma(\mathbf{h}) = \begin{cases} 0 & , \text{ se } |\mathbf{h}| = 0 \\ C_0 + C_1 \left[1 - \exp\left(-\frac{|\mathbf{h}|}{a}\right) \right] & , \text{ se } |\mathbf{h}| \neq 0 \end{cases} \quad (11)$$

$$\gamma(\mathbf{h}) = \begin{cases} 0 & , \text{ se } |\mathbf{h}| = 0 \\ C_0 + C_1 \left[1 - \exp\left(-\frac{|\mathbf{h}|}{a}\right)^2 \right] & , \text{ se } |\mathbf{h}| \neq 0 \end{cases} \quad (12)$$

Conforme citado, estes modelos são para fenômenos isotrópicos, ou seja,

aqueles que apresentam semivariogramas experimentais idênticos em todas as principais direções. Para fenômenos anisotrópicos é comum classificar a anisotropia, antes de aplicar uma transformação nos dados, ou seja:

- a) Anisotropia geométrica: variogramas com mesmo patamar e diferentes alcances;
- b) Anisotropia zonal: variogramas com mesmo alcance e diferentes patamares;
- c) Anisotropia combinada: variogramas com diferentes patamares e diferentes alcances.

2.2.4 A inferência geoestatística

Dentro do processo de modelagem encontra-se a modelagem estatística, que pode ser entendida como o esforço humano em explicar, probabilisticamente, os fenômenos. Considerando, especificamente, fenômenos onde as coordenadas espaciais são relevantes pode-se definir e subdividir a Estatística Espacial.

Para Bailey e Gatrell (1995), a Estatística Espacial é um ramo da Estatística composto por métodos científicos para a análise de dados cuja localização no espaço é levada em consideração. Cressie (1993) subdivide tais dados em: dados geoestatísticos, dados de área (*lattices*) e padrão de pontos.

A Geoestatística pode ser vista em uma linguagem de Processos Estocásticos, que tem como ideia básica prever pontos desconhecidos (não amostrados) do espaço de índices \mathbb{D} do processo estocástico fundamental, proposto como modelo ou modelar a estrutura de dependência espacial, que pode ser de interesse, independente de predição.

Os preditores da Geoestatística usam a vizinhança amostrada com o intuito de fazer valer a percepção de que a estrutura de dependência espacial do processo reduz a “incerteza” das predições.

Como já dito, devido a limitações de ordem prática, em grande parte das aplicações da Geoestatística só há uma realização do processo estocástico, então pressuposições de estacionariedade são necessárias (estacionariedade da covariância e da semivariância).

A estrutura de dependência espacial pode ser medida por uma ferramenta própria da área, denominada semivariograma, que plota as semivariâncias (definidas em situação oportuna) em função dos correspondentes vetores de distâncias (VIEIRA, 2000).

Constatada a presença de dependência espacial entre as realizações $y(\mathbf{x}_i)$ do processo estocástico fundamental, doravante, por simplicidade de notação, $Y(\mathbf{x}_i)$, pode-se fazer predições em pontos \mathbf{x} da região de estudo \mathbb{D} , por meio de um preditor não tendencioso e de variância mínima, em que \mathbf{x}_0 é um ponto de \mathbb{D} .

Conforme citado, na Geoestatística, o método de predição é chamado de *Krigagem*. Cressie (1993) afirma que a palavra krigagem é sinônimo de “predição ótima”, em outras palavras, a krigagem significa fazer inferências de valores não observados de um processo estocástico $Y(\cdot)$, dado por $\{Y(\mathbf{x}_i) : \mathbf{x}_i \in \mathbb{D} \subset \mathbb{R}^p\}$, $\forall i = 1, 2, \dots, n$, em que \mathbf{x}_i é um vetor composto pelas coordenadas da i -ésima realização do processo.

Frequentemente usa-se a krigagem do tipo linear, a qual está associada ao acrônimo BLUP (*Best Linear Unbiased Predictor*). O termo “linear” indica que as estimativas são combinações lineares dos dados observados; “unbiased” indica que o valor esperado dos resíduos seja zero; “best” porque visa minimizar a variância

dos resíduos e, finalmente, “predictor” por estimar variáveis aleatórias (ISAAKS; SRIVASTAVA, 1989).

Dessa forma, o uso mais frequente da krigagem a faz ser entendida como uma predição linear. Além disso, conforme Journel e Huijbregts (2003) e Vieira (2000), o que se deseja com a krigagem é expressar os resultados da predição em forma de mapas de isolinhas (curvas que expressam valores de uma dada variável aleatória através de contornos), conforme Figura 8, e mapas de isocores (definido neste trabalho como faixas de cores que expressam classes de valores preditos de um determinado processo estocástico fundamental), conforme Figura 9.

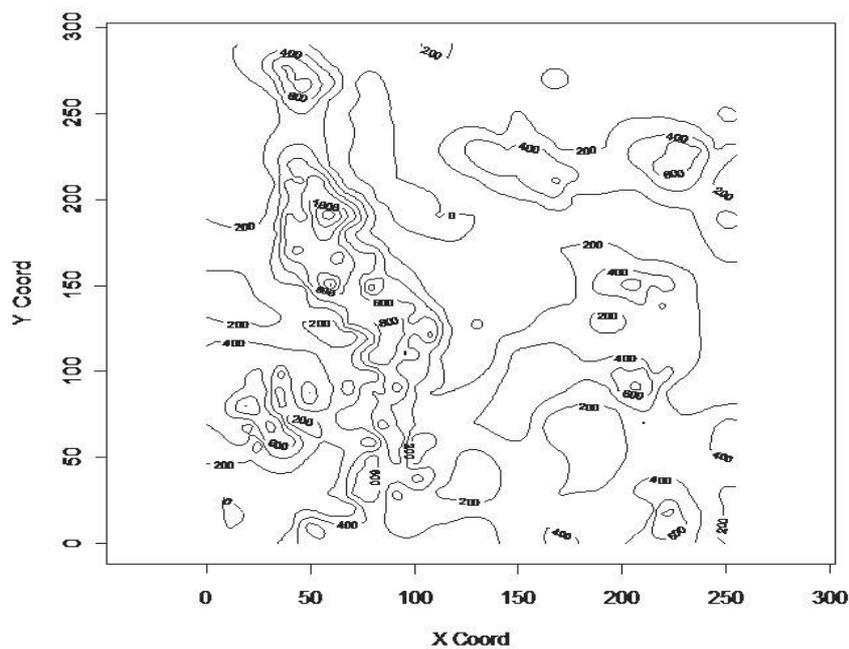


Figura 8 Exemplo de um mapa de isolinhas (krigagem)

Fonte: Landim (2006)

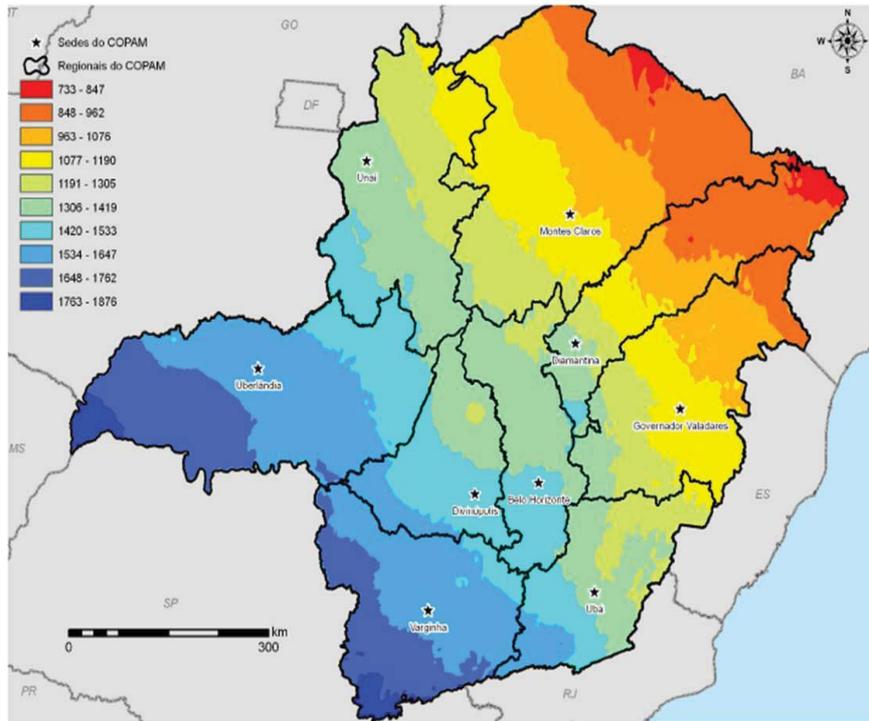


Figura 9 Mapa de isocores das classes de precipitação pluvial total média anual (mm) para o Estado de Minas Gerais - ZEE/MG

Fonte: Scolforo, Carvalho e Oliveira (2008)

Então, partindo do desejo de prever valores $y(\mathbf{x}_0)$, para qualquer localização \mathbf{x}_0 , onde não se tem realizações do processo estocástico $Y(\mathbf{x}_i)$, e assumindo a krigagem como uma combinação linear dos dados, do tipo

$$y(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i y(\mathbf{x}_i) + \lambda_0 \quad (13)$$

em que:

- a) n é o número de realizações $y(\mathbf{x}_i)$ envolvidas na estimativa;
- b) λ_i são os pesos associados a cada realização $y(\mathbf{x}_i)$;
- c) λ_0 é uma constante.

Com base na equação (13), tomando $y(\mathbf{x}_i)$ como realização do processo estocástico $Y(\mathbf{x}_i)$ e sob certas condições de regularidade o preditor da krigagem é dado por

$$\hat{Y}(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i) + \lambda_0 \quad (14)$$

Este preditor, equação (14), aparentemente não apresenta ineditismo algum em relação aos outros métodos de predição. Porém, conforme salienta Andriotti (2003), Bailey e Gatrell (1995), Chilès e Del?ner (1999) e Vieira (2000), os pesos λ_i são atribuídos de acordo com a variabilidade espacial expressa no semivariograma.

Nota: Vale ressaltar que nem todos os preditores de krigagem apresentam a forma da equação (14), porém, deseja-se mostrar neste trabalho, conforme exposto, que os preditores com este formato e os demais que se enquadram como “melhores” preditores apresentam propriedades ótimas geometricamente (não tendenciosidade e variância mínima).

Diante disso, pode-se afirmar que o preditor de krigagem linear $\hat{Y}(\mathbf{x}_0)$ é uma média ponderada, sendo a maneira de atribuição dos pesos a responsável pela otimalidade do mesmo.

Na literatura especializada, encontra-se os seguintes tipos de krigagem: krigagem simples, krigagem ordinária, krigagem universal e krigagem disjuntiva.

É possível encontrar outros tipos, porém, estes são os principais e mais importantes para este trabalho, sendo que, nesta tese, considera-se a Esperança Condicional como uma krigagem. Já a krigagem disjuntiva não enquadra nos objetivos deste trabalho. Passa-se, então, a descrever cada um:

- a) Krigagem Simples - é a krigagem em que se assume a função média do processo estocástico conhecida, logo, não é estimada a partir das realizações (BAILEY; GATRELL, 1995);
- b) Krigagem Ordinária - segundo Chilès e Delfiner (1999) é a krigagem que envolve apenas o variograma, tornando-a a mais utilizada e cuja função média não é conhecida, portanto, estimada a partir dos dados;
- c) Krigagem Universal - é a krigagem dos casos em que a variável regionalizada não seja estacionária, em que a remoção da tendência é feita no ajuste de polinômios de baixo grau, então, o restante do procedimento analítico torna-se uma análise de resíduos;
- d) Krigagem Disjuntiva - é a krigagem em que os pesos são encontrados por meio da projeção do valor desconhecido $y(\mathbf{x}_0)$ no espaço vetorial intermediário gerado pela soma das “n” funções $f(Y(\mathbf{x}_i)), \forall i = 1, 2, \dots, n$.

Cada sistema de equações da krigagem pode ser explicitado algebricamente, porém, convém apresentar a estimação geoestatística em uma forma geométrica (em uma abordagem didática) através da projeção, além de, apresentar também, as condições requeridas para que o preditor seja tomado como “ótimo”.

2.2.5 Vizinhaça de krigagem

As principais formas de determinação da vizinhaça são: vizinhaça única, distância constante, número constante de vizinhos e quadrantes.

No método da “vizinhaça única”, todos os valores medidos são considerados vizinhos, onde o tamanho da amostra pode influenciar na escolha desse método, pois, amostras grandes podem gerar um custo computacional alto. Para o método da “distância constante”, seleciona-se uma vizinhaça que é composta por todos os elementos que estão dentro de um círculo de raio especificado. O método “número constante de vizinhos” é bastante usado, pois, primeiramente, os vizinhos são procurados dentro de um raio inicial até que a quantidade de vizinhos especificada seja encontrada. Já o método dos “quadrantes” é uma alternativa que utiliza um número específico de vizinhos em cada quadrante ao redor do valor a ser estimado.

2.2.6 Autovalidação

Em qualquer método de estimação existe sempre certo grau de incerteza inerente ao processo, quer seja sobre as hipóteses assumidas ou sobre os parâmetros ajustados na modelagem.

A autovalidação ou *jackknife* é o procedimento de quantificação dessa incerteza. Este procedimento faz com que cada ponto medido “deixe de existir” (simulando um ponto perdido ou não amostrado) no processo de estimação, levando em conta a variabilidade espacial local expressa nas primeiras distâncias no variograma.

Dessa forma, tem-se dois conjuntos de dados, o conjunto dos medidos e

o conjunto dos preditos. Então, para cada um dos n locais, com o valor medido $Y(\mathbf{x}_i)$, há um valor estimado $\hat{Y}(\mathbf{x}_i)$, em que pode-se estimar uma regressão linear entre eles, dado por

$$\hat{Y}(\mathbf{x}_i) = a + b.Y(\mathbf{x}_i),$$

em que:

- a) a é o intercepto, que idealmente deve ser zero, porém, para $a > 0$ é provável que esteja ocorrendo uma superestimação e para $a < 0$, esteja ocorrendo uma subestimação;
- b) b é o coeficiente angular da reta que idealmente deve ser 1.

Essa regressão ou outros métodos de análise do erro $Y(\mathbf{x}_i) - \hat{Y}(\mathbf{x}_i)$, permitem discernir a qualidade da predição. Um deles, muito comum em recursos computacionais, é o citado por Webster e Oliver (2007). O método consiste em modelar variogramas empíricos através de dois ou mais modelos, em que para cada modelo deve-se estimar $Y(\mathbf{x}_i)$ e obter a variância. Assim, três estatísticas podem ser obtidas, a saber:

- a) O erro médio ME (*mean error*), dado por $ME = \frac{1}{N} \sum_{i=1}^N \{Y(\mathbf{x}_i) - \hat{Y}(\mathbf{x}_i)\}$;
- b) O erro quadrático médio MSE (*mean squared error*), dado por $MSE = \frac{1}{N} \sum_{i=1}^N \{Y(\mathbf{x}_i) - \hat{Y}(\mathbf{x}_i)\}^2$;
- c) O erro padrão SE (*standard error*), dado por $SE = \frac{1}{N} \sum_{i=1}^N \frac{\{Y(x_i) - \hat{Y}(x_i)\}^2}{\hat{\sigma}^2(\mathbf{x}_i)}$, em que $\hat{\sigma}^2(\mathbf{x}_i)$ é a estimativa da variância de krigagem.

O erro médio ME é, naturalmente, zero, pois a krigagem é não viesada. Mas, não apresenta acurácia em muitos casos. Quanto a isso, MSE tem acurácia se o modelo adotado no variograma também tiver, então o valor de MSE tende a ser igual à variância de krigagem e então o SE tende a 1.

Diante dessa breve apresentação sobre a visão geral da Geoestatística, ressalta-se que muitos aspectos particulares dos dados ficariam escondidos se não fossem os procedimentos dessa metodologia.

Contudo, conforme citado, a Geoestatística não tem por finalidade desencorajar o uso da Estatística Clássica, a qual em seu espaço, potencialidades e limitações. Mas, é justamente em situações em que a Clássica tem limitações que a Geoestatística tem suas maiores potencialidades.

2.3 Espaços de Hilbert

A busca pelo “melhor” preditor implica em um processo de otimização e, de acordo com Luenberger (1969), a teoria da otimização pode ser vista como relações geométricas simples e intuitivas, com grande utilidade prática. Um dos princípios utilizados por este autor para justificar tal afirmação é o *Teorema da Projeção*. Este teorema é um dos mais simples e importantes resultados da teoria da otimização.

Conforme se sabe do *Espaço Euclidiano Tridimensional*, a menor distância entre um ponto e um plano é a reta perpendicular que vai do ponto ao plano, ilustrado na Figura 10.

Então, Luenberger (1969) relaciona este resultado, aparentemente simples, com espaços de maior dimensão e até de dimensão infinita, formando a base dos procedimentos de estimação.

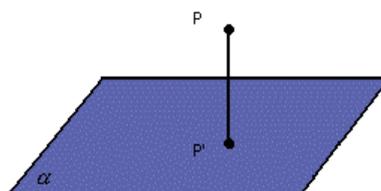


Figura 10 Ilustração da distância entre um ponto e um plano

Dessa forma, é comum encontrar na literatura que o conceito matemático de um Espaço de Hilbert é uma generalização do Espaço Euclidiano, pois amplia os métodos da álgebra vetorial e o cálculo para espaços com qualquer número finito ou infinito de dimensões. Logo, de forma simples, pode-se conceituar o Espaço de Hilbert como um espaço vetorial que possui a estrutura de um produto interno, permitindo medidas de comprimento e ângulo.

Os primeiros espaços de Hilbert foram estudados na primeira década do Século XX. Como ferramentas indispensáveis nas teorias de equações diferenciais parciais, mecânica quântica, análise de Fourier e teoria ergódica (base matemática do estudo da termodinâmica), John von Neumann cunhou o termo “Espaço de Hilbert” como uma homenagem a David Hilbert.

Destaca-se também que os exemplos de espaços de Hilbert incluem espaços de funções de quadrado-integrável e espaços de sequências. Além da necessidade de intuição geométrica, outros aspectos são pontos altos na teoria dos espaços de Hilbert, tais como, espaços vetoriais, combinação linear, produto interno, norma, subespaços vetoriais fechados, espaços de Banach, entre outros.

2.3.1 Conceitos e ilustrações

Um espaço de Hilbert pode ser visto como um espaço vetorial normado completo. Um dos exemplos mais conhecidos de um espaço de Hilbert é o Espaço Euclidiano tridimensional, denotado por \mathbb{R}^3 , dotado de um produto interno. Este produto interno toma dois vetores x e y e gera um número real $x \cdot y$.

Logo, pode-se perceber que alguns conceitos são de fundamental importância para o estudo dos espaços de Hilbert. Então, apresenta-se conceitos como *espaços vetoriais, produto interno, norma, espaços vetoriais completos*, além de outros relacionados.

Conceito 1 (Vetor). *Uma entidade matemática própria para definir grandezas não escalares, ou seja, grandezas que dependem de módulo, direção e sentido.*

Um exemplo de grandezas desse tipo, conforme Boldrini et al. (1980), é a força, pois para sua análise, o módulo, a direção e o sentido (Figura 11) são importantes.



Figura 11 Representação das 3 características de um vetor

Um vetor pode ser, basicamente, representado de duas maneiras: geomé-

trica e analiticamente. O modo geométrico facilita sua visualização, porém, não é adequado algebricamente. Simbolicamente, este método utiliza-se da representação \overrightarrow{OP} , uma vez que encontra-se na literatura especializada a convenção de que seu início é fixado na origem do espaço adotado, conforme Figura 12. O modo analítico consiste em definir um sistema de coordenadas e decompor o vetor segundo as coordenadas. Por exemplo, para o \mathbb{R}^3 , um vetor $\mathbf{u} \in \mathbb{R}^3$ pode ser denotado como $\mathbf{u} = (a, b, c)$ ou $\mathbf{u} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ ou $f = (a, b, c)$, em que f é uma função com domínio em \mathbb{R} .

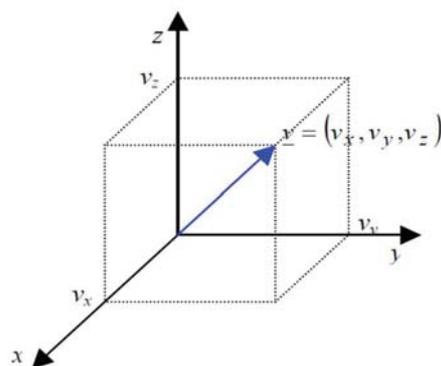


Figura 12 Representação geométrica da extremidade de um vetor

Boldrini et al. (1980) apresentam as duas operações clássicas definidas para os vetores, multiplicação de um vetor por um escalar (Figura 13) e adição de vetores (Figura 14). Multiplicar um vetor \mathbf{v} por um escalar $k > 0$ é considerar um novo vetor $\mathbf{w} = k\mathbf{v}$, que possui a mesma direção de \mathbf{v} e tem como comprimento k vezes o comprimento de \mathbf{v} . Se $k < 0$, o vetor $\mathbf{w} = k\mathbf{v}$ será igual ao oposto do vetor $|k|\cdot\mathbf{v}$ e se $k = 0$, $\mathbf{w} = k\mathbf{v}$ será o vetor nulo $\mathbf{0}$. Somar 2 ou mais vetores é considerar um novo vetor resultante da soma de suas respectivas coordenadas.

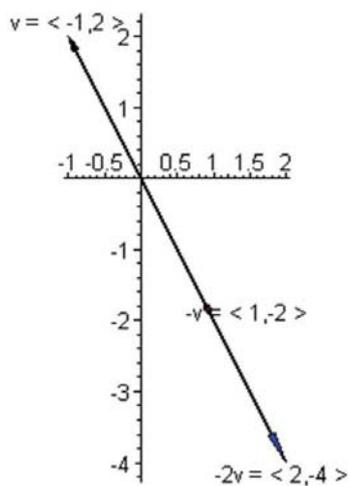


Figura 13 Representação geométrica do produto de um vetor por um escalar

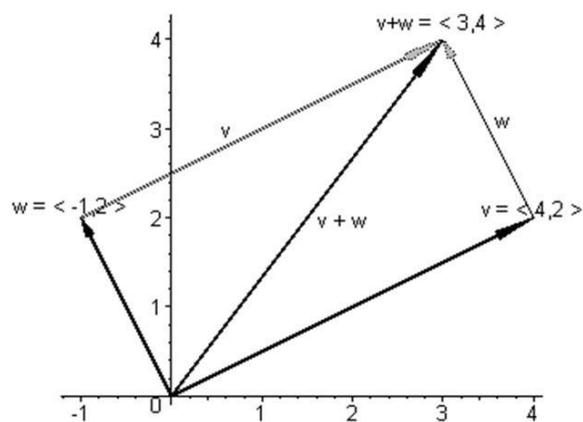


Figura 14 Representação geométrica da soma de 2 vetores

Vale salientar que o conjunto dos escalares utilizados na definição da operação produto por escalar é chamado comumente de Corpo (termo bastante utilizado na literatura).

Essas operações, juntamente com suas propriedades, caracterizam certos conjuntos cujos componentes “comportam-se” como vetores. Tais conjuntos recebem o nome de *Espaços Vetoriais*.

Os espaços vetoriais podem ser classificados como *espaços vetoriais reais* e *espaços vetoriais complexos*. Essa classificação depende do corpo de escalares adotado para a operação produto por escalar, ou seja, se o corpo adotado for o conjunto \mathbb{R} dos números reais, tem-se o primeiro tipo, caso o corpo seja o conjunto \mathbb{C} dos números complexos, tem-se o outro tipo. Adota-se neste trabalho apenas os espaços vetoriais reais, apesar de apresentar vários conceitos de forma geral, e, visando a simplicidade de notação, utiliza-se simplesmente o termo espaços vetoriais.

Conceito 2 (Espaço Vetorial). *Seja \mathbf{K} um corpo e seja \mathbf{V} um conjunto não vazio com regras de adição e multiplicação por escalar, em que a cada $\mathbf{u}, \mathbf{v} \in \mathbf{V}$ associa-se uma soma $\mathbf{u} + \mathbf{v} \in \mathbf{V}$ e a cada $\mathbf{u} \in \mathbf{V}, k \in \mathbf{K}$ associa-se um produto $k \cdot \mathbf{u} \in \mathbf{V}$. Então \mathbf{V} é chamado Espaço Vetorial sobre \mathbf{K} se prevalecer os seguintes axiomas:*

$$A_1 : (\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}), \forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbf{V};$$

$$A_2 : \mathbf{u} + \mathbf{0} = \mathbf{u}, \forall \mathbf{u} \in \mathbf{V} \text{ e } \mathbf{0} \text{ sendo um vetor nulo de } \mathbf{V};$$

$$A_3 : \mathbf{u} + (-\mathbf{u}) = \mathbf{0}, \forall \mathbf{u} \in \mathbf{V} \text{ e } -\mathbf{u} \text{ sendo seu vetor oposto pertencente a } \mathbf{V};$$

$$A_4 : \mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}, \forall \mathbf{u}, \mathbf{v} \in \mathbf{V};$$

$$M_1 : (a + b) \mathbf{u} = a\mathbf{u} + b\mathbf{u}, \forall a, b \in \mathbf{K} \text{ e } \forall \mathbf{u} \in \mathbf{V};$$

$$M_2 : k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}, \forall \mathbf{u}, \mathbf{v} \in \mathbf{V} \text{ e } \forall k \in \mathbf{K};$$

$$M_3 : (ab) \mathbf{u} = a(b\mathbf{u}), \forall a, b \in \mathbf{K} \text{ e } \forall \mathbf{u} \in \mathbf{V};$$

$$M_4 : 1\mathbf{u} = \mathbf{u}, \forall \mathbf{u} \in \mathbf{V} \text{ e } 1 \in \mathbf{K}.$$

Lipschutz (1994) afirma, sobre o conceito 2, que é perceptível a divisão dos axiomas em dois grupos, os 4 primeiros correspondem à estrutura aditiva de \mathbf{V} e os demais dizem respeito à ação do corpo \mathbf{K} sobre \mathbf{V} .

Com base nestes axiomas, pode-se verificar as propriedades de um espaço vetorial, conforme descritos no Teorema 1.

Teorema 1 (Propriedades). *Seja \mathbf{V} um espaço vetorial sobre um corpo \mathbf{K} , então*

$$(i) \quad k\mathbf{0} = \mathbf{0}, \forall k \in \mathbf{K} \text{ e } \mathbf{0} \in \mathbf{V};$$

$$(ii) \quad 0\mathbf{u} = \mathbf{0}, \forall \mathbf{u} \in \mathbf{V} \text{ e } 0 \in \mathbf{K};$$

$$(iii) \quad k\mathbf{u} = \mathbf{0} \Rightarrow k = 0 \text{ ou } \mathbf{u} = \mathbf{0}, \text{ para } k \in \mathbf{K} \text{ e } \mathbf{u} \in \mathbf{V};$$

$$(iv) \quad (-k)\mathbf{u} = k(-\mathbf{u}) = -k\mathbf{u}, \forall \mathbf{u} \in \mathbf{V} \text{ e } \forall k \in \mathbf{K}.$$

Demonstração:

- (i) pelo axioma (A_2) com $\mathbf{u} = \mathbf{0}$, tem-se $\mathbf{0} + \mathbf{0} = \mathbf{0}$. Logo, pelo axioma (M_1) , $k\mathbf{0} = k(\mathbf{0} + \mathbf{0}) = k\mathbf{0} + k\mathbf{0}$. Adicionando $-k\mathbf{0}$ a ambos os membros, obtem-se $k\mathbf{0} = \mathbf{0}$;
- (ii) sabendo que $\mathbf{0} + \mathbf{0} = \mathbf{0}$ e pelo axioma (M_2) $\mathbf{0}\mathbf{u} = (\mathbf{0} + \mathbf{0})\mathbf{u} = \mathbf{0}\mathbf{u} + \mathbf{0}\mathbf{u}$. Adicionando $-\mathbf{0}\mathbf{u}$ a ambos os membros, obtem-se $\mathbf{0}\mathbf{u} = \mathbf{0}$;
- (iii) seja $k\mathbf{u} = \mathbf{0}$ e $k \neq 0$. Então existe um escalar k^{-1} tal que $k^{-1}k = 1$, daí $\mathbf{u} = 1\mathbf{u} = (k^{-1}k)\mathbf{u} = k^{-1}(k\mathbf{u}) = k^{-1}\mathbf{0} = \mathbf{0}$;
- (iv) como $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$, obtem-se $\mathbf{0} = k\mathbf{0} = k(\mathbf{u} + (-\mathbf{u})) = k\mathbf{u} + k(-\mathbf{u})$. Adicionando $-k\mathbf{u}$ a ambos os membros, tem-se $-k\mathbf{u} = k(-\mathbf{u})$. Como $k + (-k) = 0$, obtem-se $\mathbf{0} = 0\mathbf{u} = (k + (-k))\mathbf{u} = k\mathbf{u} + (-k)\mathbf{u}$. Adicionando $-k\mathbf{u}$ a ambos os membros, tem-se $-k\mathbf{u} = (-k)\mathbf{u}$. Assim, $(-k)\mathbf{u} = k(-\mathbf{u}) = -k\mathbf{u}$. \blacklozenge

Luenberger (1969), após apresentar as propriedades de um espaço vetorial,

considera um importante conceito sobre conjuntos de espaços vetoriais combinados que produzem um espaço vetorial maior.

Conceito 3 (Espaços Vetoriais Combinados). *Sejam \mathbf{V} e \mathbf{W} espaços vetoriais sobre o mesmo corpo \mathbf{K} . Então o produto cartesiano de \mathbf{V} e \mathbf{W} , denotado por $\mathbf{V} \times \mathbf{W}$, consiste na coleção de pares ordenados (\mathbf{v}, \mathbf{w}) com $\mathbf{v} \in \mathbf{V}$ e $\mathbf{w} \in \mathbf{W}$. As operações soma e produto por escalar são definidas, respectivamente, em $\mathbf{V} \times \mathbf{W}$ por*

$$(\mathbf{v}_1, \mathbf{w}_1) + (\mathbf{v}_2, \mathbf{w}_2) = (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}_1 + \mathbf{w}_2)$$

e

$$k(\mathbf{v}, \mathbf{w}) = (k\mathbf{v}, k\mathbf{w})$$

com $k \in \mathbf{K}$.

Pode-se verificar que o produto cartesiano $\mathbf{V} \times \mathbf{W}$, assim definido, é um espaço vetorial, pois satisfaz aos axiomas do conceito 2.

Dentre os vários exemplos de espaços vetoriais encontrados na literatura, cita-se um, utilizado por Lipschutz (1994), cuja aplicação neste trabalho tem grande importância. Tal exemplo é o de *Espaço de Funções*.

Conceito 4 (Espaço de Funções). *Sejam \mathbf{S} um conjunto não-vazio e \mathbf{K} um corpo arbitrário. Considerando o conjunto $\mathbf{F}(\mathbf{S})$ de todas as funções de \mathbf{S} em \mathbf{K} , então, a soma de duas funções $f, g \in \mathbf{F}(\mathbf{S})$ é a função $f + g \in \mathbf{F}(\mathbf{S})$, definida por*

$$(f + g)(s) = f(s) + g(s), \forall s \in \mathbf{S}$$

e o produto de um escalar $k \in \mathbf{K}$ com uma função $f \in \mathbf{F}(\mathbf{S})$ é a função $kf \in \mathbf{F}(\mathbf{S})$, definida por

$$(kf)(s) = kf(s), \forall s \in \mathbf{S}.$$

Assim, $\mathbf{F}(\mathbf{S})$ com essas operações é um espaço vetorial sobre \mathbf{K} .

Lipschutz (1994) demonstra que o Conceito 4 atende ao Conceito 2, isto é, o espaço de funções é um espaço vetorial sobre \mathbf{K} .

Com muita frequência, é necessário detectar, dentro de um espaço vetorial \mathbf{V} , subconjuntos \mathbf{W} que sejam eles próprios espaços vetoriais “menores”. Assim, Boldrini et al. (1980) definem subespaços vetoriais com base em duas condições.

Conceito 5 (Subespaços Vetoriais). *Dado um espaço vetorial \mathbf{V} , um subconjunto \mathbf{W} , não-vazio, será um subespaço vetorial de \mathbf{V} se*

(i) Para quaisquer $\mathbf{u}, \mathbf{v} \in \mathbf{W}$, $\mathbf{u} + \mathbf{v} \in \mathbf{W}$;

(ii) Para quaisquer $k \in \mathbf{K}$ e $\mathbf{u} \in \mathbf{W}$, $k\mathbf{u} \in \mathbf{W}$.

Boldrini et al. (1980), comentando sobre o conceito 5, fazem 3 observações: a primeira é sobre a garantia de que realizando as operações (soma e produto

por um escalar) em \mathbf{W} não se obtem vetores fora de \mathbf{W} , logo \mathbf{W} é um espaço vetorial; a segunda é sobre a necessidade de um subespaço vetorial conter o vetor nulo, o que pode ser obtido por (ii) quando $k = 0$; e, a última observação, todo espaço vetorial admite pelo menos 2 subespaços (chamados de subespaços triviais), o conjunto formado somente pelo vetor nulo e o próprio espaço vetorial.

Um exemplo clássico de subespaço vetorial é tomar $\mathbf{V} = \mathbb{R}^3$ e $\mathbf{W} \subset \mathbf{V}$, um plano passando pela origem, pois, pode-se verificar as condições (i) e (ii) do conceito 5.

Comenta-se agora, uma das mais importantes características de um espaço vetorial, a obtenção de novos vetores a partir de vetores dados (BOLDRINI et al., 1980).

Conceito 6 (Combinação Linear). *Seja \mathbf{V} um espaço vetorial sobre o corpo \mathbf{K} , com $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in \mathbf{V}$ e $a_1, a_2, \dots, a_n \in \mathbf{K}$. Então, o vetor*

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n$$

é um elemento de \mathbf{V} , chamado de combinação linear de $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$.

Fixando os vetores $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ em \mathbf{V} , o conjunto $[\mathbf{M}]$ de todos os vetores de \mathbf{V} que são combinações lineares destes é um subespaço vetorial denominado *subespaço gerado por $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$* , denotado como

$$[\mathbf{M}] = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$$

ou, simplesmente, $[\mathbf{M}]$.

Demonstra-se, então, que $[\mathbf{M}]$, assim caracterizado, é um subespaço vetorial.

Demonstração:

Para que $[\mathbf{M}]$ seja um subespaço vetorial, conforme Conceito 5, deve-se verificar 2 propriedades:

- (i) Para quaisquer $\mathbf{u}, \mathbf{v} \in [\mathbf{M}]$, $\mathbf{u} + \mathbf{v} \in [\mathbf{M}]$;
- (ii) Para quaisquer $k \in \mathbf{K}$ e $\mathbf{u} \in [\mathbf{M}]$, $k\mathbf{u} \in [\mathbf{M}]$.

Fixando $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n \in \mathbf{V}$, tomando quaisquer $\mathbf{u}, \mathbf{v} \in [\mathbf{M}]$, tal que

$$\mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n$$

$$\mathbf{v} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \dots + b_n\mathbf{v}_n$$

$\forall a_1, a_2, \dots, a_n$ e $b_1, b_2, \dots, b_n \in \mathbf{K}$ e fazendo $\mathbf{u} + \mathbf{v}$, conforme definido para espaços vetoriais, tem-se

$$\mathbf{u} + \mathbf{v} = (a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n) + (b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \dots + b_n\mathbf{v}_n)$$

$$\mathbf{u} + \mathbf{v} = (a_1 + b_1)\mathbf{v}_1 + (a_2 + b_2)\mathbf{v}_2 + \dots + (a_n + b_n)\mathbf{v}_n \in [\mathbf{M}],$$

o que verifica (i). Para verificar (ii), pode-se tomar qualquer $\mathbf{k} = (a_1, a_2, \dots, a_n) \in \mathbf{K}$ que satisfaz $k\mathbf{u} \in [\mathbf{M}]$ pela própria definição de combinação linear. Logo, se $[\mathbf{M}]$ é o conjunto de todos os vetores de \mathbf{V} que são combinações lineares destes, $[\mathbf{M}]$ é um subespaço vetorial. \blacklozenge

Luenberger (1969) comenta sobre uma característica interessante do subespaço gerado $[\mathbf{M}]$. Como se sabe, um subespaço vetorial deve conter o ve-

tor nulo (dito “passar pela origem”) para receber essa caracterização, porém, um plano que não passa pela origem pode gerar um espaço vetorial (basta tomar $a_1 = a_2 = \dots = a_n = 0$ para resolver o problema do vetor nulo, por exemplo).

Com base nessa característica, pode-se definir um importante conceito.

Conceito 7 (Variedade Linear). *Seja \mathbf{N} um subespaço vetorial. A translação de \mathbf{N} consiste em deslocar este subespaço vetorial de forma que o conjunto resultante \mathbf{V} não contenha o vetor nulo, denotada por $\mathbf{V} = \mathbf{x}_0 + \mathbf{N}$, com $\mathbf{x}_0 \neq \mathbf{0}$. Diz-se, então, que \mathbf{V} é uma variedade linear.*

Pela Figura 15, tomando $\mathbf{N} = \mathbb{R}^2$ como exemplo, pode-se obter uma variedade linear \mathbf{V} adicionando um vetor não nulo \mathbf{x}_0 a cada vetor \mathbf{n} de \mathbf{N} , tal que $\mathbf{v} = \mathbf{x}_0 + \mathbf{n}$.

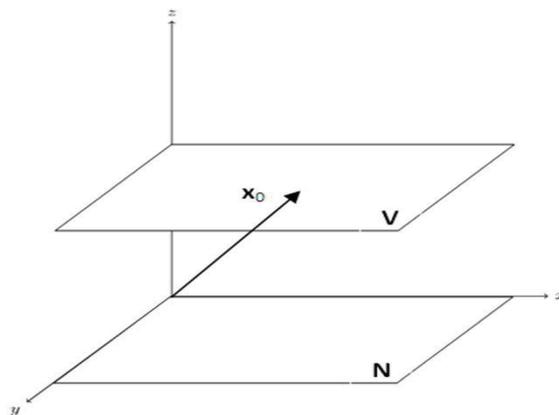


Figura 15 Representação geométrica de uma variedade linear para $\mathbf{N} = \mathbb{R}^2$

Até agora, no estudo de espaços vetoriais, não aparecem conceitos como “comprimento” e “ângulo” (importantes para discutir a ortogonalidade) de vetores,

e, finalmente, otimização. Assim, deseja-se introduzir mais uma estrutura sobre um espaço vetorial \mathbf{V} com o intuito de formalizar tais conceitos.

Conceito 8 (Produto Interno). *Seja \mathbf{V} um espaço vetorial. Um produto interno, denotado por $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle$ sobre \mathbf{V} é uma função que associa a cada par de vetores \mathbf{v}_1 e \mathbf{v}_2 de \mathbf{V} um número real, denotado por $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle : \mathbf{V} \times \mathbf{V} \rightarrow \mathbb{R}$, satisfazendo às propriedades:*

$$(i) \langle \mathbf{v}_i, \mathbf{v}_i \rangle \geq 0, \forall \mathbf{v}_i \in \mathbf{V}, \text{ e } \langle \mathbf{v}_i, \mathbf{v}_i \rangle = 0 \iff \mathbf{v}_i = \mathbf{0};$$

$$(ii) \langle k\mathbf{v}_1, \mathbf{v}_2 \rangle = k \langle \mathbf{v}_1, \mathbf{v}_2 \rangle, \forall k \in \mathbb{R} \text{ e } \forall \mathbf{v}_1, \mathbf{v}_2 \in \mathbf{V};$$

$$(iii) \langle \mathbf{v}_1 + \mathbf{v}_2, \mathbf{v}_3 \rangle = \langle \mathbf{v}_1, \mathbf{v}_3 \rangle + \langle \mathbf{v}_2, \mathbf{v}_3 \rangle, \forall \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in \mathbf{V};$$

$$(iv) \langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \langle \mathbf{v}_2, \mathbf{v}_1 \rangle, \forall \mathbf{v}_1, \mathbf{v}_2 \in \mathbf{V}.$$

Bierens (2007) afirma que um espaço vetorial \mathbf{V} com um produto interno é chamado de *espaço com produto interno*.

Entre os exemplos citados por Boldrini et al. (1980), destaca-se dois: o primeiro com $\mathbf{V} = \mathbb{R}^3$, em que $\mathbf{v}_1 = (x_1, y_1, z_1)$ e $\mathbf{v}_2 = (x_2, y_2, z_2)$. Então $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = x_1x_2 + y_1y_2 + z_1z_2$ e o segundo com \mathbf{V} sendo o espaço de funções contínuas no intervalo $[0, 1]$, com $\mathbf{v}_1 = f_1, \mathbf{v}_2 = f_2 \in V$, então $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle =$

$$\langle f_1, f_2 \rangle = \int_0^1 f_1(t)f_2(t)dt.$$

O produto interno é usado para, finalmente, caracterizar a noção de ortogonalidade de vetores. Formalmente, pode-se fazer a seguinte definição:

Conceito 9 (Ortogonalidade de Vetores). *Seja \mathbf{V} um espaço vetorial com produto interno $\langle \cdot, \cdot \rangle$. Diz-se que dois vetores $\mathbf{v}_1, \mathbf{v}_2$ de \mathbf{V} são ortogonais (em relação a este produto interno) se $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = 0$. Tal condição é denotada por $\mathbf{v}_1 \perp \mathbf{v}_2$.*

Boldrini et al. (1980) apresentam 5 propriedades da ortogonalidade de vetores. São elas:

- (i) $\mathbf{0} \perp \mathbf{v}, \forall \mathbf{v} \in \mathbf{V}$;
- (ii) $\mathbf{v}_1 \perp \mathbf{v}_2 \implies \mathbf{v}_2 \perp \mathbf{v}_1$;
- (iii) se $\mathbf{v}_1 \perp \mathbf{v}_2, \forall \mathbf{v}_2 \in \mathbf{V}$, então $\mathbf{v}_1 = \mathbf{0}$;
- (iv) se $\mathbf{v}_1 \perp \mathbf{w}$ e $\mathbf{v}_2 \perp \mathbf{w}$, então $\mathbf{v}_1 + \mathbf{v}_2 \perp \mathbf{w}$;
- (v) se $\mathbf{v}_1 \perp \mathbf{v}_2$ e k é um escalar, então $k\mathbf{v}_1 \perp \mathbf{v}_2$.

Quanto ao desejo de formalizar o conceito de comprimento de um vetor em um espaço vetorial com a estrutura de produto interno, conceitua-se *norma*.

Conceito 10 (Norma). *Seja \mathbf{V} um espaço com produto interno $\langle \cdot, \cdot \rangle$. Diz-se que a norma (ou comprimento) de um vetor \mathbf{v} de \mathbf{V} em relação a este produto interno é dado por $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$.*

Segundo Boldrini et al. (1980), se $\|\mathbf{v}\| = 1$, isto é, $\langle \mathbf{v}, \mathbf{v} \rangle = 1$, \mathbf{v} é chamado *vetor unitário*, ou, que \mathbf{v} está normalizado. Assim, todo vetor não nulo $\mathbf{v} \in \mathbf{V}$ pode ser normalizado, através de $\mathbf{u} = \frac{\mathbf{v}}{\|\mathbf{v}\|}$.

Bierens (2007), Boldrini et al. (1980) e Luenberger (1969) apresentam as propriedades da norma, generalização formal da noção de comprimento.

Conceito 11 (Propriedades da norma). *Seja \mathbf{V} um espaço vetorial com produto interno. Para quaisquer $\mathbf{v}, \mathbf{w} \in \mathbf{V}$ e $k \in \mathbb{R}$, valem:*

$$(i) \quad \|\mathbf{v}\| \geq 0 \text{ e } \|\mathbf{v}\| = 0 \iff \mathbf{v} = \mathbf{0};$$

$$(ii) \quad \|k\mathbf{v}\| = |k| \cdot \|\mathbf{v}\|;$$

$$(iii) \quad |\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \cdot \|\mathbf{w}\|;$$

$$(iv) \quad \|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|.$$

A propriedade (iii) do conceito 11 é chamada de *desigualdade de Cauchy-Schwarz*. Essa desigualdade pode ser demonstrada da seguinte forma:

Seja \mathbf{v} e \mathbf{w} dois vetores em \mathbf{V} . Se $\mathbf{v} = \mathbf{0}$, vale a igualdade de (iii). Para provar a desigualdade toma-se $\mathbf{v} \neq \mathbf{0}$. Para qualquer $t \in \mathbb{R}$,

$$\begin{aligned} \langle t\mathbf{v} + \mathbf{w}, t\mathbf{v} + \mathbf{w} \rangle &> 0 \Rightarrow \\ \Rightarrow \langle t\mathbf{v}, t\mathbf{v} \rangle + \langle t\mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{w}, t\mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle &> 0 \Rightarrow \\ \Rightarrow \langle \mathbf{v}, \mathbf{v} \rangle t^2 + \langle \mathbf{v}, \mathbf{w} \rangle t + \langle \mathbf{w}, \mathbf{v} \rangle t + \langle \mathbf{w}, \mathbf{w} \rangle &> 0 \Rightarrow \\ \Rightarrow \langle \mathbf{v}, \mathbf{v} \rangle t^2 + 2 \langle \mathbf{v}, \mathbf{w} \rangle t + \langle \mathbf{w}, \mathbf{w} \rangle &> 0 \end{aligned}$$

Como obteve-se um trinômio de 2º grau e o coeficiente $\langle \mathbf{v}, \mathbf{v} \rangle$ de t^2 é sempre positivo ($\mathbf{v} \neq \mathbf{0}$), Δ deve ser negativo, $\Delta = 4 \langle \mathbf{v}, \mathbf{w} \rangle^2 - 4 \langle \mathbf{v}, \mathbf{v} \rangle \langle \mathbf{w}, \mathbf{w} \rangle < 0$, ou seja, $4 \langle \mathbf{v}, \mathbf{w} \rangle^2 - 4 \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 < 0$. Logo $|\langle \mathbf{v}, \mathbf{w} \rangle| < \|\mathbf{v}\| \cdot \|\mathbf{w}\|$ (BOLDRINI et al., 1980). ◆

Já a propriedade (iv) é conhecida na literatura como *desigualdade triangular*, que Bierens (2007) demonstra assim:

$$\begin{aligned}
\|\mathbf{v} + \mathbf{w}\|^2 &= \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle \\
&= \langle \mathbf{v}, \mathbf{v} + \mathbf{w} \rangle + \langle \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle \\
&= \langle \mathbf{v}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle \\
&= \|\mathbf{v}\|^2 + 2\langle \mathbf{v}, \mathbf{w} \rangle + \|\mathbf{w}\|^2 \\
&\leq \|\mathbf{v}\|^2 + 2|\langle \mathbf{v}, \mathbf{w} \rangle| + \|\mathbf{w}\|^2 \\
&\leq \|\mathbf{v}\|^2 + 2\|\mathbf{v}\| \cdot \|\mathbf{w}\| + \|\mathbf{w}\|^2 \\
&\leq (\|\mathbf{v}\| + \|\mathbf{w}\|)^2 \Rightarrow \|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|. \quad \blacklozenge
\end{aligned}$$

Boldrini et al. (1980) observam que a desigualdade de Cauchy-Schwarz possibilita a definição de *ângulo* entre dois vetores não nulos em um espaço vetorial \mathbf{V} com produto interno. Então, sejam $\mathbf{v}, \mathbf{w} \in \mathbf{V}$, não-nulos. A desigualdade $|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \|\mathbf{w}\|$ pode ser escrita como

$$\frac{|\langle \mathbf{v}, \mathbf{w} \rangle|}{\|\mathbf{v}\| \cdot \|\mathbf{w}\|} = \left| \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \cdot \|\mathbf{w}\|} \right| \leq 1$$

e, portanto, existe um ângulo θ entre 0 e π radianos tal que

$$\cos \theta = \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\| \cdot \|\mathbf{w}\|}.$$

Este ângulo θ é chamado de ângulo entre \mathbf{v} e \mathbf{w} .

Bierens (2007) e Lipschutz (1994) definem *espaços vetoriais normados* com base nas propriedades da norma, conforme Conceito 12.

Conceito 12 (Espaço Vetorial Normado). *Seja \mathbf{V} um espaço vetorial em que a cada vetor de \mathbf{V} esteja associado um número real, denotado por $\|\mathbf{v}\|$. O espaço vetorial dotado com essa função $\|\cdot\| : \mathbf{V} \rightarrow [0, \infty)$, satisfazendo as 4 propriedades da norma, é chamado de espaço vetorial normado.*

Lipschutz (1994) comenta que, sendo \mathbf{V} um espaço com produto interno e a norma em \mathbf{V} definida por $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ pode-se verificar as propriedades da norma. Assim, todo espaço com produto interno é um espaço vetorial normado. Por outro lado, pode haver normas em um espaço vetorial \mathbf{V} que não vem de um produto interno em \mathbf{V} .

Para exemplificar, cita-se dois casos apresentados por Barata (2009) utilizando espaços vetoriais complexos. Seja $\mathbb{C}^n = \{(z_1, \dots, z_n), \forall z_1, \dots, z_n \in \mathbb{C}\}, n \geq 1$, o espaço vetorial complexo das n -uplas de números complexos. Para $\mathbf{z} = (z_1, \dots, z_n) \in \mathbb{C}^n$, a expressão

$$\|\mathbf{z}\|_1 := \sum_{k=1}^n |z_k|$$

define uma norma em \mathbb{C}^n . Outra expressão que também define uma norma em \mathbb{C}^n é dada por

$$\|\mathbf{z}\|_\infty := \max\{|z_1|, \dots, |z_n|\}.$$

Bierens (2007), com base neste conceito 12, mostra que a função $\|\cdot\|$ define um métrica $d(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|$ em \mathbf{V} , isto é, uma função que mede a distância entre \mathbf{v} e \mathbf{w} . Então, surge a necessidade do conceito a seguir.

Conceito 13 (Espaço Métrico). *Seja \mathbf{V} um espaço normado. Uma métrica é a função $d(\cdot, \cdot) : \mathbf{V} \times \mathbf{V} \rightarrow [0, \infty)$ que satisfaz às seguintes propriedades:*

(i) $d(\mathbf{v}, \mathbf{w}) = d(\mathbf{w}, \mathbf{v})$;

(ii) $d(\mathbf{v}, \mathbf{w}) > 0$, se $\mathbf{v} \neq \mathbf{w}$;

(iii) $d(\mathbf{v}, \mathbf{v}) = 0$;

(iv) $d(\mathbf{v}, \mathbf{u}) \leq d(\mathbf{v}, \mathbf{w}) + d(\mathbf{w}, \mathbf{u})$ (desigualdade triangular).

Diz-se, então, que um espaço normado com uma métrica definida é chamado de espaço métrico.

Domingues (1982) resume as 4 propriedades do Conceito 13 em 3 e denota um espaço métrico pelo par (\mathbf{V}, d) . Além disso, apresenta alguns exemplos de espaços métricos. Cita-se nesse trabalho apenas dois, a saber: (\mathbb{R}, d) com $d(v, w) = |v - w|$ e (\mathbb{R}^n, d) com $d(\mathbf{v}, \mathbf{w}) = \left(\sum_{i=1}^n |v_i - w_i|^2 \right)^{\frac{1}{2}}$.

Luenberger (1969) apresenta um conceito fundamental para o estudo das propriedades da topologia (ramo da matemática que estuda os espaços topológicos, sendo considerado uma extensão da geometria e onde se enquadra o estudo dos espaços de Hilbert).

Conceito 14 (Conjunto aberto). *Seja \mathbf{P} um subconjunto de um espaço normado \mathbf{V} . O vetor $\mathbf{p} \in \mathbf{P}$ é um ponto interior de \mathbf{P} se existir $\epsilon > 0$ tal que todo vetor \mathbf{v} de \mathbf{V} que satisfaz a $\|\mathbf{v} - \mathbf{p}\| < \epsilon$ pertence a \mathbf{P} . O conjunto de todos os pontos interiores é denotado por $\dot{\mathbf{P}}$ e se $\mathbf{P} = \dot{\mathbf{P}}$, então \mathbf{P} é um conjunto aberto.*

Uma notação válida para a compreensão do Conceito 14 é $S(\mathbf{v}, \epsilon)$, que simboliza uma bola “sem fronteira” de centro \mathbf{v} e raio ϵ que contem pontos (vetores) \mathbf{p} que também pertençam a \mathbf{P} , conforme exemplo representado na Figura 16, tomando $\mathbf{V} = \mathbb{R}^2$.

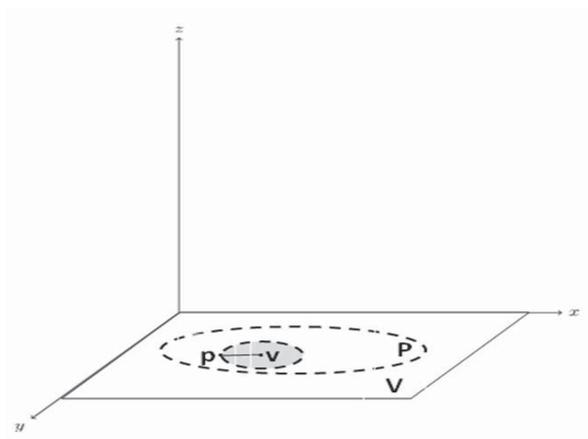


Figura 16 Representação geométrica de um conjunto aberto para $\mathbf{V} = \mathbb{R}^2$

Daí, apresenta-se o conceito de conjunto fechado, apresentado por Sallum, Murakami e Silva (1999), como um conjunto complementar do conjunto aberto.

Conceito 15 (Conjunto fechado). *Seja \mathbf{Q} um subconjunto de um espaço normado \mathbf{V} . \mathbf{Q} é fechado, se seu complementar \mathbf{Q}^c é aberto.*

Representa-se na Figura 17 um exemplo de conjunto fechado, tomando $\mathbf{V} = \mathbb{R}^2$.

Luenberger (1969) apresenta duas proposições (P_1) e (P_2) consequentes dos conceitos 14 e 15, a saber:

P_1 : a intersecção finita de conjuntos abertos é aberta;

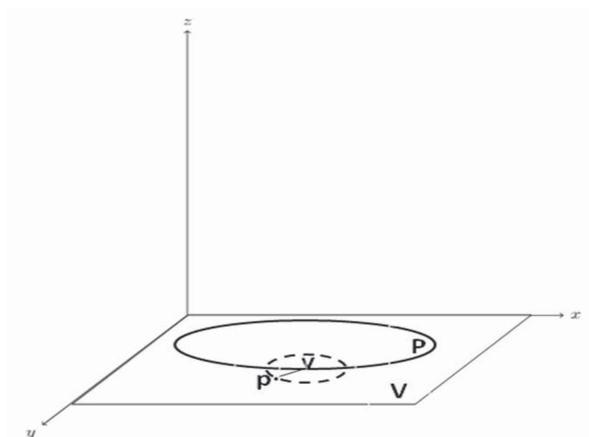


Figura 17 Representação geométrica de um conjunto fechado para $V = \mathbb{R}^2$

P_2 : a intersecção finita de conjuntos fechados é fechada.

Para chegar em um importante conceito, passa-se antes pelas definições de *sequências* e *sequências de Cauchy*, apresentadas a seguir.

Conceito 16 (Sequência). *Seja $\{a_n\}$ uma função cujo domínio é o conjunto dos números inteiros positivos e contradomínio um conjunto numérico qualquer, representada por $f : \mathbb{N}^* \rightarrow \mathbb{B}$, ou seja, cada número inteiro positivo n está associado com um número $f(n)$, ($n \mapsto f(n)$). A função, assim definida, é chamada de sequência.*

Observa-se, pelo Conceito 16, que a_n pode ser inclusive um vetor v_n .

Comumente, o conjunto $\mathbb{B} = \mathbb{R}$ e se, quando n cresce, a_n se torna cada vez mais próximo de um número real L , diz-se que a sequência $\{a_n\}$ tem limite L , ou converge para L , denotado por $\lim_{n \rightarrow \infty} a_n = L$. Observa-se que uma sequência que não é convergente, é chamada de divergente.

Luenberger (1969), porém, define assim: Seja \mathbf{V} um espaço normado. Uma sequência infinita de vetores $\{\mathbf{v}_n\}$ é dita convergir para um vetor \mathbf{v} se a sequência $\{\|\mathbf{v} - \mathbf{v}_n\|\}$ de números reais convergir para zero, denotado por $\mathbf{v}_n \rightarrow \mathbf{v}$.

Conceito 17 (Sequência de Cauchy). *Uma sequência $\{\mathbf{v}_n\}$ em um espaço normado é dita ser uma sequência de Cauchy se $\|\mathbf{v}_n - \mathbf{v}_m\| \rightarrow 0$ quando $n, m \rightarrow \infty$, isto é, dado $\epsilon > 0$, existe um inteiro positivo N tal que $\|\mathbf{v}_n - \mathbf{v}_m\| < \epsilon, \forall n, m > N$.*

Luenberger (1969) comenta que em um espaço normado, toda sequência convergente é uma sequência de Cauchy, visto que, se $\mathbf{v}_n \rightarrow \mathbf{v}$, então

$$\|\mathbf{v}_n - \mathbf{v}_m\| = \|\mathbf{v}_n - \mathbf{v} + \mathbf{v} - \mathbf{v}_m\| \leq \|\mathbf{v}_n - \mathbf{v}\| + \|\mathbf{v} - \mathbf{v}_m\| \rightarrow 0.$$

Entretanto, uma sequência de Cauchy pode não convergir.

Barata (2009) apresenta um exemplo clássico de uma sequência de \mathbb{Q} (conjunto dos números racionais) que é uma sequência de Cauchy em relação à métrica $d(\cdot, \cdot)$, mas que não converge em \mathbb{Q} . É o caso de

$$\{s_n\} = 1 + \frac{1}{1!} + \frac{1}{2!} + \dots + \frac{1}{n!}.$$

Neste trabalho, há um interesse particular em espaços normados em que as sequências de Cauchy são convergentes.

De forma geral, um espaço normado em que toda sequência de Cauchy é convergente é chamado de *espaço completo* ou de *Banach* (BARATA, 2009).

Conceito 18 (Espaço de Banach). *Um espaço vetorial normado V é completo se toda sequência de Cauchy de V tem um limite em V . Então, um espaço vetorial normado completo é chamado de espaço de Banach.*

Apesar da necessidade de certo cuidado em se utilizar espaços de Banach, por exemplo, em problemas de otimização, afirma Luenberger (1969), a principal vantagem em sua utilização é que, buscando um vetor ótimo (o que maximiza um dado objetivo), frequentemente uma sequência crescente de vetores é construída. Portanto, o vetor ótimo desejado é o limite da sequência. Quando o limite da sequência é desconhecido, verificar se o espaço vetorial é completo pode ser um teste de convergência da sequência de Cauchy.

Sabe-se dos fundamentos da geometria, conforme já citado, que a menor distância entre um ponto e uma reta é dada por um segmento de reta perpendicular que une os dois e que isso pode ser generalizado para a menor distância entre um ponto e um plano. Isso é, de fato, um caso especial de um dos mais importantes princípios da otimização, o *teorema da projeção*.

O conceito de ortogonalidade é fundamental para este objetivo (otimizar via projeção), porém, é um conceito que não está disponível em espaços normados gerais, onde a norma é definida de forma geral.

Assim, o espaço de Hilbert possibilita tal análise, pois é um caso especial dos espaços normados, aqueles com a estrutura de produto interno definida. Então, conforme já citado, dois vetores são definidos como ortogonais se seu produto interno é zero (LUENBERGER, 1969).

Este autor afirma ainda que os espaços de Hilbert, dotados com produto interno, possuem uma “riqueza” de propriedades que possibilitam a generalização

do espaço Euclidiano \mathbb{R}^n , implicando em análises em diversas áreas do conhecimento.

Conceito 19 (Espaço de Hilbert). *Um espaço de Hilbert H é um espaço vetorial dotado de um produto interno com norma e métrica associadas, tal que toda sequência de Cauchy em H tem um limite em H .*

Pode-se perceber, salienta Bierens (2007), o espaço Euclidiano \mathbb{R}^n é um espaço vetorial dotado de um produto interno $\langle \mathbf{x}, \mathbf{y} \rangle$, com norma, $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$, e métrica, $\|\mathbf{x} - \mathbf{y}\|$, associadas, tal que toda sequência de Cauchy tem um limite em \mathbb{R}^n . Então, isso faz de \mathbb{R}^n um espaço de Hilbert.

Este autor apresenta também, basicamente, qual a diferença entre um espaço de Banach e um espaço de Hilbert. Assim, a diferença é a definição de norma, ou seja, em um espaço de Hilbert a norma é definida via produto interno, ao passo que em um espaço de Banach a norma é definida de forma geral. Logo, um espaço de Hilbert é um caso particular de um espaço de Banach.

2.3.2 O teorema da projeção

A definição de ortogonalidade de vetores, apresentada no Conceito 9, tem muitas consequências diretas em espaços de Hilbert, sob o aspecto geométrico. Conforme exemplifica Luenberger (1969): o teorema de Pitágoras é uma dessas consequências porque é válido também nesses espaços, ou seja, se $\mathbf{v} \perp \mathbf{w}$, então $\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$.

Essa afirmação pode ser provada da seguinte forma:

$$\begin{aligned}
 \|\mathbf{v} + \mathbf{w}\|^2 &= \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle = \\
 &= \langle \mathbf{v}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle = \\
 &= \|\mathbf{v}\|^2 + 2 \underbrace{\langle \mathbf{w}, \mathbf{v} \rangle}_{\text{como } \mathbf{v} \perp \mathbf{w}} + \|\mathbf{w}\|^2 = \\
 &= \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2.
 \end{aligned}$$



Então, basicamente o que se deseja ao lidar com a otimização sob os aspectos de projeção é: dado um vetor \mathbf{v} que pertence a um espaço de Hilbert H e um subespaço $[\mathbf{M}] \subset \mathbf{H}$, deseja-se encontrar um vetor $\mathbf{m} \in [\mathbf{M}]$ que mais se aproxima de \mathbf{v} , ou seja, $\|\mathbf{v} - \mathbf{m}\|$ mínima.

Barata (2009) e Luenberger (1969) apresentam, de forma peculiar, 3 importantes questões que devem ser levantadas para se encontrar uma solução completa desse problema. Primeira, existe um vetor $\mathbf{m} \in [\mathbf{M}]$ que minimiza $\|\mathbf{v} - \mathbf{m}\|$ (ou, existem vetores melhores que outros)? Segunda, a solução é única? E, terceira, qual é a solução ou como é caracterizada?

Apresenta-se, então, dois teoremas essenciais para responder tais questões.

Teorema 2 (Vetor de Minimização). *Seja \mathbf{H} um espaço de Hilbert, $[\mathbf{M}]$ um subespaço de \mathbf{H} e \mathbf{v} um vetor arbitrário de \mathbf{H} . Se existir um vetor $\mathbf{m}_0 \in [\mathbf{M}]$ tal que $\|\mathbf{v} - \mathbf{m}_0\| \leq \|\mathbf{v} - \mathbf{m}\|$, $\forall \mathbf{m} \in [\mathbf{M}]$, então \mathbf{m}_0 é único. Uma condição necessária e suficiente para que $\mathbf{m}_0 \in [\mathbf{M}]$ seja o vetor único de minimização em $[\mathbf{M}]$ é que o vetor de resíduo $\mathbf{v} - \mathbf{m}_0$ seja ortogonal a $[\mathbf{M}]$.*

Demonstração:

Mostra-se, primeiro, que se \mathbf{m}_0 é o vetor desejado, então $\mathbf{v} - \mathbf{m}_0$ é ortogonal a $[\mathbf{M}]$. Para fazer isso, supõe-se que existe um $\mathbf{m} \in [\mathbf{M}]$ que não seja ortogonal a $\mathbf{v} - \mathbf{m}_0$. Então sem perda de generalidade e $\langle \mathbf{v} - \mathbf{m}_0, \mathbf{m} \rangle = \delta \neq 0$, pode-se assumir que $\|\mathbf{m}\| = 1$. Seja o vetor $\mathbf{m}_1 \in [\mathbf{M}]$ definido como $\mathbf{m}_1 := \mathbf{m}_0 + \delta \mathbf{m}$. Então

$$\begin{aligned} \|\mathbf{v} - \mathbf{m}_1\|^2 &= \|\mathbf{v} - \mathbf{m}_0 - \delta \mathbf{m}\|^2 = \\ &= \|\mathbf{v} - \mathbf{m}_0\|^2 - \langle \mathbf{v} - \mathbf{m}_0, \mathbf{m} \rangle - \underbrace{\langle \mathbf{v} - \mathbf{m}_0, \delta \mathbf{m} \rangle}_{\delta \langle \mathbf{v} - \mathbf{m}_0, \mathbf{m} \rangle} + \\ &\quad - \underbrace{\langle \delta \mathbf{m}, \mathbf{v} - \mathbf{m}_0 \rangle}_{\delta \langle \mathbf{v} - \mathbf{m}_0, \mathbf{m} \rangle} + \underbrace{\|\delta \mathbf{m}_1\|^2}_{|\delta|^2} = \\ &= \|\mathbf{v} - \mathbf{m}_0\|^2 - |\delta|^2 < \|\mathbf{v} - \mathbf{m}_0\|^2. \end{aligned}$$

Assim, se $\mathbf{v} - \mathbf{m}_0$ não é ortogonal a $[\mathbf{M}]$, \mathbf{m}_0 não é o vetor desejado.

Mostra-se, agora, que se $\mathbf{v} - \mathbf{m}_0$ é ortogonal a $[\mathbf{M}]$, então \mathbf{m}_0 é o vetor único de minimização. Para qualquer $\mathbf{m} \in [\mathbf{M}]$, o teorema de Pitágoras mostra que

$$\begin{aligned} \|\mathbf{v} - \mathbf{m}\|^2 &= \|\mathbf{v} - \mathbf{m}_0 + \mathbf{m}_0 - \mathbf{m}\|^2 = \\ &= \|\mathbf{v} - \mathbf{m}_0\|^2 + \|\mathbf{m}_0 - \mathbf{m}\|^2. \end{aligned}$$

Assim, $\|\mathbf{v} - \mathbf{m}\| > \|\mathbf{v} - \mathbf{m}_0\|$, para $\mathbf{m} \neq \mathbf{m}_0$. ◆

Esta demonstração não comprova a existência do vetor único de minimização. Mas, que, se existir, este é único e que $\mathbf{v} - \mathbf{m}_0$ é ortogonal ao subespaço $[\mathbf{M}]$. A Figura 18, ilustra o teorema para $\mathbf{H} = \mathbb{R}^3$.

Um conceito necessário à continuidade da abordagem do tema é o con-

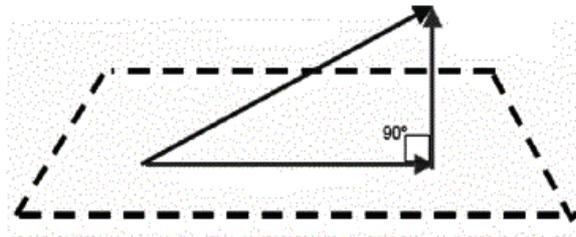


Figura 18 Representação do teorema da projeção quando $\mathbf{H} = \mathbb{R}^3$

Fonte: adaptado de Luenberger (1969)

ceito denominado *lei do paralelogramo*, que diz: em um espaço de Hilbert, vale a igualdade

$$\|\mathbf{v} + \mathbf{w}\|^2 + \|\mathbf{v} - \mathbf{w}\|^2 = 2\|\mathbf{v}\|^2 + 2\|\mathbf{w}\|^2.$$

Então, utilizando a expansão das normas em termos de produto interno, demonstra-se:

$$\begin{aligned} \|\mathbf{v} + \mathbf{w}\|^2 &= \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle = \\ &= \langle \mathbf{v}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle + \langle \mathbf{w}, \mathbf{w} \rangle = \\ &= \|\mathbf{v}\|^2 + 2\langle \mathbf{v}, \mathbf{w} \rangle + \|\mathbf{w}\|^2. \end{aligned}$$

Analogamente,

$$\|\mathbf{v} - \mathbf{w}\|^2 = \|\mathbf{v}\|^2 - 2\langle \mathbf{v}, \mathbf{w} \rangle + \|\mathbf{w}\|^2.$$

Daí

$$\|\mathbf{v} + \mathbf{w}\|^2 + \|\mathbf{v} - \mathbf{w}\|^2 = 2\|\mathbf{v}\|^2 + 2\|\mathbf{w}\|^2.$$

Teorema 3 (Projeção - Clássica). *Seja \mathbf{H} um espaço de Hilbert e $[\mathbf{M}]$ um subespaço fechado de \mathbf{H} . Para qualquer vetor $\mathbf{v} \in \mathbf{H}$, existe um único vetor $\mathbf{m}_0 \in [\mathbf{M}]$ tal que $\|\mathbf{v} - \mathbf{m}_0\| \leq \|\mathbf{v} - \mathbf{m}\|, \forall \mathbf{m} \in [\mathbf{M}]$. Além disso, uma condição necessária e suficiente para que $\mathbf{m}_0 \in [\mathbf{M}]$ seja o vetor único de minimização em $[\mathbf{M}]$ é que o vetor de resíduo $\mathbf{v} - \mathbf{m}_0$ seja ortogonal a $[\mathbf{M}]$.*

Demonstração:

A unicidade e ortogonalidade foram provadas no Teorema 2. Então, prova-se somente neste teorema a existência do vetor de minimização.

Se $\mathbf{v} \in [\mathbf{M}]$, $\mathbf{m}_0 = \mathbf{v}$ e a demonstração está completa. Assume-se, então, que $\mathbf{v} \notin [\mathbf{M}]$ e define-se $\delta = \inf_{\mathbf{m} \in [\mathbf{M}]} \|\mathbf{v} - \mathbf{m}\|$. Logo, deseja-se encontrar um $\mathbf{m}_0 \in [\mathbf{M}]$ com $\|\mathbf{v} - \mathbf{m}_0\| = \delta$.

Com este objetivo, seja $\{\mathbf{m}_i\}$ uma sequência de vetores de $[\mathbf{M}]$ tal que $\|\mathbf{v} - \mathbf{m}_i\| \rightarrow \delta$. Pela lei do paralelogramo

$$\begin{aligned} \|(\mathbf{m}_j - \mathbf{v}) + (\mathbf{v} - \mathbf{m}_i)\|^2 + \|(\mathbf{m}_j - \mathbf{v}) - (\mathbf{v} - \mathbf{m}_i)\|^2 &= \\ &= 2\|\mathbf{m}_j - \mathbf{v}\|^2 + 2\|\mathbf{v} - \mathbf{m}_i\|^2. \end{aligned}$$

Trabalhando só o primeiro membro da igualdade, tem-se

$$\|(\mathbf{m}_j - \mathbf{v}) + (\mathbf{v} - \mathbf{m}_i)\|^2 + \|(\mathbf{m}_j - \mathbf{v}) -$$

$$\begin{aligned}
&= \|(\mathbf{m}_j - \mathbf{v})^2 + 2(\mathbf{m}_j - \mathbf{v})(\mathbf{v} - \mathbf{m}_i) + \\
&+ (\mathbf{v} - \mathbf{m}_i)^2 + \|(\mathbf{m}_j - \mathbf{v})^2 - 2(\mathbf{m}_j - \mathbf{v})(\mathbf{v} - \mathbf{m}_i) + (\mathbf{v} - \mathbf{m}_i)^2\| = \\
&= \|\mathbf{m}_j^2 - 2\mathbf{v}\mathbf{m}_j + \mathbf{v}^2 + \mathbf{v}^2 - 2\mathbf{v}\mathbf{m}_i + \mathbf{m}_i^2\| + \\
&+ \|\mathbf{m}_j^2 - 2\mathbf{v}\mathbf{m}_j + \mathbf{v}^2 + \mathbf{v}^2 - 2\mathbf{v}\mathbf{m}_i + \mathbf{m}_i^2\| = \\
&= \|2\mathbf{m}_j^2 - 4\mathbf{v}\mathbf{m}_j + 4\mathbf{v}^2 - 4\mathbf{v}\mathbf{m}_i + 2\mathbf{m}_i^2\| = \\
&= \|4\mathbf{v}^2 - 4\mathbf{v}(\mathbf{m}_j + \mathbf{m}_i) + 2\mathbf{m}_j^2 + 2\mathbf{m}_i^2\| = \\
&= \|4\mathbf{v}^2 - 4\mathbf{v}(\mathbf{m}_j + \mathbf{m}_i) + \mathbf{m}_j^2 + \mathbf{m}_i^2 + \mathbf{m}_j^2 + \mathbf{m}_i^2 + 2\mathbf{m}_j\mathbf{m}_i - 2\mathbf{m}_j\mathbf{m}_i\| = \\
&= \|[2\mathbf{v} - (\mathbf{m}_j + \mathbf{m}_i)]^2 + (\mathbf{m}_j + \mathbf{m}_i)^2\| = \\
&= \left\| \left[2\mathbf{v} - 2\left(\frac{\mathbf{m}_j + \mathbf{m}_i}{2}\right) \right]^2 + (\mathbf{m}_j + \mathbf{m}_i)^2 \right\| = \\
&= 4 \left\| \mathbf{v} - \frac{\mathbf{m}_j + \mathbf{m}_i}{2} \right\|^2 + \|\mathbf{m}_j - \mathbf{m}_i\|^2.
\end{aligned}$$

Então, rearranjando a lei do paralelogramo, obtem-se

$$\|\mathbf{m}_j - \mathbf{m}_i\|^2 = 2\|\mathbf{m}_j - \mathbf{v}\|^2 + 2\|\mathbf{v} - \mathbf{m}_i\|^2 - 4 \left\| \mathbf{v} - \frac{\mathbf{m}_j + \mathbf{m}_i}{2} \right\|^2.$$

Pode-se perceber que $\frac{\mathbf{m}_j + \mathbf{m}_i}{2} \in [\mathbf{M}], \forall i, j$, desde que $[\mathbf{M}]$ seja um subespaço vetorial linear. Portanto, pela definição de δ , $\left\| \mathbf{v} - \frac{\mathbf{m}_j + \mathbf{m}_i}{2} \right\| \geq \delta$, logo

$$\|\mathbf{m}_j - \mathbf{m}_i\|^2 \leq 2\|\mathbf{m}_j - \mathbf{v}\|^2 + 2\|\mathbf{v} - \mathbf{m}_i\|^2 - 4\delta^2.$$

Uma vez que $\|\mathbf{v} - \mathbf{m}_i\|^2 \rightarrow \delta^2$ quando $i \rightarrow \infty$, conclui-se que $\|\mathbf{m}_j - \mathbf{m}_i\|^2 \rightarrow 0$, quando $i, j \rightarrow \infty$.

Portanto, $\{\mathbf{m}_i\}$ é uma sequência de Cauchy e uma vez que $[\mathbf{M}]$ é um subespaço vetorial fechado de um espaço vetorial completo, a sequência $\{\mathbf{m}_i\}$ tem um limite

\mathbf{m}_0 em $[\mathbf{M}]$, logo $\|\mathbf{v} - \mathbf{m}_0\| = \delta$. ◆

Deve-se observar, que, nem o texto do teorema nem a demonstração da existência do vetor de minimização, faz referência explícita à estrutura do produto interno, pois apenas a norma é utilizada. Porém, a demonstração faz uso frequente da lei do paralelogramo, logo, esta estrutura está implícita na existência do vetor, o que exige que \mathbf{H} é um espaço de Hilbert e não um espaço de Banach arbitrário.

O Teorema 3 é utilizado também para estabelecer algumas propriedades estruturais adicionais em espaços de Hilbert. O objetivo principal é mostrar que, dado qualquer subespaço fechado de um espaço de Hilbert, qualquer vetor pode ser escrito como uma soma de outros 2 vetores, sendo o primeiro pertencente ao subespaço dado e o outro ortogonal ao primeiro. Assim, surge o conceito de *Complementos Ortogonais*.

Conceito 20 (Complemento Ortogonal). *Dado um subconjunto \mathbf{S} de um espaço de Hilbert, o conjunto de todos os vetores ortogonais a \mathbf{S} é chamado de Complemento Ortogonal de \mathbf{S} , denotado por \mathbf{S}^\perp .*

Lipschutz (1994) e Luenberger (1969) comentam que o complemento ortogonal \mathbf{S}^\perp (lê-se: “ \mathbf{S} perp”) de um conjunto \mathbf{S} que consiste apenas do vetor nulo é todo o espaço vetorial. Além disso, para qualquer conjunto \mathbf{S} , \mathbf{S}^\perp é um subespaço fechado.

Outras propriedades são apresentadas por Luenberger (1969), resumindo as relações básicas existentes entre um conjunto \mathbf{S} e seu complemento ortogonal \mathbf{S}^\perp (Proposição 1).

Proposição 1 (Propriedades). *Sejam S e T subconjuntos de um espaço de Hilbert H . Então,*

(i) S^\perp é um subespaço fechado de H ;

(ii) $S \subseteq S^{\perp\perp}$;

(iii) Se $S \subseteq T$ então $T^\perp \subseteq S^\perp$;

(iv) $S^{\perp\perp\perp} = S^\perp$;

(v) $S^{\perp\perp}$ é o menor subespaço fechado que contém S .

Demonstração:

- (i) S^\perp é um subespaço porque $\mathbf{0} \in S^\perp$, pois $\mathbf{0}$ é ortogonal a todo vetor em H e uma combinação linear de vetores ortogonais a um conjunto também é ortogonal ao conjunto. Supondo $\mathbf{v}, \mathbf{w} \in S^\perp$, quaisquer escalares a e b e qualquer vetor $\mathbf{s} \in S$, tem-se: $\langle a\mathbf{v} + b\mathbf{w}, \mathbf{s} \rangle = \langle a\mathbf{v}, \mathbf{s} \rangle + \langle b\mathbf{w}, \mathbf{s} \rangle = a\langle \mathbf{v}, \mathbf{s} \rangle + b\langle \mathbf{w}, \mathbf{s} \rangle = a \cdot 0 + b \cdot 0 = 0$. Assim, $a\mathbf{v} + b\mathbf{w} \in S^\perp$, e, portanto, S^\perp é um subespaço de H . Além disso, S^\perp é fechado, pois se $\{\mathbf{x}_n\}$ é uma sequência convergente de S^\perp ($\mathbf{x}_n \rightarrow \mathbf{x}$), então $0 = \langle \mathbf{x}_n, \mathbf{s} \rangle \rightarrow \langle \mathbf{x}, \mathbf{s} \rangle, \forall \mathbf{s} \in S$, logo $\mathbf{x} \in S^\perp$.
- (ii) Seja $\mathbf{s} \in S$. Então $\langle \mathbf{v}, \mathbf{s} \rangle = 0, \forall \mathbf{v} \in S^\perp$, logo $\mathbf{s} \in S^{\perp\perp}$. Consequentemente, $S \subseteq S^{\perp\perp}$.
- (iii) Seja $\mathbf{w} \in T^\perp$ e \mathbf{v} qualquer vetor de T , então $\langle \mathbf{w}, \mathbf{v} \rangle = 0$. Como $S \subseteq T$, então $\langle \mathbf{w}, \mathbf{v} \rangle = 0, \forall \mathbf{v} \in S$. Assim, $\mathbf{w} \in S^\perp$ implicando em $T^\perp \subseteq S^\perp$.

(iv) De (ii) sabe-se que $S^\perp \subseteq S^{\perp\perp\perp}$ e $S \subseteq S^{\perp\perp}$ que de (iii), $S^{\perp\perp\perp} \subseteq S$.
Então, $S^\perp = S^{\perp\perp\perp}$.



Para apresentar o teorema que motiva a expressão *Complemento Ortogonal* de um conjunto de vetores ortogonais a um conjunto, deve-se antes conceituar *soma direta* de dois subespaços vetoriais. Ressalta-se, porém, que, doravante, denota-se os subespaços gerados por letras maiúsculas apenas (e não $[M]$), visando a simplicidade na notação.

Conceito 21 (Soma Direta). *Seja X um espaço vetorial. Diz-se que X é a soma direta de dois subespaços M e N se cada vetor $x \in X$ tem uma única representação dada por $x = m + n$, onde $m \in M$ e $n \in N$. A soma direta é denotada por $X = M \oplus N$.*

Teorema 4 (Projeção Ortogonal). *Se M é um subespaço linear fechado de um espaço de Hilbert H , então $H = M \oplus M^\perp$ e $M = M^{\perp\perp}$.*

Demonstração:

Seja $x \in H$. Pelo Teorema 3 sabe-se que existe um único vetor $m_0 \in M$ tal que $\|x - m_0\| \leq \|x - m\|$, $\forall m \in M$ e $n_0 = x - m_0 \in M^\perp$. Daí, $x = m_0 + n_0$ com $m_0 \in M$ e $n_0 \in M^\perp$.

Para mostrar que a representação $H = M \oplus M^\perp$ é única, supõe-se que $x = m_1 + n_1$ com $m_1 \in M$ e $n_1 \in M^\perp$. Então, fazendo $x - x$, $0 = m_1 - m_0 + n_1 - n_0$, mas $m_1 - m_0$ e $n_1 - n_0$ são ortogonais. Consequentemente, pelo teorema de Pitágoras, $\|0\|^2 = \|(m_1 - m_0) + (n_1 - n_0)\|^2 = \|m_1 - m_0\|^2 + \|n_1 - n_0\|^2$, que implica em $m_0 = m_1$ e $n_0 = n_1$. Logo, a representação é única.

Para mostrar que $M = M^{\perp\perp}$, é suficiente mostrar que $M^{\perp\perp} \subseteq M$, uma vez que, pela Proposição 1, item (ii), tem-se $M \subseteq M^{\perp\perp}$. Supondo que $x \in M^{\perp\perp}$ e $m \in M$, com $n \in M^\perp$, deseja-se mostrar que $n = \mathbf{0}$ para que $M = M^{\perp\perp}$. Se $n \in M^\perp$ então $m \in M^{\perp\perp}$. Mas, $x - m \in M^{\perp\perp}$, isto é, $n \in M^{\perp\perp}$. Portanto, $n \perp n$ porque pertence a M^\perp e a $M^{\perp\perp}$, o que implica $n = \mathbf{0}$. \blacklozenge

Luenberger (1969) faz dois comentários importantes sobre este teorema. O primeiro é: se o conjunto M é um subespaço em um espaço de Hilbert, seu complemento ortogonal M^\perp contém os demais vetores necessários para gerar o espaço. O segundo: dado um vetor $x \in H$, M um subespaço fechado de H e o vetor $m_0 \in M$, $x - m_0 \in M^\perp$ é então uma projeção ortogonal de x em M .

Lipschutz (1994) destaca também que um conjunto S de vetores em um espaço de Hilbert é chamado de ortogonal se cada dois vetores em S são ortogonais; e S é chamado ortonormal se S for ortogonal e cada vetor em S tem comprimento unitário. Isto é, $S = \{u_1, u_2, \dots, u_r\}$ é ortogonal se $\langle u_i, u_j \rangle = 0, \forall i \neq j$ e S é ortonormal se $\langle u_i, u_j \rangle = \begin{cases} 0, & \text{para } i \neq j \\ 1, & \text{para } i = j \end{cases}$

Como os conjuntos ortogonais são de grande importância neste trabalho, apresenta-se a Proposição 2.

Proposição 2. *Um conjunto ortogonal de vetores não nulos é um conjunto linearmente independente.*

Nota:

Entende-se conjunto *linearmente independente* (LI) como o conjunto cuja equação, envolvendo seus vetores, resulte no vetor nulo, ou seja, dado o conjunto $\{v_1, \dots, v_n\}$, diz-se ser LI se $a_1 v_1 + \dots + a_n v_n = \mathbf{0}$, o que implica em $a_1 =$

$a_2 = \dots = a_n = 0$. Caso existe algum $a_i \neq 0$, diz-se que $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ é linearmente dependente (LD) (BOLDRINI et al., 1980).

Demonstração:

Seja $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ um subconjunto finito de um dado conjunto ortogonal e escalares a_i , $i = 1, 2, \dots, n$, tais que $\sum_{i=1}^n a_i \mathbf{v}_i = \mathbf{0}$. Aplicando o produto interno de ambos os lados desta equação com o vetor não nulo \mathbf{v}_k , tem-se $\left\langle \sum_{i=1}^n a_i \mathbf{v}_i, \mathbf{v}_k \right\rangle = \langle \mathbf{0}, \mathbf{v}_k \rangle \Rightarrow \left\langle \sum_{i=1}^n a_i \mathbf{v}_i, \mathbf{v}_k \right\rangle = 0 \Rightarrow \langle a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \dots + a_n \mathbf{v}_n, \mathbf{v}_k \rangle = 0$. Então, de acordo com o Conceito 8, item (iii), esta igualdade só será válida se cada $a_i = 0$, $\forall i = 1, 2, \dots, n$. Logo, o conjunto é LI. \blacklozenge

Luenberger (1969) comenta que, em espaços de Hilbert, conjuntos ortonormais são desejáveis, em relação a outros conjuntos LI, principalmente porque facilmente são obtidos.

Segundo Ferreira (2008), uma forma de se obter um conjunto ortonormal a partir de qualquer conjunto ortogonal é chamado de *Ortonormalização de Gram-Schmidt*, conforme Teorema 5.

Teorema 5 (Gram-Schmidt). *Seja $\{\mathbf{x}_i\}$ uma sequência finita ou enumerável de vetores LI em um espaço de Hilbert \mathbf{H} . Então, existe uma sequência ortonormal $\{\mathbf{e}_i\}$ tal que cada espaço gerado pelos primeiros vetores \mathbf{e}_i , $i = 1, 2, \dots, n$, é igual ao espaço gerado pelos primeiros vetores \mathbf{x}_i , $i = 1, 2, \dots, n$, ou seja, a cada n , $[\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n] = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$.*

Demonstração:

Para o primeiro vetor, toma-se $\mathbf{e}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}$, que obviamente gera o mesmo espaço que \mathbf{x}_1 .

Para conceber o segundo vetor é necessário passar por duas etapas, ou seja, atribuir $\mathbf{x}_2 - \langle \mathbf{x}_2, \mathbf{e}_1 \rangle \mathbf{e}_1 = \mathbf{z}_2$ e tomar $\mathbf{e}_2 = \frac{\mathbf{z}_2}{\|\mathbf{z}_2\|}$. De forma direta, pode-se verificar que $\mathbf{z}_2 \perp \mathbf{e}_1$ e $\mathbf{e}_2 \perp \mathbf{e}_1$. O vetor \mathbf{z}_2 não pode ser nulo uma vez que \mathbf{x}_2 e \mathbf{e}_1 são LI; além disso, \mathbf{e}_1 e \mathbf{e}_2 geram o mesmo espaço que \mathbf{x}_1 e \mathbf{x}_2 , já que \mathbf{x}_2 pode ser expresso como uma combinação linear de \mathbf{e}_1 e \mathbf{e}_2 .

Os demais vetores \mathbf{e}_i 's são definidos por indução, ou seja,

$$\mathbf{x}_n - \sum_{i=1}^{n-1} \langle \mathbf{x}_n, \mathbf{e}_i \rangle \mathbf{e}_i = \mathbf{z}_n \text{ e } \mathbf{e}_n = \frac{\mathbf{z}_n}{\|\mathbf{z}_n\|}.$$

Analogamente, pode-se verificar que $\mathbf{z}_n \perp \mathbf{e}_i$, $\forall i < n$, e \mathbf{z}_n é não nulo já que é uma combinação linear de vetores LI. Também por indução \mathbf{e}_i , $i = 1, 2, \dots, n$, gera o mesmo espaço que \mathbf{x}_i , $i = 1, 2, \dots, n$.

Se $\{\mathbf{x}_i\}$ é finito, a ortonormalização termina, caso contrário o processo continua produzindo uma sequência infinita de vetores ortonormais. \blacklozenge

Segundo Luenberger (1969) este procedimento é melhor compreendido através do diagrama em \mathbb{R}^2 (Figura 19).

Adaptando um exemplo apresentado por Ferreira (2008) com $\mathbf{H} = \mathbb{R}^3$ e um conjunto LI $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} = \{(1, 0, 0), (1, 0, 1), (0, -1, 0)\}$, inicia-se, então, o processo de ortonormalização fazendo $\mathbf{e}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|} = 1 \cdot \mathbf{x}_1 = (1, 0, 0)$. Em seguida, $\mathbf{z}_2 = \mathbf{x}_2 - \langle \mathbf{x}_2, \mathbf{e}_1 \rangle \mathbf{e}_1 = (1, 0, 1) - 1 \cdot (1, 0, 0) = (0, 0, 1)$ e $\mathbf{e}_2 = \frac{\mathbf{z}_2}{\|\mathbf{z}_2\|} = 1 \cdot \mathbf{z}_2 = (0, 0, 1)$. E, finalmente, $\mathbf{z}_3 = \mathbf{x}_3 - \langle \mathbf{x}_3, \mathbf{e}_2 \rangle \mathbf{e}_2 - \langle \mathbf{x}_3, \mathbf{e}_1 \rangle \mathbf{e}_1 = (0, -1, 0) - 0 \cdot (0, 0, 1) - 0 \cdot (1, 0, 0) = (0, -1, 0)$ e $\mathbf{e}_3 = \frac{\mathbf{z}_3}{\|\mathbf{z}_3\|} = 1 \cdot \mathbf{z}_3 = (0, -1, 0)$. Assim, $[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3] = [\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$, ou seja, $[(1, 0, 0), (1, 0, 1), (0, -1, 0)] = [(1, 0, 0), (0, 0, 1), (0, -1, 0)]$, $\mathbf{z}_n \perp \mathbf{e}_i$, $\forall i < n$ e $\mathbf{e}_n \perp \mathbf{e}_i$, $\forall i < n$.

De forma natural surge um questionamento: supondo que $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$

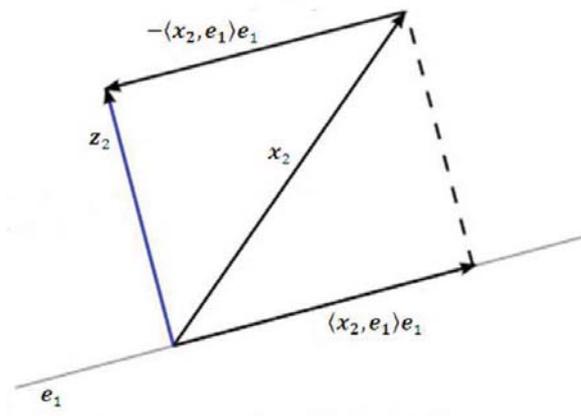


Figura 19 Diagrama em \mathbb{R}^2 da ortogonalização de Gram-Schmidt

Fonte: adaptado de Luenberger (1969)

são vetores de um espaço de Hilbert \mathbf{H} e que esses vetores geram um subespaço finito e fechado \mathbf{M} de \mathbf{H} , dado um vetor qualquer $\mathbf{x} \in \mathbf{H}$ é possível encontrar o vetor $\hat{\mathbf{x}}$ em \mathbf{M} que mais se aproxima de \mathbf{x} ?

Luenberger (1969), comentando sobre este problema, afirma que se $\hat{\mathbf{x}}$ pode ser expresso como combinação linear dos vetores \mathbf{y}_i , isto é, $\hat{\mathbf{x}} = \alpha_1 \mathbf{y}_1 + \alpha_2 \mathbf{y}_2 + \dots + \alpha_n \mathbf{y}_n$, o problema é equivalente a encontrar os escalares α_i , $\forall i = 1, 2, \dots, n$, que minimiza $\|\mathbf{x} - (\alpha_1 \mathbf{y}_1 + \alpha_2 \mathbf{y}_2 + \dots + \alpha_n \mathbf{y}_n)\|$.

Conforme apresentado nos Teoremas 2 e 3, o único vetor que resolve este problema é a projeção ortogonal de \mathbf{x} em \mathbf{M} , ou, de forma equivalente, é o vetor diferença $(\mathbf{x} - \hat{\mathbf{x}})$, ortogonal a cada vetor \mathbf{y}_i . Portanto,

$$\langle \mathbf{x} - \alpha_1 \mathbf{y}_1 - \alpha_2 \mathbf{y}_2 - \dots - \alpha_n \mathbf{y}_n, \mathbf{y}_i \rangle = 0. \quad (15)$$

A equação (15) pode ser expressa também por

$$\begin{aligned} \langle \mathbf{y}_1, \mathbf{y}_1 \rangle \alpha_1 + \langle \mathbf{y}_2, \mathbf{y}_1 \rangle \alpha_2 + \dots + \langle \mathbf{y}_n, \mathbf{y}_1 \rangle \alpha_n &= \langle \mathbf{x}, \mathbf{y}_1 \rangle \\ \langle \mathbf{y}_1, \mathbf{y}_2 \rangle \alpha_1 + \langle \mathbf{y}_2, \mathbf{y}_2 \rangle \alpha_2 + \dots + \langle \mathbf{y}_n, \mathbf{y}_2 \rangle \alpha_n &= \langle \mathbf{x}, \mathbf{y}_2 \rangle \\ \vdots & \quad \quad \quad \ddots \quad \quad \quad \vdots \\ \langle \mathbf{y}_1, \mathbf{y}_n \rangle \alpha_1 + \langle \mathbf{y}_2, \mathbf{y}_n \rangle \alpha_2 + \dots + \langle \mathbf{y}_n, \mathbf{y}_n \rangle \alpha_n &= \langle \mathbf{x}, \mathbf{y}_n \rangle. \end{aligned}$$

Essas equações são chamadas de *equações normais* de minimização.

Uma matriz de grande importância é a *Matriz de Gram-Schmidt*, dada por

$$G = G(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n) = \begin{bmatrix} \langle \mathbf{y}_1, \mathbf{y}_1 \rangle & \langle \mathbf{y}_1, \mathbf{y}_2 \rangle & \cdots & \langle \mathbf{y}_1, \mathbf{y}_n \rangle \\ \langle \mathbf{y}_2, \mathbf{y}_1 \rangle & \langle \mathbf{y}_2, \mathbf{y}_2 \rangle & \cdots & \langle \mathbf{y}_2, \mathbf{y}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{y}_n, \mathbf{y}_1 \rangle & \langle \mathbf{y}_n, \mathbf{y}_2 \rangle & \cdots & \langle \mathbf{y}_n, \mathbf{y}_n \rangle \end{bmatrix},$$

composta pelos coeficientes das equações normais, em sua forma transposta.

Ressalta-se também que o determinante $g = g(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$ dessa matriz, denominado *Determinante de Gram-Schmidt*, indica a possibilidade de o conjunto de equações ter solução única, isto é, $g \neq 0$, condição necessária e suficiente (LUENBERGER, 1969).

Proposição 3. *O determinante $g(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$ é diferente de zero, se e somente se, os vetores $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ são linearmente independentes.*

Demonstração:

Considerando a afirmação contrária, $g(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n) = 0$, se e somente se, os vetores $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ são L.D., supõe-se a princípio que \mathbf{y}_i 's sejam L.D., isto é, existem constantes α_i diferentes de zero, tal que $\sum_{i=1}^n \alpha_i \mathbf{y}_i = \mathbf{0}$. Então fica

claro que as linhas do determinante g tem uma dependência correspondente e, conseqüentemente, o determinante é zero.

Assumindo agora que o determinante é zero. Nesse caso existem constantes α_i , diferentes de zero, tal que $\sum_{i=1}^n \alpha_i \langle \mathbf{y}_i, \mathbf{y}_j \rangle = 0, \forall j$. Aplicando (ii) do Conceito 8, tem-se

$$\begin{aligned} \sum_{i=1}^n \alpha_i \langle \mathbf{y}_i, \mathbf{y}_j \rangle &= 0 \\ \left\langle \sum_{i=1}^n \alpha_i \mathbf{y}_i, \mathbf{y}_j \right\rangle &= 0 \\ \sum_{j=1}^n \bar{\alpha}_j \left\langle \sum_{i=1}^n \alpha_i \mathbf{y}_i, \mathbf{y}_j \right\rangle &= \sum_{j=1}^n \bar{\alpha}_j \cdot 0 \\ \sum_{i=1}^n \alpha_i \mathbf{y}_i \left\langle \sum_{j=1}^n \bar{\alpha}_j, \mathbf{y}_j \right\rangle &= 0 \end{aligned}$$

Como $\left\langle \sum_{j=1}^n \bar{\alpha}_j, \mathbf{y}_j \right\rangle \neq 0$ então $\sum_{i=1}^n \alpha_i \mathbf{y}_i = \mathbf{0}$ e os vetores $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ são L.D..

Diante do exposto, conclui-se que $g(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n) \neq 0 \Leftrightarrow \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ são L.I.. ◆

O Teorema 6 apresenta um importante resultado na tarefa de buscar uma distância mínima entre um vetor \mathbf{x} e o subespaço \mathbf{M} .

Teorema 6. *Seja $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ linearmente independentes e δ a menor distância entre um vetor \mathbf{x} e o subespaço gerado por \mathbf{y}_i 's, i.e.,*

$$\delta = \min \|\mathbf{x} - \alpha_1 \mathbf{y}_1 - \alpha_2 \mathbf{y}_2 - \dots - \alpha_n \mathbf{y}_n\| = \|\mathbf{x} - \hat{\mathbf{x}}\|.$$

Então

$$\delta^2 = \frac{g(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n, \mathbf{x})}{g(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)}.$$

Demonstração:

$$\begin{aligned} &\text{Por definição (Conceito 10), } \delta^2 = \|\mathbf{x} - \hat{\mathbf{x}}\|^2 = \langle \mathbf{x} - \hat{\mathbf{x}}, \mathbf{x} - \hat{\mathbf{x}} \rangle \\ &= \langle \mathbf{x} - \hat{\mathbf{x}}, \mathbf{x} \rangle - \langle \mathbf{x} - \hat{\mathbf{x}}, \hat{\mathbf{x}} \rangle. \end{aligned}$$

Pelo Teorema 3, $\mathbf{x} - \hat{\mathbf{x}}$ é ortogonal a \mathbf{M} , então, em particular, $\langle \mathbf{x} - \hat{\mathbf{x}}, \hat{\mathbf{x}} \rangle$ é zero. Portanto,

$$\begin{aligned} \delta^2 &= \langle \mathbf{x} - \hat{\mathbf{x}}, \mathbf{x} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle - \langle \hat{\mathbf{x}}, \mathbf{x} \rangle = \\ &= \langle \mathbf{x}, \mathbf{x} \rangle - \alpha_1 \langle \mathbf{y}_1, \mathbf{x} \rangle - \alpha_2 \langle \mathbf{y}_2, \mathbf{x} \rangle - \dots - \alpha_n \langle \mathbf{y}_n, \mathbf{x} \rangle \end{aligned}$$

e, ainda,

$$\alpha_1 \langle \mathbf{y}_1, \mathbf{x} \rangle + \alpha_2 \langle \mathbf{y}_2, \mathbf{x} \rangle + \dots + \alpha_n \langle \mathbf{y}_n, \mathbf{x} \rangle + \delta^2 = \langle \mathbf{x}, \mathbf{x} \rangle. \quad (16)$$

Juntando a equação (16) com as equações normais, tem-se $n + 1$ equações lineares de $n + 1$ termos desconhecidos, $\alpha_1, \alpha_2, \dots, \alpha_n, \delta^2$.

Pela regra de Cramer pode-se determinar δ^2 por

$$\delta^2 = \frac{\begin{vmatrix} \langle \mathbf{y}_1, \mathbf{y}_1 \rangle & \langle \mathbf{y}_2, \mathbf{y}_1 \rangle & \cdots & \langle \mathbf{y}_n, \mathbf{y}_1 \rangle & \langle \mathbf{x}, \mathbf{y}_1 \rangle \\ \langle \mathbf{y}_1, \mathbf{y}_2 \rangle & \langle \mathbf{y}_2, \mathbf{y}_2 \rangle & \cdots & \langle \mathbf{y}_n, \mathbf{y}_2 \rangle & \langle \mathbf{x}, \mathbf{y}_2 \rangle \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \langle \mathbf{y}_1, \mathbf{y}_n \rangle & \langle \mathbf{y}_2, \mathbf{y}_n \rangle & \cdots & \langle \mathbf{y}_n, \mathbf{y}_n \rangle & \langle \mathbf{x}, \mathbf{y}_n \rangle \\ \langle \mathbf{y}_1, \mathbf{x} \rangle & \langle \mathbf{y}_2, \mathbf{x} \rangle & \cdots & \langle \mathbf{y}_n, \mathbf{x} \rangle & \langle \mathbf{x}, \mathbf{x} \rangle \end{vmatrix}}{\begin{vmatrix} \langle \mathbf{y}_1, \mathbf{y}_1 \rangle & \langle \mathbf{y}_2, \mathbf{y}_1 \rangle & \cdots & \langle \mathbf{y}_n, \mathbf{y}_1 \rangle & 0 \\ \langle \mathbf{y}_1, \mathbf{y}_2 \rangle & \langle \mathbf{y}_2, \mathbf{y}_2 \rangle & \cdots & \langle \mathbf{y}_n, \mathbf{y}_2 \rangle & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \langle \mathbf{y}_1, \mathbf{y}_n \rangle & \langle \mathbf{y}_2, \mathbf{y}_n \rangle & \cdots & \langle \mathbf{y}_n, \mathbf{y}_n \rangle & 0 \\ \langle \mathbf{y}_1, \mathbf{x} \rangle & \langle \mathbf{y}_2, \mathbf{x} \rangle & \cdots & \langle \mathbf{y}_n, \mathbf{x} \rangle & 1 \end{vmatrix}} = \frac{g(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n, \mathbf{x})}{g(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)}$$



Avaliando este teorema, pode-se perceber que, embora o interesse teórico por δ seja grande, a prática depara-se com a inviabilidade de se calcular determinantes de matrizes muito grandes. Calcular um determinante de uma matriz de dimensão $n + 1$ é, em geral, tão difícil como inverter uma matriz de dimensão n . Portanto, a razão dada por $\delta^2 = \frac{g(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n, \mathbf{x})}{g(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)}$ não apresenta vantagens, sobretudo, computacionais.

Luenberger (1969) apresenta uma alternativa para essa dificuldade encontrada, ou seja, ao invés de se obter $\hat{\mathbf{x}}$ de uma combinação linear de \mathbf{y}_i 's pela resolução das equações normais, pode-se utilizar a ortogonalização de Gram-Schmidt juntamente com séries de Fourier.

Elucidando, aplica-se, primeiramente, o procedimento de Gram-Schmidt

ao conjunto $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$, obtendo assim um conjunto ortonormal $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$, logo, gera-se M . Em seguida, o vetor $\hat{\mathbf{x}}$ é dado por séries de Fourier,

$$\hat{\mathbf{x}} = \sum_{i=1}^n \langle \mathbf{x}, \mathbf{e}_i \rangle \mathbf{e}_i$$

já que $\mathbf{x} - \hat{\mathbf{x}}$ é ortogonal a M .

Assim, $\hat{\mathbf{x}}$ pode ser facilmente obtido, já que os vetores independentes \mathbf{y}_i 's estão ortonormalizados.

Essas ressalvas são importantes porque mostram que o procedimento de Gram-Schmidt pode ser usado para resolver o problema de obtenção de δ^2 , usando o teorema da projeção. Além disso, é interessante notar que o procedimento pode ser visto também como uma aproximação. Por exemplo, dada uma sequência $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$ de vetores independentes, o procedimento de Gram-Schmidt define

$$\mathbf{e}_1 = \frac{\mathbf{y}_1}{\|\mathbf{y}_1\|}; \dots; \mathbf{e}_k = \frac{\mathbf{y}_k - \sum_{i=1}^{k-1} \langle \mathbf{y}_k, \mathbf{e}_i \rangle \mathbf{e}_i}{\left\| \mathbf{y}_k - \sum_{i=1}^{k-1} \langle \mathbf{y}_k, \mathbf{e}_i \rangle \mathbf{e}_i \right\|},$$

em que o vetor $\mathbf{y}_k - \sum_{i=1}^{k-1} \langle \mathbf{y}_k, \mathbf{e}_i \rangle \mathbf{e}_i$ é o erro ótimo ($\min \|\mathbf{y}_k - \hat{\mathbf{x}}\|$), mínimo de todos os $\hat{\mathbf{x}}$ no subespaço $[\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{k-1}] = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{k-1}]$. O vetor \mathbf{e}_k , por sua vez, é a normalização desse erro.

Por conseguinte, a minimização de distância entre qualquer vetor \mathbf{x} e o subespaço $[\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]$ pode ser encontrada pela ortogonalização de Gram-Schmidt aplicada a uma sequência do tipo $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n, \mathbf{x}\}$, onde o erro ótimo ($\min \|\mathbf{x} - \hat{\mathbf{x}}\|$) é encontrado na última etapa do procedimento.

Contudo, apesar da obtenção do erro ótimo, ou seja, a menor distância entre um vetor qualquer de um espaço de Hilbert e um vetor de um subespaço de dimensão finita, através da solução de equações normais, em algumas situações o subespaço não é finito, tornando, geralmente, impossível reduzir as equações normais a um conjunto finito.

Porém, Luenberger (1969) apresenta, uma importante classe de problemas que podem ser reduzidos a um conjunto finito de equações lineares similares às equações normais. Dessa forma, apresenta-se, primeiramente, o Teorema 7 que, com uma modificação trivial do Teorema 3, torna-se aplicável à variedades lineares.

Teorema 7 (Projeção - Modificada). *Seja \mathbf{N} um subespaço fechado de um espaço de Hilbert \mathbf{H} , \mathbf{x} um vetor de \mathbf{H} e \mathbf{V} a variedade linear $\mathbf{x} + \mathbf{N}$. Então, há um único vetor $\mathbf{x}_0 \in \mathbf{V}$ de norma mínima. Além disso, \mathbf{x}_0 é ortogonal a \mathbf{N} .*

Demonstração:

Prova-se este teorema adicionando $-\mathbf{x}$ a \mathbf{V} , para que se torne um subespaço fechado e assim, valer o teorema da projeção clássica. ◆

Observa-se aqui, que a solução \mathbf{x}_0 não é ortogonal à variedade linear \mathbf{V} , mas, ao subespaço \mathbf{N} em que \mathbf{V} é obtido.

Além disso, são de interesse particular dois tipos de variedade linear. O primeiro tipo é aquele que consiste de elementos na forma $\mathbf{x} + \sum_{i=1}^n \alpha_i \mathbf{x}_i$, onde $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ é um conjunto LI em \mathbf{H} e \mathbf{x} é um vetor fixo em \mathbf{H} . Este tipo é interessante porque pode-se reduzir uma variedade linear n -dimensional em um sistema de equações normais n -dimensional. Já o segundo tipo consiste de todos

os vetores \mathbf{x} em um espaço de Hilbert \mathbf{H} que satisfazem às condições

$$\langle \mathbf{x}, \mathbf{y}_1 \rangle = c_1$$

$$\langle \mathbf{x}, \mathbf{y}_2 \rangle = c_2$$

...

$$\langle \mathbf{x}, \mathbf{y}_n \rangle = c_n$$

onde $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$ são vetores LI em \mathbf{H} e c_1, c_2, \dots, c_n são constantes fixas. Se \mathbf{N} for denotado como um subespaço gerado $[\mathbf{N}]$, fica evidente que cada $c_i = 0$ (Conceito 6), então a variedade linear é um subespaço, ou, mais especificamente, \mathbf{N}^\perp . Este tipo de variedade linear é interessante porque o subespaço em questão tem dimensão n e a variedade é então denominada de *variedade linear de codimensão n* (CHONG; ZAK, 2001; LUENBERGER, 1969).

Então, pode-se analisar a questão da norma mínima, neste caso, procurando por um vetor bem próximo da origem que pertença a uma variedade linear de codimensão finita.

Teorema 8 (Norma mínima em variedade linear). *Seja \mathbf{H} um espaço de Hilbert e $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$ um conjunto de vetores independentes em \mathbf{H} . Entre todos os vetores $\mathbf{x} \in \mathbf{H}$, que satisfazem*

$$\langle \mathbf{x}, \mathbf{y}_1 \rangle = c_1$$

$$\langle \mathbf{x}, \mathbf{y}_2 \rangle = c_2$$

...

$$\langle \mathbf{x}, \mathbf{y}_n \rangle = c_n$$

\mathbf{x}_0 é o que tem norma mínima. Então

$$\mathbf{x}_0 = \sum_{i=1}^n \beta_i \mathbf{y}_i$$

onde os coeficientes β_i satisfazem as equações

$$\langle \mathbf{y}_1, \mathbf{y}_1 \rangle \beta_1 + \langle \mathbf{y}_2, \mathbf{y}_1 \rangle \beta_2 + \dots + \langle \mathbf{y}_n, \mathbf{y}_1 \rangle \beta_n = c_1$$

$$\langle \mathbf{y}_1, \mathbf{y}_2 \rangle \beta_1 + \langle \mathbf{y}_2, \mathbf{y}_2 \rangle \beta_2 + \dots + \langle \mathbf{y}_n, \mathbf{y}_2 \rangle \beta_n = c_2$$

...

$$\langle \mathbf{y}_1, \mathbf{y}_n \rangle \beta_1 + \langle \mathbf{y}_2, \mathbf{y}_n \rangle \beta_2 + \dots + \langle \mathbf{y}_n, \mathbf{y}_n \rangle \beta_n = c_n.$$

Demonstração:

Seja \mathbf{N} o subespaço n -dimensional gerado pelos vetores \mathbf{y}_i e a variedade linear \mathbf{V} uma translação do subespaço \mathbf{N}^\perp . Uma vez que \mathbf{N}^\perp é fechado, a existência e a unicidade de uma solução ótima decorre do Teorema 7. Além disso, a solução ótima \mathbf{x}_0 é ortogonal a \mathbf{N}^\perp e, portanto, $\mathbf{x}_0 \in \mathbf{N}^{\perp\perp}$. Entretanto, uma vez que \mathbf{N} é fechado, $\mathbf{N}^{\perp\perp}$, $\mathbf{N}^{\perp\perp} = \mathbf{N}$. Conclui-se, então, que $\mathbf{x}_0 \in \mathbf{N}^{\perp\perp}$ ou, equivalente-

mente,

$$\mathbf{x}_0 = \sum_{i=1}^n \beta_i \mathbf{y}_i.$$

Os coeficientes β_i são escolhidos de modo que \mathbf{x}_0 satisfaça as n restrições das equações apresentadas no teorema. \blacklozenge

2.4 Espaços de Hilbert de variáveis aleatórias

Para a compreensão dessa seção, alguns conceitos elementares são necessários, tais como: *variáveis aleatórias, distribuição de probabilidade, valor esperado, matriz de covariância, etc.*

Revisando brevemente tais conceitos, destaca-se autores como Gnedenko (2008), Magalhães (2006) e Mood, Graybill e Boes (1974).

Conceito 22 (Variável Aleatória). *Para um dado espaço de probabilidade (Ω, \mathbb{A}, P) , uma variável aleatória, denotada por \mathbf{Y} , é uma função com domínio em Ω e contradomínio em \mathbb{R} (reta real) tal que o conjunto A_r , definido por $A_r = \{\omega : \mathbf{Y}(\omega) \leq r\}$, pertença a \mathbb{A} para cada número real r .*

Ressalta-se a convenção de que as variáveis aleatórias são, em geral, representadas por letras maiúsculas do fim do alfabeto, sendo que para cada elemento $\omega \in \Omega$ tem-se um número real $\mathbf{Y}(\omega)$ e que o cálculo de probabilidades é garantido ao se exigir que qualquer intervalo de \mathbb{R} tem um conjunto associado em Ω que seja um evento (matematicamente, diz-se que a variável aleatória é uma função real mensurável em \mathbb{A}).

Conceito 23 (Função de Distribuição). *Uma função de uma variável aleatória \mathbf{Y} , denotada por $\mathbf{F}_{\mathbf{Y}}(\cdot)$, tendo como domínio a reta real e contradomínio o intervalo $[0, 1]$, satisfazendo $\mathbf{F}_{\mathbf{Y}}(y) = P[\mathbf{Y} \leq y] = P[\{\omega : \mathbf{Y}(\omega) \leq y\}]$ para cada número real y , recebe o nome de função de distribuição.*

Uma função de distribuição de uma variável aleatória \mathbf{Y} em (Ω, \mathbb{A}, P) deve obedecer às seguintes propriedades:

- a) $\lim_{y \rightarrow -\infty} \mathbf{F}_{\mathbf{Y}}(y) = 0$ e $\lim_{y \rightarrow \infty} \mathbf{F}_{\mathbf{Y}}(y) = 1$;
- b) $\mathbf{F}_{\mathbf{Y}}(\cdot)$ é contínua à direita;
- c) $\mathbf{F}_{\mathbf{Y}}(\cdot)$ é não decrescente, isto é, $\mathbf{F}_{\mathbf{Y}}(y) \leq \mathbf{F}_{\mathbf{Y}}(z)$ sempre que $y \leq z, \forall y, z \in \mathbb{R}$.

Outro conceito importante e que merece destaque por ser uma quantidade frequentemente utilizada como resumo do comportamento de uma variável aleatória é o *valor esperado*, também chamado de média ou esperança de uma variável aleatória.

Conceito 24 (Valor esperado). *Seja \mathbf{Y} uma variável aleatória. O valor esperado de \mathbf{Y} , denotado por $E[\mathbf{Y}]$, é definido por*

$$E[\mathbf{Y}] = \sum_i y_i f_{\mathbf{Y}}(y_i)$$

se \mathbf{Y} é discreto com valores $y_1, y_2, \dots, y_i, \dots$;

$$E[\mathbf{Y}] = \int_{-\infty}^{\infty} y f_{\mathbf{Y}}(y) dy$$

se \mathbf{Y} é contínuo com função densidade de probabilidade $f_{\mathbf{Y}}(y)$.

Um resultado muito útil que permite calcular o valor esperado de uma função de variável aleatória, sem ser necessário obter a distribuição de probabilidade dessa função é dado pelo Teorema 9.

Teorema 9 (Esperança de função de variável aleatória). *Seja \mathbf{Y} uma variável aleatória cujo valor esperado existe e $\mathbf{Y} = g(\mathbf{X})$ uma função de uma variável aleatória \mathbf{X} no mesmo espaço de probabilidade. Então*

$$E[g(\mathbf{X})] = \int_0^{\infty} [1 - \mathbf{F}_{\mathbf{X}}(x)] dx - \int_{-\infty}^0 \mathbf{F}_{\mathbf{X}}(x) dx.$$

A demonstração desse teorema pode ser encontrada em Magalhães (2006), para os casos em que \mathbf{Y} é discreta ou contínua, e em Ash (1972) para o caso geral.

Conforme citado acima, g é uma função de uma variável \mathbf{Y} que tem função de distribuição $\mathbf{F}_{\mathbf{Y}}$, logo, interessa-se em $E[g(\mathbf{Y})]$, que poderia ser obtida

do modo usual, se a função de distribuição da variável $g(\mathbf{X})$, isto é, $\mathbf{F}_g(y) = P[g(\mathbf{X}) \leq y]$ fosse conhecida. Como tal conhecimento demandaria muito esforço, o Teorema 9 pode ser utilizado na obtenção do valor esperado.

Como exemplo dessa afirmação, cita-se o caso de uma variável aleatória \mathbf{Y} , cuja função densidade de probabilidade é dada por

$$f_X(x) = \begin{cases} \frac{1}{4}(x+2), & -2 \leq y < 0; \\ \frac{1}{2}, & 0 \leq y < 1; \\ 0, & \text{cc.} \end{cases}$$

Considerando $g(\mathbf{X}) = \mathbf{X}^2$, para obter $E[g(\mathbf{X})]$, tem-se:

$$\begin{aligned} E[g(\mathbf{X})] &= E[\mathbf{X}^2] = \int_{-\infty}^{\infty} g(x) f_{\mathbf{X}}(x) dx \\ &= \int_{-2}^0 x^2 \frac{1}{4}(x+2) dx + \int_0^1 x^2 \frac{1}{2} dx \\ &= \frac{1}{4} \left(\frac{x^4}{4} + \frac{2x^3}{3} \right) \Big|_{-2}^0 + \frac{x^3}{6} \Big|_0^1 \\ &= \frac{1}{2}. \end{aligned}$$

Portanto, o uso do Teorema 9 evitou o conhecimento da densidade de \mathbf{X}^2 , facilitando o resultado desejado.

Com base no valor esperado, pode-se conhecer outras medidas (em relação à origem) de variáveis aleatórias unidimensionais, tais como: $E[\mathbf{X}^2]$ - o segundo momento ordinário de \mathbf{X} ; $E[(\mathbf{X} - E(\mathbf{X}))^2]$ - a variância de \mathbf{Y} .

Considerando uma coleção finita de variáveis aleatórias $\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n\}$, pode-se definir sua distribuição conjunta de probabilidade.

Conceito 25 (Função de Distribuição Conjunta). *Seja $\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n\}$ variáveis aleatórias definidas no mesmo espaço de probabilidade (Ω, \mathbb{A}, P) . A função de distribuição conjunta de $\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n\}$, denotada por $\mathbf{F}_{\mathbf{Y}_1, \dots, \mathbf{Y}_n}(\cdot, \dots, \cdot)$, é definida como $P[\mathbf{Y}_1 \leq y_1, \mathbf{Y}_2 \leq y_2, \dots, \mathbf{Y}_n \leq y_n]$ para todo (y_1, \dots, y_n) .*

Pode-se perceber, então, que uma função de distribuição conjunta é uma função com domínio no espaço euclidiano n -dimensional e contradomínio igual ao intervalo $[0, 1]$.

Já o valor esperado de qualquer função g de $\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n\}$ é definida como

$$E[g(\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n\})] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(y_1, \dots, y_n) f_{\mathbf{Y}_1, \dots, \mathbf{Y}_n}(y_1, \dots, y_n) dy_1 \dots dy_n$$

e a matriz $(n \times n)$ de covariância $cov(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n)$ cujo ij -ésimo elemento é definido como

$$E\{[\mathbf{Y}_i - E[\mathbf{Y}_i]][\mathbf{Y}_j - E[\mathbf{Y}_j]]\}$$

para $i \neq j$ que, em casos de média zero, reduz-se a $E[\mathbf{Y}_i \mathbf{Y}_j]$. Lembrando também que duas variáveis aleatórias são ditas não correlacionadas se $E[\mathbf{Y}_i \mathbf{Y}_j] = E[\mathbf{Y}_i]E[\mathbf{Y}_j]$ ou se o ij -ésimo elemento da matriz de covariância é nulo.

Assim, com base nesses conceitos, pode-se abordar o tema da presente seção, *Espaços de Hilbert de variáveis aleatórias*.

Conceito 26 (Espaços de Hilbert de variáveis aleatórias). *Seja $\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_m\}$ um conjunto finito de variáveis aleatórias com $E[\mathbf{Y}_i^2] < \infty$ para cada i . Pode-se então definir um espaço de Hilbert H consistindo de todas as funções $g(\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n\}) = G$, onde o produto interno de dois elementos em H é definido como $\langle G_1, G_2 \rangle = E[G_1 G_2]$.*

Vale ressaltar também que, se nos espaços de Hilbert H cada variável aleatória tem valor esperado igual a zero, então dois vetores G_1, G_2 de H são ortogonais se forem não correlacionados, ou seja,

$$\langle G_1, G_2 \rangle = E[G_1 G_2] = E[G_1] E[G_2] = 0.$$

O estudo da otimização, pela abordagem geométrica, não deve ser vista como uma exclusividade da Matemática, pois, os conceitos relacionados a espaços vetoriais resultam da fusão da Geometria, Álgebra Linear, Análise e, por que não citar a Estatística, que, por sua vez, dá um significado prático a muitos desses conceitos.

Muitos são os problemas, em diversas áreas do conhecimento que a abordagem geométrica configura-se como solução ideal e/ou factível.

3 METODOLOGIA

Divide-se este capítulo em Materiais (descrição dos recursos utilizados para a corroboração dos resultados obtidos geometricamente) e Métodos (descrição da metodologia adotada na apresentação e discussão dos resultados).

3.1 Materiais

Visando atingir os objetivos propostos neste trabalho, utiliza-se como materiais, dados provenientes de simulações estocásticas computacionais, de acordo com a hierarquização proposta.

Os dados foram simulados em *grid's* (malhas) de vários tamanhos (144, 121, 100, 81, 64, 49, 25 e 16 realizações) em configurações regulares (conforme recomenda Oliveira (1991) e representação de exemplo da Figura 20) em 2 grandes grupos:

- a) Com distribuição conjunta de probabilidade conhecida, a saber, um processo gaussiano, dado por

$$f_{\mathbf{Y}}(\mathbf{y}) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-1} \exp \left\{ -\frac{1}{2} [\mathbf{y} - \mu(\mathbf{x}_i)]^t \Sigma^{-1} [\mathbf{y} - \mu(\mathbf{x}_i)] \right\}$$

em que: \mathbf{Y} é o vetor de dados $(\mathbf{Y}(\mathbf{x}_1), \dots, \mathbf{Y}(\mathbf{x}_n))$ em n localizações; $m(\mathbf{x}_i)$ é uma tendência de 1ª ordem em relação às coordenadas, cujo modelo é $m(\mathbf{x}_i) = f(x, y) = 100 - 2x + 0,3y$; Σ é uma matriz $n \times n$ de variâncias e covariâncias, em que as covariâncias são estabelecidas pelo modelo esférico, dado pela equação (10).

- b) Sem o conhecimento da distribuição conjunta de probabilidade, porém, com

média conhecida e dados com uma tendência em relação às coordenadas, conforme descrição acima. Para que a distribuição conjunta fosse desconhecida, primeiramente, considera-se o mesmo processo gaussiano e, em seguida, adiciona-se $\exp\{N(0, 1)\}$ aos dados.

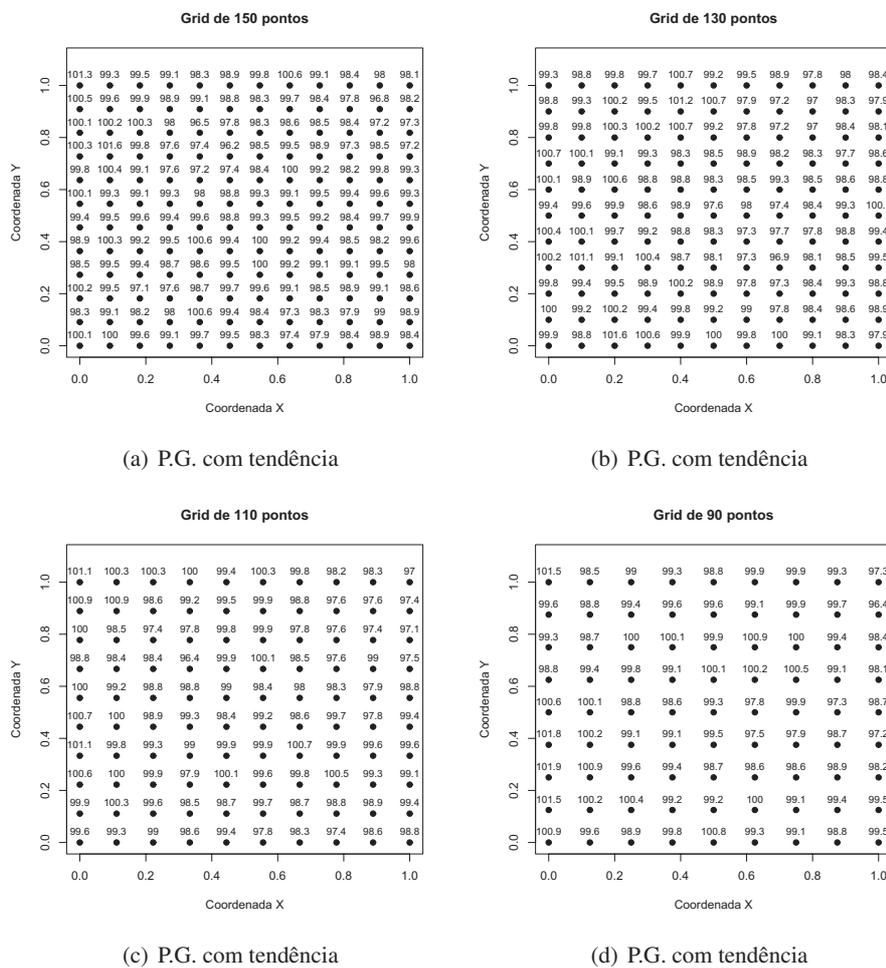


Figura 20 *grid's* regulares de dados simulados, considerando um Processo Gaussiano de média zero e com tendência

Como critério de comparação entre as krigagens, utiliza-se a variância de krigagem, tomando o valor médio dos valores krigados.

Todas as simulações e análises foram feitas no *software* R (R Development Core Team, 2010) - pacote geoR (RIBEIRO JUNIOR; DIGGLE, 2001).

3.2 Métodos

Quanto aos métodos utilizados no trabalho, adota-se as demonstrações geométricas dos preditores, visando sempre a projeção ortogonal da função-vetor $Y(\mathbf{x}_0)$ no subespaço vetorial resultante, aplicando as restrições impostas pelos mesmos.

Para cada preditor apresentado, discute-se a sua hierarquização, quanto à precisão da predição. Dessa forma, aquele preditor que se aplicar à condição imposta na hierarquização e tiver o menor subespaço de projeção deverá apresentar maior variância de krigagem. Corroborando com a hierarquização, simulações, de vários tamanhos e seguindo as restrições impostas, foram feitas.

Para cada tamanho de amostra, simulações foram feitas até que um comportamento de estrutura de dependência espacial fosse percebido para todo o procedimento exigido no estudo. Por exemplo, não era suficiente simular um processo gaussiano com parâmetros claramente estimáveis no primeiro variograma, a estrutura de dependência espacial deveria ainda existir após a remoção da tendência dos dados.

Em seguida, uma análise exploratória e variográfica foi feita. Adotando o método de ajuste de variogramas, denominado simplesmente como *wls* (mínimos quadrados ponderados pela variância), conforme recomenda Cressie (1993) e Waller e Gotway (2004), os parâmetros foram estimados para que as krigagens fossem

feitas.

Salienta-se, porém, que por falta de softwares estatísticos (ou até mesmo pacotes) implementados para a predição usando a Esperança Condicional, a apresentação do preditor ficou restrita à teoria (demonstração). Além disso, ainda pelo mesmo motivo, a Krigagem Linear Universal foi realizada conforme recomendada por Diggle e Ribeiro Junior (2007), ou seja, retira-se a tendência e analisa-se os resíduos pela Krigagem Linear Ordinária.

Tanto a hierarquização teórica quanto os resultados da simulação estão apresentados em tabelas-resumo.

4 RESULTADOS E DISCUSSÃO

Apresenta-se, basicamente, este capítulo em duas abordagens: teórica, apresentando a predição geoestatística com o enfoque geométrico, e prática, apresentando os resultados de simulações estocásticas.

4.1 Função de predição

Seja (Ω, \mathbb{A}, P) um espaço de probabilidade, e $\{Y(\mathbf{x}) : \mathbf{x} \in \mathbb{D} \subset \mathbb{R}^p\}$, um processo estocástico, em que $Y(\mathbf{x})$ é uma variável aleatória definida sobre (Ω, \mathbb{A}, P) .

Define-se uma amostra \mathbf{A} desse processo estocástico como

$$\mathbf{A} = \{Y(\mathbf{x}_1), Y(\mathbf{x}_2), \dots, Y(\mathbf{x}_n)\}$$

em que cada $Y(\mathbf{x}_i)$ pertence ao processo estocástico considerado e que pode-se observar suas realizações $\{y(\mathbf{x}_1), y(\mathbf{x}_2), \dots, y(\mathbf{x}_n)\}$.

Com base nisso, o que se pretende é predizer a realização $y(\mathbf{x}_0)$ de $Y(\mathbf{x}_0)$, usando a amostra \mathbf{A} , ou seja, construir uma função real $g(\mathbf{x}_0; \mathbf{A}) = g(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))$ que implica em

$$g(\mathbf{x}_0; \mathbf{A}) = g(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) = \hat{y}(\mathbf{x}_0).$$

A maneira como se constrói a função de predição $\hat{Y}(\mathbf{x}_0) = g(\cdot, \dots, \cdot) = g(\cdot)$, em Geoestatística, é baseada no esforço de minimizar o erro quadrático médio do preditor, ou seja,

$$\hat{Y}(\mathbf{x}_0) = g(\mathbf{x}_0; Y(\mathbf{x}_1), Y(\mathbf{x}_2), \dots, Y(\mathbf{x}_n)) \Rightarrow E \left\{ \left[Y(\mathbf{x}_0) - \hat{Y}(\mathbf{x}_0) \right]^2 \right\} \text{mínima.}$$

4.2 Predição ótima

Dado um ponto \mathbf{x}_0 e uma amostra $\mathbf{A} = \{Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)\}$, o que se deseja é prever $Y(\mathbf{x}_0)$ usando as informações de \mathbf{A} (observa-se que, para fazer a predição de $Y(\mathbf{x}_0)$ não é necessário, a princípio, usar uma amostra, pois até a opinião de um especialista pode ser utilizada, mas, conforme a objetividade buscada na ciência, adota-se a utilização de uma amostra e não outro método subjetivo).

Para a predição nesses moldes, constrói-se uma função

$$g(\mathbf{x}_0; \mathbf{A}) = g(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) = G_{\mathbf{A}}(\mathbf{x}_0) = \hat{Y}_{\mathbf{A}}(\mathbf{x}_0) = \hat{Y}(\mathbf{x}_0)$$

que preverá $Y(\mathbf{x}_0)$, isto é, $\hat{Y}(\mathbf{x}_0)$ é preditor de $Y(\mathbf{x}_0)$.

Para fazer o cálculo de probabilidade sobre $\hat{Y}(\mathbf{x}_0)$ sem inconsistências essenciais, pressupõe-se que tal função g é Borel-mensurável para o espaço (Ω, \mathbb{A}, P) , isto é, pode-se aplicar a função P desse espaço sobre eventos do tipo

$$[g(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) \leq y], \forall y \in \mathbb{R}.$$

Nota: A função criada acima é dita, matematicamente, função mensurável que, segundo Bartle (1995), é aquela que apresenta comportamento suficientemente simples para que se possa desenvolver uma teoria de integração. O entendimento de uma teoria de integração depende do entendimento dos fundamentos do cálculo. Nos cursos básicos de Cálculo, aprende-se que a integral está intimamente relacionada com a noção de área. A teoria da integração tem suas raízes nos trabalhos dos matemáticos gregos Eudoxo¹ e Arquimedes², cujos cálculos de áreas

¹Eudoxo Cnido, 390 – 338 a.C. - astrônomo, matemático e filósofo grego

²Arquimedes de Siracusa, 287 – 212 a.C. - matemático, físico e inventor grego

e volumes eram por meio do método de exaustão (decompor um todo em infinitas partes até que a área e o volume se tornassem conhecidas). No século XVII, com a descoberta do teorema fundamental do cálculo, feita de forma independente por Newton³ e Leibniz⁴, a teoria da integração teve um grande avanço. Já no século XIX, Riemann⁵ foi quem formalizou a noção de integral através do uso de limites. Porém, apesar de sua popularidade, esta teoria tem algumas deficiências que se tornam evidentes no estudo de sequências de funções integráveis. Preenchendo as lacunas da teoria de integração de Riemann, Henri Lebesgue⁶ desenvolveu em 1902 uma teoria de integração, denominada *Integração de Lebesgue*. Isnard (2007) afirma que um conceito sempre presente, quando se fala em mensurabilidade, é o de função contínua (aquela que a imagem inversa de qualquer conjunto aberto do seu contradomínio é um conjunto aberto do seu domínio). Em contrapartida, no desenvolvimento da teoria da integração, um conceito importante, e com características em comum com as funções contínuas, é o de função mensurável (aquela que considera o comportamento em relação a conjuntos abertos do contradomínio e subconjuntos específicos do domínio, conhecidos como conjuntos mensuráveis - elementos de \mathbb{A}).

Assim, diz-se que o processo estocástico é Borel-mensurável para o espaço de probabilidade (Ω, \mathbb{A}, P) quando possibilita o cálculo de probabilidades do tipo

$$P[Y_{i_1}(\mathbf{x}_1) \leq y_1, \dots, Y_{i_r}(\mathbf{x}_n) \leq y_r] = F_{Y_{i_1}(\mathbf{x}_1), \dots, Y_{i_r}(\mathbf{x}_n)}(\cdot, \dots, \cdot)$$

$\forall y_i \in \mathbb{D}, \forall r = 1, 2, 3, \dots$. Em outras palavras, o cálculo de probabilidade de

³Isaac Newton, 1643 – 1727, cientista inglês

⁴Gottfried Wilhelm von Leibniz, 1646 – 1716, cientista alemão

⁵Georg Friedrich Bernhard Riemann, 1826 – 1866, matemático alemão

⁶Henri Léon Lebesgue, 1875 – 1941, matemático francês

qualquer evento composto ou elementar E relacionado ao processo estocástico é possível, pois E pode ser construído a partir de eventos do tipo $[Y_{i_j}(\mathbf{x}_j) \leq y_j] \in \mathbb{A}$.

Dessa forma, pode-se equivaler a Borel-mensurabilidade à ideia de que os eventos podem ser obtidos (ou decompostos), por eventos de \mathbb{A} , ou seja, existe $F_{\hat{Y}(\mathbf{x}_0)}(y(\mathbf{x}_0))$ e com essa função acumulada pode-se calcular qualquer probabilidade envolvendo $\hat{Y}(\mathbf{x}_0)$. Ressalta-se que, como não se pode vislumbrar facilmente fenômenos espaciais que não tratem de variáveis aleatórias discretas e/ou (absolutamente) contínuas, então a Borel-mensurabilidade do processo estocástico assumido neste trabalho será sempre atendida para algum espaço de probabilidade (Ω, \mathbb{A}, P) , adequadamente definido.

O conjunto de todas as funções-vetores g Borel-mensuráveis sobre (Ω, \mathbb{A}, P) forma um espaço vetorial \mathbf{B} , pois se g_1 e g_2 são Borel-mensuráveis então $a_1g_1 + a_2g_2$ também é variável aleatória.

Para a predição ótima do processo estocástico $\{Y(\mathbf{x}) : \mathbf{x} \in \mathbb{D} \subset \mathbb{R}^p\}$, sob abordagem geométrica, define-se um espaço vetorial associado que consiste de todas as funções-vetores h Borel-mensuráveis para (Ω, \mathbb{A}, P) com argumento sobre os subconjuntos do tipo

$\{Y(\mathbf{x}_0), Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)\}, \forall n \in \mathbb{N}, \forall \mathbf{x}_i, i = 0, 1, 2, \dots, n$. Este espaço vetorial é denominado de \mathbf{U} (de “universo”).

De acordo com o Conceito 2 e suas propriedades, pode-se verificar que de fato \mathbf{U} é um espaço vetorial. Assim, vale observar que $Y(\mathbf{x}_0) \in \mathbf{U}$ (Conforme Figura 21 para $\mathbf{U} = \mathbb{R}^3$), isto é, $Y(\mathbf{x}_0)$ é um vetor de \mathbf{U} porque $Y(\mathbf{x}_0) = h(Y(\mathbf{x}_0), Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) = 1 \times Y(\mathbf{x}_0) + 0 \times Y(\mathbf{x}_1) + \dots + 0 \times Y(\mathbf{x}_n)$, para uma função h que suprime as variáveis aleatórias $Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)$ e identifica

$h(Y(\mathbf{x}_0), Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))$ com $Y(\mathbf{x}_0)$. Observa-se, ainda, que \mathbf{B} é um subconjunto de \mathbf{U} e que $Y(\mathbf{x}_0)$ não pertence necessariamente à \mathbf{B} , mas sempre pertence à \mathbf{U} .

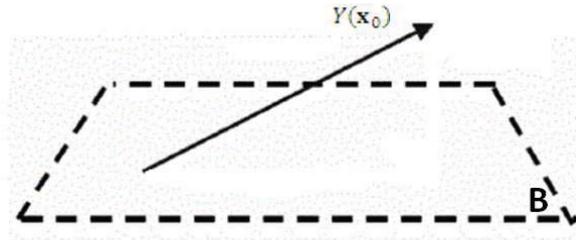


Figura 21 Representação do vetor-função $Y(\mathbf{x}_0) \in \mathbf{U}$ para $\mathbf{U} = \mathbb{R}^3$

Para evitar que sejam considerados \mathbf{U} e \mathbf{B} inadequados, limita-se a espaços vetoriais que permitam otimização segundo uma norma quadrática. Além disso, uma pressuposição adicional é suposta válida sempre: que o processo estocástico fundamental seja de 2ª ordem (existência do 2º momento estatístico de qualquer variável $Y(\mathbf{x})$ desse processo estocástico), representado por

$$E[(Y(\mathbf{x}))^2] < \infty, \forall \mathbf{x} \in \mathbb{D}$$

necessidade justificada pelas hipóteses apresentadas na visão geral da Geoestatística.

Com isso, fica implícito também que \mathbf{U} é dotado de um produto interno, dado por

$$\langle H_1, H_2 \rangle = E[H_1 H_2],$$

onde $H_i = h_i(Y(\mathbf{x}_0), Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)), \forall n$.

Em particular, $\langle Y(\mathbf{x}), Y(\mathbf{x}) \rangle = E[Y(\mathbf{x})Y(\mathbf{x})]$ sempre existe, pois, a norma do vetor $Y(\mathbf{x})$ é dada por

$$\|Y(\mathbf{x})\| = +\sqrt{\langle Y(\mathbf{x}), Y(\mathbf{x}) \rangle} = +\sqrt{E[(Y(\mathbf{x}))^2]}$$

e o processo é de 2ª ordem.

Lidando com este processo estocástico, pode-se mostrar que \mathbf{U} é um espaço de Hilbert. Como \mathbf{B} é um subespaço completo e fechado de \mathbf{U} , então \mathbf{B} é um espaço de Hilbert também. Pelo Teorema 2 e 3 sabe-se que existe um único elemento $g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) \in \mathbf{B}$ tal que

$$\begin{aligned} E\{[g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) - Y(\mathbf{x}_0)]^2\} &\leq \\ &\leq E\{[g(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) - Y(\mathbf{x}_0)]^2\} \end{aligned}$$

$\forall g \in \mathbf{B}$. A função g_{MIN} é a projeção ortogonal de $Y(\mathbf{x}_0)$ sobre \mathbf{B} , e ainda, g_{MIN} é dada por

$$g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) = E[Y(\mathbf{x}_0)|Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)]. \quad (17)$$

A equação (17) é denominada *Esperança Condicional* de $Y(\mathbf{x}_0)$ dado a amostra \mathbf{A} .

Assim, g_{MIN} é um preditor $\hat{Y}(\mathbf{x}_0)$ especial, a qual representa-se aqui por $\hat{Y}_{EC}(\mathbf{x}_0)$ (*EC* de “esperança condicional”). Portanto, pode-se afirmar que $\hat{Y}_{EC}(\mathbf{x}_0)$ é o melhor (sem aspas) preditor possível para $Y(\mathbf{x}_0)$, obtido a partir da amostra $\mathbf{A} = \{Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)\}$.

Este preditor $\hat{Y}_{EC}(\mathbf{x}_0)$ é reconhecido como melhor porque:

a) É naturalmente não tendencioso,

$$E[\hat{Y}_{EC}(\mathbf{x}_0)] = E\{E[Y(\mathbf{x}_0)|Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)]\} = E[Y(\mathbf{x}_0)];$$

b) É, por ser a projeção ortogonal, de menor erro quadrático médio possível, e, devido (i), também é de variância mínima em relação ao erro de predição.

Essa variância mínima é dada por $E[(\hat{Y}_{EC}(\mathbf{x}_0) - Y(\mathbf{x}_0))^2] = \sigma_{EC}^2(\mathbf{x}_0)$.

Geometricamente, pode-se fazer a representação através da Figura 22 para $\mathbf{U} = \mathbb{R}^3$.

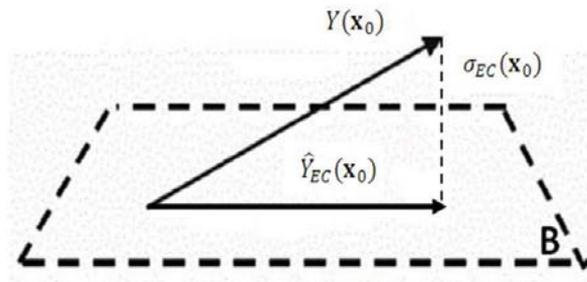


Figura 22 Representação da projeção ortogonal do melhor preditor $\hat{Y}_{EC}(\mathbf{x}_0)$ para $\mathbf{U} = \mathbb{R}^3$

Este resultado tem o potencial de resolver definitivamente o problema da predição, tornando, portanto, desnecessária a busca por outros preditores, conforme passa-se a demonstrar.

Na busca pelo melhor preditor é suficiente restringir a busca ao conjunto das funções que são de quadrado integrável, denotado por

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g^2(y(\mathbf{x}_1), y(\mathbf{x}_2), \dots, y(\mathbf{x}_n)) dx_1 \dots dx_n < \infty. \quad (18)$$

Baseando-se então em Journal (1977) e Luenberger (1969), toma-se como

critério de otimalidade a minimização de

$$E \{ [Y(\mathbf{x}_0) - g(\mathbf{x}_0; Y(\mathbf{x}_1), Y(\mathbf{x}_2), \dots, Y(\mathbf{x}_n))]^2 \} \quad (19)$$

para todas as funções $g(\cdot)$ em H , e o método para se fazer essa inferência pode ser visto como uma otimização (minimização por projeção) em um *Espaço de Hilbert*, com norma (definida em função de produto interno) adequadamente definida de tal forma que a minimização da norma ou da esperança (equação (19)) se equivalem.

Então, a formulação e demonstração de que a EC é a melhor função de predição é feita no intuito de encontrar uma forma geral para a mesma, independente de classe (como a classe das “funções lineares” ou a classe das “funções quadráticas”, entre outras), conforme passa-se a descrever.

Dada a função $g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))$ definida como a melhor função de predição de $Y(\mathbf{x}_0)$ baseada em $\mathbf{A} = \{Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)\}$, tem-se que $g_{MIN}[\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)]$ apresenta o erro quadrático médio mínimo, se e somente se,

$$\begin{aligned} E \{ [Y(\mathbf{x}_0) - g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))]^2 \} &\leq \\ &\leq E \{ [Y(\mathbf{x}_0) - g(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))]^2 \}, \end{aligned}$$

para qualquer função $g(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))^2$.

Para tanto, utiliza-se o seguinte Teorema:

Teorema 10 (Esperança Condicional). *Seja*
 $f_{Y(\mathbf{x}_0), Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)} [y(\mathbf{x}_0), y(\mathbf{x}_1), \dots, y(\mathbf{x}_n)]$ a função densidade de
 probabilidade de $n + 1$ variáveis aleatórias $Y(\mathbf{x}_0), Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)$.
 A melhor função de predição de $Y(\mathbf{x}_0)$, dado $\mathbf{A} =$
 $\{Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)\}$, é dada por $g_{MIN}[\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)] =$
 $E[Y(\mathbf{x}_0) | Y(\mathbf{x}_1) = y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n) = y(\mathbf{x}_n)]$, para todo
 $y(\mathbf{x}_1), \dots, y(\mathbf{x}_n)$.

Demonstração:

Admitindo que as funções sejam de quadrado integrável, o problema consiste em
 achar uma função $g_{MIN}[\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)]$ tal que

$E \left\{ [Y(\mathbf{x}_0) - g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))]^2 \right\}$ seja mínima. Então,

$$\begin{aligned}
 & E \left\{ [Y(\mathbf{x}_0) - g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))]^2 \right\} = \\
 & = E \left\{ [Y(\mathbf{x}_0) - E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))] + \right. \\
 & \quad \left. + E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) - \right. \\
 & \quad \left. - g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))]^2 \right\} = \\
 & = E \left\{ [Y(\mathbf{x}_0) - E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))]^2 \right\} + \\
 & \quad - E \left\{ [E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) - \right. \\
 & \quad \left. - g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))]^2 \right\}. \tag{20}
 \end{aligned}$$

A equação (20) só é válida se o termo do produto cruzado for nulo. Para
 mostrar isso, recorre-se a um resultado de Teorema apresentado e demonstrado por

Graybill (1976) onde

$$\begin{aligned} E[Y(\mathbf{x}_0)] &= E_{Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)} \times E_{Y(\mathbf{x}_0), Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)}(\mathbf{x}_0) = \\ &= E_{Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)} \times \{E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))\} \end{aligned}$$

então $E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) = E_{Y(\mathbf{x}_0), Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)}(\mathbf{x}_0)$. Logo, o produto cruzado da equação (20) fica

$$\begin{aligned} &E[\{Y(\mathbf{x}_0) - E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))\} \times \\ &\times \{E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) - g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))\}] = \\ &= E_{Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)} [E_{Y(\mathbf{x}_0), Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)} \{Y(\mathbf{x}_0) + \\ &- E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))\} \times \\ &\times \{E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) + \\ &- g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))\}]. \end{aligned} \tag{21}$$

Contudo, o segundo termo,

$$\{E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) - g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))\}$$

não depende da variável aleatória $Y(\mathbf{x}_0)$ (o que resulta em uma diferença de constantes). Assim, examina-se somente o outro termo,

$$\begin{aligned} &E_{Y(\mathbf{x}_0), Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)} \{Y(\mathbf{x}_0) - E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))\} = \\ &= E_{Y(\mathbf{x}_0), Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)} [Y(\mathbf{x}_0)] - E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) = 0. \end{aligned}$$

Então, da equação (20) pode-se afirmar que

$$\begin{aligned} E \left\{ [Y(\mathbf{x}_0) - g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))]^2 \right\} &\geq \\ &\geq E \left\{ [Y(\mathbf{x}_0) - E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))]^2 \right\} \end{aligned}$$

em que a igualdade vale se, e somente se,

$$E \left\{ [E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) - g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))]^2 \right\} = 0$$

isto é,

$$g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) = E(Y(\mathbf{x}_0) | Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n)) \quad (22)$$



Com base nestes resultados apresentados, pode-se fazer uma associação destes com os comentários geométricos da EC que Journel (1977) faz. Segundo este autor, o subespaço vetorial da EC, denotado por H_{EC} , contém todos os subespaços dos preditores lineares da Geoestatística. Então, como H_{EC} é o mais abrangente subespaço de projeção de $Y(\mathbf{x}_0)$, chegando a ser o próprio espaço H , então $\hat{Y}(\mathbf{x}_0)$ é o melhor de todos os preditores da Geoestatística.

Ressalta-se ainda, que a EC é não viesado, pois quando o subespaço H_{EC} for o próprio espaço H , onde está $Y(\mathbf{x}_0)$, $E[\hat{Y}(\mathbf{x}_0)] = E[Y(\mathbf{x}_0)]$ (demonstrado algebricamente por Graybill (1976)).

Então, dado qualquer elemento de H_{EC} , $Y_i(\mathbf{x}_0) = g_i(Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))$, a projeção $\hat{Y}(\mathbf{x}_0)$ de $Y(\mathbf{x}_0)$ em H_{EC} é caracterizada por uma das 2 relações:

- a) $\|Y(\mathbf{x}_0) - \hat{Y}_{EC}(\mathbf{x}_0)\| = \min \|Y(\mathbf{x}_0) - Y_i(\mathbf{x}_0)\|, \forall Y_i(\mathbf{x}_0) \in H_{EC};$
- b) $\langle Y(\mathbf{x}_0) - \hat{Y}_{EC}(\mathbf{x}_0), Y_i(\mathbf{x}_0) \rangle = 0, \forall Y_i(\mathbf{x}_0) \in H_{EC}.$

Ocorre, porém, que essa esperança, na quase totalidade dos casos, não pode ser obtida, visto que o cálculo exigiria o conhecimento da distribuição conjunta $(n + 1)$ -variada, necessidade verificada para expressar

$$\begin{aligned} & \langle Y(\mathbf{x}_0) - \hat{Y}_{EC}(\mathbf{x}_0), Y_i(\mathbf{x}_0) \rangle = \\ & = E \{ [Y(\mathbf{x}_0) - g_{MIN}(\mathbf{x}_0; Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))] [g_i(Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_n))] \}, \end{aligned}$$

ou seja, é uma quantidade de informação inexistente, na maioria das situações práticas.

Isso corrobora com o que Journel (1977) afirma sobre a EC, ou seja, para se obter o melhor preditor deve-se atender a pré-requisitos bem difíceis de serem cumpridos.

Então, devido a problemas de ordem prática, deve-se avançar na busca de preditores mais factíveis, ou seja, preditores restritos a funções lineares. E é neste subconjunto que está a maior parte dos preditores da geoestatística.

4.3 A krigagem em termos de projeção

Usar a vizinhança amostrada é uma característica dos preditores da Geoestatística, pois dessa forma pode-se reduzir a variância de predição. Por isso é que muitos são os estudos, e em diversas áreas do conhecimento, que têm usado os procedimentos de predição, denominados de Krigagem.

Andriotti (2003) afirma também que a forma como a ponderação é feita na krigagem vem a ser a sua principal vantagem, pois elimina em média os erros de

estimação. Isso acontece porque a krigagem leva em consideração:

- a) O número de amostras utilizadas;
- b) As posições das amostras na área a ser avaliada;
- c) As distâncias entre as amostras;
- d) A continuidade espacial da variável em estudo.

Conforme apresenta-se nessa seção, a krigagem é também um preditor linear, pois é formado por uma combinação linear das realizações de uma variável aleatória. Assim, seja $A = \{Y(\mathbf{x}), \mathbf{x} \in \mathbb{D}\}$ um conjunto de variáveis aleatórias $Y(\mathbf{x})$ e \mathbf{S} um espaço vetorial definido como o conjunto de todas as combinações lineares dos elementos de A , $\mathbf{S} = \{\sum_i \lambda_i Y(\mathbf{x}_i); Y(\mathbf{x}_i) \in A, \lambda_i \in \mathbb{R}\}$.

Destaca-se que este espaço vetorial \mathbf{S} é dotado de um produto interno definido como uma covariância não central (e não necessariamente estacionária), dado por $\langle Y(\mathbf{x}), Y(\mathbf{y}) \rangle = E[Y(\mathbf{x})Y(\mathbf{y})]$. Na prática, afirma Journel (1977), essa covariância caracteriza a correlação espacial entre duas variáveis aleatórias $Y(\mathbf{x})$ e $Y(\mathbf{y})$ localizadas em pontos \mathbf{x} e \mathbf{y} diferentes.

Define-se também $E[Y(\mathbf{x})] = m(\mathbf{x})$ como a esperança de $Y(\mathbf{x})$ e

$$E\{[Y(\mathbf{x}) - m(\mathbf{x})][Y(\mathbf{y}) - m(\mathbf{y})]\} = E[Y(\mathbf{x})Y(\mathbf{y})] - m(\mathbf{x})m(\mathbf{y}) = \sigma_{\mathbf{xy}}$$

a covariância centrada na média.

Journel (1977) define a norma de um vetor $Y(\mathbf{x})$ como a raiz quadrada positiva do produto interno $\langle Y(\mathbf{x}), Y(\mathbf{x}) \rangle$, isto é, $\|Y(\mathbf{x})\|^2 = \langle Y(\mathbf{x}), Y(\mathbf{x}) \rangle$ e a distância entre dois elementos $Y(\mathbf{x})$ e $Y(\mathbf{y})$ é definida como a norma $\|Y(\mathbf{x}) - Y(\mathbf{y})\|$ do vetor $Y(\mathbf{x}) - Y(\mathbf{y})$.

Dessa forma, pode-se trabalhar com a krigagem em termos de projeção, ou seja, dado $Y(\mathbf{x}_0) \in \mathbf{S}$ uma variável aleatória desconhecida e $\mathbf{B} \subset \mathbf{S}$ um subespaço vetorial qualquer de \mathbf{S} (ou até mesmo uma variedade linear), pode-se provar que existe um único elemento $\hat{Y}(\mathbf{x}_0) \in \mathbf{B}$ que satisfaz $\|Y(\mathbf{x}_0) - \hat{Y}(\mathbf{x}_0)\|$, onde $\hat{Y}(\mathbf{x}_0)$ é chamado de projeção de $Y(\mathbf{x}_0)$ no subespaço \mathbf{B} .

Portanto, pode-se perceber que a krigagem nada mais é do que uma projeção de um vetor desconhecido em um subespaço específico, onde a busca pelo preditor $\hat{Y}(\mathbf{x}_0)$ é realizada. Além disso, o que se espera é que a distância entre $Y(\mathbf{x}_0)$ e $\hat{Y}(\mathbf{x}_0)$ seja mínima,

$$E \left\{ \|Y(\mathbf{x}) - \hat{Y}(\mathbf{x}_0)\|^2 \right\} = E \left\{ [Y(\mathbf{x}) - \hat{Y}(\mathbf{x}_0)]^2 \right\}$$

, denominada na literatura como *variância de krigagem*.

Conforme citado anteriormente, sabe-se que existem vários preditores denominados krigagens e existem vários subespaços vetoriais que estão contidos em \mathbf{S} , e até mesmo a \mathbf{B} , ou seja, pode existir $\mathbf{B}_i \subset \mathbf{B}$, para $i = 1, 2, \dots, n$. Logo, os preditores podem pertencer a subespaços diferentes, conforme representação da Figura 23. Como $\mathbf{B}_1 \subset \mathbf{B}$ com $\hat{Y}_1(\mathbf{x}_0) \in \mathbf{B}_1$ e $\hat{Y}(\mathbf{x}_0) \in \mathbf{B}$, aceita-se que a projeção $\hat{Y}(\mathbf{x}_0)$ é mais próxima de $Y(\mathbf{x}_0)$ que $\hat{Y}_1(\mathbf{x}_0)$ porque \mathbf{B} é maior que \mathbf{B}_1 , permitindo então uma estimativa melhor.

Essa percepção cria uma possibilidade de classificação geométrica dos vários preditores existentes na Geoestatística (objetivo principal deste trabalho).

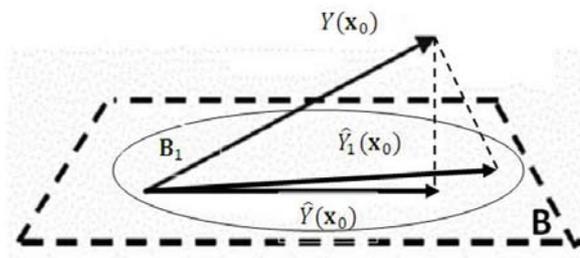


Figura 23 Diagrama em \mathbb{R}^3 de duas projeções de $Y(x_0)$

Fonte: adaptado de Journel (1977)

4.4 Preditores lineares

Considerando, primeiramente, a classe dos preditores lineares, ou seja, o subespaço vetorial de dimensão $(n + 1)$, $B_{(n+1)} \subset B$, gerado pelas combinações lineares

$$\sum_{i=1}^n \lambda_i Y(x_i) + \lambda_0 \cdot 1,$$

de n variáveis aleatórias mais uma constante, destaca-se a *Krigagem Linear Simples*, definida como a projeção do vetor $Y(x_0)$ no subespaço $B_{(n+1)}$ ou em qualquer variedade linear $V \subset B_{(n+1)}$, conforme Figura 24.

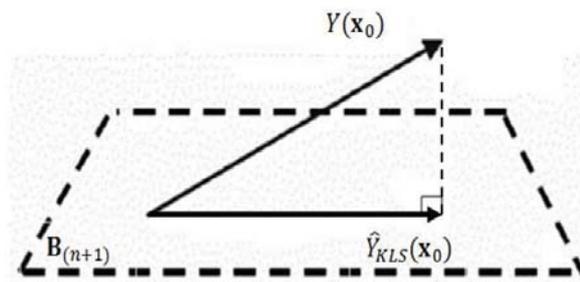


Figura 24 Diagrama em \mathbb{R}^3 da projeção de $Y(x_0)$ em $B_{(n+1)}$

Fonte: adaptado de Journel (1977)

4.4.1 Krigagem linear simples

Conforme definido, a krigagem linear simples projeta $Y(\mathbf{x}_0)$ no subespaço vetorial $\mathbf{B}_{(n+1)}$, então representa-se tal preditor como

$$\hat{Y}_{KLS}(\mathbf{x}_0) = \lambda_0 + \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i), \quad (23)$$

pois $\mathbf{B}_{(n+1)} = \{g(\cdot) : g = \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i) + \lambda_0\}$.

Obviamente, o intuito dessa projeção é minimizar $[Y(\mathbf{x}_0) - \hat{Y}_{KLS}(\mathbf{x}_0)]$, então, sabe-se que ela é única e ortogonal a $\mathbf{B}_{(n+1)} \subset \mathbf{B} \subset H$.

Como $\langle g_i, g_j \rangle = E[g_i g_j]$ e $\{1, Y(\mathbf{x}_1), Y(\mathbf{x}_2), \dots, Y(\mathbf{x}_n)\}$ é uma base geradora de $\mathbf{B}_{(n+1)}$, tem-se:

$$\Rightarrow \left\{ \begin{array}{l} \langle Y(\mathbf{x}_0) - \hat{Y}_{KLS}(\mathbf{x}_0), 1 \rangle = 0 \\ \langle Y(\mathbf{x}_0) - \hat{Y}_{KLS}(\mathbf{x}_0), Y(\mathbf{x}_1) \rangle = 0 \\ \dots \\ \langle Y(\mathbf{x}_0) - \hat{Y}_{KLS}(\mathbf{x}_0), Y(\mathbf{x}_n) \rangle = 0 \end{array} \right. \Rightarrow \left\{ \begin{array}{l} E \left\{ \left[Y(\mathbf{x}_0) - \left(\lambda_0 + \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i) \right) \right] \times 1 \right\} = 0 \\ E \left\{ \left[Y(\mathbf{x}_0) - \left(\lambda_0 + \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i) \right) \right] \times Y(\mathbf{x}_1) \right\} = 0 \\ \dots \\ E \left\{ \left[Y(\mathbf{x}_0) - \left(\lambda_0 + \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i) \right) \right] \times Y(\mathbf{x}_n) \right\} = 0 \end{array} \right. \Rightarrow$$

$$\Rightarrow \begin{cases} E[Y(\mathbf{x}_0)] - \left[\lambda_0 + \sum_{i=1}^n \lambda_i E[Y(\mathbf{x}_i)] \right] = 0 \\ E \left\{ Y(\mathbf{x}_0)Y(\mathbf{x}_1) - \left(\lambda_0 Y(\mathbf{x}_1) + \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i)Y(\mathbf{x}_1) \right) \right\} = 0 \\ \dots \\ E \left\{ Y(\mathbf{x}_0)Y(\mathbf{x}_n) - \left(\lambda_0 Y(\mathbf{x}_n) + \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i)Y(\mathbf{x}_n) \right) \right\} = 0 \end{cases}$$

Lembrando que

$$Cov[Y(\mathbf{x}_0), Y(\mathbf{x}_i)] = E[Y(\mathbf{x}_0)Y(\mathbf{x}_i)] - \mu(\mathbf{x}_0)\mu(\mathbf{x}_i) = \sigma_{0i}$$

e substituindo no sistema,

$$\begin{cases} \mu(x_0) - \lambda_0 - \sum_{i=1}^n \lambda_i \mu(\mathbf{x}_i) = 0 \\ \sigma_{01} + \mu(\mathbf{x}_0)\mu(\mathbf{x}_1) - \lambda_0 \mu(\mathbf{x}_1) - \sum_{i=1}^n \lambda_i [\sigma_{i1} + \mu(\mathbf{x}_i)\mu(\mathbf{x}_1)] = 0 \\ \dots \\ \sigma_{0n} + \mu(\mathbf{x}_0)\mu(\mathbf{x}_n) - \lambda_0 \mu(\mathbf{x}_n) - \sum_{i=1}^n \lambda_i [\sigma_{in} + \mu(\mathbf{x}_i)\mu(\mathbf{x}_n)] = 0 \end{cases}$$

Trabalhando a primeira equação e usando tal resultado com as demais equações do sistema, obtem-se

$$\begin{cases} \mu(x_0) = \lambda_0 + \sum_{i=1}^n \lambda_i \mu(x_i) \\ \sigma_{01} + \left[\lambda_0 \mu(x_1) + \sum_{i=1}^n \lambda_i \mu(x_1)\mu(x_i) \right] - \lambda_0 \mu(x_1) + \\ - \sum_{i=1}^n \lambda_i \sigma_{1i} - \sum_{i=1}^n \lambda_i \mu(x_1)\mu(x_i) = 0 \\ \dots \\ \sigma_{0n} - \sum_{i=1}^n \lambda_i \sigma_{1n} = 0 \end{cases} \quad (24)$$

A primeira equação do sistema (24) apresenta uma característica importante do preditor de krigagem simples, abordada pelo aspecto geométrico, a “não-tendenciosidade”, ou seja,

$$\begin{aligned} E[Y(\mathbf{x}_0) - \hat{Y}_{KLS}(\mathbf{x}_0)] &= 0 \Rightarrow \\ \Rightarrow E[\hat{Y}_{KLS}(\mathbf{x}_0)] &= E[Y(\mathbf{x}_0)] \Rightarrow \\ \Rightarrow \mu(\mathbf{x}_0) &= \lambda_0 + \sum_{i=1}^n \lambda_i \mu(\mathbf{x}_i), \end{aligned}$$

o que garante a precisão de $\hat{Y}_{KLS}(\mathbf{x}_0)$ em prever $Y(\mathbf{x}_0)$.

Este resultado é importante porque a grande maioria dos preditores lineares necessita da verificação dessa propriedade para que sejam considerados “ótimos”, o que é “nativo” em $\hat{Y}_{KLS}(\mathbf{x}_0)$.

O sistema de equações (24) pode ser representado também matricialmente, onde padroniza-se neste trabalho o termo *vetor* para a matriz coluna (dimensão $n \times 1$) e para a matriz linha (dimensão $1 \times n$) simplesmente o vetor transposto.

Então,

$$\begin{aligned} &\left\{ \begin{array}{l} \mu(\mathbf{x}_0) = \lambda_0 + \sum_{i=1}^n \lambda_i \mu(\mathbf{x}_i) \\ \sigma_{01} - \sum_{i=1}^n \lambda_i \sigma_{1i} = 0 \\ \dots \\ \sigma_{0n} - \sum_{i=1}^n \lambda_i \sigma_{in} = 0 \end{array} \right. \Rightarrow \\ \Rightarrow &\left\{ \begin{array}{l} \lambda_0 + \lambda_1 \mu(\mathbf{x}_1) + \lambda_2 \mu(\mathbf{x}_2) + \dots + \lambda_n \mu(\mathbf{x}_n) = \mu(\mathbf{x}_0) \\ \lambda_1 \sigma_{11} + \lambda_2 \sigma_{12} + \dots + \lambda_n \sigma_{1n} = \sigma_{01} \\ \dots \\ \lambda_1 \sigma_{n1} + \lambda_2 \sigma_{n2} + \dots + \lambda_n \sigma_{nn} = \sigma_{0n} \end{array} \right. \Rightarrow \end{aligned}$$

$$\begin{aligned}
&\Rightarrow (\lambda_0, \lambda_1, \dots, \lambda_n) \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ \mu(\mathbf{x}_1) & \sigma_{11} & \sigma_{21} & \cdots & \sigma_{n1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu(\mathbf{x}_n) & \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{nn} \end{bmatrix} = \left(\mu(\mathbf{x}_0) \quad \vdots \quad \mathbf{v}_0^t \right) \Rightarrow \\
&\Rightarrow \left(\lambda_0 \quad \vdots \quad \lambda^t \right) \begin{bmatrix} 1 & \mathbf{0}^t \\ \mu & \mathbf{V} \end{bmatrix} = \left(\mu(\mathbf{x}_0) \quad \vdots \quad \mathbf{v}_0^t \right) \Rightarrow \\
&\Rightarrow \left(\lambda_0 \quad \vdots \quad \lambda^t \right) = \left(\mu(\mathbf{x}_0) \quad \vdots \quad \mathbf{v}_0^t \right) \left[\begin{array}{c|c} 1 & \mathbf{0}^t \\ \hline \mu & \mathbf{V} \end{array} \right]^{-1}.
\end{aligned}$$

Segundo Searle (1982),

$$\left[\begin{array}{c|c} 1 & \mathbf{0}^t \\ \hline \mu & \mathbf{V} \end{array} \right]^{-1} = \left[\begin{array}{c|c} 1 & \mathbf{0}^t \\ \hline -\mathbf{V}^{-1}\mu & \mathbf{V}^{-1} \end{array} \right],$$

logo,

$$\begin{aligned}
&\left(\lambda_0 \quad \vdots \quad \lambda^t \right) = \left(\mu(x_0) \quad \vdots \quad \mathbf{v}_0^t \right) \left[\begin{array}{c|c} 1 & \mathbf{0}^t \\ \hline -\mathbf{V}^{-1}\mu & \mathbf{V}^{-1} \end{array} \right] \Rightarrow \\
&\Rightarrow \left(\lambda_0 \quad \vdots \quad \lambda^t \right) = \left(\underbrace{\mu(\mathbf{x}_0) - \mathbf{v}_0^t \mathbf{V}^{-1} \mu}_{\lambda_0} \quad \vdots \quad \underbrace{\mathbf{v}_0^t \mathbf{V}^{-1}}_{\lambda^t} \right).
\end{aligned}$$

Dessa forma, o preditor $\hat{Y}_{KLS}(\mathbf{x}_0)$ fica denotado matricialmente por

$$\hat{Y}_{KLS}(\mathbf{x}_0) = \lambda_0 + \lambda^t \mathbf{Y} = \mu(\mathbf{x}_0) - \mathbf{v}_0^t \mathbf{V}^{-1} \mu + \mathbf{v}_0^t \mathbf{V}^{-1} \mathbf{Y}. \quad (25)$$

Oliveira (1991) e Journel (1977) afirmam que o termo “krigagem” (*kri-*

ging em inglês e *krigeage* em francês) é um procedimento de predição, dentre uma determinada classe de possíveis preditores, que estima com variância mínima, também chamada de *variância de krigagem* ou *erro quadrático médio do preditor*.

Vieira (2000) ratifica esta afirmação, pois diz que o simples fato de poder conhecer também a variância da estimativa diferencia a krigagem de qualquer outro método de predição espacial. Essa propriedade é interessante porque, além de permitir a estimativa de valores sem tendência para os locais não amostrados, pode-se ainda conhecer a confiança associada a essas estimativas.

Geometricamente, a variância de krigagem para o preditor de krigagem simples $E\{[Y(\mathbf{x}_0) - \hat{Y}_{KLS}(\mathbf{x}_0)]^2\}$ não é nada além da distância mínima quadrática $\|Y(\mathbf{x}_0) - \hat{Y}_{KLS}(\mathbf{x}_0)\|^2$, denotada por $\sigma_{KLS}^2(\mathbf{x}_0)$.

Assim,

$$\begin{aligned}
\sigma_{KLS}^2(x_0) &= \text{var} \left[Y(\mathbf{x}_0) - \hat{Y}_{KLS}(\mathbf{x}_0) \right] = \\
&= E \left\{ \left[Y(\mathbf{x}_0) - \hat{Y}_{KLS}(\mathbf{x}_0) \right]^2 \right\} = \\
&= \left\| Y(\mathbf{x}_0) - \hat{Y}_{KLS}(\mathbf{x}_0) \right\|^2 = \\
&= \left\| Y(\mathbf{x}_0) - \left[\lambda_0 + \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i) \right] \right\|^2 = \\
&= \left\| Y(\mathbf{x}_0) - \left[\mu(\mathbf{x}_0) - \sum_{i=1}^n \lambda_i \mu(\mathbf{x}_i) + \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i) \right] \right\|^2 = \\
&= \left\| Y(\mathbf{x}_0) - \mu(\mathbf{x}_0) + \sum_{i=1}^n \lambda_i \mu(\mathbf{x}_i) - \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i) \right\|^2 = \\
&= \left\| Y(\mathbf{x}_0) - \mu(\mathbf{x}_0) - \sum_{i=1}^n \lambda_i [Y(\mathbf{x}_i) - \mu(\mathbf{x}_i)] \right\|^2 = \\
&= \left\| Y(\mathbf{x}_0) - \mu(\mathbf{x}_0) \right\|^2 - 2 \left\langle Y(\mathbf{x}_0) - \mu(\mathbf{x}_0), \sum_{i=1}^n \lambda_i [Y(\mathbf{x}_i) - \mu(\mathbf{x}_i)] \right\rangle + \\
&+ \left\| \sum_{i=1}^n \lambda_i [Y(\mathbf{x}_i) - \mu(\mathbf{x}_i)] \right\|^2 =
\end{aligned}$$

$$= \|Y(\mathbf{x}_0) - \mu(\mathbf{x}_0)\|^2 - 2 \sum_{i=1}^n \lambda_i \langle Y(\mathbf{x}_0) - \mu(\mathbf{x}_0), Y(\mathbf{x}_i) - \mu(\mathbf{x}_i) \rangle + \left\| \sum_{i=1}^n \lambda_i [Y(\mathbf{x}_i) - \mu(\mathbf{x}_i)] \right\|^2.$$

Adotando a notação $\|Y(\mathbf{x}_0) - \mu(\mathbf{x}_0)\|^2 = \sigma_{00}$ e lembrando que $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$, $\sigma_{KLS}^2(\mathbf{x}_0)$ pode ser denotada por

$$\begin{aligned} \sigma_{KLS}^2(\mathbf{x}_0) &= \\ &= \sigma_{00} - 2 \sum_{i=1}^n \lambda_i \sigma_{0i} + \sum_{i=1}^n \lambda_i \sum_{j=1}^n \lambda_j \langle Y(\mathbf{x}_i) - \mu(\mathbf{x}_i), Y(\mathbf{x}_j) - \mu(\mathbf{x}_j) \rangle = \\ &= \sigma_{00} - 2 \sum_{i=1}^n \lambda_i \sigma_{0i} + \sum_{i=1}^n \lambda_i \sum_{j=1}^n \lambda_j \sigma_{ij}. \end{aligned}$$

Sabe-se, do sistema de equações (24), que $\sum_{i=1}^n \lambda_i \sum_{j=1}^n \lambda_j \sigma_{ij} = \sigma_{0i}$, logo $\sigma_{KLS}^2(\mathbf{x}_0)$ fica

$$\sigma_{KLS}^2(\mathbf{x}_0) = \sigma_{00} - 2 \sum_{i=1}^n \lambda_i \sigma_{0i} + \sum_{i=1}^n \lambda_i \sigma_{0i} = \sigma_{00} - \sum_{i=1}^n \lambda_i \sigma_{0i}. \quad (26)$$

Analogamente ao preditor, pode-se representar a variância de krigagem $\sigma_{KLS}^2(\mathbf{x}_0)$ matricialmente,

$$\sigma_{KLS}^2(\mathbf{x}_0) = \sigma_{00} - \lambda^t \mathbf{v}^t(\mathbf{x}_0) = \text{var}[Y(\mathbf{x}_0)] - \mathbf{v}^t(\mathbf{x}_0) \mathbf{V}^{-1} \mathbf{v}(\mathbf{x}_0). \quad (27)$$

Percebe-se que até aqui, nenhuma pressuposição foi imposta além de o processo estocástico ser de 2^a ordem. Outrossim, na dedução da variância de krigagem, $\sigma_{KLS}^2(\mathbf{x}_0) = \left\| Y(\mathbf{x}_0) - \hat{Y}_{KLS}(\mathbf{x}_0) \right\|^2$, pode parecer estranho quando se fala em diferenças entre valor estimado e medido, pois o propósito da krigagem é justamente estimar valores para locais em que estes não foram medidos. Contudo,

o raciocínio é que, caso o valor em \mathbf{x}_0 fosse conhecido, o que poderia acontecer com a variância, ou, em outras palavras, o objetivo é que a estimativa da variância represente da melhor forma possível tal diferença. Para tanto, um mecanismo bastante eficaz pautado neste princípio é o *jackknife* (também conhecido como autovalidação), discutido na fundamentação teórica.

Conforme apresentado, uma variável aleatória é intrínseca quando, além de satisfazer a estacionariedade de 1ª ordem (primeiro momento estatístico), a diferença $\{Y(\mathbf{x}_i) - Y(\mathbf{x}_j)\}$ tem variância finita e não depende de \mathbf{x}_i , para qualquer distância entre \mathbf{x}_i e \mathbf{x}_j . Algebricamente, denota-se por

$$\begin{aligned} \text{var}[Y(\mathbf{x}_i) - Y(\mathbf{x}_j)] &= E\left\{[Y(\mathbf{x}_i) - Y(\mathbf{x}_j)]^2\right\} = \\ &= \text{var}[Y(\mathbf{x}_i)] + \text{var}[Y(\mathbf{x}_j)] - 2\text{cov}[Y(\mathbf{x}_i), Y(\mathbf{x}_j)] = \\ &= \sigma^2 + \sigma^2 - 2\sigma_{ij} = 2\sigma^2 - 2\sigma_{ij} = 2\gamma_{ij} \Rightarrow \\ &\Rightarrow \sigma_{ij} = \sigma^2 - \gamma_{ij}, \end{aligned}$$

em que $\sigma^2 = \gamma_0$ e $\gamma_{ij} = \gamma(\mathbf{x}_i - \mathbf{x}_j)$.

Aplicando, então, a hipótese intrínseca ao Sistema de Equações (24), tem-se um Sistema de Equações em função de semivariâncias,

$$\left\{ \begin{array}{l} \lambda_0 + \sum_{j=1}^n \lambda_j \mu(\mathbf{x}_j) = \mu(\mathbf{x}_0) \\ \sum_{j=1}^n \lambda_j \gamma_{ij} - \sigma^2 \left(\sum_{j=1}^n \lambda_j - 1 \right) = \gamma_{01} \\ \vdots \\ \sum_{j=1}^n \lambda_j \gamma_{in} - \sigma^2 \left(\sum_{j=1}^n \lambda_j - 1 \right) = \gamma_{0n} \end{array} \right. \quad (28)$$

Aplicando apenas na segunda equação do Sistema (24), ressalta-se que a

aplicação da hipótese se dá da seguinte forma:

$$\begin{aligned}
 \sum_{j=1}^n \lambda_j \sigma_{1j} = \sigma_{01} &\Rightarrow \sum_{j=1}^n \lambda_j (\sigma^2 - \gamma_{ij}) = \sigma^2 - \gamma_{01} \Rightarrow \\
 \Rightarrow \sum_{j=1}^n \lambda_j \sigma^2 - \sum_{j=1}^n \lambda_j \gamma_{ij} &= \sigma^2 - \gamma_{01} \Rightarrow \\
 \Rightarrow \sigma^2 \left(\sum_{j=1}^n \lambda_j - 1 \right) - \sum_{j=1}^n \lambda_j \gamma_{ij} &= -\gamma_{01} \Rightarrow \\
 \Rightarrow \sum_{j=1}^n \lambda_j \gamma_{ij} - \sigma^2 \left(\sum_{j=1}^n \lambda_j - 1 \right) &= \gamma_{01}.
 \end{aligned}$$

4.4.2 Krigagem linear ordinária

Segundo Journel (1977), o termo “ordinária” indica que, na prática, é o tipo de krigagem mais utilizado, pois, não exige o conhecimento da função média do processo, mas, em contrapartida, deve-se impor a necessidade de aceitá-la como constante,

$$E[Y(\mathbf{x}_i)] = \mu(\mathbf{x}_i) = \mu. \quad (29)$$

Partindo do preditor de krigagem linear simples e considerando a estacionariedade de primeira ordem e a não tendenciosidade, o que impõe uma restrição aos pesos $\lambda_0, \lambda_1, \dots, \lambda_n$ ao preditor $\hat{Y}_{KLO}(\mathbf{x}_0)$, tem-se

$$\begin{aligned}
 E[\hat{Y}_{KLO}(\mathbf{x}_0)] &= E\left[\lambda_0 + \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i)\right] = \mu \Rightarrow \\
 \Rightarrow \lambda_0 + \sum_{i=1}^n \lambda_i \mu &= \mu \Rightarrow \lambda_0 + \mu \sum_{i=1}^n \lambda_i = \mu \Rightarrow \\
 \Rightarrow \lambda_0 = \mu - \mu \sum_{i=1}^n \lambda_i &\Rightarrow \lambda_0 = \mu \left(1 - \sum_{i=1}^n \lambda_i\right) = 0.
 \end{aligned}$$

Como μ é desconhecida, esta igualdade é válida para todo $\mu \in \mathbb{R}$, se

$$\begin{cases} \lambda_0 = 0 \\ \sum_{i=1}^n \lambda_i = 1 \end{cases}$$

Então, diante disso, o preditor fica

$$\hat{Y}_{KLO}(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i), \quad (30)$$

com a restrição de $\sum_{i=1}^n \lambda_i = 1$.

Dessa forma, acontece a introdução de duas restrições ao subespaço onde $Y(\mathbf{x}_0)$ é projetado:

- a) $\lambda_0 = 0 \Rightarrow$ o subespaço de projeção B_{n+1} perde uma dimensão, tornando-o B_n , ou seja, o subespaço linear que consiste de todas as funções lineares de $\{Y(\mathbf{x}_1), Y(\mathbf{x}_2), \dots, Y(\mathbf{x}_n)\}$ do tipo $\sum_{i=1}^n \lambda_i Y(\mathbf{x}_i)$;
- b) $\sum_{i=1}^n \lambda_i = 1 \Rightarrow$ o subespaço de projeção B_n perde uma dimensão porque acontece a variedade linear de dimensão $n - 1$, ou seja, qualquer vetor de B_n pode ser escrito como $\lambda_1 Y(\mathbf{x}_1) + \lambda_2 Y(\mathbf{x}_2) + \dots + \lambda_n Y(\mathbf{x}_n)$, desde que $\{Y(\mathbf{x}_1), Y(\mathbf{x}_2), \dots, Y(\mathbf{x}_n)\}$ seja uma base de B_n . Como no subespaço, $\lambda_1, \dots, \lambda_n$ podem ser quaisquer reais, então os vetores de B_n em que os vetores somam 1 formam um subconjunto. Assim, pegando um vetor de B_n

e impondo a restrição, tem-se

$$\begin{aligned}
 & \lambda_1 Y(\mathbf{x}_1) + \dots + \lambda_n Y(\mathbf{x}_n) - \bar{Y} + \bar{Y} = \\
 & = \lambda_1 Y(\mathbf{x}_1) + \dots + \lambda_n Y(\mathbf{x}_n) - (\lambda_1 + \dots + \lambda_n) \bar{Y} + \bar{Y} = \\
 & = \lambda_1 Y(\mathbf{x}_1) + \dots + \lambda_n Y(\mathbf{x}_n) - (\lambda_1 \bar{Y} + \dots + \lambda_n \bar{Y}) + \bar{Y} = \\
 & = \lambda_1 [Y(\mathbf{x}_1) - \bar{Y}] + \dots + \lambda_n [Y(\mathbf{x}_n) - \bar{Y}] + \bar{Y},
 \end{aligned}$$

em que \bar{Y} é um vetor de B_n . Logo, $\lambda_1 Y(\mathbf{x}_1) + \lambda_2 Y(\mathbf{x}_2) + \dots + \lambda_n Y(\mathbf{x}_n)$ sob a restrição citada, pode ser entendido como um vetor pertencente ao espaço gerado por $\{[Y(\mathbf{x}_1) - \bar{Y}], \dots, [Y(\mathbf{x}_n) - \bar{Y}]\}$ que está contido em B_n , mas não tem dimensão n , pois conta apenas com $n - 1$ vetores.

A segunda restrição acima, apresenta $\{[Y(\mathbf{x}_1) - \bar{Y}], \dots, [Y(\mathbf{x}_n) - \bar{Y}]\}$ como um espaço gerado, mas, por definição, B_{n-1} deve conter o vetor nulo (ausente em uma variedade linear), o que pode ser resolvido por $a_1 [Y(\mathbf{x}_1) - \bar{Y}] + \dots + a_{n-1} [Y(\mathbf{x}_{n-1}) - \bar{Y}] = 0$, com $a_1 = a_2 = \dots = a_n = 0$. Consequentemente, percebe-se também que $\{[Y(\mathbf{x}_1) - \bar{Y}], \dots, [Y(\mathbf{x}_n) - \bar{Y}]\}$ é L.I..

Então, pela Proposição 2 do Teorema 4, $\hat{Y}_{KLO}(\mathbf{x}_0)$ é a projeção ortogonal de $Y(\mathbf{x}_0)$ em B_{n-1} . De forma similar à krigagem linear simples, essa projeção minimiza a variância de krigagem.

Sendo assim, essa projeção é caracterizada pelas condições:

- a) $\hat{Y}_{KLO}(\mathbf{x}_0) = \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i)$, com $\sum_{i=1}^n \lambda_i = 1$, resultado vastamente encontrado na literatura sob abordagem algébrica;
- b) $Y(\mathbf{x}_0) - \hat{Y}_{KLO}(\mathbf{x}_0)$ é ortogonal a todas as diferenças $Y(\mathbf{x}_i) - \bar{Y}, \forall i = 1, \dots, n$ e $\bar{Y} \in B_{n-1}$, tomando, por exemplo, $\bar{Y} = \frac{1}{N} \sum_{i=1}^n Y(\mathbf{x}_i)$, corroborando com o que Journel (1977) apresenta.

Assim,

$$\begin{cases} \langle \hat{Y}_{KLO}(\mathbf{x}_0) - Y(\mathbf{x}_0), Y(\mathbf{x}_1) - \bar{Y} \rangle = 0 \\ \vdots \\ \langle \hat{Y}_{KLO}(\mathbf{x}_0) - Y(\mathbf{x}_0), Y(\mathbf{x}_n) - \bar{Y} \rangle = 0 \end{cases}$$

Fazendo $\psi = \langle \hat{Y}_{KLO}(\mathbf{x}_0) - Y(\mathbf{x}_0), \bar{Y} \rangle$, a condição (ii) torna-se

$$\begin{aligned} & \langle \hat{Y}_{KLO}(\mathbf{x}_0) - Y(\mathbf{x}_0), \bar{Y} \rangle = \psi \Rightarrow \\ & \Rightarrow \langle \hat{Y}_{KLO}(\mathbf{x}_0), \bar{Y} \rangle - \langle Y(\mathbf{x}_0), \bar{Y} \rangle = \psi \Rightarrow \\ & \Rightarrow \begin{cases} \langle \hat{Y}_{KLO}(\mathbf{x}_0), Y(\mathbf{x}_1) \rangle = \langle Y(\mathbf{x}_0), Y(\mathbf{x}_1) \rangle + \psi \\ \vdots \\ \langle \hat{Y}_{KLO}(\mathbf{x}_0), Y(\mathbf{x}_n) \rangle = \langle Y(\mathbf{x}_0), Y(\mathbf{x}_n) \rangle + \psi \end{cases} \end{aligned} \quad (31)$$

Como $\langle \mathbf{g}_i, \mathbf{g}_j \rangle = E[\mathbf{g}_i \mathbf{g}_j]$ e operando de forma análoga à krigagem linear simples, o sistema (31) fica

$$\begin{cases} \sum_{j=1}^n \lambda_j = 1 \\ \sum_{j=1}^n \lambda_j \sigma_{1j} = \sigma_{01} + \psi \\ \vdots \\ \sum_{j=1}^n \lambda_j \sigma_{nj} = \sigma_{0n} + \psi \end{cases} \quad (32)$$

O sistema de equações (32) pode ser representado matricialmente, con-

forme Petenate (1983) demonstra,

$$\begin{aligned}\hat{Y}_{KLO}(\mathbf{x}_0) &= \sum_{i=1}^n \lambda_i Y(\mathbf{x}_i) = \lambda^t \mathbf{Y} = \\ &= \mathbf{v}^t(\mathbf{x}_0) \mathbf{V}^{-1} \mathbf{Y} + \frac{[\mathbf{1} - \mathbf{1}^t \mathbf{V}^{-1} \mathbf{v}^t(\mathbf{x}_0)] \times \mathbf{1}^t \mathbf{V}^{-1} \mathbf{Y}}{\mathbf{1}^t \mathbf{V}^{-1} \mathbf{1}}.\end{aligned}\quad (33)$$

Adequando o processo estocástico para a aplicação da krigagem linear ordinária, pode-se obter a variância de krigagem,

$$\begin{aligned}\text{var} [\hat{Y}_{KLO}(\mathbf{x}_0) - Y(\mathbf{x}_0)] &= E \left\{ \left[\hat{Y}_{KLO}(\mathbf{x}_0) - Y(\mathbf{x}_0) \right]^2 \right\} = \\ &= E \left\{ [\lambda^t \mathbf{Y} - Y(\mathbf{x}_0)]^2 \right\} = E \left\{ \lambda^t \mathbf{Y} \mathbf{Y}^t \lambda - 2\lambda^t \mathbf{Y} Y(\mathbf{x}_0) + [Y(\mathbf{x}_0)]^2 \right\} = \\ &= E \left\{ \lambda^t \begin{bmatrix} Y(\mathbf{x}_1)Y(\mathbf{x}_1) & \cdots & Y(\mathbf{x}_1)Y(\mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ Y(\mathbf{x}_n)Y(\mathbf{x}_1) & \cdots & Y(\mathbf{x}_n)Y(\mathbf{x}_n) \end{bmatrix} \lambda + \right. \\ &\quad \left. -2\lambda^t \begin{bmatrix} Y(\mathbf{x}_1)Y(\mathbf{x}_0) \\ \vdots \\ Y(\mathbf{x}_n)Y(\mathbf{x}_0) \end{bmatrix} + [Y(\mathbf{x}_0)]^2 \right\} = \\ &= E \left\{ \lambda_1 Y(\mathbf{x}_1)Y(\mathbf{x}_1)\lambda_1 + \dots + \lambda_n Y(\mathbf{x}_n)Y(\mathbf{x}_1)\lambda_1 + \dots + \right. \\ &\quad \left. + \lambda_1 Y(\mathbf{x}_1)Y(\mathbf{x}_n)\lambda_n + \dots + \lambda_n Y(\mathbf{x}_n)Y(\mathbf{x}_n)\lambda_n + \right. \\ &\quad \left. -2[\lambda_1 Y(\mathbf{x}_1)Y(\mathbf{x}_0) + \dots + \lambda_n Y(\mathbf{x}_n)Y(\mathbf{x}_0)] + [Y(\mathbf{x}_0)]^2 \right\}.\end{aligned}$$

Aplicando $\text{cov} [Y(x_i), Y(x_j)] = E [Y(x_i)Y(x_j)] - \mu^2 = \sigma_{ij}$ ao resultado

acima, tem-se

$$\begin{aligned}
& \lambda_1 (\sigma_{11} + \mu^2) \lambda_1 + \dots + \lambda_n (\sigma_{n1} + \mu^2) \lambda_1 + \dots + \\
& + \lambda_1 (\sigma_{1n} + \mu^2) \lambda_1 + \dots + \lambda_n (\sigma_{nn} + \mu^2) \lambda_n + \\
& - 2 [\lambda_1 (\sigma_{01} + \mu^2) + \dots + \lambda_n (\sigma_{0n} + \mu^2)] + (\sigma_{00} + \mu^2) \Rightarrow \\
& \Rightarrow \lambda^t V \lambda + \mu^2 - 2 \lambda^t v^t(x_0) - 2 \mu^2 + \sigma_{00} + \mu^2
\end{aligned}$$

que resulta em

$$\lambda^t V \lambda - 2 \lambda^t v^t(x_0) + \sigma_{00}. \quad (34)$$

Fazendo com que a equação (34) fique em termos do sistema de equações lineares (sistema (31)), tem-se

$$\sigma_{KLO}^2(\mathbf{x}_0) = \sigma_{00} - \sum_{j=1}^n \lambda_j \sigma_{0j} + \psi \quad (35)$$

em que ψ é o multiplicador de Lagrange, pois, por se tratar de uma variância e, ainda, conforme a condição de variância mínima imposta pela projeção ortogonal, faz-se necessário este artifício para se obter o mínimo da função, cuja restrição é $\sum_{j=1}^n \lambda_j = 1$. Pode-se notar ainda que ψ é positivo, tornando a variância de krigagem deste preditor, $\sigma_{KLO}^2(\mathbf{x}_0)$, maior que $\sigma_{KLS}^2(\mathbf{x}_0)$.

Geometricamente, o preditor $\hat{Y}_{KLO}(\mathbf{x}_0)$ é menos preciso (isto é, variância de krigagem maior) do que $\hat{Y}_{KLS}(\mathbf{x}_0)$, pois a projeção acontece em um subespaço menor (B_{n-1} vs B_{n+1}). Então,

$$\left\| \hat{Y}_{KLO}(\mathbf{x}_0) - Y(\mathbf{x}_0) \right\| \geq \left\| \hat{Y}_{KLS}(\mathbf{x}_0) - Y(\mathbf{x}_0) \right\| \quad (36)$$

mostrando que $Y(\mathbf{x}_0)$ está mais distante de $\hat{Y}_{KLO}(\mathbf{x}_0)$ do que de $\hat{Y}_{KLS}(\mathbf{x}_0)$. Por-

tanto, algebricamente,

$$\text{var} \left[\hat{Y}_{KLO}(\mathbf{x}_0) - Y(\mathbf{x}_0) \right] \geq \text{var} \left[\hat{Y}_{KLS}(\mathbf{x}_0) - Y(\mathbf{x}_0) \right] \quad (37)$$

em média, validando a percepção prática de que “menos informação resulta em menos certeza”.

4.4.3 Krigagem linear universal

O preditor da KLU é uma proposta para situações onde a função média do processo é desconhecida e, ainda, não estacionária. Encontra-se na literatura especializada uma hipótese para esses casos, denominada “hipótese de tendência”. Nessa hipótese, a função aleatória $Y(\mathbf{x}_i)$, em qualquer posição \mathbf{x}_i , consiste de três componentes:

$$Y(\mathbf{x}_i) = m(\mathbf{x}_i) + \varepsilon'(\mathbf{x}_i) + \varepsilon''$$

em que $m(\mathbf{x}_i)$ apresenta uma tendência que pode ser modelada por uma função polinomial.

Dessa forma, pode-se atribuir uma estrutura linear do tipo

$$m(\mathbf{x}_i) = \sum_{l=1}^L a_l f_l(\mathbf{x}_i)$$

em que $f_l(\cdot)$ são L funções conhecidas e a_l são L desconhecidos coeficientes.

Um exemplo dessa combinação linear, é considerar a região de estudo $\mathbb{D} = \mathbb{R}^2$, em que $f_l(\cdot)$ podem ser as funções lineares

$$f_1(x, y) = 1, \quad f_2(x, y) = x \quad e \quad f_3(x, y) = y.$$

Sempre que possível, essas funções devem ser parcimoniosas, onde neste caso são em relação às coordenadas onde as realizações foram feitas. Assim,

$$m(\mathbf{x}_i) = a_1 f_1(x, y) + a_2 f_2(x, y) + a_3 f_3(x, y) = a_1 + a_2 x + a_3 y.$$

Ressalta-se que foi uma função como essa ($m(\mathbf{x}_i) = 100 - 2x + 0,3y$) que foi usada para perturbar o processo gaussiano de média zero, gerado na simulação deste trabalho.

Com as pressuposições citadas acima, o subespaço de projeção fica mais restrito ainda, pois, uma variedade linear, assim definida, tem dimensão $n - L$.

Sob a hipótese de tendência e em busca da não tendenciosidade, parte-se do preditor da KLS,

$$\begin{aligned} E \left[Y(\mathbf{x}_0) - \hat{Y}_{KLU}(\mathbf{x}_0) \right] &= 0 \Rightarrow \\ \Rightarrow \mu(\mathbf{x}_0) - E \left[\lambda_0 + \sum_{j=1}^n \lambda_j Y(\mathbf{x}_j) \right] &= 0 \Rightarrow \\ \Rightarrow \sum_{l=1}^L a_l f_l(\mathbf{x}_0) - \lambda_0 - \sum_{j=1}^n \lambda_j \sum_{l=1}^L a_l f_l(\mathbf{x}_j) &= 0 \Rightarrow \\ \Rightarrow \lambda_0 + \sum_{l=1}^L a_l \left[\sum_{j=1}^n \lambda_j f_l(\mathbf{x}_j) - f_l(\mathbf{x}_0) \right] &= 0 \end{aligned} \quad (38)$$

A expressão (38) só será verdadeira se

- a) $\lambda_0 = 0$; e
- b) $f_l(\mathbf{x}_0) = \sum_{j=1}^n \lambda_j f_l(\mathbf{x}_j)$, $\forall l = 1, 2, \dots, L$.

Isso introduz $L + 1$ restrições sobre o sistema, restringindo então o subespaço vetorial de B_{n+1} para B_{n-L} .

Como $\hat{Y}_{KLU}(\mathbf{x}_0)$ é uma combinação linear que pertence a subespaços de um espaço de Hilbert, pode-se afirmar que $\hat{Y}_{KLU}(\mathbf{x}_0)$ é a projeção ortogonal de $Y(\mathbf{x}_0)$ sobre B_{n-L} . Assim, $\hat{Y}_{KLU}(\mathbf{x}_0)$ é o elemento de B_{n-L} que está mais próximo de $Y(\mathbf{x}_0)$ então $\sigma_{KLU}^2(\mathbf{x}_0) = \left\| Y(\mathbf{x}_0) - \hat{Y}_{KLU}(\mathbf{x}_0) \right\|^2$ é mínima.

De forma análoga aos outros preditores lineares, o preditor da KLU, dado por

$$\hat{Y}_{KLU}(\mathbf{x}_0) = \sum_{j=1}^n \lambda_j Y(\mathbf{x}_j), \quad (39)$$

com $Y(\mathbf{x}_j) = \left[\sum_{l=1}^L a_l f_l(\mathbf{x}_j) \right] + \varepsilon'(\mathbf{x}_j) + \varepsilon''$, surge da solução do sistema de equações lineares, dado por

$$\left\{ \begin{array}{l} \lambda_0 = 0 \\ \sum_{j=1}^n \lambda_j f_1(\mathbf{x}_j) = f_1(\mathbf{x}_0) \\ \vdots \\ \sum_{j=1}^n \lambda_j f_L(\mathbf{x}_j) = f_L(\mathbf{x}_0) \\ \sum_{j=1}^n \lambda_j \sigma_{1j} - \sum_{l=1}^L \psi_l f_l(\mathbf{x}_1) = \sigma_{01} \\ \vdots \\ \sum_{j=1}^n \lambda_j \sigma_{nj} - \sum_{l=1}^L \psi_l f_l(\mathbf{x}_n) = \sigma_{0n} \end{array} \right. \quad (40)$$

em que ψ_l são L multiplicadores de Lagrange.

Para obter a variância da KLU, $\sigma_{KLU}^2(\mathbf{x}_0)$, pode-se recorrer às demonstrações algébricas, disponíveis na literatura. Geometricamente, entretanto, deve-se

partir de

$$\begin{aligned}\sigma_{KLU}^2(\mathbf{x}_0) &= \left\| Y(\mathbf{x}_0) - \hat{Y}_{KLU}(\mathbf{x}_0) \right\|^2 = \\ &= \|Y(\mathbf{x}_0)\|^2 + \left\| \hat{Y}_{KLU}(\mathbf{x}_0) \right\|^2 - 2 \langle Y(\mathbf{x}_0), \hat{Y}_{KLU}(\mathbf{x}_0) \rangle.\end{aligned}\quad (41)$$

Da equação (41), conforme mostra Journel (1977), chega-se na expressão

$$\sigma_{KLU}^2(\mathbf{x}_0) = \sigma_{00} - \sum_{j=1}^n \lambda_j \sigma_{0j} + \sum_{l=1}^L \psi_l f_l(\mathbf{x}_0).\quad (42)$$

Como em todos os outros preditores da Geoestatística, pode-se substituir a função de covariância pela função de semivariância.

Diante do exposto sobre os preditores lineares geoestatísticos, pode-se verificar que, considerando um processo estocástico com a função média conhecida ou com a f.d.p. (função densidade de probabilidade) conjunta conhecida, o preditor da KLU é menos preciso do que KLS. Tal conclusão é justificada pela dimensão do subespaço de projeção de $Y(\mathbf{x}_0)$, o que limita a busca por projeções $\hat{Y}(\mathbf{x}_0)$ mais próximas de $Y(\mathbf{x}_0)$.

4.5 Hierarquização

Na Tabela 1 apresenta-se o resumo da hierarquização geométrica dos preditores lineares geoestatísticos, sendo que:

- a) EC: é a esperança condicional, cuja dimensão do subespaço de projeção é ∞ ;
- b) KLS: é a krigagem linear simples, cuja dimensão do subespaço de projeção é $n + 1$;

- c) KLU: é a krigagem linear universal, cuja dimensão do subespaço de projeção é $n - L$ (L é o número de parâmetros do modelo de tendência);
- d) KLO: é a krigagem linear ordinária, cuja dimensão do subespaço de projeção é $n - 1$;
- e) NAD: significa “não se aplica por definição”.

Tabela 1 Hierarquização dos principais preditores geoestatísticos considerando aspectos geométricos.

Principais Preditores Geoestatísticos	$E [Y(\mathbf{x}_i)]$ com tendência			$E [Y(\mathbf{x}_i)]$ constante
	Dist. F(.) conhecida	$E [Y(\mathbf{x}_i)]$ conhecida	$E [Y(\mathbf{x}_i)]$ desconhecida	$E [Y(\mathbf{x}_i)]$ Desconhecida
EC	1 ^o	NAD	NAD	NAD
KLS	2 ^o	1 ^o	NAD	NAD
KLU	3 ^o	2 ^o	1 ^o	NAD
KLO	NAD	NAD	NAD	1 ^o

Pode-se perceber que, uma vez que as pressuposições são consideradas, a adoção de um preditor cujo subespaço de projeção tem dimensão inferior leva a predições menos precisas, ou seja, as predições terão variância de krigagem maior, do que a adoção do preditor correto.

No intuito de validar tal hierarquização, procede-se simulações computacionais, em que considera-se a classificação adotada quanto ao conhecimento da distribuição conjunta de probabilidade com a adição de uma tendência parcimoniosa em relação às coordenadas.

4.6 Simulação

A corroboração da hierarquização geométrica dos preditores lineares se deu em forma de simulações computacionais (conforme *script* apresentado nos Anexos) divididas em 2 grandes grupos, conforme apresentado na Metodologia:

- a) Com distribuição conjunta de probabilidade conhecida, a saber, um processo gaussiano, com média conhecida, com o acréscimo de uma tendência em relação às coordenadas, cujo modelo é dado por

$$m(\mathbf{x}_i) = f(x, y) = 100 - 2x + 0,3y;$$

- b) Sem o conhecimento da distribuição conjunta de probabilidade, porém, com média conhecida e dados perturbados com uma tendência em relação às coordenadas, cujo modelo é o mesmo descrito no item acima.

Como critério de comparação entre as krigagens, estabelece-se a variância de krigagem, $\sigma_{KLS}^2(\mathbf{x}_0)$ e $\sigma_{KLU}^2(\mathbf{x}_0)$, conforme resumo na Tabela 2.

Graficamente, pode-se avaliar as diferenças entre as simulações através das figuras dos Anexos.

Analisando os resultados apresentados na Tabela 2, pode-se verificar que as diferenças entre as variâncias de krigagens foram razoavelmente pequenas, o que pode ser explicado pela parcimônia do modelo de tendência, pois, quanto maior o número de parâmetros do modelo de tendência (associado ao número L), menor será a dimensão do subespaço de predição ($n - L$), conseqüentemente, maior será a variância de krigagem, porque a projeção se dá em subespaços de dimensões cada vez menores.

Tabela 2 Resultados das simulações que ratificam numericamente a hierarquização geométrica dos principais preditores geoestatísticos.

	Dist. F(.) conhecida				Dist. F(.) desconhecida	
	KLS		KLU		KLS	KLU
<i>grid</i>	dim	$\sigma_{KLS}^2(\mathbf{x}_0)$	dim	$\sigma_{KLU}^2(\mathbf{x}_0)$	$\sigma_{KLS}^2(\mathbf{x}_0)$	$\sigma_{KLU}^2(\mathbf{x}_0)$
144	145	0,1427	141	0,3369	0,1955	0,3606
121	122	0,1047	118	0,3229	0,4639	0,6723
100	101	0,1836	97	0,3945	0,1841	0,4086
81	82	0,0941	78	0,3610	0,7670	2,2942
64	65	0,2077	61	0,4496	0,2143	1,9756
49	50	0,2708	46	0,5944	0,4076	1,3983
25	26	0,3566	22	0,6137	0,2343	1,4474
16	17	0,4586	13	0,8022	0,4398	0,7452

Outrossim, como as simulações, em cada um dos grandes grupos, foram repetidas para tamanhos diferentes de amostras (144, 121, 100, 81, 64, 49, 25 e 16 realizações), percebe-se um crescimento acentuado na variância de krigagem ao reduzir do tamanho amostral, o que ratifica a influência do número de informações em uma predição (ou estimação) estatística.

Vale salientar que as krigagens EC e KLO não foram apresentadas por questões de limitações práticas (como é o caso da EC, pois na prática dificilmente se conhece a distribuição conjunta do fenômeno em estudo) e por não haver outra krigagem linear com as mesmas restrições para uma comparação de precisão.

5 CONCLUSÃO

Problemas envolvendo a teoria da predição espacial apresentam naturalmente uma complexidade que analiticamente é de difícil compreensão, gerando assim sucessivos e grotescos erros de utilização de metodologias criadas para serem ótimas (como é o caso dos preditores geoestatísticos).

A abordagem feita neste trabalho de tese apresenta uma grande vantagem: a de possibilitar a “visualização” dos métodos geoestatísticos de predição, favorecendo o entendimento dos mesmos, além de classificá-los quanto às restrições e pressuposições impostas.

Acredita-se que, conhecendo a perda de precisão existente na escolha equivocada dos preditores geoestatísticos, devido a não observância dessas restrições e pressuposições impostas ao processo estocástico (o que reduz a dimensão dos subespaços de projeção), a comunidade científica ficará alerta em verificar de antemão tais fatores limitantes.

Com isso, recomenda-se como trabalhos futuros, a criação de uma metodologia que, a partir da hierarquização geométrica (por ter se mostrado uma abordagem apropriada) dos preditores geoestatísticos, possa testar as restrições e pressuposições que o processo estocástico venha estar submetido.

Quanto à hipótese adotada neste trabalho de tese, conclui-se que a abordagem geométrica dos preditores geoestatísticos se mostrou apropriada para a hierarquização dos mesmos quanto à precisão na predição espacial, desde que os fatores limitantes sejam observados. Em outras palavras:

- a) A esperança condicional (apresentada como uma krigagem) apresenta-se como a melhor predição geoestatística possível devido a dimensão infinita

do subespaço de projeção, porém, com exigências de difícil cumprimento na prática;

- b) A krigagem linear simples apresenta-se como o melhor preditor linear geoestatístico devido a dimensão $(n + 1)$ do subespaço de projeção, porém, exige o conhecimento pontual das médias do processo;
- c) A krigagem linear ordinária apresenta-se como o segundo melhor preditor linear geoestatístico devido a dimensão $(n - 1)$ do subespaço de projeção, porém, exige a estacionariedade da média (sem necessidade de conhecê-la);
- d) A krigagem linear universal apresenta-se como o pior dos preditores lineares geoestatísticos devido a dimensão $(n - L)$ do subespaço de projeção, em que L é o número de funções utilizadas na modelagem da tendência existente na parte estrutural do modelo, logo, a média do processo estocástico é desconhecida e não estacionária.

Ressalta-se que tal hierarquização foi demonstrada e corroborada computacionalmente.

REFERÊNCIAS

- ALVES, R.; DELGADO, C. **Texto acadêmico: processos estocásticos**. Porto: Universidade do Porto, 1997. Disponível em: <<http://pdfdatabase.com/download/>>. Acesso em: 20 fev. 2010.
- ANDRIOTTI, J. L. S. **Fundamentos de estatística e geoestatística**. São Leopoldo: UNISINOS, 2003. 165 p.
- ASH, R. B. **Real analysis and probability**. New York: Academic, 1972. 476 p.
- BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. London: Longman Scientific and Technical, 1995. 413 p.
- BARATA, J. C. A. **Curso de física-matemática**. São Paulo: USP, 2009. Disponível em: <<http://denebola.if.usp.br/~jbarata>>. Acesso em: 12 fev. 2010.
- BARTLE, R. G. **The elements of integration and Lebesgue measure**. New York: J. Wiley, 1995. 192 p.
- BIERENS, H. J. **Introduction to Hilbert spaces**. State College: Universidade do Estado da Pensilvânia, 2007. Apostila. Disponível em: <<http://econ.la.psu.edu/~hbierens>>. Acesso em: 27 maio 2010.
- BOLDRINI, J. L. et al. **Álgebra linear**. 2 ed. São Paulo: Harper & Row do Brasil, 1980. 411 p.
- BURROUGH, P. A. **Principles of geographical information systems for land resources assessment**. Oxford: Clarendon, 1987. 193 p.

CAMARGO, E. C. G. **Geoestatística**: fundamentos e aplicações. [S. l.: s. n.], 1998. cap. 5. Disponível em:
<http://www.dpi.inpe.br/gilberto/tutoriais/gis_ambiente/5geoest.pdf>. Acesso em: 17 nov. 2008.

CHILÈS, J. P.; DELFINER, P. **Geostatistics**: modeling spatial uncertainty. New York: Wiley-Interscience, 1999. 695 p.

CHONG, E. K. P.; ZAK, S. H. **An introduction to optimization**. 2nd ed. New York: J. Wiley, 2001. 476 p.

CLARKE, A. B.; DISNEY, R. L. **Probabilidade e processos estocásticos**. Tradução de Gildasio Amado Filho. Rio de Janeiro: Livros Técnicos e Científicos, 1979. 338 p.

CRESSIE, N. **Statistics for spatial data**. New York: J. Wiley, 1993. 900 p.

DIGGLE, P. J.; RIBEIRO JUNIOR, P. J. **Model-based geostatistics**. New York: Springer, 2007. 228 p.

DOMINGUES, H. H. **Espaços métricos e introdução à topologia**. São Paulo: Atual, 1982. 184 p.

FERREIRA, D. F. **Estatística multivariada**. Lavras: UFLA, 2008. 662 p.

GNEDENKO, B. V. **A teoria da probabilidade**. Rio de Janeiro: Ciência Moderna, 2008. 692 p.

GRAYBILL, F. A. **Theory and application of the linear model**. Massachusetts: Ouxburg, 1976. 704 p.

ISAAKS, E. H.; SRIVASTAVA R. M. **An introduction to applied geostatistics**. New York: University, 1989. 560 p.

ISNARD, C. **Introdução à medida e integração**. Rio de Janeiro: IMPA, 2007. 314 p.

JOURNEL, A. G. Kriging in terms of projections. **Mathematical Geology**, New York, v. 9, n. 6, p. 563-586, Nov. 1977.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining geostatistics**. London: Academic, 2003. 600 p.

KANNAN, D. **An introduction to stochastic process**. New York: North Holland, 1979. 296 p.

LANDIM, P. M. B. Sobre geoestatística e mapas. **Terræ Didática**, v. 2, n. 1, p. 19-33, 2006. Disponível em: <<http://www.ige.unicamp.br/terraedidatica/>>. Acesso em: 06 maio 2010.

LIPSCHUTZ, S. **Álgebra linear**. 3. ed. São Paulo: Makron Books, 1994. 403 p.

LUENBERGER, D. G. **Optimization by vector space methods**. New York: J. Wiley, 1969. 348 p.

MAGALHÃES, M. N. **Probabilidade e variáveis aleatórias**. 2. ed. São Paulo: USP, 2006. 428 p.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3. ed. New York: McGraw-Hill, 1974. 564 p.

OLIVEIRA, M. S. et al. **Introdução à estatística**. Lavras: UFLA, 2009. 334 p.

OLIVEIRA, M. S. **Planos amostrais para variáveis espaciais utilizando geoestatística**. 1991. 100 p. Dissertação (Mestrado em Estatística) - Universidade Estadual de Campinas, Campinas, 1991.

PETENATE, A. J. **Optimal allocation and other aspects of kriging on a line.** 1983. 93 p. Tese (Doutorado em Estatística) - Iowa State University, Ames, 1983.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing.** Vienna: R Foundation for Statistical Computing, 2010. Disponível em: <<http://www.r-project.org>>. Acesso em: 30 Apr. 2010.

RIBEIRO JUNIOR, P. J.; DIGGLE, P. J. geoR: a package for geostatistical analysis. **R-NEWS**, London, v. 1, n. 2, p. 15-18, June 2001.

SALLUM, E. M.; MURAKAMI, L. S. I.; SILVA, J. P. **Cálculo diferencial geométrico no \mathbb{R}^n .** São Paulo: IME-USP, 1999. 56 p.

SCOLFORO, J. R. S.; OLIVEIRA, A. D.; CARVALHO, L. M. T. **Zoneamento ecológico-econômico do estado de Minas Gerais: componentes geofísico e biótico.** Lavras: UFLA, 2008. 161 p.

SEARLE, S. R. **Matrix algebra useful for statistics.** New York: J. Wiley, 1982. 438 p.

VIEIRA, S.R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R.F. et al. (Ed.). **Tópicos em ciência do solo.** Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 2000. v. 1, p. 1-53.

WALLER, L. A.; GOTWAY, C. A. **Applied spatial statistics for public health data.** New Jersey: Wiley-Interscience, 2004. 494 p.

WEBSTER, R.; OLIVER, M. A. **Geostatistics for environmental scientists.** 2nd ed. Chichester: J. Wiley, 2007. 315 p.

ANEXOS

Processo gaussiano

```
#####
### SIMULAÇÃO E ANÁLISE DOS GEODADOS PARA A TESE - HIERARQUIZAÇÃO ###
#####

require(geoR)

#####
#####

## Simulação 1:

simu <- grf(150,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)

plot(variogram(simu,max.dist=.8*max(dist(simu$coords))))

simu$data <- 100 - 2*simu$coords[,1] + .3*simu$coords[,2]+ simu$data
v1 <- variogram(simu,max.dist=.8*max(dist(simu$coords)),trend="cte")
plot(v1)

sink("simulado_1.txt",type="output", append=T)

simu
sink()

#####
#####

## Simulação 2:

simu <- grf(130,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)

plot(variogram(simu,max.dist=dist.max))

simu$data <- 100 - 2*simu$coords[,1] + .3*simu$coords[,2]+ simu$data
v1 <- variogram(simu,max.dist=dist.max,trend="cte")

plot(v1)
```

```
sink("simulado_2.txt",type="output", append=T)
simu
sink()
#####
#####
## Simulação 3:
simu <- grf(110,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)
plot(variog(simu,max.dist=dist.max))
simu$data <- 100 - 2*simu$coords[,1] + .3*simu$coords[,2]+ simu$data
v1 <- variog(simu,max.dist=dist.max,trend="cte")
plot(v1)
sink("simulado_3.txt",type="output", append=T)
simu
sink()
#####
#####
## Simulação 4:
simu <- grf(90,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)
plot(variog(simu,max.dist=dist.max))
simu$data <- 100 - 2*simu$coords[,1] + .3*simu$coords[,2]+ simu$data
v1 <- variog(simu,max.dist=dist.max,trend="cte")
plot(v1)
sink("simulado_4.txt",type="output", append=T)
simu
sink()
#####
#####
## Simulação 5:
```

```

simu <- grf(70,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)
plot(variogram(simu,max.dist=dist.max))
simu$data <- 100 - 2*simu$coords[,1] + .3*simu$coords[,2]+ simu$data
v1 <- variogram(simu,max.dist=dist.max,trend="cte")
plot(v1)
sink("simulado_5.txt",type="output", append=T)
simu
sink()
#####
#####
## Simulação 6:
simu <- grf(50,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)
plot(variogram(simu,max.dist=dist.max))
simu$data <- 100 - 2*simu$coords[,1] + .3*simu$coords[,2]+ simu$data
v1 <- variogram(simu,max.dist=dist.max,trend="cte")
plot(v1)
sink("simulado_6.txt",type="output", append=T)
simu
sink()
#####
#####
## Simulação 7:
simu <- grf(30,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)
plot(variogram(simu,max.dist=dist.max))
simu$data <- 100 - 2*simu$coords[,1] + .3*simu$coords[,2]+ simu$data
v1 <- variogram(simu,max.dist=dist.max,trend="cte")
plot(v1)

```

```

write.table(simu,"simulado7.csv",append=FALSE,
sep="\t",quote=F,row.names=T,col.names=T,
na="NA",qmethod = "double")
d30a=read.csv2("simulado7.csv",sep="\t")
attach(d30a)
is.data.frame(d30a)
d30=read.geodata("simulado77.txt",head=T)
sink("simulado_7.txt",type="output", append=T)
simu
sink()
#####
#####
## Simulação 8:
simu <- grf(20,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)
plot(variog(simu,max.dist=dist.max))
simu$data <- 100 - 2*simu$coords[,1] + .3*simu$coords[,2]+ simu$data
v1 <- variog(simu,max.dist=dist.max,trend="cte")
plot(v1)
write.table(simu,"simulado8.csv",append=FALSE,
sep="\t",quote=F,row.names=T,col.names=T,
na="NA",qmethod = "double")
d20a=read.csv2("simulado8.csv",sep="\t")
attach(d20a)
is.data.frame(d20a)
d20=read.geodata("simulado8.txt",head=T)
sink("simulado_8.txt",type="output", append=T)
simu
sink()
#####

```

```
#####
#####
# Análise Simulação 1 - KLS
#library(geoR)
d150 <- read.geodata("simulado1.txt",head=T)
d150
plot(d150$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 150 pontos",pch=19,ylim=c(0,1.1))
text(d150$coords[,1],d150$coords[,2],round(d150$data,1),adj=0,
cex=0.8,pos=3)
summary(d150)
mean(d150$data)
var(d150$data)
dist.max = .8*max(dist(d150$coords))
v150=variog(d150,max.dist=dist.max)
plot(v150,main="Variograma - 150 pontos",
xlab="Distâncias",ylab="Semivariâncias")
wls150 = variofit(v150,cov.model="sph",fix.nug=T, nug=0.1)
lines(wls150)
grid150 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig150 <- krige.conv(d150,loc=grid150,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls150$cov.pars))
image(krig150, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte
mean(krig150$krige.var)
#####
# Análise Simulação 1 - KLU
```

```

dist.max = .8*max(dist(d150$coords))
v150u = variog(d150,max.dist=dist.max,
trend="1st")
plot(v150u,main="Variograma - 150 pontos - U",
xlab="Distâncias",ylab="Semivariâncias")
wls150u = variofit(v150u,cov.model="sph",fix.nug=T, nug=0.1)
lines(wls150u)
grid150u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig150u <- krige.conv(d150,loc=grid150u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls150u$cov.pars))
image(krig150u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte
mean(krig150u$krige.var)
var(d150$coords[,1])
var(d150$coords[,2])
(var.result=mean(krig150u$krige.var)+var(d150$coords[,1])+
var(d150$coords[,2]))
#####
#####
#####
# Análise Simulação 2 - KLS
#library(geoR)
d130 <- read.geodata("simulado2.txt",head=T)
d130
plot(d130$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 130 pontos",pch=19,ylim=c(0,1.1))
text(d130$coords[,1],d130$coords[,2],round(d130$data,1),adj=0,

```

```

cex=0.8,pos=3)
summary(d130)
mean(d130$data)
var(d130$data)
dist.max = .8*max(dist(d130$coords))
v130=variog(d130,max.dist=dist.max)
plot(v130,main="Variograma - 130 pontos",
xlab="Distâncias",ylab="Semivariâncias")
wls130 = variofit(v130,cov.model="sph",fix.nug=T, nug=0.1)
lines(wls130)
grid130 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig130 <- krige.conv(d130,loc=grid130,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls130$cov.pars))
image(krig130, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte
mean(krig130$krige.var)
#####
# Análise Simulação 2 - KLU
dist.max = .8*max(dist(d130$coords))
v130u = variog(d130,max.dist=dist.max,
trend="1st")
plot(v130u,main="Variograma - 130 pontos - U",
xlab="Distâncias",ylab="Semivariâncias")
wls130u = variofit(v130u,cov.model="sph",fix.nug=T, nug=0.1)
lines(wls130u)
grid130u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig130u <- krige.conv(d130,loc=grid130u,

```

```

krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls130u$cov.pars))
image(krig130u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte
mean(krig130u$krige.var)
var(d130$coords[,1])
var(d130$coords[,2])
(var.result=mean(krig130u$krige.var)+var(d130$coords[,1])+
var(d130$coords[,2]))
#####
#####
#####
# Análise Simulação 3 - KLS
#library(geoR)
d110 <- read.geodata("simulado3.txt",head=T)
d110
plot(d110$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 110 pontos",pch=19,ylim=c(0,1.1))
text(d110$coords[,1],d110$coords[,2],round(d110$data,1),adj=0,
cex=0.8,pos=3)
summary(d110)
mean(d110$data)
var(d110$data)
dist.max = .8*max(dist(d110$coords))
v110=variog(d110,max.dist=dist.max)
plot(v110,main="Variograma - 110 pontos",
xlab="Distâncias",ylab="Semivariâncias")
wls110 = variofit(v110,cov.model="sph",fix.nug=T, nug=0.1)

```

```

lines(wls110)
grid110 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig110 <- krige.conv(d110,loc=grid110,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls110$cov.pars))
image(krig110, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte
mean(krig110$krige.var)
#####
# Análise Simulação 3 - KLU
dist.max = .8*max(dist(d110$coords))
v110u = variog(d110,max.dist=dist.max,
trend="1st")
plot(v110u,main="Variograma - 110 pontos - U",
xlab="Distâncias",ylab="Semivariâncias")
wls110u = variofit(v110u,cov.model="sph",fix.nug=T, nug=0.1)
lines(wls110u)
grid110u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig110u <- krige.conv(d110,loc=grid110u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls110u$cov.pars))
image(krig110u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte
mean(krig110u$krige.var)
var(d110$coords[,1])
var(d110$coords[,2])

```

```

(var.result=mean(krig110u$krige.var)+var(d110$coords[,1])+
var(d110$coords[,2]))
#####
#####
#####
# Análise Simulação 4 - KLS
#library(geoR)
d90 <- read.geodata("simulado4.txt",head=T)
d90
plot(d90$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 90 pontos",pch=19,ylim=c(0,1.1))
text(d90$coords[,1],d90$coords[,2],round(d90$data,1),adj=0,
cex=0.8,pos=3)
summary(d90)
mean(d90$data)
var(d90$data)
dist.max = .8*max(dist(d90$coords))
v90=variog(d90,max.dist=dist.max)
plot(v90,main="Variograma - 90 pontos",
xlab="Distâncias",ylab="Semivariâncias")
wls90 = variofit(v90,cov.model="sph",fix.nug=T, nug=0.1)
lines(wls90)
grid90 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig90 <- krige.conv(d90,loc=grid90,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls90$cov.pars))
image(krig90, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte

```

```

mean(krig90$krige.var)
#####
# Análise Simulação 4 - KLU
dist.max = .8*max(dist(d90$coords))
v90u = variog(d90,max.dist=dist.max,
trend="1st")
plot(v90u,main="Variograma - 90 pontos - U",
xlab="Distâncias",ylab="Semivariâncias")
wls90u = variofit(v90u,cov.model="sph",fix.nug=T, nug=0.1)
lines(wls90u)
grid90u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig90u <- krige.conv(d90,loc=grid90u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls90u$cov.pars))
image(krig90u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte
mean(krig90u$krige.var)
var(d90$coords[,1])
var(d90$coords[,2])
(var.result=mean(krig90u$krige.var)+var(d90$coords[,1])+
var(d90$coords[,2]))
#####
#####
#####
# Análise Simulação 5 - KLS
#library(geoR)
d70 <- read.geodata("simulado5.txt",head=T)
d70

```

```

plot(d70$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 70 pontos",pch=19,ylim=c(0,1.1))
text(d70$coords[,1],d70$coords[,2],round(d70$data,1),adj=0,
cex=0.8,pos=3)
summary(d70)
mean(d70$data)
var(d70$data)
dist.max = .8*max(dist(d70$coords))
v70=variog(d70,max.dist=dist.max)
plot(v70,main="Variograma - 70 pontos",
xlab="Distâncias",ylab="Semivariâncias")
wls70 = variofit(v70,cov.model="sph",fix.nug=T, nug=0.1)
lines(wls70)
grid70 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig70 <- krige.conv(d70,loc=grid70,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls70$cov.pars))
image(krig70, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte
mean(krig70$krige.var)
#####
# Análise Simulação 5 - KLU
dist.max = .8*max(dist(d70$coords))
v70u = variog(d70,max.dist=dist.max,
trend="1st")
plot(v70u,main="Variograma - 70 pontos - U",
xlab="Distâncias",ylab="Semivariâncias")
wls70u = variofit(v70u,cov.model="sph",fix.nug=T, nug=0.1)

```

```

lines(wls70u)
grid70u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig70u <- krige.conv(d70,loc=grid70u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls70u$cov.pars))
image(krig70u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte
mean(krig70u$krige.var)
var(d70$coords[,1])
var(d70$coords[,2])
(var.result=mean(krig70u$krige.var)+var(d70$coords[,1])+
var(d70$coords[,2]))
#####
#####
#####
# Análise Simulação 6 - KLS
#library(geoR)
d50 <- read.geodata("simulado6.txt",head=T)
d50
plot(d50$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 50 pontos",pch=19,ylim=c(0,1.1))
text(d50$coords[,1],d50$coords[,2],round(d50$data,1),adj=0,
cex=0.8,pos=3)
summary(d50)
mean(d50$data)
var(d50$data)
dist.max = .8*max(dist(d50$coords))
v50=variog(d50,max.dist=dist.max)

```

```

plot(v50,main="Variograma - 50 pontos",
     xlab="Distâncias",ylab="Semivariâncias")
wls50 = variofit(v50,cov.model="sph",fix.nug=T, nug=0.1)
lines(wls50)
grid50 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig50 <- krige.conv(d50,loc=grid50,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls50$cov.pars))
image(krig50, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte
mean(krig50$krige.var)
#####
# Análise Simulação 6 - KLU
dist.max = .8*max(dist(d50$coords))
v50u = variog(d50,max.dist=dist.max,
trend="1st")
plot(v50u,main="Variograma - 50 pontos - U",
     xlab="Distâncias",ylab="Semivariâncias")
wls50u = variofit(v50u,cov.model="sph",fix.nug=T, nug=0.1)
lines(wls50u)
grid50u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig50u <- krige.conv(d50,loc=grid50u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls50u$cov.pars))
image(krig50u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte

```

```

mean(krig50u$krige.var)
var(d50$coords[,1])
var(d50$coords[,2])
(var.result=mean(krig50u$krige.var)+var(d50$coords[,1])+
var(d50$coords[,2]))
#####
#####
#####
# Análise Simulação 7 - KLS
#library(geoR)
d30 <- read.geodata("simulado7.txt",head=T)
d30
plot(d30$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 30 pontos",pch=19,ylim=c(0,1.1))
text(d30$coords[,1],d30$coords[,2],round(d30$data,1),adj=0,
cex=0.8,pos=3)
summary(d30)
mean(d30$data)
var(d30$data)
dist.max = .8*max(dist(d30$coords))
v30=variog(d30,max.dist=dist.max)
plot(v30,main="Variograma - 30 pontos",
xlab="Distâncias",ylab="Semivariâncias")
wls30 = variofit(v30,cov.model="sph",fix.nug=T, nug=0.1)
lines(wls30)
grid30 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig30 <- krige.conv(d30,loc=grid30,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls30$cov.pars))

```

```

image(krig30, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte
mean(krig30$krige.var)
#####
# Análise Simulação 7 - KLU
dist.max = .8*max(dist(d30$coords))
v30u = variog(d30,max.dist=dist.max,
trend="1st")
plot(v30u,main="Variograma - 30 pontos - U",
xlab="Distâncias",ylab="Semivariâncias")
wls30u = variofit(v30u,cov.model="sph",fix.nug=T, nug=0.1)
lines(wls30u)
grid30u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig30u <- krige.conv(d30,loc=grid30u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls30u$cov.pars))
image(krig30u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte
mean(krig30u$krige.var)
var(d30$coords[,1])
var(d30$coords[,2])
(var.result=mean(krig30u$krige.var)+var(d30$coords[,1])+
var(d30$coords[,2]))
#####
#####
#####
# Análise Simulação 8 - KLS

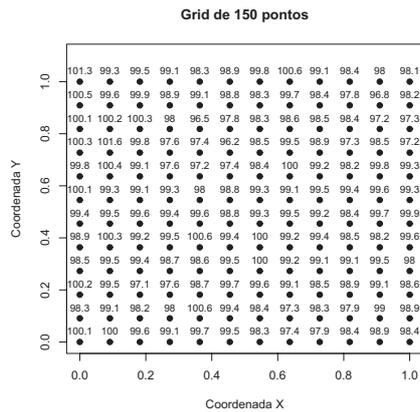
```

```

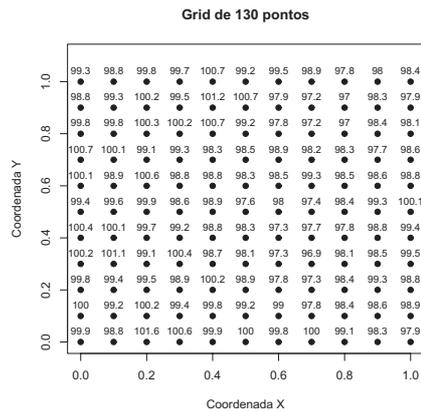
#library(geoR)
d20 <- read.geodata("simulado8.txt",head=T)
d20
plot(d20$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 20 pontos",pch=19,ylim=c(0,1.1))
text(d20$coords[,1],d20$coords[,2],round(d20$data,1),adj=0,
cex=0.8,pos=3)
summary(d20)
mean(d20$data)
var(d20$data)
dist.max = .8*max(dist(d20$coords))
v20=variog(d20,max.dist=dist.max)
plot(v20,main="Variograma - 20 pontos",
xlab="Distâncias",ylab="Semivariâncias")
wls20 = variofit(v20,cov.model="sph",fix.nug=T, nug=0.1)
lines(wls20)
grid20 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig20 <- krige.conv(d20,loc=grid20,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls20$cov.pars))
image(krig20, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte
mean(krig20$krige.var)
#####
# Análise Simulação 8 - KLU
dist.max = .8*max(dist(d20$coords))
v20u = variog(d20,max.dist=dist.max,
trend="1st")

```

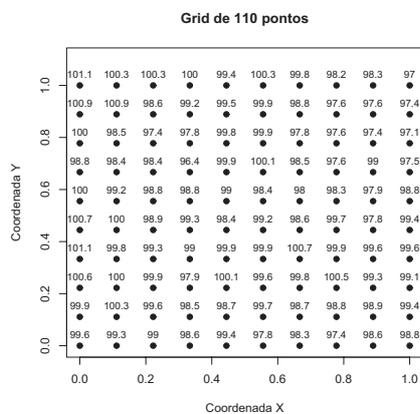
```
plot(v20u,main="Variograma - 20 pontos - U",
     xlab="Distâncias",ylab="Semivariâncias")
wls20u = variofit(v20u,cov.model="sph",fix.nug=T, nug=0.1)
lines(wls20u)
grid20u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig20u <- krige.conv(d20,loc=grid20u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls20u$cov.pars))
image(krig20u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte
mean(krig20u$krige.var)
var(d20$coords[,1])
var(d20$coords[,2])
(var.result=mean(krig20u$krige.var)+var(d20$coords[,1])+
var(d20$coords[,2]))
```



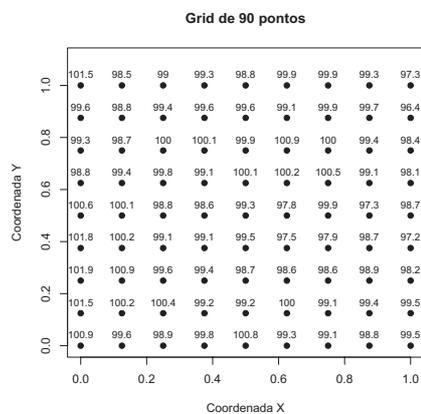
(a) P.G. com tendência



(b) P.G. com tendência



(c) P.G. com tendência



(d) P.G. com tendência

Figura 25 *Grid's* regulares de dados simulados, considerando um Processo Gaussiano de média zero e com tendência

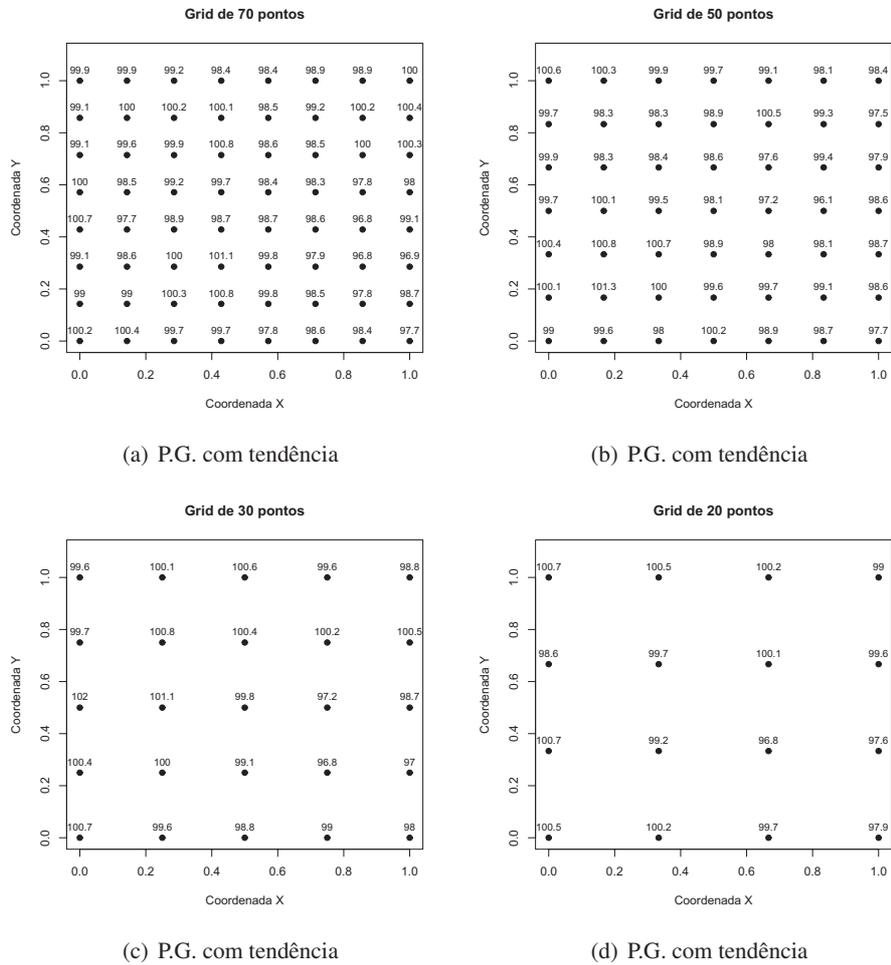
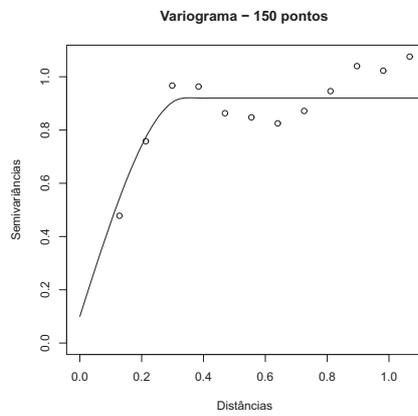
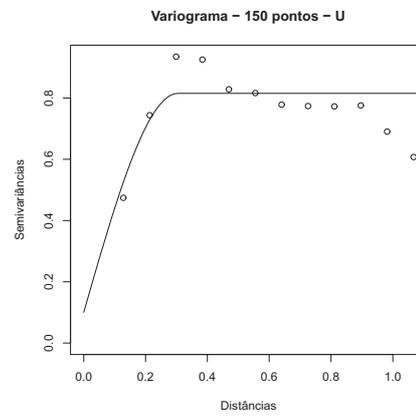


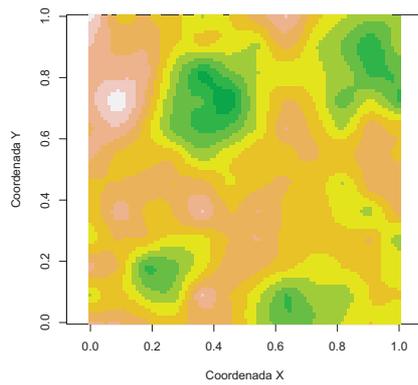
Figura 26 *Grid's* regulares de dados simulados, considerando um Processo Gaussiano de média zero e com tendência



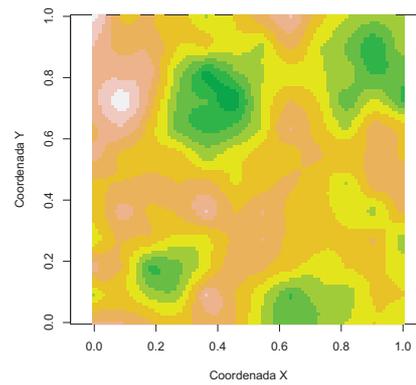
(a) Variograma para KLS



(b) Variograma para KLU

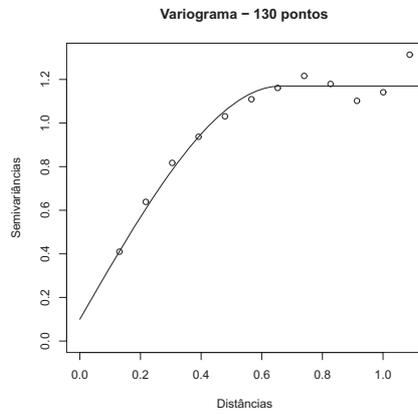


(c) KLS de 150 pontos

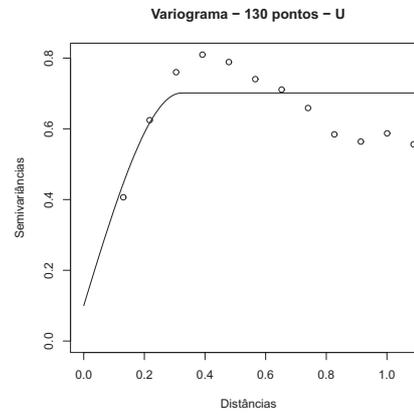


(d) KLU de 150 pontos

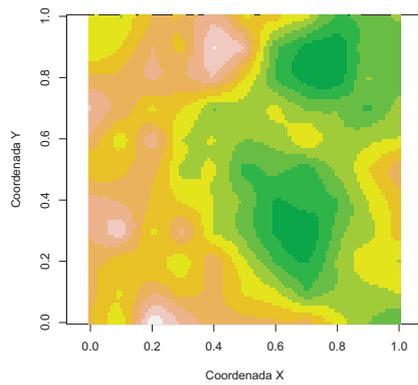
Figura 27 Análise geoestatística de um *grid* regular de 150 pontos simulados, considerando o modelo esférico de semivariograma



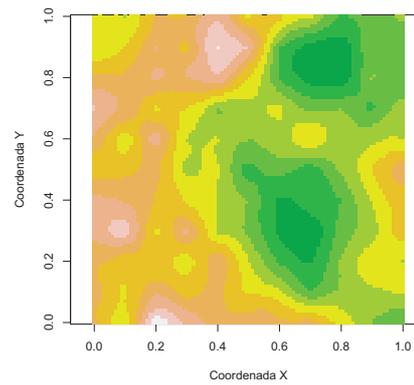
(a) Variograma para KLS



(b) Variograma para KLU

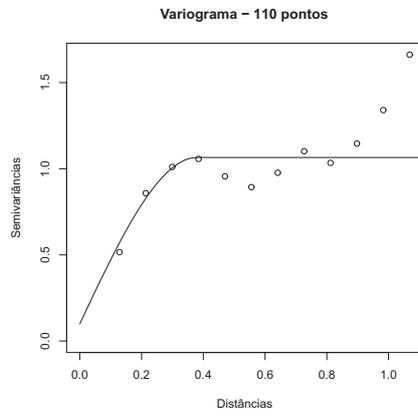


(c) KLS de 130 pontos

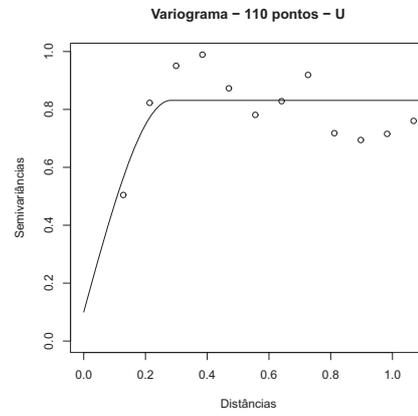


(d) KLU de 130 pontos

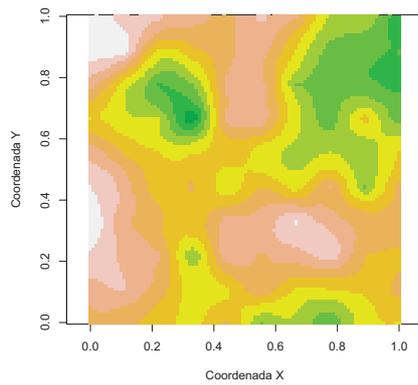
Figura 28 Análise geoestatística de um *grid* regular de 130 pontos simulados, considerando o modelo esférico de semivariograma



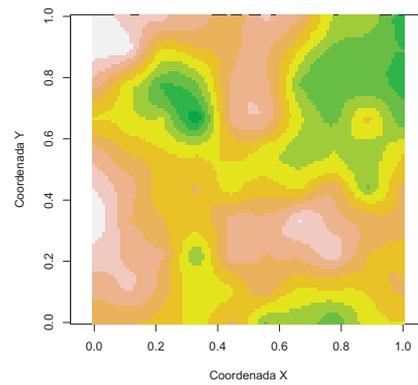
(a) Variograma para KLS



(b) Variograma para KLU

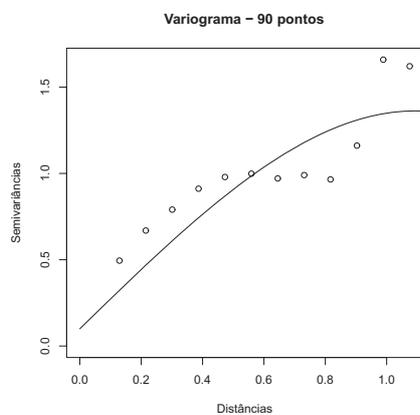


(c) KLS de 110 pontos

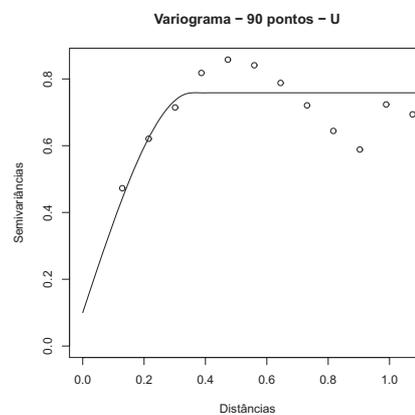


(d) KLU de 110 pontos

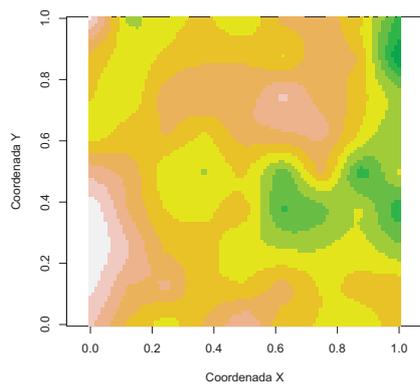
Figura 29 Análise geoestatística de um *grid* regular de 110 pontos simulados, considerando o modelo esférico de semivariograma



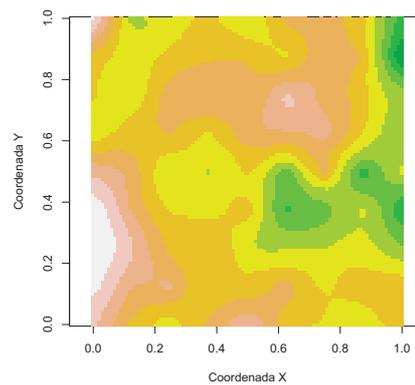
(a) Variograma para KLS



(b) Variograma para KLU

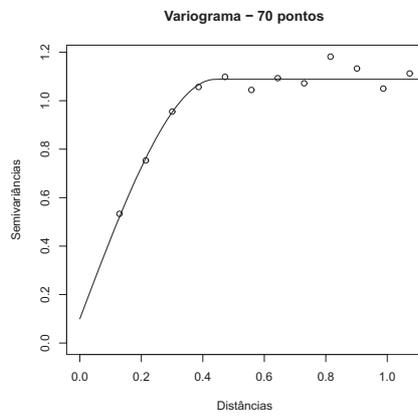


(c) KLS de 90 pontos

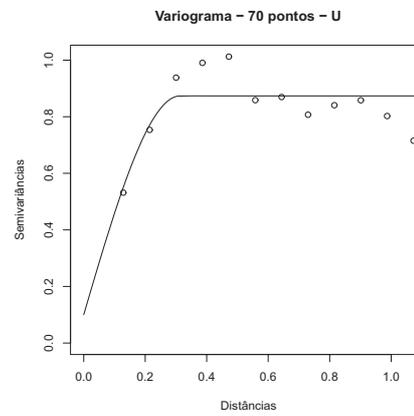


(d) KLU de 90 pontos

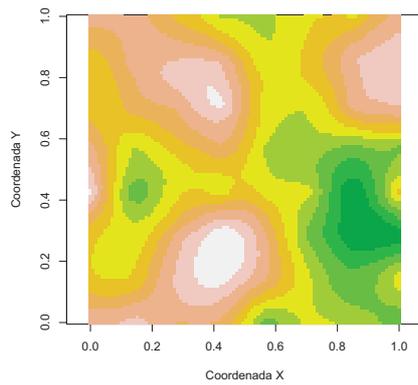
Figura 30 Análise geoestatística de um *grid* regular de 90 pontos simulados, considerando o modelo esférico de semivariograma



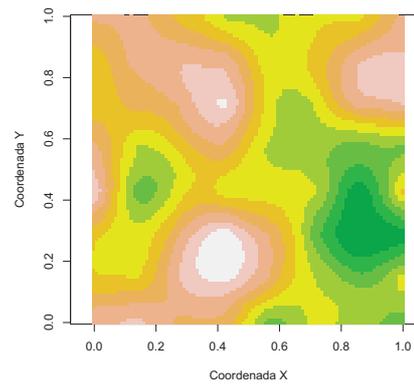
(a) Variograma para KLS



(b) Variograma para KLU

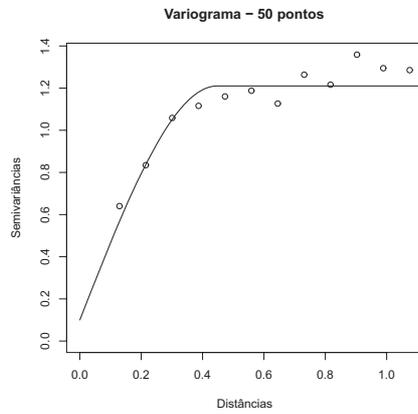


(c) KLS de 70 pontos

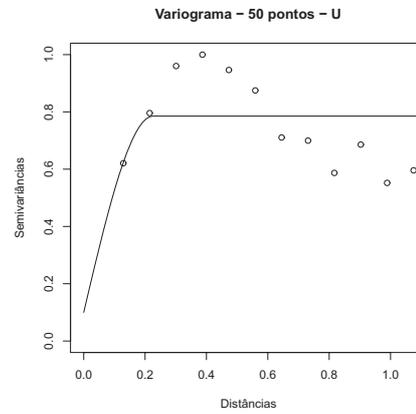


(d) KLU de 70 pontos

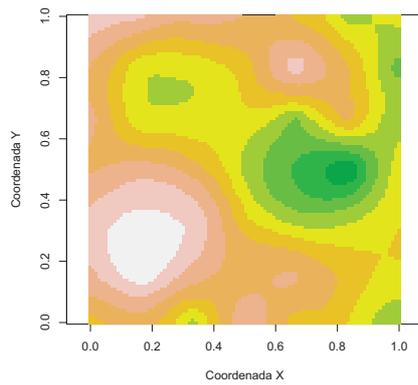
Figura 31 Análise geoestatística de um *grid* regular de 70 pontos simulados, considerando o modelo esférico de semivariograma



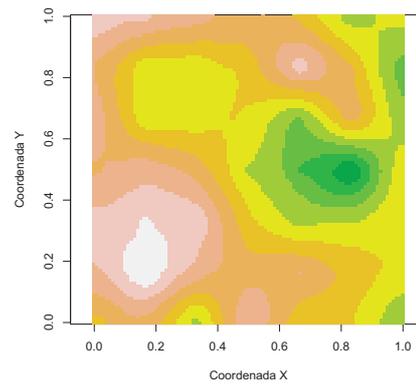
(a) Variograma para KLS



(b) Variograma para KLU

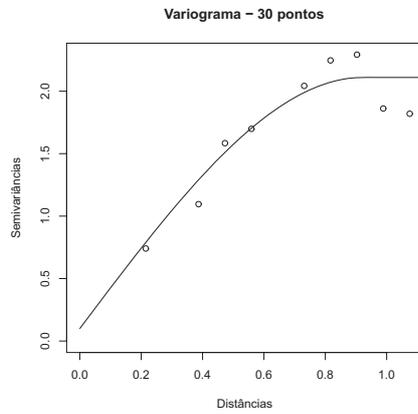


(c) KLS de 50 pontos

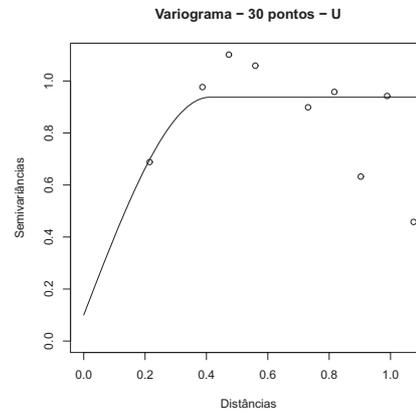


(d) KLU de 50 pontos

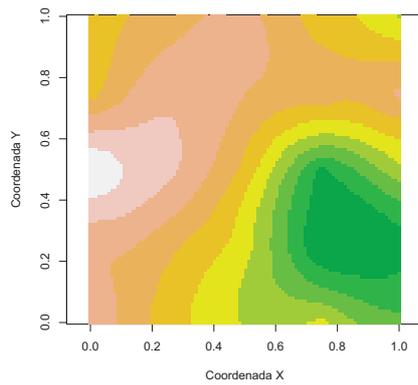
Figura 32 Análise geoestatística de um *grid* regular de 50 pontos simulados, considerando o modelo esférico de semivariograma



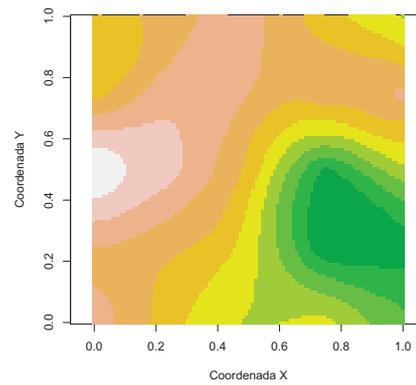
(a) Variograma para KLS



(b) Variograma para KLU

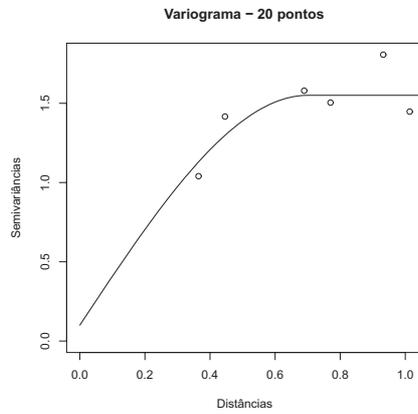


(c) KLS de 30 pontos

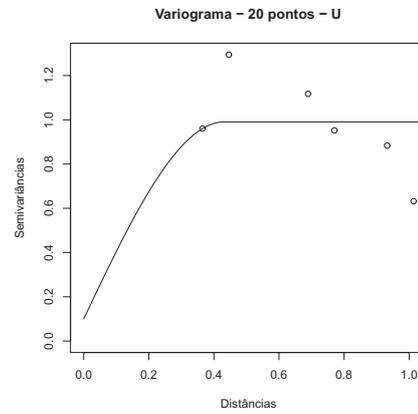


(d) KLU de 30 pontos

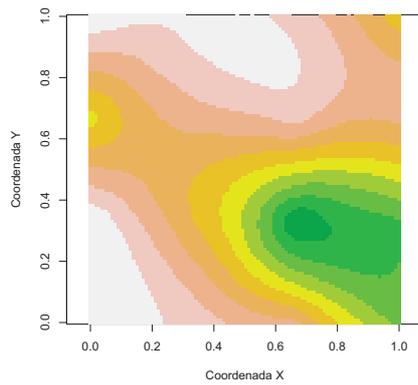
Figura 33 Análise geoestatística de um *grid* regular de 30 pontos simulados, considerando o modelo esférico de semivariograma



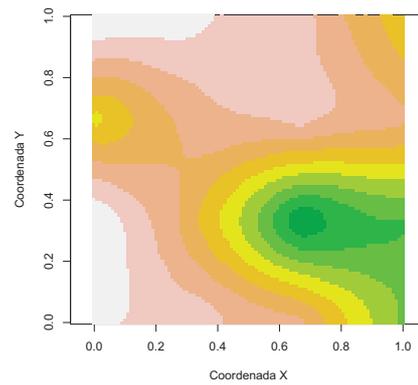
(a) Variograma para KLS



(b) Variograma para KLU



(c) KLS de 20 pontos



(d) KLU de 20 pontos

Figura 34 Análise geoestatística de um *grid* regular de 20 pontos simulados, considerando o modelo esférico de semivariograma

Processo desconhecido

```
#####
## Simulação 1 - Perturbação:
simu <- grf(150,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)
result = vector(mode="numeric")
norm=exp(rnorm(length(simu$data),0,1))
for (i in 1:length(simu$data))
{result[i]=simu$data[[i]]
}
total=as.geodata(data.frame(X=simu$coords[,1],Y=simu$coords[,2],
Dados=result+norm))
plot(variog(total,max.dist=.8*max(dist(total$coords))))
total$data <- 100 - 2*total$coords[,1] + .3*total$coords[,2]+
total$data
v1 <- variog(total,max.dist=.8*max(dist(total$coords)),trend="cte")
plot(v1)
wls = variofit(v1,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls)
write.table(total,"perturbal.csv",append=FALSE,
sep="\t",quote=F,row.names=T,col.names=T,
na="NA",qmethod = "double")
p150=read.csv2("perturbal.csv",sep="\t")
attach(p150)
is.data.frame(p150)
p150=read.geodata("perturbal.txt",head=T)
p150
#####
#####
## Simulação 2 - Perturbação:
```

```

simu <- grf(130,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)
result = vector(mode="numeric")
norm=exp(rnorm(length(simu$data),0,1))
for (i in 1:length(simu$data))
{result[i]=simu$data[[i]]
}
total=as.geodata(data.frame(X=simu$coords[,1],Y=simu$coords[,2],
Dados=result+norm))
plot(variog(total,max.dist=.8*max(dist(total$coords))))
total$data <- 100 - 2*total$coords[,1] + .3*total$coords[,2]+
total$data
v1 <- variog(total,max.dist=.8*max(dist(total$coords)),trend="cte")
plot(v1)
wls = variofit(v1,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls)
v2 = variog(total,max.dist=dist.max,trend="1st")
plot(v2,main="Variograma - 130 pontos - U",
xlab="Distâncias",ylab="Semivariâncias")
wlsu = variofit(v2,cov.model="sph",fix.nug=F, nug=0.1)
lines(wlsu)
write.table(total,"perturba2.csv",append=FALSE,
sep="\t",quote=F,row.names=T,col.names=T,
na="NA",qmethod = "double")
p130=read.csv2("perturba2.csv",sep="\t")
attach(p130)
is.data.frame(p130)
p130=read.geodata("perturba2.txt",head=T)
p130
#####

```

```
#####
## Simulação 3 - Perturbação:
simu <- grf(110,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)
result = vector(mode="numeric")
norm=exp(rnorm(length(simu$data),0,1))
for (i in 1:length(simu$data))
{result[i]=simu$data[[i]]
}
total=as.geodata(data.frame(X=simu$coords[,1],Y=simu$coords[,2],
Dados=result+norm))
plot(variog(total,max.dist=.8*max(dist(total$coords))))
total$data <- 100 - 2*total$coords[,1] + .3*total$coords[,2]+
total$data
v1 <- variog(total,max.dist=.8*max(dist(total$coords)),trend="cte")
plot(v1)
wls = variofit(v1,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls)
write.table(total,"perturba3.csv",append=FALSE,
sep="\t",quote=F,row.names=T,col.names=T,
na="NA",qmethod = "double")
p110=read.csv2("perturba3.csv",sep="\t")
attach(p110)
is.data.frame(p110)
p110=read.geodata("perturba3.txt",head=T)
p110
#####
#####
## Simulação 4 - Perturbação:
simu <- grf(90,grid="reg",cov.model="sph",cov.par=c(.65,.35),
```

```

nug=.1,mean=0,RF=T)
result = vector(mode="numeric")
norm=exp(rnorm(length(simu$data),0,1))
for (i in 1:length(simu$data))
{result[i]=simu$data[[i]]
}
total=as.geodata(data.frame(X=simu$coords[,1],Y=simu$coords[,2],
Dados=result+norm))
plot(variog(total,max.dist=.8*max(dist(total$coords))))
total$data <- 100 - 2*total$coords[,1] + .3*total$coords[,2]+
total$data
v1 <- variog(total,max.dist=.8*max(dist(total$coords)),trend="cte")
plot(v1)
wls = variofit(v1,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls)
write.table(total,"perturba4.csv",append=FALSE,
sep="\t",quote=F,row.names=T,col.names=T,
na="NA",qmethod = "double")
p90=read.csv2("perturba4.csv",sep="\t")
attach(p90)
is.data.frame(p90)
p90=read.geodata("perturba4.txt",head=T)
p90
#####
#####
## Simulação 5 - Perturbação:
simu <- grf(70,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)
result = vector(mode="numeric")
norm=exp(rnorm(length(simu$data),0,1))

```

```

for (i in 1:length(simu$data))
{result[i]=simu$data[[i]]
}
total=as.geodata(data.frame(X=simu$coords[,1],Y=simu$coords[,2],
Dados=result+norm))
plot(variog(total,max.dist=.8*max(dist(total$coords))))
total$data <- 100 - 2*total$coords[,1] + .3*total$coords[,2]+
total$data
v1 <- variog(total,max.dist=.8*max(dist(total$coords)),trend="cte")
plot(v1)
wls = variofit(v1,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls)
write.table(total,"perturba5.csv",append=FALSE,
sep="\t",quote=F,row.names=T,col.names=T,
na="NA",qmethod = "double")
p70=read.csv2("perturba5.csv",sep="\t")
attach(p70)
is.data.frame(p70)
p70=read.geodata("perturba5.txt",head=T)
p70
#####
#####
## Simulação 6 - Perturbação:
simu <- grf(50,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)
result = vector(mode="numeric")
norm=exp(rnorm(length(simu$data),0,1))
for (i in 1:length(simu$data))
{result[i]=simu$data[[i]]
}

```

```

total=as.geodata(data.frame(X=simu$coords[,1],Y=simu$coords[,2],
Dados=result+norm))
plot(variog(total,max.dist=.8*max(dist(total$coords))))
total$data <- 100 - 2*total$coords[,1] + .3*total$coords[,2]+
total$data
v1 <- variog(total,max.dist=.8*max(dist(total$coords)),trend="cte")
plot(v1)
wls = variofit(v1,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls)
write.table(total,"perturba6.csv",append=FALSE,
sep="\t",quote=F,row.names=T,col.names=T,
na="NA",qmethod = "double")
p50=read.csv2("perturba6.csv",sep="\t")
attach(p50)
is.data.frame(p50)
p50=read.geodata("perturba6.txt",head=T)
p50
#####
#####
## Simulação 7 - Perturbação:
simu <- grf(30,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)
result = vector(mode="numeric")
norm=exp(rnorm(length(simu$data),0,1))
for (i in 1:length(simu$data))
{result[i]=simu$data[[i]]
}
total=as.geodata(data.frame(X=simu$coords[,1],Y=simu$coords[,2],
Dados=result+norm))
plot(variog(total,max.dist=.8*max(dist(total$coords))))

```

```

total$data <- 100 - 2*total$coords[,1] + .3*total$coords[,2]+
total$data
v1 <- variog(total,max.dist=.8*max(dist(total$coords)),trend="cte")
plot(v1)
wls = variofit(v1,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls)
write.table(total,"perturba7.csv",append=FALSE,
sep="\t",quote=F,row.names=T,col.names=T,
na="NA",qmethod = "double")
p30=read.csv2("perturba7.csv",sep="\t")
attach(p30)
is.data.frame(p30)
p30=read.geodata("perturba7.txt",head=T)
p30
#####
#####
## Simulação 8 - Perturbação:
simu <- grf(20,grid="reg",cov.model="sph",cov.par=c(.65,.35),
nug=.1,mean=0,RF=T)
result = vector(mode="numeric")
norm=exp(rnorm(length(simu$data),0,1))
for (i in 1:length(simu$data))
{result[i]=simu$data[[i]]
}
total=as.geodata(data.frame(X=simu$coords[,1],Y=simu$coords[,2],
Dados=result+norm))
plot(variog(total,max.dist=.8*max(dist(total$coords))))
total$data <- 100 - 2*total$coords[,1] + .3*total$coords[,2]+
total$data
v1 <- variog(total,max.dist=.8*max(dist(total$coords)),trend="cte")

```

```

plot(v1)
wls = variofit(v1,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls)
write.table(total,"perturba8.csv",append=FALSE,
sep="\t",quote=F,row.names=T,col.names=T,
na="NA",qmethod = "double")
p20=read.csv2("perturba8.csv",sep="\t")
attach(p20)
is.data.frame(p20)
p20=read.geodata("perturba8.txt",head=T)
p20
#####
#####
#####
# Análise Simulação 1 - KLS - perturbação
#library(geoR)
p150 <- read.geodata("perturba1.txt",head=T)
p150
plot(p150$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 150 pontos",pch=19,ylim=c(0,1.1))
text(p150$coords[,1],p150$coords[,2],round(p150$data,1),adj=0,
cex=0.8,pos=3)
summary(p150)
mean(p150$data)
var(p150$data)
dist.max = .8*max(dist(p150$coords))
v150=variog(p150,max.dist=dist.max)
plot(v150,main="Variograma - 150 pontos",
xlab="Distâncias",ylab="Semivariâncias")
wls150 = variofit(v150,cov.model="sph",fix.nug=F, nug=0.1)

```

```

lines(wls150)
grid150 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig150 <- krige.conv(p150,loc=grid150,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls150$cov.pars))
image(krig150, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte
mean(krig150$krige.var)
#####
# Análise Simulação 1 - KLU - perturbação
dist.max = .8*max(dist(p150$coords))
v150u = variog(p150,max.dist=dist.max,
trend="1st")
plot(v150u,main="Variograma - 150 pontos - U",
xlab="Distâncias",ylab="Semivariâncias")
wls150u = variofit(v150u,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls150u)
grid150u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig150u <- krige.conv(p150,loc=grid150u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls150u$cov.pars))
image(krig150u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte
mean(krig150u$krige.var)
var(p150$coords[,1])
var(p150$coords[,2])

```

```

(var.result=mean(krig150u$krige.var)+var(p150$coords[,1])+
var(p150$coords[,2]))
#####
#####
#####
# Análise Simulação 2 - KLS - perturbação
#library(geoR)
p130 <- read.geodata("perturba2.txt",head=T)
p130
plot(p130$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 130 pontos",pch=19,ylim=c(0,1.1))
text(p130$coords[,1],p130$coords[,2],round(p130$data,1),adj=0,
cex=0.8,pos=3)
summary(p130)
mean(p130$data)
var(p130$data)
dist.max = .8*max(dist(p130$coords))
v130=variog(p130,max.dist=dist.max)
plot(v130,main="Variograma - 130 pontos",
xlab="Distâncias",ylab="Semivariâncias")
wls130 = variofit(v130,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls130)
grid130 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig130 <- krige.conv(p130,loc=grid130,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls130$cov.pars))
image(krig130, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte

```

```

mean(krig130$krige.var)
#####
# Análise Simulação 2 - KLU - perturbação
dist.max = .8*max(dist(p130$coords))
v130u = variog(p130,max.dist=dist.max,
trend="1st")
plot(v130u,main="Variograma - 130 pontos - U",
xlab="Distâncias",ylab="Semivariâncias")
wls130u = variofit(v130u,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls130u)
grid130u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig130u <- krige.conv(p130,loc=grid130u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls130u$cov.pars))
image(krig130u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte
mean(krig130u$krige.var)
var(p130$coords[,1])
var(p130$coords[,2])
(var.result=mean(krig130u$krige.var)+var(p130$coords[,1])+
var(p130$coords[,2]))
#####
#####
#####
# Análise Simulação 3 - KLS - perturbação
#library(geoR)
p110 <- read.geodata("perturba3.txt",head=T)
p110

```

```

plot(p110$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 110 pontos",pch=19,ylim=c(0,1.1))
text(p110$coords[,1],p110$coords[,2],round(p110$data,1),adj=0,
cex=0.8,pos=3)
summary(p110)
mean(p110$data)
var(p110$data)
dist.max = .8*max(dist(p110$coords))
v110=variog(p110,max.dist=dist.max)
plot(v110,main="Variograma - 110 pontos",
xlab="Distâncias",ylab="Semivariâncias")
wls110 = variofit(v110,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls110)
grid110 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig110 <- krige.conv(p110,loc=grid110,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls110$cov.pars))
image(krig110, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte
mean(krig110$krige.var)
#####
# Análise Simulação 3 - KLU - perturbação
dist.max = .8*max(dist(p110$coords))
v110u = variog(p110,max.dist=dist.max,
trend="1st")
plot(v110u,main="Variograma - 110 pontos - U",
xlab="Distâncias",ylab="Semivariâncias")
wls110u = variofit(v110u,cov.model="sph",fix.nug=F, nug=0.1)

```

```

lines(wls110u)
grid110u <- expand.grid(seq(0,1,1=100), seq(0,1,1=100))
krig110u <- krige.conv(p110,loc=grid110u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls110u$cov.pars))
image(krig110u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte
mean(krig110u$krige.var)
var(p110$coords[,1])
var(p110$coords[,2])
(var.result=mean(krig110u$krige.var)+var(p110$coords[,1])+
var(p110$coords[,2]))
#####
#####
#####
# Análise Simulação 4 - KLS - perturbação
#library(geoR)
p90 <- read.geodata("perturba4.txt",head=T)
p90
plot(p90$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 90 pontos",pch=19,ylim=c(0,1.1))
text(p90$coords[,1],p90$coords[,2],round(p90$data,1),adj=0,
cex=0.8,pos=3)
summary(p90)
mean(p90$data)
var(p90$data)
dist.max = .8*max(dist(p90$coords))
v90=variog(p90,max.dist=dist.max)

```

```

plot(v90,main="Variograma - 90 pontos",
     xlab="Distâncias",ylab="Semivariâncias")
wls90 = variofit(v90,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls90)
grid90 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig90 <- krige.conv(p90,loc=grid90,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls90$cov.pars))
image(krig90, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte
mean(krig90$krige.var)
#####
# Análise Simulação 4 - KLU - perturbação
dist.max = .8*max(dist(p90$coords))
v90u = variog(p90,max.dist=dist.max,
trend="1st")
plot(v90u,main="Variograma - 90 pontos - U",
     xlab="Distâncias",ylab="Semivariâncias")
wls90u = variofit(v90u,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls90u)
grid90u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig90u <- krige.conv(p90,loc=grid90u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls90u$cov.pars))
image(krig90u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte

```

```

mean(krig90u$krige.var)
var(p90$coords[,1])
var(p90$coords[,2])
(var.result=mean(krig90u$krige.var)+var(p90$coords[,1])+
var(p90$coords[,2]))
#####
#####
#####
# Análise Simulação 5 - KLS - perturbação
#library(geoR)
p70 <- read.geodata("perturba5.txt",head=T)
p70
plot(p70$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 70 pontos",pch=19,ylim=c(0,1.1))
text(p70$coords[,1],p70$coords[,2],round(p70$data,1),adj=0,
cex=0.8,pos=3)
summary(p70)
mean(p70$data)
var(p70$data)
dist.max = .8*max(dist(p70$coords))
v70=variog(p70,max.dist=dist.max)
plot(v70,main="Variograma - 70 pontos",
xlab="Distâncias",ylab="Semivariâncias")
wls70 = variofit(v70,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls70)
grid70 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig70 <- krige.conv(p70,loc=grid70,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls70$cov.pars))

```

```

image(krig70, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte
mean(krig70$krige.var)
#####
# Análise Simulação 5 - KLU - perturbação
dist.max = .8*max(dist(p70$coords))
v70u = variog(p70,max.dist=dist.max,
trend="1st")
plot(v70u,main="Variograma - 70 pontos - U",
xlab="Distâncias",ylab="Semivariâncias")
wls70u = variofit(v70u,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls70u)
grid70u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig70u <- krige.conv(p70,loc=grid70u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls70u$cov.pars))
image(krig70u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte
mean(krig70u$krige.var)
var(p70$coords[,1])
var(p70$coords[,2])
(var.result=mean(krig70u$krige.var)+var(p70$coords[,1])+
var(p70$coords[,2]))
#####
#####
#####
# Análise Simulação 6 - KLS - perturbação

```

```

#library(geoR)
p50 <- read.geodata("perturba6.txt",head=T)
p50
plot(p50$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 50 pontos",pch=19,ylim=c(0,1.1))
text(p50$coords[,1],p50$coords[,2],round(p50$data,1),adj=0,
cex=0.8,pos=3)
summary(p50)
mean(p50$data)
var(p50$data)
dist.max = .8*max(dist(p50$coords))
v50=variog(p50,max.dist=dist.max)
plot(v50,main="Variograma - 50 pontos",
xlab="Distâncias",ylab="Semivariâncias")
wls50 = variofit(v50,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls50)
grid50 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig50 <- krige.conv(p50,loc=grid50,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls50$cov.pars))
image(krig50, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte
mean(krig50$krige.var)
#####
# Análise Simulação 6 - KLU - perturbação
dist.max = .8*max(dist(p50$coords))
v50u = variog(p50,max.dist=dist.max,
trend="1st")

```

```

plot(v50u,main="Variograma - 50 pontos - U",
     xlab="Distâncias",ylab="Semivariâncias")
wls50u = variofit(v50u,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls50u)
grid50u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig50u <- krige.conv(p50,loc=grid50u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls50u$cov.pars))
image(krig50u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte
mean(krig50u$krige.var)
var(p50$coords[,1])
var(p50$coords[,2])
(var.result=mean(krig50u$krige.var)+var(p50$coords[,1])+
var(p50$coords[,2]))
#####
#####
#####
# Análise Simulação 7 - KLS - perturbação
#library(geoR)
p30 <- read.geodata("perturba7.txt",head=T)
p30
plot(p30$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 30 pontos",pch=19,ylim=c(0,1.1))
text(p30$coords[,1],p30$coords[,2],round(p30$data,1),adj=0,
cex=0.8,pos=3)
summary(p30)
mean(p30$data)

```

```

var(p30$data)
dist.max = .8*max(dist(p30$coords))
v30=variog(p30,max.dist=dist.max)
plot(v30,main="Variograma - 30 pontos",
xlab="Distâncias",ylab="Semivariâncias")
wls30 = variofit(v30,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls30)
grid30 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig30 <- krige.conv(p30,loc=grid30,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls30$cov.pars))
image(krig30, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte
mean(krig30$krige.var)
#####
# Análise Simulação 7 - KLU - perturbação
dist.max = .8*max(dist(p30$coords))
v30u = variog(p30,max.dist=dist.max,
trend="1st")
plot(v30u,main="Variograma - 30 pontos - U",
xlab="Distâncias",ylab="Semivariâncias")
wls30u = variofit(v30u,cov.model="sph",fix.nug=F, nug=0.1,
ini=c(1.5,0.2))
lines(wls30u)
grid30u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig30u <- krige.conv(p30,loc=grid30u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",

```

```

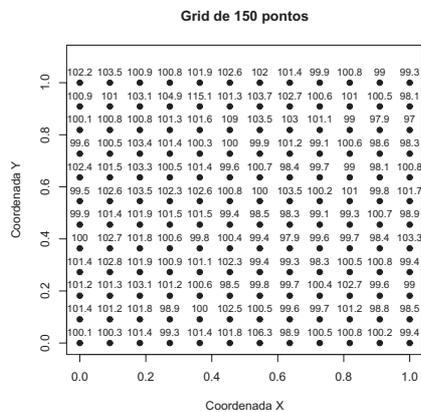
cov.model="sph", cov.pars=wls30u$cov.pars))
image(krig30u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte
mean(krig30u$krige.var)
var(p30$coords[,1])
var(p30$coords[,2])
(var.result=mean(krig30u$krige.var)+var(p30$coords[,1])+
var(p30$coords[,2]))
#####
#####
#####
# Análise Simulação 8 - KLS - perturbação
#library(geoR)
p20 <- read.geodata("perturba8.txt",head=T)
p20
plot(p20$coords,xlab="Coordenada X", ylab="Coordenada Y",
main="Grid de 20 pontos",pch=19,ylim=c(0,1.1))
text(p20$coords[,1],p20$coords[,2],round(p20$data,1),adj=0,
cex=0.8,pos=3)
summary(p20)
mean(p20$data)
var(p20$data)
dist.max = .8*max(dist(p20$coords))
v20=variog(p20,max.dist=dist.max)
plot(v20,main="Variograma - 20 pontos",
xlab="Distâncias",ylab="Semivariâncias")
wls20 = variofit(v20,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls20)
grid20 <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))

```

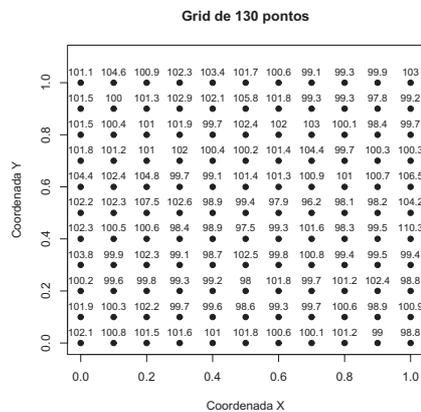
```

krig20 <- krige.conv(p20,loc=grid20,
krige=krige.control(type.krige = "SK", trend.d = "cte",
trend.l = "cte",
beta=100, cov.model="sph", cov.pars=wls20$cov.pars))
image(krig20, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y", main = "Krigagem Simples")
# O Branco é mais forte
mean(krig20$krige.var)
#####
# Análise Simulação 8 - KLU - perturbação
dist.max = .8*max(dist(p20$coords))
v20u = variog(p20,max.dist=dist.max,
trend="1st")
plot(v20u,main="Variograma - 20 pontos - U",
xlab="Distâncias",ylab="Semivariâncias")
wls20u = variofit(v20u,cov.model="sph",fix.nug=F, nug=0.1)
lines(wls20u)
grid20u <- expand.grid(seq(0,1,l=100), seq(0,1,l=100))
krig20u <- krige.conv(p20,loc=grid20u,
krige=krige.control(type.krige = "OK",
trend.d = "1st", trend.l = "1st",
cov.model="sph", cov.pars=wls20u$cov.pars))
image(krig20u, col=terrain.colors(10), xlab="Coordenada X",
ylab="Coordenada Y")
# O Branco é mais forte
mean(krig20u$krige.var)
var(p20$coords[,1])
var(p20$coords[,2])
(var.result=mean(krig20u$krige.var)+var(p20$coords[,1])+
var(p20$coords[,2]))

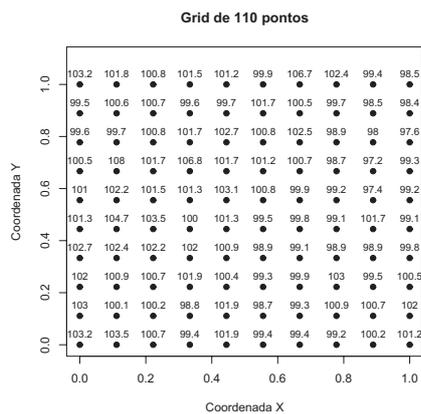
```



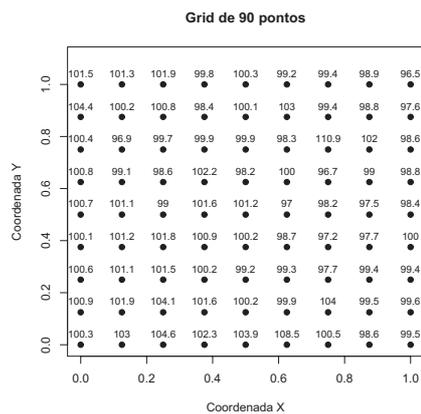
(a) P. Desconhecido com tendência



(b) P. Desconhecido com tendência



(c) P. Desconhecido com tendência



(d) P. Desconhecido com tendência

Figura 35 *Grid's* regulares de dados simulados, considerando um Processo Gaussiano com perturbação e tendência

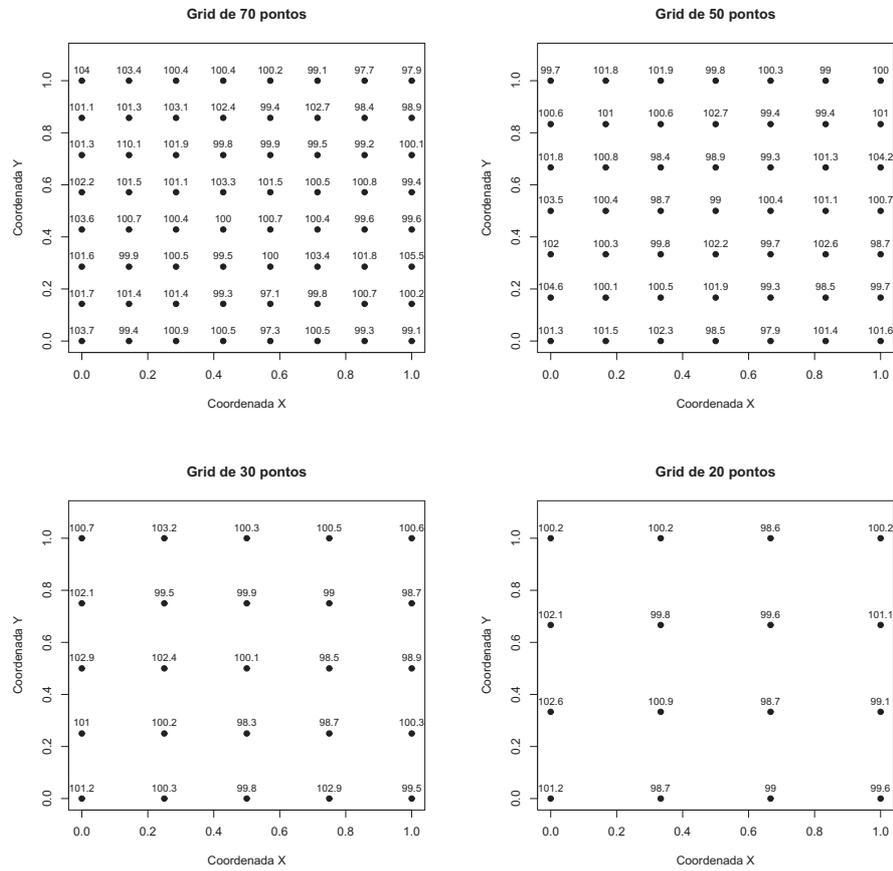
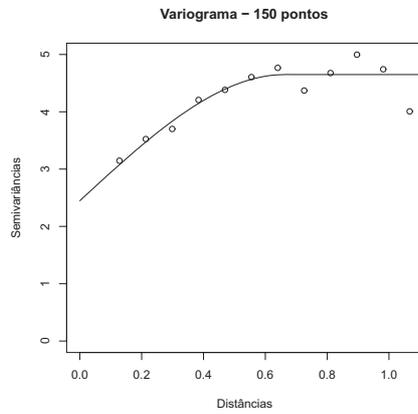
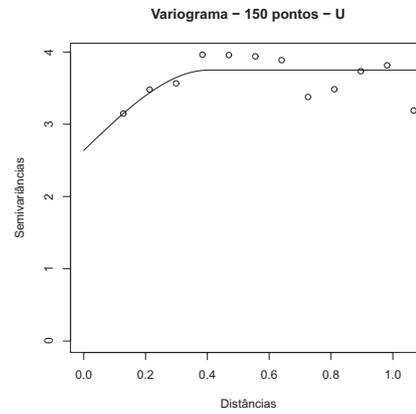


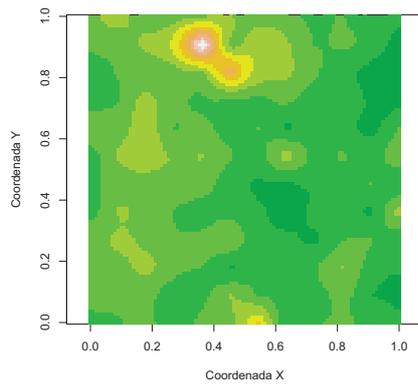
Figura 36 *Grid's* regulares de dados simulados, considerando um Processo Gaussiano com perturbação e tendência



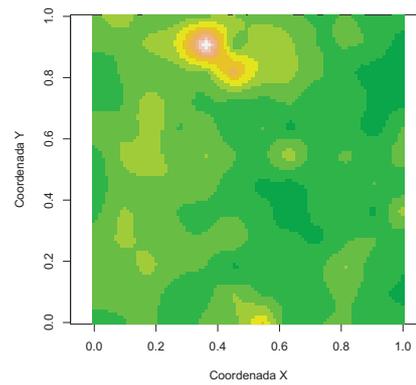
(a) Variograma para KLS



(b) Variograma para KLU

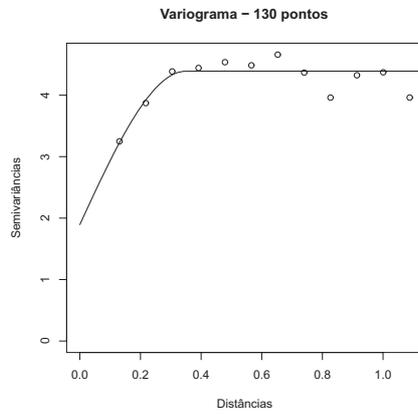


(c) KLS de 150 pontos

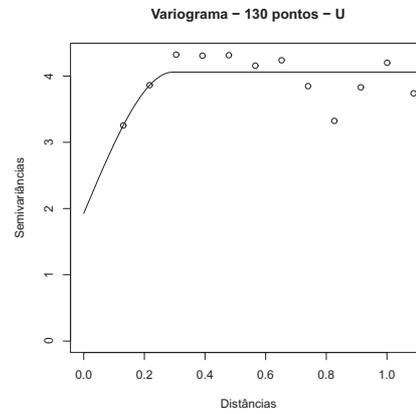


(d) KLU de 150 pontos

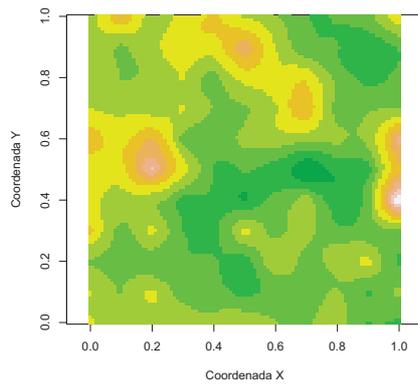
Figura 37 Análise geoestatística de um *grid* regular de 150 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano



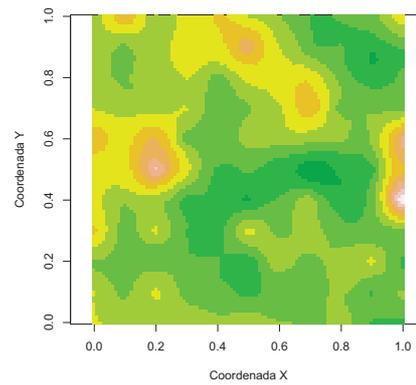
(a) Variograma para KLS



(b) Variograma para KLU

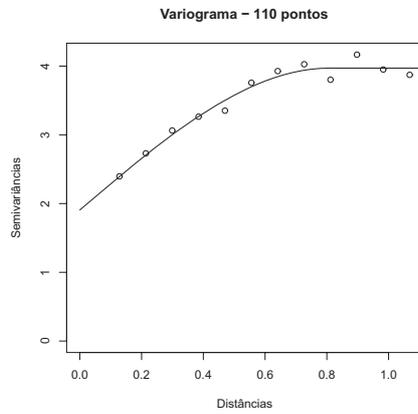


(c) KLS de 130 pontos

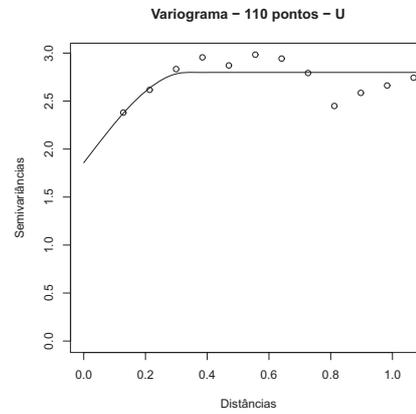


(d) KLU de 130 pontos

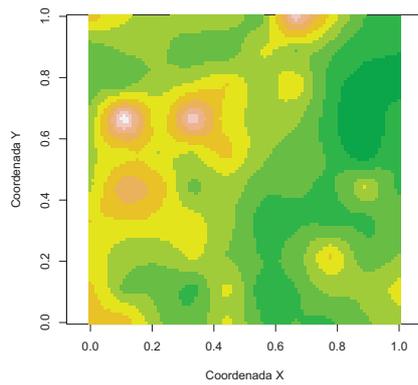
Figura 38 Análise geoestatística de um *grid* regular de 130 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano



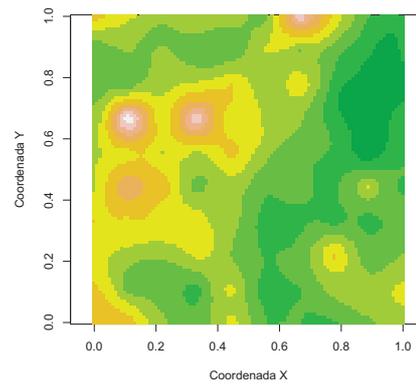
(a) Variograma para KLS



(b) Variograma para KLU

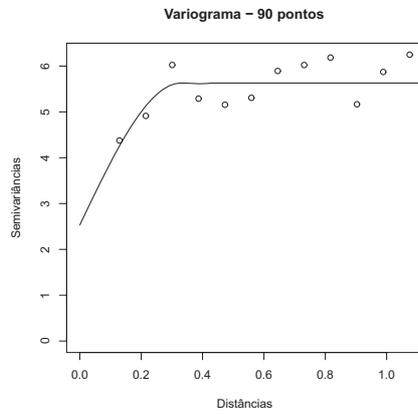


(c) KLS de 110 pontos

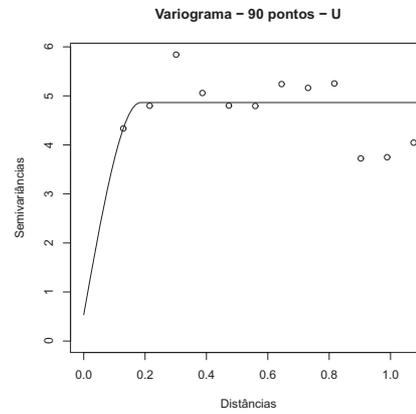


(d) KLU de 110 pontos

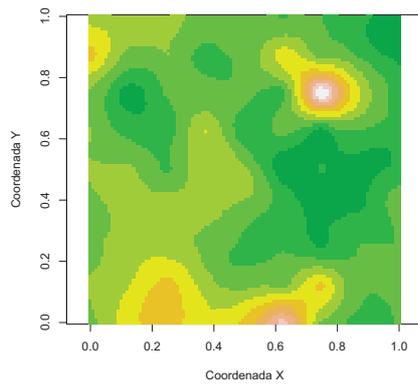
Figura 39 Análise geoestatística de um *grid* regular de 110 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano



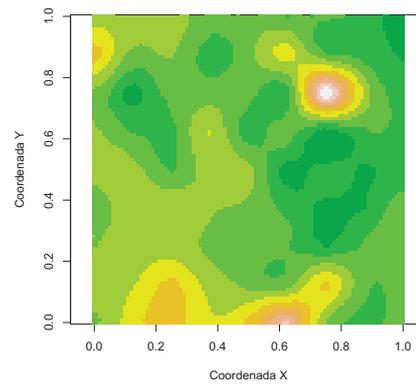
(a) Variograma para KLS



(b) Variograma para KLU

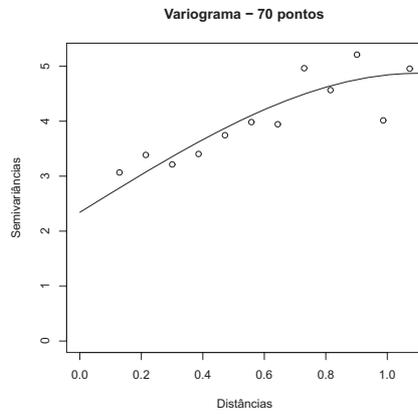


(c) KLS de 90 pontos

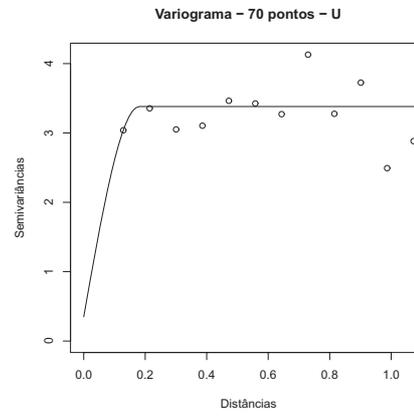


(d) KLU de 90 pontos

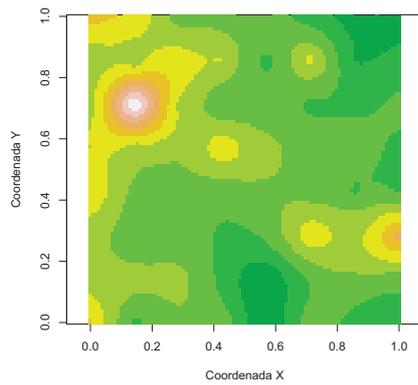
Figura 40 Análise geoestatística de um *grid* regular de 90 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano



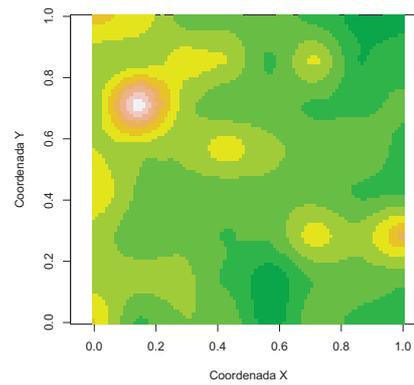
(a) Variograma para KLS



(b) Variograma para KLU

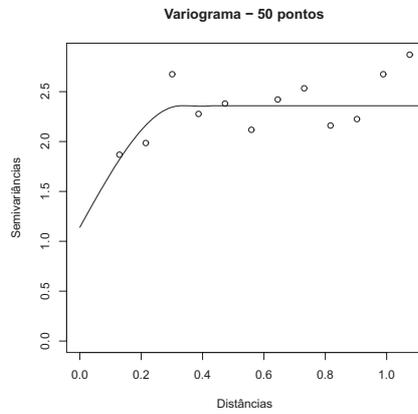


(c) KLS de 70 pontos

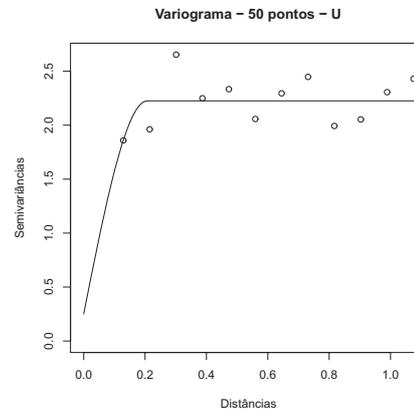


(d) KLU de 70 pontos

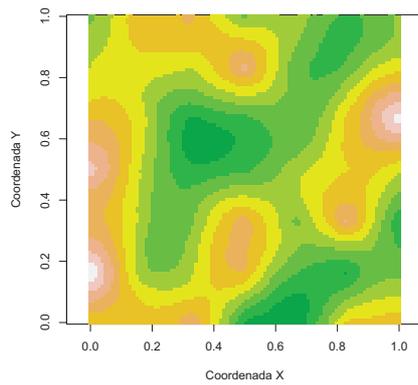
Figura 41 Análise geoestatística de um *grid* regular de 70 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano



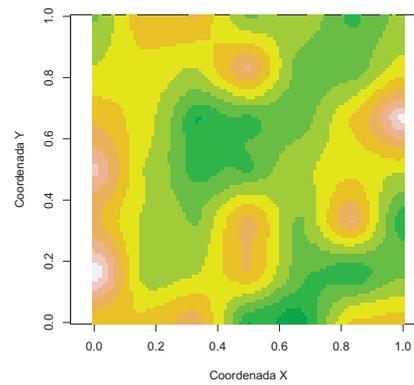
(a) Variograma para KLS



(b) Variograma para KLU

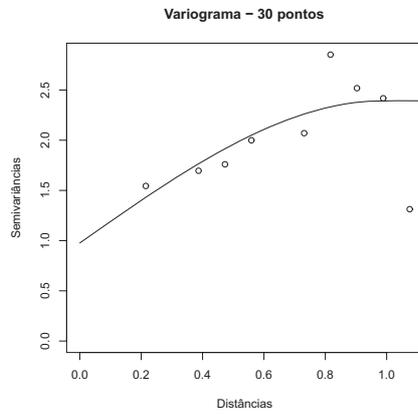


(c) KLS de 50 pontos

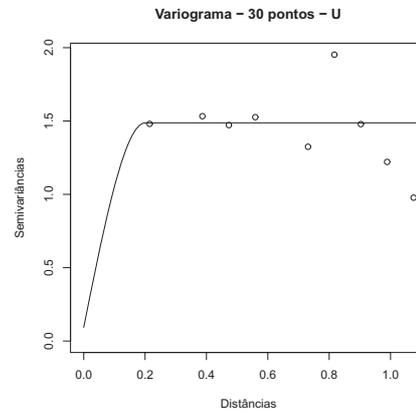


(d) KLU de 50 pontos

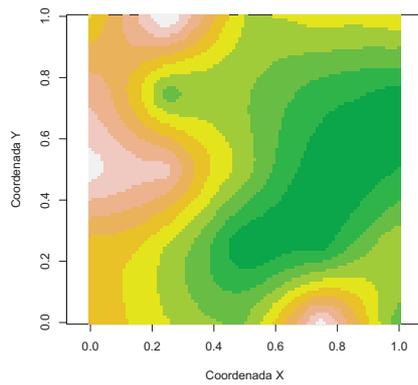
Figura 42 Análise geoestatística de um *grid* regular de 50 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano



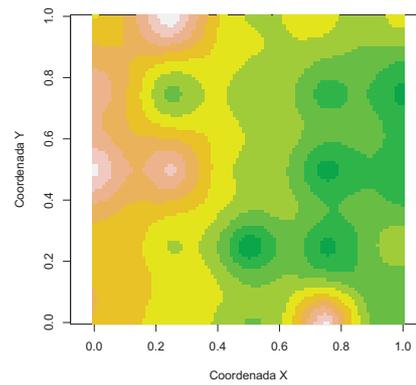
(a) Variograma para KLS



(b) Variograma para KLU

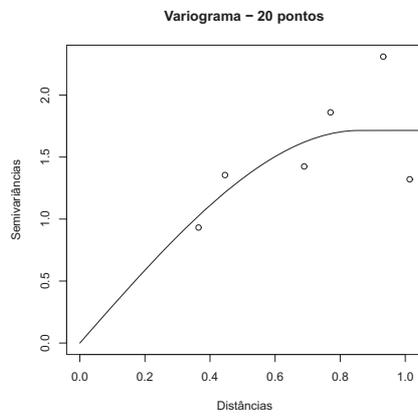


(c) KLS de 30 pontos

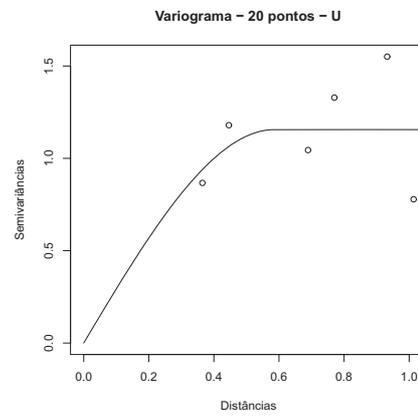


(d) KLU de 30 pontos

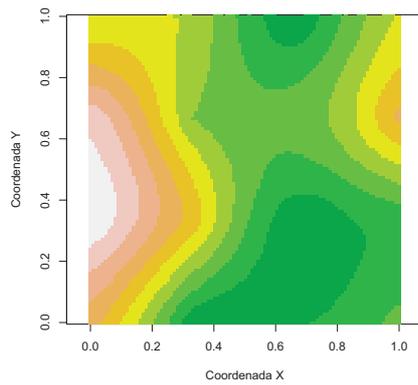
Figura 43 Análise geoestatística de um *grid* regular de 30 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano



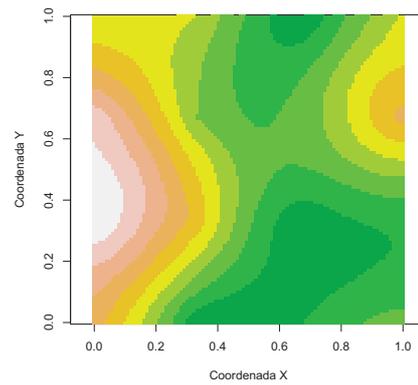
(a) Variograma para KLS



(b) Variograma para KLU



(c) KLS de 20 pontos



(d) KLU de 20 pontos

Figura 44 Análise geoestatística de um *grid* regular de 20 pontos simulados, considerando o modelo esférico de semivariograma e perturbação e tendência em um Processo Gaussiano