Carlos Alexandre Santana Oliveira

Estimativa multivariada alicerçada em tabelas de covariância

Porto Alegre

2020

Estimativa multivariada alicerçada em tabelas de covariância

Esta tese foi analisada e julgada adequada para a obtenção do título de Doutor em Engenharia, área de concentração de Tecnologia Mineral, e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais da Universidade Federal do Rio Grande do Sul.

Universidade Federal do Rio Grande do Sul - UFRGS

Escola de Engenharia

Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e Materiais

Orientador: Prof. Dr. João Felipe Coimbra Leite Costa Coorientador: Prof. Dr. Marcel Antônio Arcari Bassani

> Porto Alegre 2020

Carlos Alexandre Santana Oliveira

Estimativa multivariada alicerçada em tabelas de covariância/ Carlos Alexandre Santana Oliveira. – Porto Alegre, 2020-

 $100~\mathrm{p.:}$ il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. João Felipe Coimbra Leite Costa

Tese (Doutorado) – Universidade Federal do Rio Grande do Sul - UFRGS Escola de Engenharia

Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e Materiais, 2020.

1. Estimativa multivariada. 2. Transformada de Fourier. 3. Tabelas de covariância. 4. Correção da tabela de covariância. I. Prof. Dr. João Felipe Coimbra Leite Costa. II. Universidade Federal do Rio Grande do Sul. III. Faculdade de Engenharia. IV. Estimativa multivariada alicerçada em tabelas de covariância.

CDU 02:141:005.7

Estimativa multivariada alicerçada em tabelas de covariância

Esta tese foi analisada e julgada adequada para a obtenção do título de Doutor em Engenharia, área de concentração de Tecnologia Mineral, e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais da Universidade Federal do Rio Grande do Sul.

Trabalho aprovado. Porto Alegre, 04 de março de 2020:

Prof^a.Dr^a. Vanessa Cerqueira Koppe PPGE3M - UFRGS

> Dr. Áttila Leães Rodrigues IF - UFRGS

Dr. Luiz Eduardo Seabra Varella PETROBRAS

> Porto Alegre 2020

Agradecimentos

A Deus, por ser a fonte de todo conhecimento.

Ao Professor Dr. João Felipe, por me acolher no LPM, por acreditar em minha capacidade de aprender, pela orientação, paciência e incentivo.

Ao Professor Dr. Marcel Bassani, pela efetiva atenção e colaboração no desenvolvimento deste trabalho e ainda, pela presteza nos auxílios na programação em *python*.

Aos colegas do LPM, Aline Aurich, Cristina Araújo, Áttila Rodrigues, Augusto Torres, David Drumond, Jonas Kloeckner e Péricles Machado, pelos inúmeros auxílios diante das dificuldades.

Ao Instituto Federal do Amapá-IFAP, pelo investimento realizado.

Aos meus pais Antônio Oliveira e Ester Oliveira (*in memoriam*), pelo amor e incentivos.

A minha família, especialmente minha esposa Erica Oliveira e meu filho Samuel Oliveira, pelas alegrias proporcionadas e por serem presentes em minha vida, me incentivando e acreditando na minha capacidade.

"Planeje seu progresso, cuidadosamente, cada hora, cada dia, cada mês. A ação organizada, unida ao entusiasmo, produz uma força irresistível". P. Meyer

Resumo

Projetos de mineração geralmente contêm dados secundários espacialmente correlacionados com a principal variável de interesse (primária). Esses dados secundários são geralmente mais densamente amostrados do que os primários (conjunto de dados heterotópicos), pois são mais baratos e rápidos de serem obtidos. Nessa situação, o uso de dados secundários em modelagem geoestatística melhora a qualidade dos modelos (estimados ou simulados) finais. A principal metodologia geoestatística utilizada para integrar esses dois tipos de dados é a cokrigagem, que requer a modelagem conjunta de variogramas diretos e cruzados usando o Modelo Linear de Corregionalização (Linear Model of Coregionalization - LMC). O problema é que modelar o LMC se torna impraticável à medida que o número de variáveis aumenta. Além disso, o LMC é bastante restritivo, pois todos os modelos de variogramas devem ser combinações lineares das mesmas estruturas básicas. Essa restrição faz com que o LMC não mostre um bom ajuste à maioria dos variogramas experimentais. Esta tese mostra uma metodologia para estimativa e simulação com dados secundários heterotópicos que não requer a modelagem do LMC. A continuidade espacial será descrita por tabelas de covariância (diretas e cruzadas), que são obtidas diretamente dos dados. As tabelas de covariância são corrigidas para garantir que as matrizes de covariância usadas na krigagem sejam positivas definidas. Dois estudos de caso são apresentados, o primeiro para comparar a metodologia proposta com as estimativas/simulações feitas usando o LMC e o segundo, aborda alternativas para se obter a continuidade espacial da variável primária e realizar estimativas. Os resultados foram satisfatórios, pois os modelos estimados e simulados com tabelas de covariância foram devidamente validados.

Palavras-chaves: estimativa multivariada, transformada de Fourier, tabelas de covariância, correção da tabela de covariância.

Abstract

Mining projects often contain secondary data spatially correlated with the main variable of interest (primary). These secondary data are usually more densely sampled than the primary (heterotopic dataset), as they are cheaper and faster to obtain. In this situation, the use of secondary data in geostatistical modeling improves the quality of the final (estimated or simulated) models. The main geostatistical methodology used to integrate these two types of data is cokriging, which requires the joint modeling of direct and cross variograms using the Linear Model of Coregionalization (LMC). The problem is that modeling the LMC becomes impractical as the number of variables increases. In addition, the LMC is very restrictive as all variogram models must be linear combinations of the same basic structures. This restriction causes the LMC not to show a good fit for most experimental variograms. This thesis shows a methodology for estimation and simulation with heterotopic secondary data that does not require LMC modeling. The spatial continuity will be described by covariance tables (direct and cross), which are obtained directly from the data. Covariance tables are corrected to ensure that the covariance matrices used in kriging are positively defined. Two case studies are presented, the first to compare the proposed methodology with the estimates/simulations made using the LMC and the second, approaches alternatives to obtain the spatial continuity of the primary variable and to make estimates. The results were satisfactory, as the estimated and simulated models with covariance tables were properly validated.

Key-words: multivariate estimation, Fourier transform, covariance tables, correction of the covariance table.

Lista de ilustrações

Figura 1.1 –	- Principais direções de anisotropia determinadas por vetores em <i>grid</i> regular: $N0^{\circ}E(0, 1, 0), N22, 5^{\circ}E(2, 5, 0), N45^{\circ}E(1, 1, 0), N67, 5^{\circ}E(5, 2, 0), N90^{\circ}E(1, 0, 0), N112, 5^{\circ}E(5, -2, 0), N135^{\circ}E(1, -1, 0) \in N157, 5^{\circ}E(2, -5, 0)$). 31
Figura 2.1 –	- Comparativo entre a rapidez da FFT e a abordagem espacial. Adaptado de Marcotte (1996).	44
Figura 3.1 –	- Interface do <i>plug-in</i> do software <i>AR2GeMS</i> para geração das tabelas de covariância (diretas e cruzadas). Figura adaptada de Kloeckner (2018).	57
Figura 3.2 –	- Fluxo de trabalho da metodologia. Imagem ilustrativa	59
Figura 4.1 –	- Mapa de localização das amostras: (a) variável primária e (b) variável secundária.	61
Figura 4.2 –	- Histograma da variável primária (a) e histograma da variável secundária (b)	62
Figura 4.3 –	- Gráfico de dispersão entre a variável primária (ppm) e variável secundá- ria (ppm).	62
Figura 4.4 –	- Modelos para extrair as tabelas de covariância: (a) BMEC da variável primária e (b) BMEC da variável secundária.	63
Figura 4.5 –	- Tabelas de covariância extraídas dos BMECs: (a) Tabela de covariância da variável primária, (b) Tabela de covariância da variável secundária e (c) Tabela de covariância média entre as tabelas de covariância cruzadas.	64
Figura 4.6 –	- Parte da tabela de covariâncias que deve ser corrigida: (a) parte central da tabela de covariância da variável primária, (b) parte central da tabela de covariância da variável secundária e (c) parte central da tabela de	64
Figura 4.7 –	- Variogramas diretos nas principais direções da variável primária. Direção de maior continuidade: $N157^{\circ}E$ (a); direção de menor continuidade: $N67^{\circ}E$ (b).	66
Figura 4.8 –	-Variogramas diretos nas principais direções da variável secundária. Dire- ção de maior continuidade: $N157^{\circ}E$ (a); direção de menor continuidade: $N67^{\circ}E$ (b).	66
Figura 4.9 –	- Variogramas cruzados nas principais direções. Direção de maior conti- nuidade: $N157^{\circ}E$ (a); direção de menor continuidade: $N67^{\circ}E$ (b)	67
Figura 4.10	-Gráfico de dispersão entre as estimativas através de variogramas ajusta- dos (SOCK_CL) e através de tabelas de covariância corrigidas (SOCK	-
	_CT(BMEC) _COR)	69

Figura 4.11–Comparação visual entre os m	odelos de estimativa: (a) estimativa através
de variogramas ajustados (SOCK_CL) e (b) através de tabelas de
covariância corrigidas (SOCI	$K_CT(BMEC) _COR).$ 70
Figura 4.12–Histogramas associados ao n	nodelo de referência da variável primária
(a), a estimativa via variogra	amas ajustados (b) e a estimativa através
de tabelas de covariância con	$\operatorname{rrigidas}(\mathbf{c})71$
Figura 4.13–Análise de deriva entre o mo	delo de referência e estimados ao longo da
direção leste (X) : modelo de r	eferência (NN), estimativa via variogramas
ajustados (SOCK_CL) e e	stimativa usando tabelas de covariância
$corrigidas$ (SOCK_CT(BME	CC)_COR)
Figura 4.14–Análise de deriva entre o mo	delo de referência e estimados ao longo da
direção norte (Y): modelo de s	referência (NN), estimativa via variogramas
ajustados (SOCK_CL) e e	stimativa usando tabelas de covariância
$\operatorname{corrigidas}$ (SOCK_CT(BME	$CC)_COR$)
Figura 4.15–Variogramas diretos dos dad	os transformados (normal score) nas prin-
cipais direções da variável _l	primária. Direção de maior continuidade:
$N157^{\rm o}E$ (a); direção de men	or continuidade: $N67^{\circ}E$ (b) 73
Figura 4.16–Variogramas diretos dos dad	os transformados (normal score) nas prin-
cipais direções da variável se	ecundária. Direção de maior continuidade:
$N157^{\rm o}E$ (a); direção de men	or continuidade: $N67^{\circ}E$ (b) 74
Figura 4.17–Variogramas cruzados dos	dados transformados (normal score) nas
principais direções. Direção d	e maior continuidade: $N157^{\circ}E$ (a); direção
de menor continuidade: $N67$	$^{\circ}E$ (b)
Figura 4.18–Simulações: realizações via v	rariograma (a) e tabelas de covariância (b). 75
Figura 4.19–Gráfico de dispersão entre u	ma realização via tabela de covariância e
uma realização via variogram	na ajustado
Figura 4.20–Histograma acumulado dos	modelos simulados (Simulações) e histo-
grama acumulado do BMEC	da variável primária (NN) 76
Figura 4.21–Variogramas das simulações	via tabelas de covariância orientados ao
longo das direções de maior	(a) e menor (b) continuidade espacial e
variograma.	
Figura 4.22–Análise de deriva entre o mod	lelo de referência (NN) e simulado ao longo
da direção leste (X): simula	ção usando modelo de variograma (a) e
simulação usando tabelas de	covariância corrigidas (b)
Figura 4.23–Análise de deriva entre o mod	lelo de referência (NN) e simulado ao longo
da direção norte (Y): simula	ação usando modelo de variograma (a) e
simulação usando tabelas de	covariância corrigidas (b)
Figura 5.1 – Mapa de localização das ame	ostras: (a) variável primária e (b) variável
secundária exaustiva	

Figura 5.2 –	- Histograma da variável primária (a) e histograma da variável secundária	
	(b)	80
Figura 5.3 –	- Espaçamento dos dados da variável primária (a) e histograma associado	
	(b)	80
Figura 5.4 –	- Gráfico de dispersão entre os valores colocados da variável primária e	Q1
Figure 5.5	Verierrenzes emperimentois de veriével primérie pas direcées Omni	01
r igura 5.5 -	$N0^{\circ}E, N22, 5^{\circ}E, N45^{\circ}E, N67, 5^{\circ}E, N90^{\circ}E, N112, 5^{\circ}E, N135^{\circ}E \in N157, 5^{\circ}E$.	83
Figura 5.6 -	- BMEC da variável primária. Dimensões em km	84
Figura 5.7 -	- Gráfico de dispersão entre os valores reais da variável primária e os	
	valores colocados da variável secundária (esquerda) e gráfico de dispersão	
	entre os valores nos nós gerados com o BMEC da primária e a secundária	
	exaustiva (direita).	84
Figura 5.8 -	- Mapa de covariância extraído da variável secundária exaustiva (a) e sua	
	parte central (b). Dimensões em km. \ldots	85
Figura 5.9 -	- Variogramas obtidos através da tabela de covariância da variável secun-	
	dária (dados estandar dizados). Direção de maior continuidade: $N135^{\rm o}E$	
	(a); direção de menor continuidade: $N45^{\circ}E$ (b)	86
Figura 5.10	-Variogramas obtidos através da tabela de covariância da variável secun-	
	dária (dados transformados). Direção de maior continuidade: $N135^{\rm o}E$	
	(a); direção de menor continuidade: $N45^{\circ}E$	86
Figura 5.11	-Mapa de localização das estimativas: (a) cokrigagem colocada via tabela	
	de covariância corrigida e (\mathbf{b}) cokrigagem colocada através de variograma	
	ajustado .Dimensões em km	87
Figura 5.12	-Gráfico de dispersão entre as estimativas via variograma ajustado e via	
	tabela de covariância corrigida.	87
Figura 5.13	–Histogramas associados às estimativas por cokrigagem colocada via	
	tabela de covariância corrigida (esquerda) e via variograma (direita).	88
Figura 5.14	-Análise de deriva entre o modelo de referência e estimados ao longo da	
	direção leste (X): modelo de referência (NN) e estimado por cokrigagem	
	colocada via variograma e via tabela de covariância corrigida.	88
Figura 5.15	-Análise de deriva entre o modelo de referência e estimados ao longo da	
	direçao norte (Y): modelo de referência (NN) e estimado por cokrigagem	0.0
T . F < F	colocada via variograma e via tabela de covariância corrigida.	89
Figura 5.16	-Realizações da simulação sequencial Gaussiana colocada da variável	
	primaria e mapa de localização das amostras da variável primária:	
	(a) realização via tabela de covariância corrigida e (b) realização via	0.0
	variograma. Dimensoes em km.	-90

Figura 5.17–Histograma acumulado dos modelos simulados (simulações) e histograma	
acumulado do modelo de referência da variável primária (NN).	90
Figura 5.18–Variogramas das simulações via tabela de covariância orientados ao	
longo das direções de maior (a) e menor (b) continuidade espacial e	
variograma modelado a partir dos dados secundários	91

Lista de tabelas

Tabela 4.1 – Estatística básica dos dados	62
Tabela 4.2 – Variogramas gerados via tabelas de covariância na principal direção. $\ .$	68
Tabela 5.1 – Estatística descritiva dos dados	81
Tabela 5.2 – Estatística descritiva das estimativas por cokrigagem colocada	88

Lista de abreviaturas e siglas

BMEC	Modelo Base para Extrair a Covariância (Base Model to Extract the Covariance)
ccdf	Função de Distribuição Cumulativa Condicional (Conditional Cumula- tive Distribution Function)
\mathcal{CM}	Matriz de Covariância (Covariance Matrix)
COCK	Cokrigagem Colocada Ordinária (Colocated Ordinary Cokriging)
DFT	Transformada Discreta de Fourier (Discrete Fourier Transform)
\mathbf{FFT}	Transformada Rápida de Fourier (Fast Fourier Transform)
LMC	Modelo Linear de Corregionalização (Linear Model of Coregionalization)
MAF	Mínimo-Máximo Fatores de Autocorrelação (Minimum/Maximum Au- tocorrelation Factors)
MM1	Modelo de Markov I
MM2	Modelo de Markov II
NN	Vizinho mais Próximo (Nearest Neighbor)
OCK	Cokrigagem Ordinária (Ordinary Cokriging)
PCA	Análise de Componentes Principais (Principal Component Analysis)
PPMT	Projection Pursuit Multivariate Transform
SCT	Stepwise Conditional Transform
SGCS	Cossimulação Sequencial Gaussiana (Sequential Gaussian Cosimulation)
SGS	Simulação Sequencial Gaussiana (Sequential Gaussian Simulation)
SOCK	Cokrigagem Ordinária Estandardizada (Standardized Ordinary Cokriging)

Sumário

1	INTRODUÇÃO	23
1.1	O problema	26
1.1.1	Modelagem conjunta da continuidade espacial via LMC	26
1.1.2	Inferência da continuidade espacial de uma variável primária na presença de	
	uma variável secundária exaustiva	30
1.2	Meta e objetivos	32
1.3	Organização da tese	32
2	REVISÃO BIBLIOGRÁFICA E FUNDAMENTAÇÃO TEÓRICA	35
2.1	Noções elementares de álgebra linear	35
2.1.1	Autovalores e autovetores	35
2.1.2	Diagonalização de matrizes reais simétricas e decomposição espectral	36
2.2	Análise de Fourier	37
2.2.1	Série de Fourier	38
2.2.2	A integral de Fourier	39
2.2.3	A transformada de Fourier	39
2.2.3.1	Propriedades da transformada de Fourier	40
2.2.4	A transformada discreta de Fourier	41
2.2.5	A transformada rápida de Fourier	42
2.3	Análise espectral	42
2.4	Modelagem da continuidade espacial via transformada de Fourier	43
2.4.1	Computação rápida de variogramas via FFT	43
2.4.2	Modelagem automática de tabelas de covariância (cruzadas) utilizando a FFT	44
2.4.3	Tabela de covariância: um mapeamento rápido e automático da continuidade	
	espacial	46
2.5	Estimativa multivariada	47
2.5.1	Cokrigagem ordinária e cokrigagem ordinária estandardizada	48
2.5.2	Cokrigagem colocada ordinária	50
2.5.3	Modelos de Markov	51
2.6	Simulação sequencial gaussiana em conjuntos de dados multivaria-	
	dos heterotópicos	52
2.7	Considerações finais	53
3	METODOLOGIA	55
3.1	Modelo base para extrair a tabela de covariânica	56
3.2	Efeito pepita	56

3.3	Obtenção das tabelas de covariância	56
3.4	Imputação do efeito pepita à tabela de covariância	57
3.5	Correção conjunta das tabelas de covariância	57
3.6	Fluxo de trabalho da metodologia	59
4	USO DE DADOS SECUNDÁRIOS HETEROTÓPICOS EM	
	ESTIMATIVA/SIMULAÇÃO COM TABELAS DE COVARIÂNCIA .	61
4.1	Apresentação do banco de dados	61
4.2	Estimativa	62
4.2.1	Resultados e discussões	69
4.3	Simulação com dados secundários	72
4.3.1	Resultados e discussões	74
4.4	Conclusão	77
5	APLICAÇÃO DA TABELA DE COVARIÂNCIA PARA INFERIR A	
	CONTINUIDADE ESPACIAL DA PRIMÁRIA NO CASO DE UMA	
	VARIÁVEL SECUNDÁRIA EXAUSTIVA	79
5.1	Análise e apresentação do banco de dados	79
5.2	Análise da continuidade espacial da variável primária	81
5.3	Análise da continuidade espacial da variável secundária	85
5.4	Análise e discussão dos resultados da estimativa com dados colocados	86
5.5	Análise e discussão dos resultados da simulação com dados colocados	89
5.6	Conclusão	91
6	CONCLUSÕES, CONTRIBUIÇÕES E RECOMENDAÇÕES	93
6.1	Conclusões da tese	93
6.2	Contribuições da tese	95
6.3	Recomendações para trabalhos futuros	96
	REFERÊNCIAS	97

1 Introdução

Em geoestatística, as estimativas/simulações precisam de funções que descrevam a dependência espacial entre as amostras de uma ou mais variáveis disponíveis em um banco de dados. A dependência espacial considera que o valor de uma variável em cada ponto no espaço se relaciona com valores obtidos em pontos situados a uma dada distância e que a influência entre pontos é tanto maior quanto menor for a distância que os separa. Essa relação entre amostras que estão separadas a uma dada distância é chamada de continuidade ou correlação espacial. A geoestatística faz uso da correlação espacial para fazer as estimativas e simulações.

Para calcular o modelo de correlação espacial e estimar valores de uma variável distribuídos no espaço, a geoestatística fornece um conjunto de técnicas. A correlação espacial é geralmente verificada através da função variograma (ou semivariograma) e a estimativa, através da krigagem (MATHERON, 1963). Usa-se o variograma para calcular a variância entre pontos separados pela distância h e, a partir dele, se realiza a estimativa.

A estimativa por krigagem leva em consideração apenas um atributo a ser analisado e, portanto, necessita de apenas um modelo de variograma. Mas existem situações em que o objeto de estudo apresenta uma ou mais variáveis secundárias correlacionadas ao atributo de interesse. Nesse caso, o uso dessas variáveis secundárias nas estimativas melhora a qualidade dos resultados. A técnica mais empregada para usar amostras secundárias em geoestatística, nesse caso, é a cokrigagem (MARECHAL, 1970). Para que a cokrigagem seja implementada em um estudo com n variáveis, é necessária a modelagem em conjunto de $\frac{n(n+1)}{2}$ variogramas diretos e cruzados. Além disso, a modelagem desses variogramas deve ser feita de modo que as matrizes de cokrigagem sejam positivas definidas¹. A maneira mais utilizada para modelar esses variogramas diretos e cruzados é usando o Modelo Linear de Corregionalização (*Linear Model of Coregionalization* - LMC) (GOOVAERTS, 1997).

O LMC modela todos os variogramas diretos e cruzados como combinações lineares das mesmas estruturas de variograma básicas. Sabe-se que essas estruturas básicas levam a matrizes positivas definidas. As matrizes de covariância obtidas via LMC são positivas

- 2. Todos os autovalores de C são positivos.
- 3. Os determinantes das submatrizes de ${\cal C}$ são positivos.
- 4. Todos os pivôs (sem trocas de linhas) são positivos.

¹ Cada um dos testes a seguir é uma condição necessária e suficiente para que uma matriz simétrica real C seja positiva definida (ISAAKS; SRIVASTAVA, 1989):

^{1.} $w^T C w > 0$ para todos os vetores não nulos w. No sistema de krigagem, C é a matriz de covariância entre os dados e w é qualquer vetor de pesos associado à estimativa.

definidas, garantindo que o sistema de cokrigagem tenha solução única.

Existem métodos para ajustar o LMC automaticamente (LARRONDO; NEUFELD; DEUTSCH, 2003; EMERY, 2010; DESASSIS; RENARD, 2013). No entanto, esses métodos dependem do cálculo de variogramas experimentais, o que não é automático. O cálculo dos variogramas experimentais requer que o geomodelador escolha parâmetros apropriados (tamanho do *lag*, azimutes de direção e tolerâncias). A escolha adequada desses parâmetros requer uma quantidade considerável de tempo e esforço quando o conjunto de dados não segue um espaçamento regular (KAPAGERIDIS, 2015).

Para evitar o LMC, vários trabalhos usando métodos de fatorização foram publicados. A fatorização envolve a transformação das variáveis correlacionadas originais em fatores não correlacionados. Então os fatores não correlacionados são estimados/simulados independentemente e transformados de volta. As técnicas de fatorização usadas na geoestatística incluem a Análise de Componentes Principais (Principal Component Analysis-PCA) (HOTELLING, 1933; GOOVAERTS, 1993), Mínimo-Máximo Fatores de Autocorrelação (Minimum/Maximum Autocorrelation Factors - MAF) (DESBARATS; DIMITRAKOPOU-LOS, 2000; SWITZER; GREEN, 1984), Stepwise Conditional Transform (SCT) (LEU-ANGTHONG; DEUTSCH, 2003) e Projection Pursuit Multivariate Transform (PPMT) (BARNETT; MANCHUK; DEUTSCH, 2014). Esses fluxos de trabalho da fatorização são adequados quando várias variáveis primárias precisam ser estimadas/simuladas e o banco de dados for isotópico, ou seja, todas as variáveis forem amostradas conjuntamente nos mesmos locais. Aqui, estamos interessados em estimar/simular somente uma variável primária usando informações auxiliares de dados secundários mais densamente amostrados (banco de dados heterotópico). Nesse caso, o LMC é necessário para modelar a correlação espacial entre os dados primários e secundários.

Como alternativa à modelagem do LMC, Deutsch e Journel (1992) e Chu (1993) propuseram construir um mapa discreto de covariância a partir dos dados originais. Esse mapa de covariância é chamado de tabela de covariância e contém os valores de covariância para as distâncias e direções definidas por um *grid*, que é posteriormente usado para a estimativa/simulação geoestatística. Marcotte (1996) também apresentou um método para calcular mapas variográficos (de variograma e/ou covariância) discretos (diretos e cruzados) com uso da transformada rápida de Fourier (*Fast Fourier Transform*- FFT). A principal limitação desses mapas de variograma é que as matrizes de covariância obtidas a partir deles não são necessariamente positivas definidas. Dessa forma, essas matrizes não podem ser aplicadas diretamente nos algoritmos geoestatísticos de estimativa e simulação.

Nesse contexto, Yao e Journel (1998) apresentaram um método para obter uma tabela de covariância que gera matrizes de covariância positivas definidas. Os autores primeiro calculam a covariância experimental das amostras ao longo das direções e separações, que correspondem às direções e separações dos dados. Segundo, esses valores de covariância amostral são interpolados em um *grid* para se obter uma tabela de covariância preliminar. Por fim, essa tabela de covariância é corrigida usando o teorema de Bochner (BOCHNER, 1939), garantindo que a tabela de covariância corrigida seja positiva definida.

A correção usada por Yao e Journel (1998) consiste em aplicar primeiro a transformada rápida de Fourier à tabela de covariância e obter o mapa de densidade espectral. Em seguida, suavizar a tabela de densidade espectral (usando média móvel), restringindo todos os valores a números reais e não negativos, de tal forma que os valores estejam estandardizados e com variância unitária. Yao e Journel (1998) advertem que uma estandardização ingênua (isto é, dividindo todos os coeficientes por sua soma) pode resultar em um efeito pepita muito alto quando muitos valores negativos são redefinidos para zero. Como a correlação espacial a curta distância desempenha o papel mais importante na estimativa/simulação, esse resultado é indesejável. Nesse contexto, Yao e Journel (1998) utilizam pesos na estandardização para que os valores de baixa frequência sejam os menos afetados. O resultado é que o efeito pepita da tabela de covariância corrigida é pouco afetado.

Pyrcz e Deutsch (2006) propuseram um método semelhante ao de Yao e Journel (1998) para obter uma tabela de covariância positiva definida. Primeiro, os autores calculam e modelam a covariância das amostras em um conjunto de direções representativas, com qualquer função (um *spline*, por exemplo). Segundo, essas covariâncias são usadas para construir uma tabela de covariância preliminar. Essa tabela de covariância é obtida pela interpolação dos valores de covariância obtidos para as diferentes direções modeladas. Terceiro, esta tabela de covariância é corrigida aplicando-se a transformada rápida de Fourier, redefinindo os valores negativos para zero e estandardizando para que sua soma seja igual a um (YAO; JOURNEL, 1998).

Na mesma linha que Yao e Journel (1998), Kloeckner et al. (2019) propuseram obter tabelas de covariância diretamente dos dados. O método consiste em (i) estimar valores a partir dos dados originais para preencher todo o *grid* e obter um modelo base para extrair a covariância (*Base Model to Extract the Covariance* - BMEC), (ii) usar a FFT e calcular a covariância experimental (MARCOTTE, 1996), equivalente ao espectro de densidade, (iii) corrigir o espectro de densidade (PYRCZ; DEUTSCH, 2006) e (iv) retro-transformar (inverso da FFT) o espectro de densidade corrigido para obter uma tabela de covariância positiva definida. O BMEC constitui um *grid* exaustivamente amostrado e, por ser gerado através da interpolação dos valores originais, carrega a continuidade espacial da variável e exerce forte influência na qualidade da tabela de covariância gerada pelo método de Kloeckner et al. (2019).

1.1 O problema

Nesta tese serão considerados dois problemas. O primeiro trata da modelagem conjunta da continuidade espacial de múltiplas variáveis correlacionadas e, o segundo aborda a inferência da continuidade espacial de variáveis primárias com poucas amostras baseadas em dados secundários exaustivos.

1.1.1 Modelagem conjunta da continuidade espacial via LMC

Quando avaliamos um depósito mineral, várias variáveis podem ser amostradas simultaneamente em um mesmo ponto ou em pontos distintos. É comum encontrarmos nos resultados da pesquisa mineral, variáveis que estão mais densamente amostradas que outras, mas que apresentam alguma correlação espacial entre si. Existindo um significativo grau de correlação, podemos estimar uma variável primária (que apresenta um número menor de amostras no grid) a partir de variáveis secundárias (que apresentam um número maior de amostras no grid) e diante disso, a geoestatística proporciona um conjunto de ferramentas que facilitam esse tipo de estimativa.

Um procedimento geoestatístico muito utilizado para estimar diversas variáveis regionalizadas, com base na correlação espacial, é a cokrigagem. Na cokrigagem, os variogramas diretos e cruzados são modelados conjuntamente, pois precisam satisfazer o LMC. O ajuste do LMC é geralmente iterativo. O usuário escolhe um modelo inicial de LMC e o ajusta até que o modelo de variograma se adapte bem aos variogramas experimentais. Por exemplo, considere duas variáveis, $U \in V$. Os variogramas diretos e cruzados de $U \in V$ devem ser modelados, a partir das estruturas básicas dos modelos, do seguinte modo (GOOVAERTS, 1997):

$$\gamma_U(h) = u_0 \gamma_0(h) + u_1 \gamma_1(h) + \dots + u_m \gamma_m(h)$$

$$(1.1)$$

$$\gamma_V(h) = v_0 \gamma_0(h) + v_1 \gamma_1(h) + \dots + v_m \gamma_m(h)$$

$$(1.2)$$

$$\gamma_{UV}(h) = \omega_0 \gamma_0(h) + \omega_1 \gamma_1(h) + \dots + \omega_m \gamma_m(h)$$
(1.3)

onde $\gamma_U(h)$, $\gamma_V(h)$ e $\gamma_{UV}(h)$ são os modelos de variogramas diretos e cruzado para U e V, respectivamente; os modelos básicos de variogramas são indicados por $\gamma_0(h)$, $\gamma_1(h)$, ..., $\gamma_m(h)$; $u, v \in \omega$ são os coeficientes. Podemos escrever as equações 1.1, 1.2 e 1.3 na forma de matrizes, que representam as combinações de cada modelo básico.

• Combinações do primeiro modelo básico, $\gamma_0(h)$.

$$\begin{bmatrix} \gamma_{U,0}(h) & \gamma_{UV,0}(h) \\ \gamma_{VU,0}(h) & \gamma_{V,0}(h) \end{bmatrix} = \begin{bmatrix} u_0 & \omega_0 \\ \omega_0 & v_0 \end{bmatrix} \cdot \begin{bmatrix} \gamma_0(h) & 0 \\ 0 & \gamma_0(h) \end{bmatrix}$$
(1.4)

• Combinações do segundo modelo básico, $\gamma_1(h)$.

$$\begin{bmatrix} \gamma_{U,1}(h) & \gamma_{UV,1}(h) \\ \gamma_{VU,1}(h) & \gamma_{V,1}(h) \end{bmatrix} = \begin{bmatrix} u_1 & \omega_1 \\ \omega_1 & v_1 \end{bmatrix} \cdot \begin{bmatrix} \gamma_1(h) & 0 \\ 0 & \gamma_1(h) \end{bmatrix}$$
(1.5)

• Combinações do m-ésimo modelo básico, $\gamma_m(h)$.

$$\begin{bmatrix} \gamma_{U,m}(h) & \gamma_{UV,m}(h) \\ \gamma_{VU,m}(h) & \gamma_{V,m}(h) \end{bmatrix} = \begin{bmatrix} u_m & \omega_m \\ \omega_m & v_m \end{bmatrix} \cdot \begin{bmatrix} \gamma_m(h) & 0 \\ 0 & \gamma_m(h) \end{bmatrix}$$
(1.6)

Para garantir que o modelo linear, nas equações 1.1, 1.2 e 1.3, seja positivo definido, basta que as matrizes dos coeficientes $u, v \in \omega$ nas equações 1.4, 1.5 e 1.6 sejam positivas definidas ou seja, tenham determinantes maiores que zero (condição necessária e suficiente para que uma matriz simétrica seja positiva definida). Mas modelar as matrizes dos coeficientes para que sejam positivas definidas é subjetivo e exige muito tempo quando três ou mais variáveis são consideradas. Além disso, o LMC é bastante restritivo, pois combinações lineares das mesmas estruturas devem ser usadas. Como consequência, o LMC, em geral, não tem um bom ajuste à maioria dos variogramas experimentais diretos e cruzados. Em geral, as variáveis mais importantes são privilegiadas e tem um ajuste melhor.

Semelhante ao variograma, a covariância também descreve a continuidade espacial de uma variável ou entre várias variáveis. Para uma variável, a covariância entre os dados separados por uma distância h é calculada a partir da equação:

$$C_{Z}(h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} Z(u_{\alpha}) \cdot Z(u_{\alpha}+h) - m_{Z-h} \cdot m_{Z+h}, \qquad (1.7)$$

em que:

N(h): é o número de pares que podem ser encontrados na distância h; m_{Z-h} : é a média dos valores que formam a cauda do vetor de dados:

$$m_{Z-h} = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} Z(u_{\alpha}).$$
(1.8)

 m_{Z+h} : é a média dos valores que formam o extremo do vetor de dados:

$$m_{Z+h} = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} Z(u_{\alpha} + h).$$
(1.9)

As propriedades associadas à função covariância são as seguintes:

- (i) Para h = 0 a covariância é igual a variância máxima dos dados, ou seja $C(0) = \sigma^2$.
- (ii) A função covariância é simétrica, ou seja C(h) = C(-h).
- (iii) Nenhum valor de covariância é maior do que a covariância para h = 0, ou seja $|C(h)| \leq C(0)$.
- (iv) O variograma se relaciona com a covariância através da expressão $\gamma(h) = C(0) C(h)$.

Nos sistemas de krigagem ordinária, por exemplo, os elementos podem ser dados em função da covariância entre os dados. A forma matricial desse sistema é representada por:

$$\begin{bmatrix} C(x_1, x_1) & C(x_1, x_2) & \cdots & C(x_1, x_N) & 1 \\ C(x_2, x_1) & C(x_2, x_2) & \cdots & C(x_2, x_N) & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ C(x_N, x_1) & C(x_N, x_2) & \cdots & C(x_N, x_N) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} B & C \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_2 \\ \vdots \\ \lambda_N \\ 1 \end{bmatrix} = \begin{bmatrix} C(x_1, x_0) \\ C(x_2, x_0) \\ \vdots \\ C(x_N, x_0) \\ 1 \end{bmatrix},$$
(1.10)

em que as matrizes do lado esquerdo da igualdade são, respectivamente, a matriz de covariância entre amostras (A) e a matriz dos ponderadores (B) (μ é o multiplicador de Lagrange) e a do lado direito da igualdade, é a matriz de covariância entre amostras e o ponto a ser estimado (C). A matriz de covariância entre amostras precisa respeitar a condição de ser positiva definida. Ao respeitar a condição de ser positiva definida, não só garante que a solução do sistema seja única, mas também que a solução seja estável. A solução é estável quando os ponderadores de krigagem não são maiores que 1 e nem menores que zero (ISAAKS; SRIVASTAVA, 1989).

No caso de duas variáveis, x e y, separadas pela distância h, a covariância cruzada é dada por:

$$C_{xy}(h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} x(u_{\alpha}) \cdot y(u_{\alpha} + h) - m_{x-h} \cdot m_{y+h}, \qquad (1.11)$$

com

$$m_{x-h} = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} x(u_{\alpha})$$
(1.12)

е

$$m_{y+h} = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} y(u_{\alpha+h}), \qquad (1.13)$$

onde m_{x-h} e m_{y+h} são, respectivamente, a média dos valores que formam a cauda e o extremo do vetor formado pelo par de dados $x \in y$.

As propriedades da função covariância cruzada são:

- (i) A função covariância cruzada não é simétrica, ou seja $C_{xy}(h) \neq C_{xy}(-h)$.
- (ii) A função covariância cruzada não tem máximo para h = 0.
- (iii) O variograma cruzado se relaciona com a covariância cruzada através da expressão $\gamma_{xy} = C_{xy} - \frac{1}{2}[C_{xy}(h) + C_{xy}(-h)].$

O sistema de cokrigagem, por ser uma extensão do método de krigagem, utiliza as matrizes de covariância no processo de estimativa. Para que o modelo de estimativa por cokrigagem tenha seu uso justificável, deve-se ter uma alta correlação espacial entre as variáveis primária e secundária (o que garante que as estimativas serão consistentes) e ainda, precisa que a matriz de covariância conjunta seja positiva definida (utiliza-se o LMC na verificação do modelo).

Nesse contexto, pretende-se investigar uma alternativa ao LMC, baseada na correção conjunta das tabelas de covariância geradas pelo método de Kloeckner et al. (2019), para estimar/simular múltiplas variáveis.

Quanto ao método de Kloeckner et al. (2019) há que se ponderar o seguinte:

- (i) O método gera tabelas de covariância diretas e cruzadas, mas somente as tabelas de covariância diretas são corrigidas para serem positivas definidas;
- (ii) Como as tabelas de covariância cruzadas não são simétricas, a correção para garantir a positividade é de difícil implementação. Para se obter uma tabela de covariância cruzada positiva definida, os espectros de densidade precisam ser corrigidos simultaneamente (YAO; JOURNEL, 1998);
- (iii) Estimar usando um modelo base para extrair a covariância, gerado a partir de poucos dados, pode levar à problemas devido a extrapolação;
- (iv) Dependendo do BMEC, o efeito pepita é incorporado ou não à tabela de covariância;
- (v) As simulações geradas a partir de tabelas de covariância tendem a não reproduzir variogramas, devido a não incorporação do efeito pepita.

Contornada as limitações do método, espera-se que as tabelas de covariância sejam eficazes na solução do sistema de cokrigagem, bem como, tornem os métodos sequenciais de simulação mais rápidos e mais eficientes.

1.1.2 Inferência da continuidade espacial de uma variável primária na presença de uma variável secundária exaustiva

A qualidade do variograma experimental se relaciona com quantidade de amostras presentes em um banco de dados e com o espaçamento entre elas. Se o banco de dados é bem amostrado e apresenta espaçamento regular, os variogramas experimentais facilmente são gerados. Mas em situações em que os dados são escassos e com espaçamento irregular, dificilmente o geomodelador chegará a um variograma experimental bem definido.

Nas situações em que se pretende realizar a cokrigagem colocada (XU et al., 1992) da variável primária, baseada na covariância da primária que não tem como ser inferida por apresentar poucas amostras e espaçamento irregular, o geomodelador em geral usa alguma informação auxiliar para inferir a continuidade espacial dessa variável. Uma alternativa é utilizar modelos de variogramas de depósitos minerais (ou reservatórios) análogos. Outra maneira é quando há uma variável secundária exaustiva e correlacionada com a primária, pode-se usá-la para inferir a continuidade espacial da primária. A alta correlação entre as variáveis indica que elas apresentam provavelmente as mesmas direções principais de continuidade espacial e, portanto, usa-se a informação secundária para inferir a anisotropia da primária.

O uso de informação secundária exaustiva para inferir a continuidade espacial da primária é comum na modelagem de reservatórios de óleo e gás. Nessa situação, a variável primária está presente em poucos poços, enquanto que a variável secundária proveniente de sísmica está exaustivamente amostrada. Embora o variograma vertical da variável primária possa ser inferido, o variograma horizontal da primária não fica bem definido e é necessário usar informação auxiliar proveniente da sísmica. Dessa forma, o variograma usado nas estimativas de simulação é ajustado ao variograma experimental vertical da primária e aos variogramas experimentais horizontais da variável secundária.

Uma particularidade do cálculo dos variogramas experimentais de uma variável exaustiva é que os dados estão em um *grid*, o que demanda usar vetores na determinação da direção. A definição desses vetores (Figura 1.1) demanda tempo do geomodelador. Além disso, os ângulos determinados por vetores apresentam, quase sempre, valores aproximados, ou seja, existe um erro de precisão no ângulo que descreve a continuidade espacial da variável em análise. Por exemplo, na direção $N22, 5^{\circ}E$ (ver Figura 1.1) usa-se o vetor (2,5,0), correspondente à $21,80^{\circ}$ (tan⁻¹ $\left(\frac{2}{5}\right) = 21,80^{\circ}$) e não ao usual 22,5°.



Figura 1.1 – Principais direções de anisotropia determinadas por vetores em grid regular: $N0^{\circ}E(0,1,0), N22, 5^{\circ}E(2,5,0), N45^{\circ}E(1,1,0), N67, 5^{\circ}E(5,2,0), N90^{\circ}E(1,0,0), N112, 5^{\circ}E(5,-2,0), N135^{\circ}E(1,-1,0) e N157, 5^{\circ}E(2,-5,0).$

Outro problema de calcular os variogramas experimentais é que o tempo de processamento é alto para *grids* grandes, como é o caso de cubos sísmicos usados para caracterizar reservatórios petrolíferos. Nesse tipo de situação, o variograma experimental é feito geralmente apenas nas direções principais de anisotropia. O modelo de variograma é então ajustado nessas direções principais. Essa maneira limita o modelo de continuidade espacial a combinações lineares de estruturas básicas e não explora totalmente a informação que a variável secundária fornece.

Nessa tese, pretende-se investigar uma alternativa ao cálculo do variograma experimental, fundamentada no uso de tabela de covariância obtida a partir dos dados exaustivos e método de Kloeckner et al. (2019). Os dados exaustivos substituirão o modelo de referência usado para extrair as covariâncias e, por não apresentarem problema de extrapolação, tendem a melhorar a qualidade das estimativas/simulações decorrentes da tabela de covariância. Além disso, a modelagem da continuidade espacial via método de Kloeckner et al. (2019) é bem mais rápida que abordagem convencional devido ao uso da FFT (MARCOTTE, 1996).

A maior rapidez com uso da tabela de covariância permite calcular variogramas experimentais em todas as possíveis direções, não apenas nas direções principais. Embora a tabela de covariância tenha grande potencial de uso em situações de dados secundários exaustivos, a aplicação dela para a realização de estimativa/simulação com cokrigagem colocada ainda não foi investigada.

1.2 Meta e objetivos

A presente pesquisa se propõe a explorar e desenvolver uma metodologia para estimar/simular múltiplas variáveis correlacionadas em conjuntos de dados heterotópicos com uso de tabelas de covariância, que dispense o modelo linear de corregionalização na modelagem da covariância conjunta.

Hipótese Investigativa: "Poder-se-ia ter uma metodologia que utilizasse tabelas de covariância diretas e cruzadas, geradas a partir de modelos base para extrair a covariância (BMECs ou variável secundária exaustiva) e transformadas de Fourier, para estimar/simular múltiplas variáveis com o auxílio de variáveis secundárias?"

Para que essa hipótese seja investigada são considerados os seguintes objetivos:

- 1. Investigar o uso de tabelas de covariância experimentais, obtidas a partir de BMECs e transformadas de Fourier no caso de múltiplas variáveis;
- Checar em quais situações as tabelas de covariância experimentais, obtidas a partir de BMECs, são positivas definidas;
- 3. Investigar métodos de correção das tabelas de covariância para torná-las positivas definidas, sem lesar a continuidade espacial;
- 4. Investigar como incorporar o efeito pepita à tabela de covariância;
- Desenvolver uma metodologia para estimar/simular múltiplas variáveis, que dispense a modelagem do LMC, baseada na correção conjunta das tabelas de covariância (diretas e cruzadas);
- 6. Verificar se as tabelas de covariância, após correção conjunta, podem ser usadas para estimar e simular múltiplas variáveis, em substituição a variografia convencional;
- 7. Investigar o uso da tabela de covariância de uma variável secundária exaustiva para inferir (ou ajudar a inferir) a continuidade espacial de uma variável primária que tem dados insuficientes para ter um variograma experimental confiável.

1.3 Organização da tese

A tese se organiza da seguinte maneira:

• No capítulo 2 temos uma breve revisão dos principais tópicos necessários à abordagem do tema, incluindo conceitos e definições da álgebra linear, da análise de Fourier,

de tabelas de covariância, de cokrigagem e da simulação sequencial Gaussiana em conjuntos multivariados heterotópicos.

- No capítulo 3 temos a apresentação da metodologia proposta. É exposto neste capítulo o procedimento utilizado para se obter uma matriz de covariância conjunta positiva definida, a partir de tabelas de covariância, sem a utilização do LMC.
- No capítulo 4 emprega-se a metologia proposta em um estudo de caso 2D. Além de se aplicar a metodologia desenvolvida, este capitulo expõe a análise e discussão dos resultados.
- O capítulo 5 mostra um estudo de caso do uso da tabela de covariância construída a
 partir dos dados de uma variável secundária exaustiva para inferir a continuidade
 espacial de uma variável primária. Após a inferência da continuidade espacial da
 primária, é realizada a estimativa/simulação da variável primária utilizando a variável
 secundária e cokrigagem colocada.
- O capítulo 6, apresenta as conclusões da tese e ainda, as contribuições e recomendações para trabalhos futuros.
2 Revisão bibliográfica e fundamentação teórica

Faz-se, a seguir, uma síntese dos tópicos importantes para o embasamento teórico da pesquisa.

2.1 Noções elementares de álgebra linear

Nesta seção, será realizado uma revisão sucinta dos conceitos elementares de álgebra linear, com intuito de colaborar com a compreensão da correção conjunta das matrizes de covariância.

2.1.1 Autovalores e autovetores

Definição 2.1.1. Se A é uma matriz $n \times n$, então um vetor não-nulo \boldsymbol{x} em \mathbb{R}^n é chamado de autovetor de A se $A\boldsymbol{x}$ é um múltiplo escalar de \boldsymbol{x} , ou seja (ANTON; RORRES, 2001):

$$A\boldsymbol{x} = \lambda \boldsymbol{x},\tag{2.1}$$

para algum escalar λ . O escalar λ é chamado de um autovalor de A e dizemos que \boldsymbol{x} é um autovetor associado a λ .

Para encontrar os autovalores de uma matriz A de tamanho $n \times n$, basta resolver a equação:

$$det(\lambda I - A) = 0, \tag{2.2}$$

onde I é a matriz identidade e det é o determinante associado à matriz resultante. Quando expandido, o determinante da equação 2.2 é um polinômio p característico de A, da forma:

$$p(\lambda) = det(\lambda I - A) = \lambda^n + c_1 \lambda^{n-1} + \dots + c_n$$
(2.3)

e equação característica:

$$\lambda^n + c_1 \lambda^{n-1} + \dots + c_n = 0, \tag{2.4}$$

sendo $c_1,..., c_n$ os coeficientes do polinômio. A equação 2.4 tem, no máximo, n autovalores distintos.

Calculados os autovalores, obtém-se os autove
tores associados à matriz A, resolvendo a equação matricial:

$$\begin{bmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \cdots & A_{nn} \end{bmatrix} \cdot \begin{bmatrix} x_{i1} \\ \vdots \\ x_{ij} \end{bmatrix} = \lambda_i \begin{bmatrix} x_{i1} \\ \vdots \\ x_{ij} \end{bmatrix}, i, j = 1, \cdots, n$$
(2.5)

em que λ_i é o autovalor associado ao autovetor x_i e x_{ij} são as componentes do autovetor x_i .

2.1.2 Diagonalização de matrizes reais simétricas e decomposição espectral

Em álgebra linear, uma matriz quadrada A é dita diagonalizável se existir uma matriz invertível P tal que $P^{-1}AP$ é uma matriz diagonal. Nesse caso, dizemos que a matriz P diagonaliza A (ANTON; RORRES, 2001).

Existem muitas matrizes que não são diagonalizáveis. De fato, algumas matrizes reais podem sequer possuir algum autovalor real. Contudo, com matrizes reais simétricas, esse problema não ocorre. A saber, temos os seguintes teoremas (LIPSCHUTZ; LIPSON, 2011):

Teorema 2.1.1. Se A é uma matriz real simétrica, então seus autovalores são reais.

Teorema 2.1.2. Sejam A uma matriz real simétrica e u e v autovetores de A associados a autovalores distintos λ_1 e λ_2 . Então u e v são ortogonais, isto é, o produto escalar entre eles é igual a zero.

Teorema 2.1.3. Se A é uma matriz real simétrica, então existe uma matriz ortogonal P tal que $D = P^{-1}AP$ é diagonal.

Teorema 2.1.4. Seja A uma matriz real $n \times n$. Então, A será simétrica se, e somente se, for ortogonalmente diagonalizável.

No teorema 2.1.3, a matriz ortogonal P é obtida normalizando uma base de autovetores ortogonais de A. Lembre-se de que uma matriz P é dita ortogonal se sua inversa P^{-1} é igual a sua transposta P^T e de que autovetores ortogonais normalizados apresentam comprimento unitário e produto escalar, entre dois autovetores distintos, igual a zero.

O teorema 2.1.4, chamado de **teorema espectral**, mostra que toda matriz simétrica e real é sempre diagonalizável. Através dele podemos escrever uma matriz simétrica e real A na forma: em que P é ortogonal e D é diagonal. Os elementos na diagonal D são exatamente os autovalores de A e as colunas de P, são os autovetores ortonormais de A (POOLE, 2004). A fatoração da matriz A em termos de autovalores e autovetores é chamada de **decomposição espectral**.

Exemplo: Vamos decompor a matriz simétrica $M = \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix}$ em termos de seus autovalores e autovetores.

Os autovalores da matriz M, calculados a partir da equação 2.2, são $\lambda_1 = 5$ e $\lambda_2 = -1$. Os autovetores de M associados aos autovalores λ_1 e λ_2 são, respectivamente, $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

e $v_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$. Os vetores $v_1 \in v_2$ são ortogonais e quando normalizados obtém-se os vetores ortonormais $u_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ e $u_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$. As matrizes ortonormal dos autovetores e diagonal dos autovalores são, respectivamente, $\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ e $\begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix}$. Assim, a decomposição espectral da matriz M é:

$$M = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$
 (2.7)

O resultado mais importante da álgebra linear, tratado neste estudo, diz respeito a decomposição espectral de uma matriz simétrica e real (ver equação 2.6). Ele será empregado na correção conjunta da matriz de covariância, obtida através de tabelas de covariância, com intuito de torná-la positiva definida, e com isso, dispensar a modelagem tradicional da continuidade espacial via LMC. Para tanto, a matriz de covariância conjunta necessita ser simétrica e real, ou seja, deve atender os critérios empregados na decomposição espectral e ainda, a definição de matriz positiva definida (ver nota de rodapé 1 na página 23).

2.2 Análise de Fourier

A análise de Fourier estuda, basicamente, a aproximação de qualquer função por uma soma de senos e cossenos. Os tópicos relacionados a análise de Fourier incluem a série de Fourier e as transformadas de Fourier, tratados a seguir.

2.2.1 Série de Fourier

Os valores de uma função f(x), definida no intervalo -L < x < L, podem ser aproximados a partir da soma de funções trigonométricas, ou seja:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right).$$
(2.8)

A soma de funções trigonométricas, descrita na equação (2.8), é chamada de série de Fourier de f no intervalo -L < x < L, ou seja, $f : \mathbb{R} \to \mathbb{R}$ é uma função periódica de período P igual a 2L. Os termos a_0 , $a_n \in b_n$ são chamados de coeficientes de Fourier e são determinados por:

$$a_0 = \frac{1}{L} \int_{-L}^{L} f(x) d(x), \qquad (2.9)$$

$$a_n = \frac{1}{L} \int_{-L}^{L} f(x) \cos \frac{n\pi x}{L} d(x)$$
(2.10)

е

$$b_n = \frac{1}{L} \int_{-L}^{L} f(x) \sin \frac{n\pi x}{L} d(x).$$
 (2.11)

A série de Fourier tem uma ampla gama de aplicação em ciências e engenharia, no estudo de condução de calor, no fenômeno ondulatório e em outros fenômenos físicos.

A série de Fourier consegue aproximar a curva em toda a parte de continuidade do gráfico de f, em vez de aproximar perto de um ponto isolado, mas não fornecem boas aproximações em torno dos pontos de descontinuidade de f (FINNEY; WEIR; GIORDANO, 2003).

Além de ser escrita na forma trigonométrica, a série de Fourier também pode ser escrita em termos de funções exponenciais complexas. Uma maneira de fazer isso é usar a fórmula de Euler (FIGUEIREDO, 2014):

$$e^{i\theta} = \cos\theta + i\sin\theta \tag{2.12}$$

e suas consequências:

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} \tag{2.13}$$

 $\sin \theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}.$ (2.14)

 \mathbf{e}

A forma trigonométrica da série de Fourier, em termos complexos, é representada pela equação (2.15):

$$f(x) = \sum_{n = -\infty}^{\infty} c_n e^{\frac{-in\pi x}{L}},$$
(2.15)

em que:

$$c_n = \frac{1}{2L} \int_{-L}^{L} f(x) e^{\frac{-in\pi x}{L}} d(x), \qquad (2.16)$$

 $\operatorname{com} n \in \mathbb{Z}.$

2.2.2 A integral de Fourier

Como exposto, a série de Fourier serve para representar, em termos de funções senos e cossenos, funções periódicas ou funções contínuas definidas em um intervalo finito de interesse. A representação da função ocorre através da soma das magnitudes dos harmônicos que compõem a série de Fourier.

A integral de Fourier, por ser uma generalização da série de Fourier, se propõe a fazer a representação harmônica de funções não periódicas. A ideia fundamental é tomar uma função periódica e fazer o seu período tender ao infinito. A representação, em forma de integral, da série de Fourier (ver equação 2.8) é:

$$f(x) = \int_0^\infty \left[A(\omega) \cos x\omega + B(\omega) \sin x\omega \right] d\omega, \qquad (2.17)$$

em que:

$$A(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(x) \cos \omega x dx, \quad \omega \ge 0$$
(2.18)

е

$$B(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(x) \sin \omega x dx, \quad \omega \ge 0, \quad (2.19)$$

onde ω é a frequência angular.

2.2.3 A transformada de Fourier

A transformada de Fourier da função f, obtida a partir da integral de Fourier, é definida como sendo a função \mathcal{F} que associa a cada função absolutamente integrável $f: \mathbb{R} \to \mathbb{R}$ a função $\hat{f}: \mathbb{R} \to \mathbb{C}$ definida pela equação:

$$\widehat{f}(\omega) = \mathcal{F}[f](\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx.$$
(2.20)

A inversa da transformada Fourier é a função \mathcal{F}^{-1} que associa a cada função $\hat{f}: \mathbb{R} \to \mathbb{C}$ a função absolutamente integrável $f: \mathbb{R} \to \mathbb{R}$ definida pela equação:

$$f(x) = \mathcal{F}^{-1}[\hat{f}](x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i\omega x} d\omega.$$
(2.21)

2.2.3.1 Propriedades da transformada de Fourier

A transformada de Fourier se baseia nas seguintes operações que envolvem funções (FIGUEIREDO, 2014) :

Propriedade 1 (Linearidade): A transformada de Fourier é um operador linear, isto é:

$$\mathcal{F}(af + bg) = a\mathcal{F}(f) + b\mathcal{F}(g),$$

em que $f \in g$ são funções em \mathcal{L}^1 (conjunto de todas as funções complexas absolutamente integráveis) e $a \in b$ números complexos quaisquer. Essa propriedade é decorrente da linearidade da integral.

Propriedade 2 (Transformadas de derivadas): Se f é uma função absolutamente integrável, tal que as suas derivadas até a ordem k também são funções absolutamente integráveis, então:

$$\mathcal{F}(f^{(k)})(\omega) = (i\omega)^k \mathcal{F}(f)(\omega).$$

Propriedade 3 (Derivadas de Transformadas de Fourier): Se f é uma função absolutamente integrável, tal que $x^k f(x)$ também são funções absolutamente integráveis, então:

$$\mathcal{F}(x^k f(x))(\omega) = i^k \mathcal{F}(f)^k(\omega).$$

Propriedade 4 (Transformada de Fourier de uma Translação): Se f é uma função absolutamente integrável, então:

$$\mathcal{F}(f(x-a))(\omega) = e^{-i\omega a} \mathcal{F}(f(x))(\omega)$$

е

$$\mathcal{F}\left(e^{iax}f(x)\right)(\omega) = \mathcal{F}\left(f(x)\right)(\omega).$$

Propriedade 5 (Transformada de Fourier de uma Dilatação): Se f é uma função absolutamente integrável e $a \neq 0$, então:

$$\mathcal{F}(f(ax))(\omega) = \frac{1}{|a|}\mathcal{F}(f)\left(\frac{\omega}{a}\right).$$

Tratarei agora de definir a convolução entre funções para então apresentar a propriedade da transformada de Fourier de uma convolução.

Sejam $f \in g$ duas funções absolutamente integráveis. Definimos a operação de convolução de $f \in g$ como sendo a função $f * g : \mathbb{R} \to \mathbb{C}$, com domínio real (\mathbb{R}) e imagem complexa (\mathbb{C}), dada por (IÓRIO, 2012):

$$(f * g)(x) = \int_{-\infty}^{\infty} f(y)g(x - y)dy, x \in \mathbb{R}.$$

Decorrem da definição, as seguintes propriedades para as funções absolutamente integráveis $f,g \in h \in \alpha \in \mathbb{R}$:

- (i) Associatividade: (f * g) * h = f * (g * h)
- (ii) Comutatividade: f * g = g * f
- (iii) **Distributividade:** (f+g) * h = f * h + g * h, $(\alpha f) * g = \alpha (f * g)$

Propriedade 6 (Transformada de Fourier de uma Convolução): Sejam $f \in g$ duas funções absolutamente integráveis, então:

$$\mathcal{F}(f * g) = \sqrt{2\pi} \mathcal{F}(f) \mathcal{F}(g).$$

2.2.4 A transformada discreta de Fourier

Em geoestatística, os dados frequentemente são amostrados de forma que não podemos utilizar o cálculo da transformada contínua. Para esse tipo de amostragem, precisamos entender como é o procedimento da transformada para um conjunto discreto e igualmente espaçado de dados, que é a DFT (*Discrete Fourier Transform* - DFT).

Considere x_n como sendo uma sequência discreta, periódica e com período N. A transformada discreta de Fourier dessa sequência se define por:

$$X_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n e^{-ik\frac{2\pi n}{N}}, k = 0, 1, 2, \cdots, N-1$$
(2.22)

e sua inversa

$$x_n = \sum_{k=0}^{N-1} X_k e^{ik\frac{2\pi n}{N}}, n = 0, 1, 2, \cdots, N-1.$$
 (2.23)

Na transformada de Fourier contínua, descrita na seção 2.2.3, calculamos a integral de menos infinito a mais infinito e agora, no caso discreto, calculamos uma soma da amostra n = 0 até a amostra N - 1.

Para um determinado valor k, são necessárias N^2 multiplicações complexas e N(N-1) somas complexas para se obter os N valores da série transformada, o que demanda um grande esforço computacional.

2.2.5 A transformada rápida de Fourier

A transformada rápida de Fourier (*Fast Fourier Transform* - FFT) ou algoritmo de Cooley-Tukey (COOLEY; TUKEY, 1965) é um algoritmo que calcula a transformada discreta de Fourier e sua inversa. A ideia básica do algoritmo é dividir a transformada discreta em transformadas menores recursivamente, buscando diminuir o esforço computacional (YAO, 1998).

A FFT pode ser entendida como sendo um algoritmo que calcula a DFT de forma rápida e com menor custo. A FFT considera que N é potência de 2 e decompõe a DFT segundo a paridade dos termos. Através da decomposição, que explora a simetria do conjugado complexo e periodicidade na DFT, a FFT consegue reduzir a ordem dos cálculos da DFT de $O(N^2)$ para $O(N \log_2 N)$, resultando na diminuição do tempo de processamento (YOO, 2001).

2.3 Análise espectral

A análise espectral (ou aplicações dos métodos da análise de Fourier) tem por finalidade procurar a periodicidade nos dados analisados, estejam eles em função do tempo ou do espaço. São vários os campos de aplicação, dentre eles, engenharia elétrica, comunicações, física e medicina.

Poucos são os fenômenos que possuem componentes periódicos detectáveis, a partir de um banco de dados, no domínio do tempo (ou espacial), mas isso não invalida a periodicidade do fenômeno no domínio da frequência e subsequente análise do espectro (PRASAD, 1991). A transformada do domínio espacial para o domínio de frequência ocorre mediante a transformada de Fourier.

Um importante resultado da análise espectral nos diz que a covariância de z(u), descrita pela equação 1.7, é definida em uma dimensão como o produto convolução (BRACEWELL, 1986 apud KLOECKNER et al., 2019):

$$C_z(h) = \int_{-\infty}^{\infty} z(u)z(u+h)du = z(u) * z(-u).$$
(2.24)

Aplicando a transformada de Fourier na equação 2.24, obtemos a densidade espectral de z(u) no domínio frequência da seguinte maneira:

$$s_{z}(\omega) = \mathcal{F}[C_{z}(h)] = \mathcal{F}[z(u)] * \overline{\mathcal{F}}[z(u)] = |Z(\omega)|^{2}, \qquad (2.25)$$

onde $\mathcal{F}[\cdot]$ é o operador da transformada de Fourier. Consequentemente, $Z(\omega) = \mathcal{F}[z(u)] = \int z(u)e^{-i\omega u}du$, $\overline{\mathcal{F}}[z(u)]$ é o conjugado complexo e $s_z(\omega)$ representa a densidade espectral. A densidade espectral pode ser usada na simulação espectral (PARDO-IGUZQUIZA; CHICA-OLMO, 1993; YAO, 2004) ou pode ser transformada novamente em tabela de covariância no domínio espacial.

A aplicação prática da transformada de Fourier é obtida através da DFT. A DFT substitui a integral por um somatório (YOO, 2001 apud KLOECKNER et al., 2019), isto é, dado N amostras discretas de z(u), a transformada de Fourier $Z(\omega)$ de z(u) é:

$$Z(\omega) = \frac{1}{N} \sum_{u=0}^{N-1} z(u) e^{-i\omega \frac{2\pi u}{N}}, \quad \omega = 0, 1, 2, \cdots, N-1$$
(2.26)

e a transformada inversa de $Z(\omega)$ é:

$$z(u) = \sum_{\omega=0}^{N-1} Z(\omega) e^{i\omega \frac{2\pi u}{N}}, u = 0, 1, 2, \cdots, N-1.$$
 (2.27)

2.4 Modelagem da continuidade espacial via transformada de Fourier

Esta seção trata da revisão das principais metodologias que usam transformadas de Fourier na modelagem da continuidade espacial.

2.4.1 Computação rápida de variogramas via FFT

A FFT foi introduzida na modelagem da continuidade espacial por Denis Marcotte no ano de 1996. Em seu artigo, Marcotte (1996) apresentou dois programas para computação rápida das funções de continuidade espacial, dentre elas variogramas diretos e cruzados, via FFT e dados originais. Os programas são executados no ambiente MATLAB e nesse caso, o acesso a eles não é gratuito.

Para computar a covariância cruzada não centrada entre duas funções randômicas contínuas $f(x, y) \in g(x, y)$, separadas pelas distâncias $h_x \in h_y$, Marcotte (1996) considerou a seguinte expressão:

$$C(h_x, h_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)(x + h_x, y + h_y) dx dy$$
 (2.28)

Usando a transformada contínua de Fourier em (2.28), obtemos:

$$TF[C] = F^*(u, v)G(u, v),$$
 (2.29)

onde $F^*(u, v)$ é o conjugado complexo de F(u, v), F(u, v) é a transformada de Fourier de f(x, y) e G(u, v) é a transformada de Fourier de g(x, y).

Marcotte (1996) salienta que a covariância não centrada pode ser obtida diretamente da transformada inversa de Fourier do produto das transformadas de f(x, y) e g(x, y). Os resultados da transformada inversa de Fourier são diretamente deslocados para a forma discreta, dado que as funções discretas f(x, y) e g(x, y) são preenchidas por zeros até o tamanho de 2N - 1 por 2M - 1, em que N e M são as dimensões do grid regular. O preenchimento é necessário, justamente para permitir a utilização da FFT no processo.

A eficiência do algoritmo da FFT é significativa, pois reduz o tempo e a complexidade computacional utilizados no cálculo da covariância. Na figura 2.1, que compara a rapidez da FFT em relação a abordagem espacial com oito direções e espaçamento máximo até a metade do grid (N/2), para um grid regular de dimensão 128 a FFT é 7,99 vezes mais rápida que a abordagem espacial, por exemplo.



Figura 2.1 – Comparativo entre a rapidez da FFT e a abordagem espacial. Adaptado de Marcotte (1996).

2.4.2 Modelagem automática de tabelas de covariância (cruzadas) utilizando a FFT

Em 1998, Yao propõe uma metodologia para a obtenção "automática" de uma tabela de covariância em seu artigo intitulado Modelagem Automática de Tabelas de Covariâncias (Cruzadas) utilizando Transformada Rápida de Fourier, um dos produtos de sua tese de doutorado. O termo "automático" foi posteriormente retratado, pois os autores reconheceram seu mau uso (KLOECKNER, 2018).

O procedimento utilizado na obtenção da covariância se baseia no teorema de Bochner (BOCHNER, 1939). O teorema de Bochner afirma que uma função C(h) é definida positiva se, e somente se, ela pode ser representada como uma transformada de Fourier positiva, não decrescente e limitada $S(\omega)$ (YAO; JOURNEL, 1998), ou seja:

$$C(h) = \int_{\Re^d} e^{-2\pi i h \cdot \omega} dS(\omega), \qquad (2.30)$$

ou no caso discreto:

$$C(h) = \sum_{k=1}^{K} e^{2\pi i h \omega} s(\omega_k), k = 1, \cdots, K,$$
(2.31)

onde h é o vetor distância de dimensão-d no espaço \Re^d ; ω é o vetor frequência no correspondente domínio de frequência e $s(\omega_k)$ é uma série de espectros.

Se o modelo da função estacionário aleatória (RF) Z(u) tiver variância unitária (C(0) = VarZ(u) = 1), então $S(\omega)$ pode ser visto como uma distribuição acumulada da função (cdf), com $dS(\omega) = s(\omega)d\omega$. O teorema de Bochner requer que a densidade $s(\omega) = (dS(\omega))/d\omega$ seja uma função densidade de probabilidade (pdf))(YAO; JOURNEL, 1998), ou seja:

$$s(\omega) \ge 0, \forall \omega \tag{2.32}$$

е

$$\int_{\Re^d} s(\omega) d\omega = 1, \tag{2.33}$$

em que $s(\omega)$ é a densidade espectral ou densidade do espectro da função aleatória Z(u).

O método de Yao e Journel (1998) se baseou no teorema de Bochner para garantir a positividade da tabela de covariância, sem utilizar um modelo analítico de covariância. As etapas do método consistiram em:

- (i) Calcular o mapa de covariância amostral Nesta etapa, os valores de covariância são calculados, considerando todas as direções e separações disponíveis no grid. A seguir, os valores de covariância das amostras são interpolados de tal forma que os locais faltantes sejam preenchidos e, ainda, remova os valores de covariância mal informados. O tamanho e a discretização da tabela de covariância após a interpolação deve ser tal que permita encontrar todos os valores de covariância necessários para um futuro sistema de krigagem. As equações que calculam a covariância direta e a covariância cruzada, são, respectivamente, 1.7 e 1.11.
- (ii) FFT e suavização da tabela de espectro Depois que a tabela de covariância amostral é concluída, uma FFT fornece um mapa preliminar de espectros. Usando média móvel,

os valores preliminares de espectro são suavizados, sob restrições de positividade e soma unitária (condições descritas nas equações 2.32 e 2.33). A tabela de espectros suavizados é lícita, ou seja, mesmo no domínio frequência ela pode ser usada nas simulações espectrais.

(iii) Retrotransformar (inverso da FFT) os valores suavizados da densidade espectral
- A transformada inversa de Fourier dos espectros suavizados gera uma tabela de covariância positiva definida, ou seja, lícita para os sistemas de krigagem/simulações.

Yao e Journel (1998) comentam que o método desenvolvido gera tabelas de covariância positivas definidas em menor tempo que a modelagem tradicional de variogramas. Além disso, as estimativas/simulações obtidas via tabelas de covariância se assemelham às obtidas via variogramas ajustados, se não melhores.

2.4.3 Tabela de covariância: um mapeamento rápido e automático da continuidade espacial

Kloeckner et al. (2019) apresentaram um método para computar tabelas de covariância via grid amostral e FFT. O método de Kloeckner et al. (2019) é resultado da dissertação de mestrado de Kloeckner (KLOECKNER, 2018) e se baseia no preenchimento do grid amostral a partir de um modelo base para extrair a covariância (BMEC), dispensando o cálculo da covariância amostral e suavizações, previstos no método de Yao e Journel (1998).

As etapas do método de Kloeckner et al. (2019) consistem em:

- (i) Preencher o grid amostral usando um modelo base para extrair a covariância (BMEC).
 O BMEC é construído através da interpolação dos valores amostrais e ele, por carregar a continuidade espacial da variável analisada, é usado para extrair a tabela de covariância.
- (ii) Usar a transformada de Fourier e o teorema da convolução, conforme equações 2.34 e 2.35, para obter as matrizes de covariância próximas e admissíveis. As matrizes de covariância são aproximadas através da densidade espectral, resultante da transformada de Fourier dos valores do BMEC. A equação 2.34 determina a matriz de covariância direta e a equação 2.35, a matriz de covariância cruzada.

$$F[C_Z(h)] = |F[Z]|^2$$
(2.34)

е

$$F[C_{ZY}(h)] = F[Z] \cdot \overline{F}[Y], \qquad (2.35)$$

onde $F[C_Z(h)]$ é a transformada de Fourier da função covariância $C_Z(h)$, |F[Z]|é o módulo da transformada de Fourier dos valores da variável Z, $F[C_{ZY}(h)]$ é a transformada de Fourier da covariância cruzada (amostral) entre as variáveis $z \in y$, F[Z] é a transformada de Fourier dos valores da variável $Z \in \overline{F}[Y]$ é o conjugado complexo da transformada de Fourier dos valores da variável Y.

Para garantir a positividade da tabela de covariância direta, os valores obtidos, que representam a densidade espectral, são corrigidos para serem positivos e terem soma unitária (YAO; JOURNEL, 1998; PYRCZ; DEUTSCH, 2006).

(iii) Retrotransformar as equações 2.34 e 2.35 para obter as tabelas de covariância desejadas. As equações 2.36 e 2.37 determinam as tabelas de covariância, respectivamente, direta e cruzada.

$$C_Z(h) = F^{-1}\left[\left|F\left[Z\right]^2\right|\right]$$
 (2.36)

е

$$C_{ZY}(h) = F^{-1}\left[F\left[Z\right] \cdot \overline{F}\left[Y\right]\right], \qquad (2.37)$$

onde F^{-1} indica a transformada inversa de Fourier.

O método aqui descrito, apesar de ser útil para gerar tabelas de covariância diretas e cruzadas a partir de um BMEC e FFT, só garante a positividade das tabelas de covariância diretas, ou seja, as tabelas de covariância geradas pelo método de Kloeckner et al. (2019) são úteis para as aplicações univariadas. Para a multivariada, tais tabelas de covariância (diretas e cruzadas) precisam ser corrigidas conjuntamente. A correção conjunta das tabelas de covariância geradas pelo Kloeckner et al. (2019) é objeto da tese e será descrita no capítulo 3.

2.5 Estimativa multivariada

Em conjuntos multivariados, é comum encontrarmos além da variável de interesse que se deseja estimar (variável primária), uma ou mais variáveis secundárias mais densamente amostradas, que podem ser utilizadas para melhorar a estimativa, bastando apenas que essas variáveis estejam espacialmente correlacionadas entre si. A principal metodologia da geoestatística usada para incorporar as informações das variáveis secundárias nas estimativas é a cokrigagem (MARECHAL, 1970), que requer a modelagem conjunta de variogramas diretos e cruzados. A modelagem conjunta de variogramas normalmente é caracterizada pelo LMC. O LMC é empregado para garantir que a matriz de covariância conjunta, utilizada na cokrigagem, seja positiva definida. O processo é eficiente, mas torna-se trabalhoso em conjuntos com mais de duas variáveis secundárias.

Dentro de um domínio, as variáveis primária e secundária podem ser encontradas no mesmo local ou em locais distintos, configurando três situações: heterotopia total, heterotopia parcial e isotopia. A heteropia total ocorre quando os valores das variáveis primária e secundária são verificados em diferentes posições e não apresentam pontos em comum e em casos de heteropia parcial, as variáveis primária e secundária apresentam alguns locais em comum. A isotopia se configura quando as variáveis primária e secundária estão disponíveis nos mesmos locais amostrados (WACKERNAGEL, 1996).

Nos casos de heteretopia total, devido as variáveis primária e secundária não serem amostradas nos mesmos locais, os variogramas cruzados não podem ser modelados diretamente. Por outro lado, na isotopia, os variogramas cruzados são facilmente calculados, porém as estimativas por cokrigagem produzem resultados semelhantes aos da krigagem. Na heterotopia parcial, quando as variáveis secundárias estão disponíveis em mais pontos que a variável primária, a cokrigagem é significativamente vantajosa.

A seguir, será apresentado as principais metodologias da cokrigagem para banco de dados heterotópicos e ainda, as alternativas da cokrigagem para contornar a modelagem de variogramas via LMC.

2.5.1 Cokrigagem ordinária e cokrigagem ordinária estandardizada

A cokrigagem ordinária (*Ordinary Cokriging* - OCK) é similar a krigagem ordinária e seu estimador é uma combinação linear de atributos primários e secundários. O estimador de cokrigagem ordinária é definido por:

$$Z_{OCK}^*(u) = \sum_{\alpha i=1}^{n_i} \lambda_{\alpha i}^{OCK} Z_i(u_{\alpha i}) + \sum_{\alpha j=1}^{n_j} \lambda_{\alpha j}^{OCK} Z_j(u_{\alpha j}), \qquad (2.38)$$

em que $Z_{OCK}^*(u)$ é a estimativa por cokrigagem ordinária da variável $Z_i(u)$, no local $u; Z_i(u_{\alpha i})$ são os dados da variável primária em n_i locais $u_{\alpha i}; Z_j(u_{\alpha j})$ são os dados da variável secundária em n_j locais $u_{\alpha j}; \sum_{\alpha i=1}^{n_i} \lambda_{\alpha i}^{OCK}$ é a soma dos pesos da variável primária e $\sum_{\alpha j=1}^{n_j} \lambda_{\alpha j}^{OCK}$ é a soma dos pesos da variável secundária.

O estimador de cokrigagem minimiza a esperança matemática do erro, ou seja:

$$E[Z_{OCK}^{*}(u) - Z_{i}(u)] = 0,$$

tornando às condições dos pesos da cokrigagem ordinária dados por:

$$\sum_{\alpha i=1}^{n_i} \lambda_{\alpha i}^{OCK} = 1$$

е

$$\sum_{\alpha j=1}^{n_j} \lambda_{\alpha j}^{OCK} = 0.$$

Para evitar pesos de krigagem negativos ou muito pequenos para as amostras da variável secundária, Isaaks e Srivastava (1989) e Goovaerts (1997) propõem que o somatório total de todos os pesos da variável de interesse e das variáveis secundárias sejam iguais a 1, ou seja:

$$\sum_{\alpha i=1}^{n_i} \lambda_{\alpha i}^{OCK} + \sum_{\alpha j=1}^{n_j} \lambda_{\alpha j}^{OCK} = 1.$$
(2.39)

Com base na condição sobre os pesos, o estimador de cokrigagem torna-se expresso como:

$$Z_{OCK}^{*}(u) = \sum_{\alpha i=1}^{n_i} \lambda_{\alpha i}^{OCK} Z_i(u_{\alpha i}) + \sum_{\alpha j=1}^{n_j} \lambda_{\alpha j}^{OCK} [Z_j(u_{\alpha j}) - m_j + m_i], \qquad (2.40)$$

onde m_i e m_j são as médias desagrupadas das variáveis envolvidas. Essa cokrigagem é chamada de cokrigagem ordinária estandardizada (*Standardized Ordinary Cokriging* – SOCK).

Usando as limitações dos pesos, o sistema de cokrigagem ordinária é escrito como:

$$\begin{cases} \sum_{\alpha i=1}^{n_{i}} \lambda_{\alpha i}^{OCK} C_{ii}(u_{\alpha i} - u_{\beta i}) + \sum_{\alpha j=1}^{n_{j}} \lambda_{\alpha j}^{OCK} C_{ij}(u_{\alpha i} - u_{\alpha j}) + \mu(u) = C_{ii}(u_{\alpha i} - u) \\ \sum_{\alpha i=1}^{n_{i}} \lambda_{\alpha i}^{OCK} C_{ji}(u_{\alpha j} - u_{\alpha i}) + \sum_{\alpha j=1}^{n_{j}} \lambda_{\alpha j}^{OCK} C_{jj}(u_{\alpha j} - u_{\beta j}) + \mu(u) = C_{ji}(u_{\alpha j} - u), \\ \sum_{\alpha i=1}^{n_{i}} \lambda_{\alpha i}^{OCK} + \sum_{\alpha j=1}^{n_{j}} \lambda_{\alpha j}^{OCK} = 1 \end{cases}$$

$$(2.41)$$

onde $\sum_{\alpha i=1}^{n_i} \lambda_{\alpha i}^{OCK}$ é o somatório dos pesos de cokrigagem atribuídos aos dados da variável primária $Z_i(u_{\alpha i})$; $\sum_{\alpha j=1}^{n_j} \lambda_{\alpha j}^{OCK}$ é o somatório dos pesos de cokrigagem atribuídos aos dados da variável secundária $Z_j(u_{\alpha j})$; $C_{ii}(u_{\alpha i} - u_{\beta i})$ é a covariância direta da variável primária entre os locais $u_{\alpha i}$ e $u_{\beta i}$; $C_{ij}(u_{\alpha i} - u_{\alpha j})$ é a covariância cruzada da variável primária nos locais $u_{\alpha i}$ e da variável secundária nos locais $u_{\alpha j}$; $C_{ii}(u_{\alpha i} - u)$ é a covariância direta da variável primária entre os locais $u_{\alpha i}$ e o local u onde se estima; $C_{ji}(u_{\alpha j} - u_{\alpha i})$ é a covariância cruzada da variável secundária nos locais $u_{\alpha j}$ e da variável primária nos locais $u_{\alpha i}$, que é igual a covariância $C_{ij}(u_{\alpha i} - u_{\alpha j})$; $C_{jj}(u_{\alpha j} - u_{\beta j})$ é a covariância direta da variável secundária entre os locais $u_{\alpha j}$ e $u_{\beta j}$; $C_{ji}(u_{\alpha j} - u)$ é a covariância cruzada da variável secundária entre locais $u_{\alpha j}$ e u onde se estima e μ é o multiplicador de Lagrange.

O sistema de cokrigagem também pode ser escrito em termos do semivariograma, bastando apenas que a covariância cruzada seja simétrica. Embora a covariância cruzada não seja simétrica, na prática ela modelada como assim o fosse (fazendo a média entre C_{ji} e C_{ij}). A modelagem da continuidade espacial é modelada através de semivariogramas e depois os valores são lidos para diversos vetores de separação h para contruir as matrizes de covariância, presentes no sistema de cokrigagem. Para que a solução do sistema exista e seja única, os variogramas diretos e cruzados devem ser permissíveis, o que usualmente é verificado através do modelo linear de corregionalização (ISAAKS; SRIVASTAVA, 1989).

Quando as variáveis primária e secundária apresentam diferentes magnitudes no valor da variância, é comum trabalhar com as variáveis estandardizadas. Dessa forma, o estimador da cokrigagem ordinária standardizada é dado pela equação 2.42 (GOOVAERTS, 1997):

$$\frac{Z_{SOCK}^*(u) - m_i}{\sigma_i} = \sum_{\alpha i=1}^{n_i} \lambda_{\alpha i}^{SOCK} \frac{Z_i(u_{\alpha i}) - m_i}{\sigma_i} + \sum_{\alpha j=1}^{n_j} \lambda_{\alpha j}^{SOCK} \frac{Z_j(u_{\alpha j}) - m_j}{\sigma_j}, \qquad (2.42)$$

em que $Z^*_{SOCK}(u)$ é o estimador da cokrigagem ordinária estandardizada; m_i é a média da variável primária; σ_i é o desvio padrão da variável primária; σ_j é o desvio padrão da variável secundária e m_j é a média da variável secundária.

2.5.2 Cokrigagem colocada ordinária

A cokrigagem colocada ordinária (*Colocated Ordinary Cokriging* - COCK) (XU et al., 1992) é um método multivariado de estimativa que visa minimizar a influência de dados secundários que estão fora do local onde a variável primária deve ser estimada. Ou seja, considera apenas o valor da variável secundária que está localizado no ponto a ser estimado a primária.

O estimador da COCK é dado pela equação (XU et al., 1992; GOOVAERTS, 1997):

$$Z_{COCK}^{*}(u) = \sum_{\alpha i=1}^{n_{i}} \lambda_{\alpha i}^{COCK}(u) Z_{1}(u_{\alpha 1}) + \lambda_{2}^{COCK}(u) \left[Z_{2}(u) - m_{2} + m_{1} \right], \qquad (2.43)$$

onde $Z^*_{COCK}(u)$ é o estimador da cokrigagem colocada ordinária da variável $Z_1(u)$ no local $u; \sum_{\alpha i=1}^{n_i} \lambda_{\alpha i}^{COCK}$ são os pesos da cokrigagem colocada ordinária para os dados da variável $Z_1(u_{\alpha i}); \lambda_2^{COCK}(u)$ é o peso da cokrigagem colocada ordinária da variável secundária no

local u; $Z_1(u_{\alpha 1})$ são os valores da variável primária em n_1 locais $u_{\alpha 1}$; $Z_2(u)$ é o valor da variável secundária no local u; m_1 é a média da variável primária $Z_1(u)em_2$ é a média da variável secundária $Z_2(u)$.

A única restrição sobre os pesos na cokrigagem colocada ordinária é que a soma deles seja igual a um, ou seja (GOOVAERTS, 1997):

$$\sum_{\alpha i=1}^{n_i} \lambda_{\alpha i}^{COCK}(u) + \lambda_2^{COCK}(u) = 1.$$
(2.44)

Os pesos da COCK são obtidos resolvendo o sistema de equações lineares (XU et al., 1992; GOOVAERTS, 1997):

$$\begin{cases} \sum_{\beta=1}^{n_{i}} \lambda_{\beta1}^{COCK}(u) C_{11}(u_{\alpha1} - u_{\beta1}) + \lambda_{2}^{COCK}(u) C_{12}(u_{\alpha1} - u) + \mu^{COCK}(u) = C_{11}(u_{\alpha1} - u) \\ \sum_{\beta=1}^{n_{i}} \lambda_{\beta1}^{COCK}(u) C_{21}(u - u_{\beta1}) + \lambda_{2}^{COCK}(u) C_{22}(0) + \mu^{COCK}(u) = C_{21}(0) \\ \sum_{\beta=1}^{n_{i}} \lambda_{\beta1}^{COCK}(u) + \lambda_{2}^{COCK}(u) = 1 \end{cases}$$

$$(2.45)$$

em que $C_{11}(.)$ e $C_{22}(.)$ são as funções covariância direta das variáveis primária e secundária, respectivamente, $C_{12}(.)$ e $C_{21}(.)$ são as funções covariância cruzadas e μ é o multiplicador de Lagrange.

Quando a variância da variável primária é significativamente diferente da variância da variável secundária, Goovaerts (1997) sugere que os valores das variáveis sejam estandardizados para terem média zero e variância unitária. Nesse caso, os pesos da COCK seriam solução do sistema 2.45 expresso em termos de correlogramas.

Tanto no sistema de cokrigagem ordinária, quanto no sistema de cokrigagem colocada as funções de covariância conjunta devem ser modeladas, respeitando as condições do modelo linear de corregionalização. Uma alternativa que dispensa a modelagem conjunta da função covariância se baseia no modelo de corregionalização de Markov I, o qual requer apenas a inferência da covariância da variável primária e o coeficiente de correlação colocado $\rho_{12}(0)$ entre as variáveis primária e secundária (JOURNEL, 1999).

2.5.3 Modelos de Markov

Os modelos MM1 (ALMEIDA, 1993) e MM2 (JOURNEL, 1999), chamados de modelos de Markov, foram introduzidos para simplificar a utilização do LMC no sistema de cokrigagem colocada. No modelo MM1, a covariância cruzada entre as variáveis é proporcional a covariância da variável primária e no MM2, a covariância cruzada é proporcional a covariância da variável secundária, mais densamente amostrada (BOEZIO, 2010). No modelo MM1, a função de covariância da variável secundária não é utilizada e a covariância cruzada $(C_{12}(h) = C_{21}(h))$ é calculada pela equação (XU et al., 1992):

$$C_{12}(h) = \frac{C_{12}(0)}{C_1(0)} \cdot C_1(h)$$
(2.46)

ou de forma equivalente:

$$\rho_{12}(h) = \rho_{12}(0) \cdot \rho_1(h), \qquad (2.47)$$

em que $C_{12}(h)$ é a função covariância cruzada entre as variáveis primária e secundária para $h \neq 0, C_1(0)$ é a covariância direta da variável primária para $h = 0, C_{12}(0)$ é a covariância cruzada entre as variáveis primária e secundária para $h = 0, C_1(h)$ é a função covariância direta da variável primária para $h \neq 0, \rho_{12}(h) = \frac{C_{12}(h)}{\sqrt{C_1(0) \cdot C_2(0)}}$ é a função correlograma cruzado para $h \neq 0, \rho_{12}(0)$ é a função correlograma cruzado para h = 0 e $\rho_1(h) = \frac{C_1(h)}{C_1(0)}$ é a função correlograma direto da variável primária para $h \neq 0$.

Além de assumir que a covariância cruzada entre as variáveis primária e secundária é proporcional à covariância da variável primária, o MM1 necessita que todas as covariâncias cruzadas compartilhem a mesma forma de continuidade espacial que a variável primária apresenta. Mas em muitas aplicações práticas, os dados secundárias são mais densamente amostrados que os dados primários e nesse caso, é muito mais simples obter a continuidade espacial das variáveis secundárias, ou seja, a inferência do $\rho_2(h)$ é mais simples que a do $\rho_1(h)$. Nesse caso, A substituição do $\rho_1(h)$ por $\rho_2(h)$ na equação 2.47 é razoável e constitui o MM2 (JOURNEL, 1999).

No modelo MM2, a função correlograma cruzado $\rho_{12}(h)$ é determinada pela equação:

$$\rho_{12}(h) = \rho_{12}(0) \cdot \rho_2(h). \tag{2.48}$$

Silva (2013) comenta que no MM2, apesar dos dados secundários serem abundantes e, portanto, mais fácil de se determinar a sua continuidade espacial, ainda é necessária a modelagem das covariâncias diretas e cruzadas, mesmo assumindo a proporcionalidade entre elas, ou seja, a dificuldade encontrada ao modelar o LMC não é completamente contornada, pois o MM2 deve atender a condição de positiva definida.

2.6 Simulação sequencial gaussiana em conjuntos de dados multivariados heterotópicos

Em conjuntos de dados multivariados heterotópicos correlacionados, dependendo do interesse do geomodelador, pode-se simular somente a variável primária condicionando informações primária e secundária ou simular mais de uma variável condicionando informações primárias e secundárias. A simulação de uma variável ocorre através da simulação sequencial gaussiana (Sequential Gaussian Simulation - SGS) (ISAAKS, 1990) e no caso de múltiplas variáveis, emprega-se a cossimulação sequencial gaussiana (Sequential Gaussian Cosimulation - SGCS) (VERLY, 1993). Independentemente do número de variáveis a serem simuladas, os algoritmos da SGS e SGCS exigem que os dados sejam multigaussianos e que os parâmetros (média e variância) da função gaussiana de distribuição acumulativa condicional (conditional cumulative distribution function - ccdf) sejam determinados pela cokrigagem simples. Maiores informações sobre a SGS e SGCS, com uso de informação secundária, podem ser consultadas em Goovaerts (1997) e Soares (2006).

Neste estudo, estamos interessados em simular uma variável primária condicionada às informações desta e da secundária. As variáveis são heterotópicas e estão correlacionadas e nesse caso, os variogramas (diretos e cruzados) utilizados na SGS devem ser modelados conjuntamente através do LMC.

2.7 Considerações finais

Os métodos de estimativa/simulação multivariada aqui tratados são aplicados quando lidamos com conjuntos de dados multivariados heterotópicos e, nesse caso, com exceção dos modelos de Markov, faz-se necessário a modelagem e ajuste de variogramas diretos e cruzados via LMC. Sabe-se que esse processo é bastante laborioso e por vezes, não se pode pôr em prática. Como alternativa à modelagem clássica da continuidade espacial é sugerido no capítulo 3 uma metodologia tecnologicamente inovadora, que dispensa o uso de variogramas e seus ajustes na geração da matriz de covariância conjunta, necessária no sistema de cokrigagem.

3 Metodologia

Conforme apresentado, os métodos convencionais de estimativa multivariada necessitam do ajuste conjunto de variogramas (diretos e cruzados) via LMC. A modelagem conjunta tem por finalidade garantir que a matriz de covariância conjunta, gerada através dos variogramas ajustados, seja positiva definida, ou seja, quando utilizada no sistema de cokrigagem garanta a solução única do sistema.

Neste estudo, é proposto o uso de tabelas de covariância (diretas e cruzadas), obtidas pelo método de Kloeckner et al. (2019), para descrever a continuidade espacial e realizar estimativas/simulações do fenômeno multivariado, dispensando a modelagem tradicional de variogramas diretos e cruzados.

Como descrito na seção (2.4.3), a metodologia de Kloeckner et al. (2019) consiste no preenchimento do grid amostral através de um BMEC, na convolução BMEC via *FFT* para obter a densidade espectral e por fim, na retrotransformada da densidade espectral para gerar a tabela de covariância desejada. A geração de tabelas de covariância (diretas e cruzadas) através do método de (KLOECKNER et al., 2019), ocorre de modo automático e dispensa a imputação de dados por parte do usuário do programa. Entretanto, somente a tabela de covariância direta é corrigida. Apesar disso, o método de Kloeckner et al. (2019) representou grandes avanços no mapeamento automático da continuidade espacial e mostrou-se eficaz nas estimativas por krigagem e simulações.

A seguir, é apresentada a metodologia para utilização das tabelas de covariância em estimativas/simulações de conjuntos multivariados heterotópicos. As fases do método proposto consistem em obter os BMECs para as variáveis primárias e secundárias e em seguida, calcular as covariâncias das amostras via FFT. As covariâncias das amostras são calculadas sobre todas as direções definidas pelo grid do BMEC. O resultado é uma tabela de covariância preliminar. Essas etapas são uma adaptação do método de Kloeckner et al. (2019) para o caso de múltiplas variáveis. Então, corrigir essas tabelas preliminares de covariância para tabelas de covariância positivas definidas. A correção consiste em primeiro construir uma grande matriz de covariância, que contenha as covariâncias diretas e cruzadas para todas as distâncias e direções usadas na cokrigagem. Em segundo lugar, essa matriz de covariância é decomposta em produto de autovetores e autovalores. Então, todos os autovalores negativos são redefinidos para um número pequeno. Os autovetores originais e os autovalores corrigidos são usados para construir uma nova matriz de covariância que é positiva definida. Então, essa matriz de covariância é usada para construir uma nova tabela de covariância positiva definida.

3.1 Modelo base para extrair a tabela de covariânica

Para obtermos tabelas de covariância pelo método de Kloeckner et al. (2019) devese inicialmente construir os BMECs. O BMEC é semelhante ao conceito de imagem de treinamento usada pela geoestatística de múltiplos pontos (GUARDIANO; SRIVASTAVA, 1993). O objetivo deste primeiro passo é preparar os dados para a transformada de Fourier presente nas etapas subsequentes. A transformada de Fourier necessita de um BMEC em grid completamente estimado e regular construído a partir dos dados experimentais (KLOECKNER, 2018).

Nesta tese, usaremos a interpolação com valor do vizinho mais próximo para obter o BMEC. A estimativa por vizinho mais próximo é uma escolha razoável quando há uma grande quantidade de dados. No caso de poucos dados, a interpolação por vizinho mais próximo deve ser evitada, pois pode resultar em uma extrapolação excessiva dos valores estimados.

3.2 Efeito pepita

Uma limitação importante do BMEC obtido pela estimativa por vizinho mais próximo é que o efeito pepita não é incorporado à tabela de covariância. Como alternativa a essa limitação, sugere-se extrair o efeito pepita do variograma experimental omnidirecional e adicioná-lo à tabela de covariância, semelhante à soma das estruturas que compõem um modelo de variograma. Inferir o efeito pepita em um variograma omnidirecional ou *down the hole* é muito mais fácil do que modelar os variogramas via LMC.

3.3 Obtenção das tabelas de covariância

Na seção 2.4.3, foi apresentado o método gerador das tabelas de covariância (diretas e cruzadas), a partir do BMEC e transformadas de Fourier. O plug-in, representando o código do método, foi implementado por Kloeckner (2018) no AR2GeMS e consiste em executar a convolução via FFT dos valores presentes no grid amostral (devidamente preenchida pelo BMEC), resultando em uma tabela de covariância associada. A figura 3.1 exibe o plug-in e revela que ele é de fácil utilização e necessita de poucos parâmetros para determinar as tabelas de covariância (diretas e cruzadas).



Figura 3.1 – Interface do *plug-in* do software *AR2GeMS* para geração das tabelas de covariância (diretas e cruzadas). Figura adaptada de Kloeckner (2018).

O *plug-in* gera em poucos segundos um mapeamento rápido e automático da continuidade espacial, representando um ganho expressivo de tempo quando comparado com a variografia convencional.

3.4 Imputação do efeito pepita à tabela de covariância

As tabelas de covariância obtidas são combinações lineares do efeito pepita e das tabelas de covariância construídas anteriormente. Os pesos são determinados pela proporção do efeito pepita em relação a variância total. Por exemplo, se o efeito pepita for 20% da variância total, as novas tabelas de covariância serão iguais a 0,80 vezes as tabelas de covariância construídas anteriormente, além de uma constante que é 0,20 vezes o efeito pepita. Para explicar o efeito pepita, Kloeckner et al. (2019) sugeriram a criação de um BMEC com o *Ensemble Simulation*. O problema é que o *Ensemble Simulation* exige que os dados primários estejam densamente amostrados, pois o erro local é obtido pela validação cruzada de k-fold (BURMAN, 1989). Neste estudo, estamos interessados em analisar dados primários escassamente amostrados e dados secundários, densamente amostrados. Além disso, o ajuste para o efeito pepita realizado neste estudo é mais fácil de ser implementado do que construir um BMEC com o *Ensemble Simulation*.

3.5 Correção conjunta das tabelas de covariância

O procedimento descrito não gera matrizes de covariância, obtidas a partir das tabelas de covariâncias, positivas definidas. Para superar essa limitação e as condições do sistema de cokrigagem, sugere-se uma correção conjunta nas matrizes de covariância (diretas e cruzadas) para torná-las positivas definidas. A correção foi extraída de (DAVIS; CURRIERO, 2019) e consiste em escrever a matriz de covariância (*covariance matrix* - CM) conjunta em função de seus autovetores e autovalores, respeitada a condição de simetria. Como as tabelas de covariância cruzadas originais não são simétricas, optou-se em usar a média simples entre elas (SOARES, 2006). A matriz de covariância conjunta é então escrita como:

$$CM = \begin{bmatrix} C_{ZZ} & C_{ZY} \\ C_{YZ} & C_{YY} \end{bmatrix} \approx \begin{bmatrix} C_{ZZ} & C_m \\ C_m & C_{YY} \end{bmatrix} = V\Lambda V', \tag{3.1}$$

onde C_{ZZ} é a matriz de covariância direta da variável primária Z, C_{YY} é a matriz de covariância direta da variável secundária Y, C_{ZY} e C_{YZ} são as matrizes de covariância cruzadas entre as variáveis $Z \in Y$, C_m é a matriz de covariância média entre as matrizes de covariância cruzadas ($C_m = \frac{C_{ZY}+C_{YZ}}{2}$), V é a matriz quadrada dos autovetores v_i , V' é a transposta da matriz quadrada de autovetores v_i e Λ é a matriz diagonal dos autovalores ($\Lambda_{i,i} = \lambda_i$), ordenados de forma decrescente.

Para que CM seja positiva definida, é suficiente que os autovalores associados à matriz sejam maiores que zero, e na negativa, um valor pequeno ε seja escolhido. No conjunto dos autovalores positivos da matriz, o valor de ε será o menor dentre eles, de tal forma que:

$$\varepsilon = \frac{\lambda_1}{\tau},\tag{3.2}$$

onde τ é um valor de tolerância predefinido e λ_1 é o autovalor de maior valor positivo. A correção é aplicada a todos os autovalores, considerando o seguinte:

$$\begin{cases} \lambda_i, & \text{se } \lambda_i \ge \varepsilon \\ \varepsilon, & \text{se } \lambda_i < \varepsilon, \end{cases}$$
(3.3)

e uma nova matriz de autovalores ($\tilde{\Lambda}_{j,j}$) é usada para reescrever a matriz de covariância conjunta:

$$\widetilde{CM} = V\widetilde{\Lambda}V', \tag{3.4}$$

resultando em uma aproximação da matriz de covariância conjunta, que é positiva definida. Para finalizar o método, incorporam-se os valores de covariância corrigidos às respectivas tabelas de covariância originais. As tabelas de covariância corrigidas são positivas definidas.

A correção realizada considerou apenas duas variáveis $(Z \in Y)$, mas pode ser estendida, sem perda de generalidade, para bancos de dados com três ou mais variáveis.

Ainda em relação a correção, o executável FIXCORRMAT.EXE (*software* GSLIB) é uma alternativa válida na correção de matriz de covariância. Assim como o método utilizado, o programa corrigi os autovalores negativos associados a matriz de covariância a

partir de um único parâmetro subjetivo, o valor mínimo do autovalor que será considerado. Se o autovalor escolhido for muito baixo, os valores corrigidos de covariância não serão confiáveis e se definido muito alto, a correção da matriz levará muito tempo para se processar, apesar de se ter uma boa matriz corrigida. Outras informações sobre o programa podem ser consultadas em Kumar e Deutsch (2009).

3.6 Fluxo de trabalho da metodologia

A figura 3.2 representa o fluxo de trabalho da metodologia proposta, iniciando com a geração de tabelas de covariância através do método de Kloeckner et al. (2019) até a geração de tabelas de covariância positivas definidas.



Figura 3.2 – Fluxo de trabalho da metodologia. Imagem ilustrativa.

4 Uso de dados secundários heterotópicos em estimativa/simulação com tabelas de cova-riância

Este capítulo retrata um estudo de caso de estimativa/simulação aplicada a um banco de dados multivariado, onde as estimativas/simulações serão realizadas através de tabelas de covariância (diretas e cruzadas) em substituição a variografia tradicional, seguindo a metodologia descrita no capítulo 3.

A primeira parte da seção do estudo de caso será dedicada à estimativa, enquanto a segunda parte será dedicada à simulação com dados secundários heterotópicos.

4.1 Apresentação do banco de dados

O conjunto de dados é sintético e contém duas variáveis: primária (Z) e secundária (Y). A variável primária contém amostras separadas de 20 x 20 metros, enquanto a secundária, apresenta espaçamento amostral de 5 x 5 metros ao longo das direções X e Y (Figura 4.1). Os histogramas mostram que as distribuições são assimétricas positivas (ver Figura 4.2), o gráfico de dispersão indica que as variáveis estão correlacionadas ($\rho = 0, 62$) (ver Figura 4.3) e a tabela 4.1 mostra a estatística básica dos dados, revelando que não há valores extremos que possam influenciar nos resultados(CV < 1).



Figura 4.1 – Mapa de localização das amostras: (a) variável primária e (b) variável secundária.



Figura 4.2 – Histograma da variável primária (a) e histograma da variável secundária (b).



Figura 4.3 – Gráfico de dispersão entre a variável primária (ppm) e variável secundária (ppm).

Tabela 4.1 – Estatística básica dos dados.

Variável	nº de amostras	Média	Variância	CV	Mínimo	Máximo
Primária	195	2,72	5,92	0,89	0	10,13
Secundária	3120	3,43	9,73	0,90	0	18,92

4.2 Estimativa

A metodologia usada para estimar é a cokrigagem ordinária estandardizada (GO-OVAERTS, 1997). Neste estudo, comparam-se duas estimativas feitas por cokrigagem ordinária estandardizada: uma utiliza o modelo linear de corregionalização para caracterizar a continuidade espacial, enquanto a outra utiliza tabelas de covariância obtidas a partir dos dados.

O LMC ajustou os variogramas experimentais (diretos e cruzados). As equações 4.1, 4.2 e 4.3 descrevem os modelos da variável primária (Z), da secundária (Y) e o variograma cruzado.

$$\gamma_Z(h) = 1, 1+2, 2Sph\left(\frac{N157^{\circ}E}{46}, \frac{N67^{\circ}E}{40}\right) + 2, 62Sph\left(\frac{N157^{\circ}E}{92}, \frac{N67^{\circ}E}{50}\right).$$
(4.1)

$$\gamma_Y(h) = 1, 5 + 3, 0Sph\left(\frac{N157^{\circ}E}{46}, \frac{N67^{\circ}E}{40}\right) + 5, 23Sph\left(\frac{N157^{\circ}E}{92}, \frac{N67^{\circ}E}{50}\right).$$
(4.2)

$$\gamma_{ZY}(h) = 0,61+1,03Sph\left(\frac{N157^{\circ}E}{46},\frac{N67^{\circ}E}{40}\right) + 3,06Sph\left(\frac{N157^{\circ}E}{92},\frac{N67^{\circ}E}{50}\right).$$
 (4.3)

Para estimar via tabelas de covariância, estimou-se, inicialmente, os modelos base para extrair a covariância (BMECs). Utilizou-se o vizinho mais próximo (NN) (veja figura 4.4) devido à sua simplicidade e a grande quantidade de dados. O estudo de caso ilustra um caso típico de estimativa de teor para planejamento de mina a curto prazo, uma vez que os dados secundários são densamente amostrados. Quando há poucas amostras, o vizinho mais próximo pode gerar um BMEC pobre em termos de capacidade de extrair a continuidade espacial.



Figura 4.4 – Modelos para extrair as tabelas de covariância: (a) BMEC da variável primária e (b) BMEC da variável secundária.

Os BMECs foram usados para se construir as tabelas de covariância (Figura 4.5), empregando a metodologia de Kloeckner et al. (2019). As tabelas de covariância cruzadas $ZY \ e \ YZ$, por não serem simétricas, foram substituídas pela média simples entre elas. O uso das covariâncias cruzadas simétricas facilita o processo de correção. Se a matriz de covariância é simétrica, a matriz dos autovetores é ortogonal, isto é, a matriz inversa é igual à sua transposta.





Figura 4.5 – Tabelas de covariância extraídas dos BMECs: (a) Tabela de covariância da variável primária, (b) Tabela de covariância da variável secundária e (c) Tabela de covariância média entre as tabelas de covariância cruzadas.

Cada tabela de covariância foi usada para construir uma grande matriz de covariância com 10.000 elementos, o que corresponde aos lags de separações entre as distâncias de -50 a 50 ao longo da direção X combinadas com as distâncias de -50 a 50 ao longo da direção Y. A correção se aplica apenas à parte central da tabela de covariância, que tem o maior impacto na estimativa e simulação (as distâncias estão dentro da vizinhança de busca). Esta parte central é mostrada na Figura 4.6. Idealmente, a matriz de covariância deve conter todas as distâncias e separações até o intervalo esperado. No entanto, o tempo de processamento para essa correção aumenta substancialmente conforme aumenta o tamanho da matriz de covariância. Os autovalores e autovetores da matriz de covariância foram calculados.



Figura 4.6 – Parte da tabela de covariâncias que deve ser corrigida: (a) parte central da tabela de covariância da variável primária, (b) parte central da tabela de covariância da variável secundária e (c) parte central da tabela de covariância média entre tabelas de covariância cruzadas.

Os autovalores negativos obtidos mostraram que as tabelas de covariância, geradas pelo método de Kloeckner et al. (2019), não são positivas definidas e, nesse caso, a correção

é obrigatória. A correção foi realizada através da decomposição da matriz de covariância conjunta (com 40.000 elementos) em função de seus autovalores e autovetores (vide equação 4.4), considerando uma tolerância $\tau = 10^4$.

$$\begin{bmatrix} C_{ZZ} & C_m \\ C_m & C_{YY} \end{bmatrix} = V\Lambda V'.$$
(4.4)

Após correções e incorporação dos valores corrigidos, o efeito pepita foi imputado à tabela de covariância. Neste caso, usamos o variograma experimental omnidirecional. Para um estudo de caso 3D, o efeito pepita pode ser rapidamente avaliado pela verificação do variograma vertical (ou *down the hole*). O variograma vertical é geralmente bem comportado, uma vez que ao longo desta direção há muitas amostras próximas dentro do espaçamento. Inferir o efeito pepita individualmente é mais fácil do que modelar o LMC. Além disso, permite ao geomodelador colocar conhecimento especializado no processo de modelagem.

Os variogramas gerados através das tabelas de covariância corrigidas CT (Variograma (BMEC)_COR) foram comparados com os obtidos pelas tabelas de covariância originais (Variograma CT(BMEC)) e com os ajustados pelo LMC (Variograma modelado a partir dos dados) (Figuras 4.7, 4.8 e 4.9), revelando que eles têm forma equivalente. Além de forma equivalente, os valores de $\gamma(h)$ (ver Tabela 4.2) gerados através da tabela de covariância corrigida são próximos aos valores de $\gamma(h)$ gerados via tabela de covariância original, apesar do erro elevado para os primeiros *lags*, influenciado pelo acréscimo do efeito pepita à tabela de covariância corrigida.



Figura 4.7 – Variogramas diretos nas principais direções da variável primária. Direção de maior continuidade: $N157^{\circ}E$ (a); direção de menor continuidade: $N67^{\circ}E$ (b).



Figura 4.8 – Variogramas diretos nas principais direções da variável secundária. Direção de maior continuidade: $N157^{\circ}E$ (a); direção de menor continuidade: $N67^{\circ}E$ (b).



Figura 4.9 – Variogramas cruzados nas principais direções. Direção de maior continuidade: $N157^{\circ}E$ (a); direção de menor continuidade: $N67^{\circ}E$ (b).

	Erro	Relativo	1000%	168%	67%	36%	22%	15%	12%	10%	9%	8%	7%	6%	5%	5%	4%	4%	3%
Variogramas cruzados		Absoluto	0.6	0.57	0.51	0.45	0.39	0.33	0.3	0.26	0.24	0.23	0.22	0.2	0.18	0.16	0.14	0.13	0.13
	CT(BMEC)_COR	$\lambda(h)$	0.66	0.91	1.27	1.7	2.13	2.47	2.74	2.94	3.04	3.13	3.24	3.36	3.51	3.63	3.75	3.83	3.85
	CT(BMEC)	$\lambda(h)$	0.06	0.34	0.76	1.25	1.74	2.14	2.44	2.68	2.8	2.9	3.02	3.16	3.33	3.47	3.61	3.7	3.72
Idária	0	Relativo	15%	6%	6%	5%	3%	3%	3%	2%	2%	2%	2%	2%	2%	2%	2%	1%	1%
	Em	Absoluto	0.51	0.41	0.34	0.3	0.23	0.21	0.19	0.17	0.16	0.15	0.15	0.14	0.15	0.14	0.14	0.12	0.11
Variogramas da secu	CT(BMEC)_COR	$\lambda(h)$	3.99	4.96	5.65	6.39	7.02	7.36	7.61	7.78	7.87	7.94	8.02	8.07	8.11	8.17	8.22	8.43	8.55
	CT(BMEC)	$\lambda(h)$	3.48	4.55	5.31	6.09	6.79	7.15	7.42	7.61	7.71	7.79	7.87	7.93	7.96	8.03	8.08	8.31	8.44
iária	Erro	Relativo	29%	21%	16%	12%	9%	7%	6%	5%	5%	5%	5%	5%	5%	4%	3%	3%	2%
		Absoluto	0.5	0.49	0.48	0.44	0.36	0.31	0.27	0.23	0.23	0.23	0.23	0.23	0.22	0.19	0.16	0.13	0.11
Variogramas da prin	CT(BMEC)_COR	$\lambda(h)$	2.2	2.77	3.47	3.98	4.34	4.56	4.75	4.87	4.91	4.89	4.87	4.89	4.96	5.1	5.22	5.33	5.42
	CT(BMEC)	$\lambda(h)$	1.7	2.28	2.99	3.54	3.98	4.25	4.48	4.64	4.68	4.66	4.64	4.66	4.74	4.91	5.06	5.2	5.31
Lag	<u> </u>	<u> </u>	5.38	10.77	16.15	21.54	26.92	32.31	37.69	43.08	48.46	53.85	59.23	64.62	70.01	75.39	80.77	86.16	91.54

ão.
ireç
l d
cipa
prin
na
ncia
ariâ
COV
de
elas
tabc
via
los
gera
las
ran
riog
Va
$\overline{2}$
4.
bela
Tal

Após a obtenção das covariâncias, as estimativas foram realizadas por cokrigagem ordinária estandardizada, alternando o modelo de continuidade espacial entre variogramas ajustados e tabelas de covariância corrigidas. As estimativas ocorreram no suporte de ponto (grid de 1 x 1 metros). Em seguida, as estimativas foram ampliadas para o suporte de bloco de 5 x 5 metros ao longo das direções X e Y. As estimativas de bloco são iguais à média das estimativas de ponto dentro do volume do bloco. A estimativa usando tabelas de covariância requer um grid com a mesma resolução da tabela de covariância requerem a construção direta das estimativas de bloco a bloco e ponto a bloco. Uma solução mais simples é executar as estimativas no suporte de ponto e fazer um upscale para o suporte de bloco final, como demonstrado.

4.2.1 Resultados e discussões

O gráfico de dispersão entre os dois métodos de estimativa, Figura 4.10, revela uma forte correlação entre eles ($\rho = 0, 98$) e um espalhamento ao redor da linha de 45° (em vermelho), ou seja, o gráfico de dispersão mostra que as estimativas realizadas através de tabelas de covariância corrigidas assemelham-se àquelas obtidas através de variogramas ajustados. Fato também confirmado através da análise visual, Figura 4.11, onde é possível perceber a existência de similaridade entre os modelos estimados, seja nas regiões de alto teor quanto nas regiões de baixo teor.



Figura 4.10 – Gráfico de dispersão entre as estimativas através de variogramas ajustados (SOCK_CL) e através de tabelas de covariância corrigidas (SOCK _CT(BMEC) _COR).



Figura 4.11 – Comparação visual entre os modelos de estimativa: (a) estimativa através de variogramas ajustados (SOCK_CL) e (b) através de tabelas de covariância corrigidas (SOCK_CT(BMEC) _COR).

Os histogramas associados às diferentes estimativas e aos dados desagrupados são apresentados na Figura 4.12 e mostram que a média global foi reproduzida. Eles também mostram que as estimativas por meio de tabelas de covariância são menos suavizadas do que as estimativas por meio de modelos de variogramas ajustados (na Figura 4.12, a variância das estimativas com variogramas é menor do que a variância das estimativas obtidas com tabelas de covariância). Esse resultado é devido à estimativa/simulação com tabelas de covariância usarem os dados que foram migrados para o nó mais próximo do grid. Essa migração é necessária porque a tabela de covariância contém valores de covariância somente para as distâncias definidas pelo grid. O resultado dessa operação é que os valores dos dados são copiados para o nó do grid mais próximo. Em outras palavras, a estimativa de um nó do grid que está mais próximo de uma referência com o valor de zero é exatamente zero quando as tabelas de covariância são usadas. No caso de estimativa usando o LMC, a estimativa é uma combinação linear (média local) dos dados circundantes dentro da vizinhança de pesquisa.


Figura 4.12 – Histogramas associados ao modelo de referência da variável primária (a), a estimativa via variogramas ajustados (b) e a estimativa através de tabelas de covariância corrigidas (c).

A análise de deriva (Figuras 4.13 e 4.14) mostra que há aderência entre as estimativas e os dados desagrupados (NN). A aderência ocorre em ambas as direções (leste e norte) e indicam que as estimativas estão reproduzindo as médias locais.



Figura 4.13 – Análise de deriva entre o modelo de referência e estimados ao longo da direção leste (X): modelo de referência (NN), estimativa via variogramas ajustados (SOCK_CL) e estimativa usando tabelas de covariância corrigidas (SOCK_CT(BMEC)_COR).



Figura 4.14 – Análise de deriva entre o modelo de referência e estimados ao longo da direção norte (Y): modelo de referência (NN), estimativa via variogramas ajustados (SOCK_CL) e estimativa usando tabelas de covariância corrigidas (SOCK_CT(BMEC)_COR).

4.3 Simulação com dados secundários

As simulações foram realizadas através da simulação sequencial Gaussiana, que permite simular a variável primária condicionada as informações desta e da variável secundária. Para realizar as simulações, inicialmente os dados originais de cada variável foram transformados em distribuições gaussianas padrão, com média zero e variância unitária. Em seguida, os variogramas (diretos e cruzado) foram calculados e os modelos ajustados, sob as restrições do LMC. As equações 4.5, 4.6 e 4.7 descrevem os modelos de variograma dos dados transformados da variável primária (Z), secundária (Y) e o variograma cruzado:

$$\gamma_Z(h) = 0, 10 + 0, 45Sph\left(\frac{N157^{\circ}E}{46}, \frac{N67^{\circ}E}{40}\right) + 0, 45Sph\left(\frac{N157^{\circ}E}{92}, \frac{N67^{\circ}E}{50}\right).$$
(4.5)

$$\gamma_Y(h) = 0, 10 + 0, 40Sph\left(\frac{N157^{\circ}E}{46}, \frac{N67^{\circ}E}{40}\right) + 0, 5Sph\left(\frac{N157^{\circ}E}{92}, \frac{N67^{\circ}E}{50}\right).$$
(4.6)

$$\gamma_{ZY}(h) = 0,08 + 0,14Sph\left(\frac{N157^{\circ}E}{46},\frac{N67^{\circ}E}{40}\right) + 0,38Sph\left(\frac{N157^{\circ}E}{92},\frac{N67^{\circ}E}{50}\right).$$
 (4.7)

As tabelas de covariância (diretas e cruzada) dos valores transformados foram geradas e corrigidas de maneira similar àquelas dos dados originais, ou seja, correção para torná-las positivas definidas e imputação do efeito pepita. As etapas do processo consistiram em (i) utilizar as estimativas NN como modelo base para extrair a covariância (BMEC_NS), (ii) usar os BMECs para calcular as tabelas de covariância (diretas e cruzada), empregando o método de Kloeckner et al. (2019), (iii) construir uma grande matriz de covariância conjunta que contenha as covariâncias diretas e cruzada associada às respectivas tabelas de covariância, (iv) corrigir esta matriz de covariância conjunta e (v) incorporar os valores de covariância corrigidos de volta às tabelas de covariância iniciais.

As Figuras 4.15, 4.16 e 4.17 mostram os variogramas gerados a partir das tabelas de covariância dos dados transformados antes (linha vermelha nas Figuras 4.15, 4.16 e 4.17) e depois da correção (linha azul nas Figuras 4.15, 4.16 e 4.17). Os dois variogramas são semelhantes e, portanto, a correção não perturba a continuidade espacial das variáveis.



Figura 4.15 – Variogramas diretos dos dados transformados (normal score) nas principais direções da variável primária. Direção de maior continuidade: $N157^{\circ}E$ (a); direção de menor continuidade: $N67^{\circ}E$ (b).



Figura 4.16 – Variogramas diretos dos dados transformados (normal score) nas principais direções da variável secundária. Direção de maior continuidade: $N157^{\circ}E$ (a); direção de menor continuidade: $N67^{\circ}E$ (b).



Figura 4.17 – Variogramas cruzados dos dados transformados (normal score) nas principais direções. Direção de maior continuidade: $N157^{\circ}E$ (a); direção de menor continuidade: $N67^{\circ}E$ (b).

4.3.1 Resultados e discussões

As simulações foram geradas com os mesmos parâmetros, exceto para o modelo de covariância utilizado (um conjunto de realizações usou o LMC e o outro, tabelas de covariância). Cinquenta realizações foram geradas. A Figura 4.18 mostra a primeira realizaçõe construída com o LMC e com tabelas de covariância, revelando que não há diferença significativa entre elas. A similaridade entre as diferentes realizações também se confirma através do gráfico de dispersão (Figura 4.19), indicando forte correlação entre eles ($\rho = 0, 98$) e espalhamento em torno da reta de 45° (em vermelho).



Figura 4.18 – Simulações: realizações via variograma (a) e tabelas de covariância (b).



Figura 4.19 – Gráfico de dispersão entre uma realização via tabela de covariância e uma realização via variograma ajustado.

Os histogramas e variogramas dos modelos simulados foram comparados com os histogramas e variogramas dos dados. As figuras 4.20 e 4.21 mostram que as simulações com tabelas de covariância (diretas e cruzada) reproduzem tanto os histogramas quanto os variogramas dos dados.



Figura 4.20 – Histograma acumulado dos modelos simulados (Simulações) e histograma acumulado do BMEC da variável primária (NN).



Figura 4.21 – Variogramas das simulações via tabelas de covariância orientados ao longo das direções de maior (a) e menor (b) continuidade espacial e variograma.

Outra validação consistiu em comparar a média local das realizações com a média local dos dados desagrupados (obtidos pelo método do polígono de influência) usando a análise de deriva. A análise de deriva (Figuras 4.22 e 4.23) mostra que as médias locais das realizações flutuam em torno das médias locais dos dados agrupados. Esse resultado indica que, em média, as realizações reproduzem a média local dos dados. As Figuras 4.22 e 4.23 também indicam que as médias locais das realizações obtidas com o LMC são semelhantes às das realizações obtidas com tabelas de covariância.



Figura 4.22 – Análise de deriva entre o modelo de referência (NN) e simulado ao longo da direção leste (X): simulação usando modelo de variograma (a) e simulação usando tabelas de covariância corrigidas (b).



Figura 4.23 – Análise de deriva entre o modelo de referência (NN) e simulado ao longo da direção norte (Y): simulação usando modelo de variograma (a) e simulação usando tabelas de covariância corrigidas (b).

4.4 Conclusão

Neste estudo, foi apresentado uma metodologia para a estimativa/simulação de variáveis com dados secundários heterotópicos e tabelas de covariância. As tabelas de covariância (diretas e cruzadas) foram inicialmente obtidas através de uma transformada do BMEC e Fourier. Essas tabelas iniciais de covariância foram ajustadas para considerar o efeito pepita, inferido através do variograma experimental. Em seguida, as tabelas de covariância foram corrigidas em conjunto para serem positivas definidas. As tabelas de covariância corrigidas podem ser usadas diretamente em algoritmos geoestatísticos.

O estudo de caso mostrou que as estimativas obtidas com a metodologia apresentada são semelhantes às obtidas com o modelo linear de corregionalização (LMC). Além disso, as simulações obtidas com as tabelas de covariância apresentaram reprodução satisfatória de histogramas e variogramas. Estudos futuros devem se concentrar na aplicação da metodologia em estudos de caso com dados reais. A grande vantagem do método é que o geomodelador não precisa ajustar o LMC. Embora o ajuste automático do LMC seja uma alternativa válida, o profissional ainda precisa ajustar cuidadosamente os parâmetros (tamanho do lag, azimutes de direção e tolerâncias) dos variogramas experimentais. Por outro lado, a metodologia proposta modela a continuidade espacial usando tabelas de covariância obtidas a partir do modelo base para extrair a covariância (BMEC). Na presença de dados densamente amostrados, o BMEC pode ser facilmente obtido a partir de uma estimativa por vizinho mais próximo. Assim, o método proposto é mais prático do que o ajuste automático do LMC.

5 Aplicação da tabela de covariância para inferir a continuidade espacial da primária no caso de uma variável secundária exaustiva

Neste capítulo, é apresentado um estudo de caso de cokrigagem colocada em um banco de dados da indústria petrolífera, analisando a aplicabilidade da tabela de covariância da variável secundária para extrair os padrões de continuidade espacial e usá-lo na estimativa/simulação da variável primária. A ideia é, devido à dificuldade em se ajustar o variograma da variável primária, fazer uso da tabela de covariância da variável secundária que está exaustivamente amostrada para estimar/simular a variável primária.

5.1 Análise e apresentação do banco de dados

O banco de dados contém amostras de duas variáveis amostradas em reservatórios de óleo e gás: primária (NCB – Net Continuous Bitumen Thickness) e secundária (Seis Thick). A variável NCB representa a espessura da camada de interesse e foi obtida através de perfis de poços. A variável Seis Thick também representa a espessura da camada de interesse, mas foi obtida a partir da interpretação de dados de sísmica. O banco de dados não fornece as unidades dos valores dessas variáveis. As amostras da primária estão disponíveis em 50 locais e situadas dentro de um domínio de 10 000 km^2 , enquanto que a variável secundária é exaustivamente amostrada (40 000 locais) e encontra-se no mesmo domínio da primária. Os mapas amostrais dessas variáveis podem ser vistos na figura 5.1 e os histogramas dos dados, na figura 5.2. Os histogramas indicam que as distribuições são assimétricas positivas, ou seja, com alta concentração de valores baixos.

Capítulo 5. Aplicação da tabela de covariância para inferir a continuidade espacial da primária no caso 80 de uma variável secundária exaustiva



Figura 5.1 – Mapa de localização das amostras: (a) variável primária e (b) variável secundária exaustiva.



Figura 5.2 – Histograma da variável primária (a) e histograma da variável secundária (b).

A figura 5.3 mostra o mapa de localização do *collar* dos poços e o histograma do espaçamento amostral, respectivamente. Essa figura mostra que o espaçamento amostral, calculado através da distância média entre os dados, é irregular e varia de mil até trinta mil metros, aproximadamente.



Figura 5.3 – Espaçamento dos dados da variável primária (a) e histograma associado (b).

O gráfico de dispersão entre as duas variáveis e a tabela contendo a estatística descritiva delas são apresentados, respectivamente, na figura 5.4 e na tabela 5.1. As variáveis estão correlacionadas ($\rho = 0, 8$) e não há valores extremos evidentes que possam influenciar na estimativa (CV < 1). Como a correlação é alta para h = 0, se assume que a continuidade espacial da variável secundária é similar a continuidade espacial da variável primária (continuam correlacionadas para $h \neq 0$).



Figura 5.4 – Gráfico de dispersão entre os valores colocados da variável primária e secundária.

Tab	ela	5	1 -		Esta	tíst	ica	de	scr	iti	va	d	\mathbf{OS}	d٤	ade	ЭS	•
-----	-----	---	-----	--	------	------	-----	----	-----	-----	----	---	---------------	----	-----	----	---

Variável	n° de amostras	Média	Desvio Padrão	CV	Mínimo	Máximo
Primária	50	16,73	8,56	0,51	0	44,06
Secundária	40000	$0,\!83$	0,56	$0,\!68$	0	$3,\!35$

5.2 Análise da continuidade espacial da variável primária

Na cokrigagem colocada da variável primária, apenas o variograma da variável primária precisa ser modelado e ainda, o variograma cruzado entre as variáveis é proporcional ao variograma da variável primária. Mas a escassez e o espaçamento irregular dos dados da variável primária causam dúvidas na anisotropia, impedindo um bom ajuste de seu variograma.

Usando 20 espaçamentos com tamanho de 2500 m e tolerância de 1250 m, calculou-se 9 variogramas experimentais para a variável primária. A figura 5.5 apresenta os variogramas experimentais obtidos para as 9 direções analisadas (Omnidirecional, $N0^{\circ}E$, N22, $5^{\circ}E$, $N45^{\circ}E$, N67, $5^{\circ}E$, $N90^{\circ}E$, N112, $5^{\circ}E$, $N135^{\circ}E$ e N157, $5^{\circ}E$). Com exceção do variograma experimental omnidirecional, os demais ficaram ruidosos e não apresentaram clara estruturação espacial. Apesar do variograma omnidirecional ser o mais bem definido, ele não indica a anisotropia da variável, mas permite inferir o efeito pepita de 0,20 (veja linha Capítulo 5. Aplicação da tabela de covariância para inferir a continuidade espacial da primária no caso 82 de uma variável secundária exaustiva

vermelha no variograma omnidirecional). O efeito pepita, gerado através do variograma omnidirecional, será posteriormente incorporado à tabela de covariância (OLIVEIRA et al., 2019).



Figura 5.5 – Variogramas experimentais da variável primária nas direções: Omni, $N0^{\circ}E$, $N22, 5^{\circ}E, N45^{\circ}E, N67, 5^{\circ}E, N90^{\circ}E, N112, 5^{\circ}E, N135^{\circ}E$ e $N157, 5^{\circ}E$.

Capítulo 5. Aplicação da tabela de covariância para inferir a continuidade espacial da primária no caso 84 de uma variável secundária exaustiva

Também se verificou a possibilidade de utilizar a tabela de covariância da variável primária em substituição ao ajuste do variograma. Para tanto, construiu-se um BMEC (usou-se uma estimativa feita pelo método do vizinho mais próximo, *Nearest Neighbor*-NN) da variável primária e verificou-se a correlação deste com a variável secundária exaustiva. A figura 5.6 representa o BMEC da primária e a figura 5.7, o gráfico de dispersão entre o BMEC da primária e a secundária exaustiva. O gráfico de dispersão indica que a correlação entre as variáveis passou de 0,8 para 0,53, ou seja, teve uma queda de 33,75% no valor da correlação.

Como a correlação é essencial na cokrigagem colocada da variável primária, a metodologia empregada na geração da tabela de covariância da variável primária apresentada no estudo de caso do capítulo 4 é inadequada para esse estudo de caso. A metodologia aplicada no estudo de caso do capítulo 4, que gera a tabela de covariância a partir de um BMEC construído por NN, não serve quando o banco de dados apresenta poucas amostras. Nesse caso, o BMEC distorce tanto o efeito pepita quanto a correlação entre as variáveis.



Figura 5.6 – BMEC da variável primária. Dimensões em km.



Figura 5.7 – Gráfico de dispersão entre os valores reais da variável primária e os valores colocados da variável secundária (esquerda) e gráfico de dispersão entre os valores nos nós gerados com o BMEC da primária e a secundária exaustiva (direita).

5.3 Análise da continuidade espacial da variável secundária

Pelo fato de as variáveis apresentarem uma alta correlação entre si (coeficiente de correlação de 0,8), assume-se que elas apresentam continuidade espacial parecidas e, a partir dessa informação, será inferida a direção de anisotropia da primária a partir da informação secundária exaustiva. Mas pensar a direção da primária a partir dos dados secundários que estão em *grid*, demanda usar vetores no cálculo dos variogramas experimentais. Uma alternativa que dispensa o uso de vetores e gera valores de covariância para todas as direções e distâncias possíveis é a tabela de covariância. Nesse caso, os dados exaustivos substituirão o BMEC e com isso, teremos uma tabela de covariância mais fidedigna (se a continuidade da primária seguir a secundária) e sem efeitos de suavização.

O mapa de covariância da variável secundária e sua parte central constam na figura 5.8. A parte central tem uma estrutura de curto alcance isotrópica e uma segunda estrutura de longo alcance com maior continuidade na direção de azimute 135°. Essas duas estruturas não foram observadas nos variogramas experimentais da primária (Figura 5.5). As direções principais, definidas pela tabela de covariância e ainda, o efeito pepita da variável primária (0,2), foram utilizados no ajuste do variograma da variável secundária (vide equação 5.1). A tabela de covariância e o variograma da variável secundária substituirão a continuidade espacial da primária nas estimativas e simulações com dados colocados.



Figura 5.8 – Mapa de covariância extraído da variável secundária exaustiva (a) e sua parte central (b). Dimensões em km.

$$\gamma(h) = 0, 2 + 0, 5Sph\left(\frac{N135^{\circ}E}{92000}, \frac{N45^{\circ}E}{40000}\right) + 0, 3Sph\left(\frac{N135^{\circ}E}{7500}, \frac{N45^{\circ}E}{7500}\right).$$
(5.1)

Foram construídas as tabelas de covariância da variável secundária, tanto dos dados estandardizados quanto dos dados transformados (normal *score*). A estandardização dos dados foi realizada para garantir que a variância da variável secundária seja equivalente a variância dos dados estandardizados da variável primária e a normalização, atende uma das condições da simulação sequencial Gaussiana. A parte central dessas tabelas, com

alcance de 20000 m, gerou matrizes de covariância com autovalores negativos e, nesse caso, a correção para torná-las positivas definidas é necessária. A correção foi realizada seguindo a metodologia proposta, incluindo a imputação do efeito pepita (usou-se o efeito pepita da variável primária obtido a partir do variograma experimental omnidirecional). As figuras 5.9 e 5.10 mostram que a correção realizada nas tabelas de covariância não perturbou a continuidade espacial da variável secundária.



Figura 5.9 – Variogramas obtidos através da tabela de covariância da variável secundária (dados estandardizados). Direção de maior continuidade: $N135^{\circ}E$ (a); direção de menor continuidade: $N45^{\circ}E$ (b).



Figura 5.10 – Variogramas obtidos através da tabela de covariância da variável secundária (dados transformados). Direção de maior continuidade: $N135^{\circ}E$ (a); direção de menor continuidade: $N45^{\circ}E$.

5.4 Análise e discussão dos resultados da estimativa com dados colocados

A estimativa foi realizada por meio da cokrigagem colocada (cokrigagem ordinária estandardizada) da variável primária. A correlação espacial foi obtida através da tabela de covariância e do variograma da variável secundária acrescidos do efeito pepita da primária (0,20). Utilizou-se o modelo de Markov I (ALMEIDA, 1993) para obter as covariâncias cruzadas. Os mapas de localização e o gráfico de dispersão entre os diferentes métodos de estimativa são apresentados nas figuras 5.11 e 5.12, respectivamente. Essas figuras indicam que as estimativas realizadas via tabela de covariância corrigida assemelham-se àquelas obtidas através de variograma ajustado devido à forte correlação, ao baixo espalhamento em torno da reta de 45° e à similaridade dos mapas de teores.



Figura 5.11 – Mapa de localização das estimativas: (a) cokrigagem colocada via tabela de covariância corrigida e (b) cokrigagem colocada através de variograma ajustado .Dimensões em km.



Figura 5.12 – Gráfico de dispersão entre as estimativas via variograma ajustado e via tabela de covariância corrigida.

Os histogramas e a estatística descritiva associados às estimativas constam, respectivamente, na Figura 5.13 e na tabela 5.2. A Figura 5.13 e a tabela 5.2 mostram que as estimativas reproduziram a média global. Além disso, as estimativas não provocaram a extrapolação dos valores de mínimo e nem de máximo.

Capítulo 5. Aplicação da tabela de covariância para inferir a continuidade espacial da primária no caso 88 de uma variável secundária exaustiva



Figura 5.13 – Histogramas associados às estimativas por cokrigagem colocada via tabela de covariância corrigida (esquerda) e via variograma (direita).

Tabela 5.2 – Estatística descritiva das estimativas por cokrigagem colocada.

Variável	Média	Desvio-Padrão	Mínimo	Máximo
Primária	16,73	8,56	0	44,06
Cokrigagem colocada via tabela de covariância	16,07	7,07	0,68	43,39
Cokrigagem colocada via variograma	16,09	6,78	1,02	41,61

A análise de deriva (Figuras 5.14 e 5.15) foi usada para validar a reprodução da média local das estimativas, utilizando o desagrupamento da variável primária (NN) como referência. Nota-se que há aderência entre as estimativas e o modelo de referência nas direções analisadas (leste e norte), ou seja, as estimativas apresentam tendência análoga ao modelo de referência e nesse caso, reproduzem a média local.



Figura 5.14 – Análise de deriva entre o modelo de referência e estimados ao longo da direção leste (X): modelo de referência (NN) e estimado por cokrigagem colocada via variograma e via tabela de covariância corrigida.



Figura 5.15 – Análise de deriva entre o modelo de referência e estimados ao longo da direção norte (Y): modelo de referência (NN) e estimado por cokrigagem colocada via variograma e via tabela de covariância corrigida.

5.5 Análise e discussão dos resultados da simulação com dados colocados

Na simulação, empregou-se a simulação sequencial Gaussiana da variável primária, utilizando a secundária exaustiva, a cokrigagem colocada (cokrigagem simples) e o modelo de Markov I. Nas simulações, usou-se a tabela de covariância e o variograma dos dados transformados da variável secundária (equação 5.1) acrescidos do efeito pepita (0,20), semelhante ao realizado na estimativa que consta na seção 5.4. Cinquenta realizações foram geradas para cada método. A Figura 5.16 mostra que as realizações honram as amostras da variável primária.

Capítulo 5. Aplicação da tabela de covariância para inferir a continuidade espacial da primária no caso 90 de uma variável secundária exaustiva



Figura 5.16 – Realizações da simulação sequencial Gaussiana colocada da variável primária e mapa de localização das amostras da variável primária: (a) realização via tabela de covariância corrigida e (b) realização via variograma. Dimensões em km.

Os histogramas e variogramas dos modelos simulados foram comparados com os histogramas e variogramas dos dados. As figuras 5.17 e 5.18 mostram que as simulações com o uso da tabela de covariância da secundária para simular a primária reproduzem os histogramas e variogramas dos dados, respectivamente.



Figura 5.17 – Histograma acumulado dos modelos simulados (simulações) e histograma acumulado do modelo de referência da variável primária (NN).



Figura 5.18 – Variogramas das simulações via tabela de covariância orientados ao longo das direções de maior (a) e menor (b) continuidade espacial e variograma modelado a partir dos dados secundários.

5.6 Conclusão

Neste estudo, foi proposto um método para a estimativa/simulação com dados primários escassos e dados secundários exaustivos com o uso da tabela de covariância inferida a partir dos dados secundários. Inicialmente buscou-se inferir a continuidade espacial da variável primária através do variograma experimental e, devido ao espaçamento irregular e escassez dos dados, isso não foi possível. A metodologia empregada no capítulo 4 para gerar tabela de covariância através do BMEC da primária também foi investigada, mas usar o NN como BMEC da primária corrompeu a correlação entre as variáveis. Recorreu-se então à variável secundária exaustiva, assumindo que ela apresenta a mesma direção de anisotropia da primária, devido a elas exibirem uma alta correlação. Optouse em utilizar a tabela de covariância da secundária exaustiva para estimar/simular a variável primária colocada, após realizar-se a correção da parte central da tabela de covariância e ser imputado o efeito pepita a ela. O efeito pepita foi obtido através do variograma experimental omnidirecional da variável primária. Os resultados mostraram que as estimativas reproduziram a média global e as médias locais. Além disso, as estimativas não extrapolaram os valores de mínimo e máximo dos dados. Em relação as simulações, elas reproduziram os histogramas e variogramas de maneira satisfatória.

6 Conclusões, contribuições e recomendações

Nesta tese, foram investigados a solução para dois assuntos: (1) modelagem conjunta da continuidade espacial de múltiplas variáveis correlacionadas sem usar o LMC e (2) a inferência da continuidade espacial de variáveis primárias com poucas amostras baseadas em dados secundários exaustivos. As seções 6.1 e 6.2 abordam, respectivamente, as conclusões e contribuições da tese para cada assunto tratado e a seção 6.3, indica algumas sugestões para trabalhos futuros.

6.1 Conclusões da tese

Sobre a modelagem conjunta da continuidade espacial, foi proposto desenvolver uma metodologia para a estimativa/simulação de dados multivariados heterotópicos, com emprego de tabelas de covariância em substituição a modelagem e ajuste de variogramas diretos e cruzados via modelo linear de corregionalização (*Linear Model of Coregionalization* – LMC). A principal motivação de desenvolver uma nova metodologia para a modelagem das covariâncias conjuntas é pelo fato de ser laborioso e complexo modelar e ajustar variogramas para que se obtenha modelos de covariância permissíveis. Além disso, o LMC limita o ajuste do variograma à combinações lineares de estruturas básicas. Os resultados obtidos no estudo de caso realizado (ver capítulo 4) mostraram que as estimativas/simulações com uso de tabelas de covariância geram resultados similares aos obtidos pela modelagem convencional.

No que se refere a inferência da continuidade espacial de uma variável primária quando há uma variável secundária exaustiva, foi proposto desenvolver uma metodologia para inferir a continuidade espacial da variável primária, alicerçada no uso da tabela de covariância da variável secundária exaustiva. Essa metodologia é empregada quando a quantidade de dados da variável primária é insuficiente para obter um variograma experimental estável. O uso da tabela de covariância da variável secundária exaustiva substitui o cálculo e ajuste do variograma da variável primária, fato que motivou o desenvolvimento da metodologia. Os resultados do estudo de caso (ver capítulo 5) revelaram que as estimativas/simulações da variável primária usando a tabela de covariância da variável secundária exaustiva ajustada produziram resultados satisfatórios.

Quanto aos objetivos 1 e 2, que trataram de investigar o uso e a aplicabilidade da tabela de covariância, eles foram atingidos. Foi apresentado na seção 2.4.3 o método gerador das tabelas covariância, seu uso e suas limitações. As tabelas de covariância geradas pelo BMEC e FFT não são necessariamente positivas definidas e ainda, não basta obter as tabelas de covariância cruzadas, e sim a média entre as duas (+h e -h).

Em relação ao objetivo 3, que investigou métodos de correção das tabelas de covariância para torná-las positivas definidas, ele também foi atingido. Na seção 3.5, foi apresentado um método para a correção das tabelas de covariância, sendo de fácil implementação computacional e eficaz na obtenção de tabelas de covariância positivas definidas. Porém, devido ao tempo de processamento, a correção é feita apenas na parte central da tabela de covariância. Foi observado que conforme se aumenta o tamanho da área a ser corrigida da tabela de covariância, o tempo de processamento também aumenta significativamente. A parte central da tabela de covariância é a mais importante, pois contém as covariâncias para as amostras que estão mais próximas do local a ser estimado/simulado (*lags* iniciais). Essas amostras são as que recebem maior peso de krigagem e, portanto, mais influenciam nos resultados.

Assim como os anteriores, o objetivo 4, que investigou a obtenção e incorporação do efeito pepita à tabela de covariância, também foi alcançado. As tabelas de covariância geradas de um BMEC obtido a partir de uma estimativa pelo método do vizinho mais próximo não incorporam o efeito pepita. Para solucionar essa limitação, nas seções 3.2 e 3.4, foi sugerido como obter e incorporar o efeito pepita à tabela de covariância. Adicionar o efeito pepita à tabela de covariância é bem mais fácil que ajustar variogramas diretos e cruzados respeitando o LMC. O efeito pepita é adicionado diretamente nos algoritmos de estimativa/simulação. As covariâncias usadas nas estimativas/simulações são obtidas pela soma do efeito pepita e da covariância obtida pela tabela de covariância.

Acerca dos objetivos 5 e 6, que trataram de desenvolver uma metodologia para a modelagem da covariância conjunta em substituição ao LMC, eles foram atingidos. A correção apresentada na seção 3.5 é uma alternativa ao LMC, pois permitiu obter uma matriz de covariância conjunta (que contém as covariâncias diretas e cruzadas) positiva definida. Os resultados apresentados validaram a utilização das tabelas de covariância tanto na cokrigagem como na simulação utilizando variáveis secundárias.

Quanto ao objetivo 7, que investigou o uso da tabela de covariância da variável secundária exaustiva para inferir a continuidade espacial de uma variável primária com dados insuficientes para ter um variograma experimental confiável, ele também foi alcançado. O estudo de caso apresentado no capítulo 5 mostrou que, em bancos de dados heterotópicos com variável secundária exaustiva e variável primária com poucas amostras, é possível inferir a anisotropia da variável primária através da tabela de covariância da variável secundária e com isso, realizar a estimativa e simulação da primária. Os resultados mostraram-se satisfatórios e dispensaram a modelagem de variogramas. A tabela de covariância da variável secundária exaustiva torna-se, assim, uma importante ferramenta na geração da continuidade espacial da primária que, por apresentar poucos dados, impossibilita ao

geomodelador um modelo de variograma confiável.

O capítulo 5 mostra uma situação típica da geoestatística aplicada a reservatórios de óleo e gás e enfatiza a aplicabilidade da tabela de covariância para essa situação. Em petróleo, principalmente na fase de exploração, a variável primária não está suficientemente amostrada para se calcular o variograma experimental. Independente de haver dados suficientes, o cálculo da tabela de covariância calculada via FFT é bem mais rápido que o do variograma experimental tradicional. Essa vantagem é relevante na geoestatística aplicada ao petróleo, onde os grids com informação secundária são grandes (mais de 150 milhões de nós). Devido ao tempo de processamento, é comum trabalhar com grids mais grosseiros (menor resolução) do que o grid da variável secundária. O uso da tabela de covariância diminui o processamento dos variogramas e, portanto, abre a possibilidade de se trabalhar com grids de maior resolução.

6.2 Contribuições da tese

Foram alcançadas as seguintes contribuições:

- Descrever a continuidade espacial dos dados através de tabelas de covariância em substituição a variografia convencional. Na abordagem convencional, usam-se funções permissíveis predefinidas para descrever a continuidade espacial da variável de interesse e, a partir dessas funções, se realiza a estimativa/simulação. A tabela de covariância é gerada de modo semi-automático, dispensando o ajuste da covariância através de funções e ainda, contempla todos os valores de covariância para todas as direções e distâncias possíveis;
- Estabeleceu-se uma metodologia para tornar a matriz de covariância conjunta positiva definida em substituição ao LMC. Na modelagem tradicional, os variogramas (diretos e cruzados) são ajustados respeitando o LMC. Esse ajuste garante que a matriz de covariância conjunta seja positiva definida, mas a modelagem do LMC é extremamente laboriosa, pois é necessário combinar as mesmas estruturas básicas para os variogramas diretos e cruzados. A metodologia estabelecida é de fácil implementação computacional e gera de forma automática a matriz de covariância conjunta positiva definida, embora calcular e corrigir autovalores requeira um tempo considerável de processamento computacional;
- Cokrigar e simular múltiplas variáveis via tabelas de covariância. As tabelas de covariância (diretas e cruzadas), obtidas pela metodologia proposta, foram implementadas nos algoritmos de cokrigagem e de simulação sequencial Gaussiana com dados secundários. Os resultados gerados com uso de tabela de covariância mostraram-se similares aos obtidos através da variografia tradicional;

• Inferir a anisotropia e realizar estimativas do fenômeno multivariado com auxílio de tabelas de covariância. As tabelas de covariância, por serem obtidas a partir dos dados, tendem a reproduzir valores fidedignos de covariância e ainda, contemplam as direções de anisotropia e reproduzem resultados confiáveis de estimativa/simulação.

6.3 Recomendações para trabalhos futuros

Para trabalhos futuros, recomendo:

- Aplicar a metodologia em conjuntos de dados multivariados heterotópicos de um depósito mineral polimetálico. A metodologia estabelecida nesta tese foi verificada apenas em banco de dados sintético de modo que se faz necessário a verificação em banco de dados multivariados reais (2D ou 3D);
- Encontrar uma maneira de incorporar o efeito pepita automaticamente à tabela de covariância, sem a necessidade de variogramas experimentais. Na metodologia apresentada, o efeito pepita é obtido através do variograma omnidirecional e em seguida, adicionado à tabela de covariância;
- Investigar maneiras mais rápidas de se obter tabelas de covariância positiva definidas. A metodologia utilizada demanda um tempo de processamento razoável;
- Adaptar a metodologia de tabelas de covariância para técnicas de simulação espectral. Uma das vantagens da tabela de covariância via FFT é a diminuição do tempo de processamento. Entretanto, essa vantagem não é plenamente aproveitada quando a tabela de covariância é usada na simulação sequencial Gaussiana, que foi usada nessa tese. As técnicas de simulação espectral são mais rápidas e, portanto, vão de encontro ao objetivo de diminuir o tempo de processamento de um *workflow* de simulação geoestatística.

Referências

ALMEIDA, A. S. Joint simulation of multiple variables with a Markov-type coregionalization model. Tese (Doutorado) — Stanford University, 1993. Citado 2 vezes nas páginas 51 e 86.

ANTON, H.; RORRES, C. *Álgebra linear com aplicações.* 8. ed. Porto Alegre: Bookman, 2001. Citado 2 vezes nas páginas 35 e 36.

BARNETT, R. M.; MANCHUK, J. G.; DEUTSCH, C. V. Projection pursuit multivariate transform. *Mathematical Geosciences*, Springer, v. 46, n. 3, p. 337–359, 2014. Citado na página 24.

BOCHNER, S. Additive set functions on groups. Annals of Mathematics, JSTOR, p. 769–799, 1939. Citado 2 vezes nas páginas 25 e 44.

BOEZIO, M. N. M. Estudo das metodologias alternativas da geoestatística multivariada aplicadas a estimativa de teores de depósitos de ferro. Tese (Doutorado) — Universidade Federal do Rio Grande do Sul, 2010. Citado na página 51.

BRACEWELL, R. N. *The Fourier transform and its applications*. New York: McGraw-Hill, 1986. Citado na página 42.

BURMAN, P. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*, Oxford University Press, v. 76, n. 3, p. 503–514, 1989. Citado na página 57.

CHU, J. Conditional fractal simulation, fast sequential indicator simulation, and interactive variogram modeling. Tese (Doutorado) — Stanford University, 1993. Citado na página 24.

COOLEY, J. W.; TUKEY, J. W. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, JSTOR, v. 19, n. 90, p. 297–301, 1965. Citado na página 42.

DAVIS, B. J.; CURRIERO, F. C. Development and evaluation of geostatistical methods for non-euclidean-based spatial covariance matrices. *Mathematical Geosciences*, Springer, v. 51, n. 6, p. 767–791, 2019. Citado na página 58.

DESASSIS, N.; RENARD, D. Automatic variogram modeling by iterative least squares: univariate and multivariate cases. *Mathematical geosciences*, Springer, v. 45, n. 4, p. 453–470, 2013. Citado na página 24.

DESBARATS, A.; DIMITRAKOPOULOS, R. Geostatistical simulation of regionalized pore-size distributions using min/max autocorrelation factors. *Mathematical Geology*, Springer, v. 32, n. 8, p. 919–942, 2000. Citado na página 24.

DEUTSCH, C. V.; JOURNEL, A. G. *Geostatistical software library and user's guide*. New York: Oxford University Press, 1992. Citado na página 24.

EMERY, X. Iterative algorithms for fitting a linear model of coregionalization. *Computers* & *Geosciences*, Elsevier, v. 36, n. 9, p. 1150–1160, 2010. Citado na página 24.

FIGUEIREDO, D. G. de. *Análise de Fourier e equações diferenciais parciais.* 4. ed. Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 2014. Citado 2 vezes nas páginas 38 e 40.

FINNEY, R. L.; WEIR, M. D.; GIORDANO, F. R. *Cálculo de George B. Thomas Jr.* 10. ed. São Paulo: Pearson Addison Wesley, 2003. v. 2. Citado na página 38.

GOOVAERTS, P. Spatial orthogonality of the principal components computed from coregionalized variables. *Mathematical Geology*, Springer, v. 25, n. 3, p. 281–302, 1993. Citado na página 24.

GOOVAERTS, P. Geostatistics for natural resources evaluation. New York: Oxford University Press, 1997. Citado 7 vezes nas páginas 23, 26, 49, 50, 51, 53 e 62.

GUARDIANO, F. B.; SRIVASTAVA, R. M. Multivariate geostatistics: beyond bivariate moments. In: *Geostatistics Troia*'92. Dordrecht: Springer, 1993. p. 133–144. Citado na página 56.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, Warwick & York, v. 24, n. 6, p. 417, 1933. Citado na página 24.

IÓRIO, V. de M. *EDP: Um curso de graduação*. 3. ed. Rio de Janeiro: Instituto de Matemática Pura e Aplicada, 2012. Citado na página 41.

ISAAKS, E. The application of Monte Carlo methods to the analysis of spatially correlated data. Tese (Doutorado) — Stanford University, 1990. Citado na página 53.

ISAAKS, E. H.; SRIVASTAVA, R. M. An introduction to applied geostatistics. New York: Oxford University Press, 1989. Citado 4 vezes nas páginas 23, 28, 49 e 50.

JOURNEL, A. Markov models for cross-covariances. *Mathematical Geology*, Springer, v. 31, n. 8, p. 955–964, 1999. Citado 2 vezes nas páginas 51 e 52.

KAPAGERIDIS, I. K. Variable lag variography using k-means clustering. Computers & geosciences, Elsevier, v. 85, p. 49–63, 2015. Citado na página 24.

KLOECKNER, J. Tabela de covariância: um mapeamento rápido e automático de continuidade espacial. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, 2018. Citado 5 vezes nas páginas 13, 44, 46, 56 e 57.

KLOECKNER, J.; MACHADO, P. L.; RODRIGUES, Á. L.; COSTA, J. F. C. L. Covariance table: A fast automatic spatial continuity mapping. *Computers & Geosciences*, Elsevier, v. 130, p. 94–104, 2019. Citado 14 vezes nas páginas 25, 29, 31, 42, 43, 46, 47, 55, 56, 57, 59, 63, 64 e 73.

KUMAR, A.; DEUTSCH, C. V. Optimal correction of indefinite correlation matrices. *CCG Annual Report*, v. 11, 2009. Citado na página 59.

LARRONDO, P. F.; NEUFELD, C. T.; DEUTSCH, C. V. Varfit: a program for semiautomatic variogram modelling. *Fifth Annual Report of the Centre for Computational Geostatistics, University of Alberta, Edmonton*, 2003. Citado na página 24. LEUANGTHONG, O.; DEUTSCH, C. V. Stepwise conditional transformation for simulation of multiple variables. *Mathematical Geology*, Springer, v. 35, n. 2, p. 155–173, 2003. Citado na página 24.

LIPSCHUTZ, S.; LIPSON, M. *Algebra Linear: Coleção Schaum.* 8. ed. Porto Alegre: Bookman, 2011. Citado na página 36.

MARCOTTE, D. Fast variogram computation with fft. Computers & Geosciences, Elsevier, v. 22, n. 10, p. 1175–1186, 1996. Citado 6 vezes nas páginas 13, 24, 25, 31, 43 e 44.

MARECHAL, A. *Cokrigeage et regression en correlation intrinsique*. Fontainebleau: Centre de Geostatistique de Fontainebleau, 1970. Citado 2 vezes nas páginas 23 e 48.

MATHERON, G. Principles of geostatistics. *Economic Geology*, Society of Economic Geologists, v. 58, n. 8, p. 1246–1266, 1963. Citado na página 23.

OLIVEIRA, C. A. S.; KLOECKNER, J.; RODRIGUES, Á. L.; BASSANI, M. A. A.; COSTA, J. F. C. L. Use of heterotopic secondary data in geostatistics using covariance tables. *Applied Earth Science*, Taylor & Francis, p. 1–12, 2019. Citado na página 82.

PARDO-IGUZQUIZA, E.; CHICA-OLMO, M. The fourier integral method: an efficient spectral method for simulation of random fields. *Mathematical Geology*, Springer, v. 25, n. 2, p. 177–217, 1993. Citado na página 43.

POOLE, D. *Álgebra linear*. São Paulo: Pioneira Thomson Learning, 2004. Citado na página 37.

PRASAD, K. Variability forescasting: geoestatistical, spectral and fractal approaches. Tese (Doutorado) — University of Pennsylvania, 1991. Citado na página 42.

PYRCZ, M. J.; DEUTSCH, C. V. Semivariogram models based on geometric offsets. *Mathematical geology*, Springer, v. 38, n. 4, p. 475–488, 2006. Citado 2 vezes nas páginas 25 e 47.

SILVA, C. Z. d. A descorrelação de variáveis com fatorização MAF em estimativa de teores. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, 2013. Citado na página 52.

SOARES, A. *Geoestatística para as ciências da terra e do ambiente.* 3. ed. Lisboa: IST Press, 2006. Citado 2 vezes nas páginas 53 e 58.

SWITZER, P.; GREEN, A. *Min/max autocorrelation factors for multivariate spatial imagery*. Tech. Rep. 6, Stanford University, 1984. Citado na página 24.

VERLY, G. Sequential gaussian cosimulation: a simulation method integrating several types of information. In: *Geostatistics Troia*'92. Dordrecht: Springer, 1993. p. 543–554. Citado na página 53.

WACKERNAGEL, H. *Multivariate geostatistics: an introduction with applications.* 2. ed. Berlin: Springer-Verlag, 1996. Citado na página 48.

XU, W.; TRAN, T.; SRIVASTAVA, R.; JOURNEL, A. G. Integrating seismic data in reservoir modeling: the collocated cokriging alternative. In: SOCIETY OF PETROLEUM ENGINEERS. *SPE annual technical conference and exhibition*. Washington, 1992. Citado 4 vezes nas páginas 30, 50, 51 e 52.

YAO, T. Automatic covariance modeling and conditional spectral simulation with fast Fourier transform. Tese (Doutorado) — Stanford University, 1998. Citado na página 42.

YAO, T. Reproduction of the mean, variance, and variogram model in spectral simulation. *Mathematical Geology*, Springer, v. 36, n. 4, p. 487–506, 2004. Citado na página 43.

YAO, T.; JOURNEL, A. G. Automatic modeling of (cross) covariance tables using fast fourier transform. *Mathematical Geology*, Springer, v. 30, n. 6, p. 589–615, 1998. Citado 6 vezes nas páginas 24, 25, 29, 45, 46 e 47.

YOO, Y. *Tutorial on Fourier theory*. 2001. Disponível em: http://www.cs.otago.ac.nz/cosc453/student_tutorials/fourier_analysis.pdf>. Citado 2 vezes nas páginas 42 e 43.