

Jonas Kloeckner

**Tabela de covariância: um mapeamento rápido
e automático de continuidade espacial**

Porto Alegre

2018

Jonas Kloeckner

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

Esta dissertação foi analisada e julgada adequada para a obtenção do título de Mestre 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 de Materiais

Orientador: Prof. Dr. João Felipe Coimbra Leite Costa

Porto Alegre

2018

Jonas Kloeckner

Tabela de covariância: um mapeamento rápido e automático de continuidade espacial/ Jonas Kloeckner. – Porto Alegre, 2018-
70 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. João Felipe Coimbra Leite Costa

Dissertação (Mestrado) – Universidade Federal do Rio Grande do Sul - UFRGS
Escola de Engenharia
Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais,
2018.

1. Tabela de covariância. 2. Transformada rápida de Fourier. 3. Teorema de Convolação. 4. Modelo base para extrair covariância. I. Prof. Dr. João Felipe Coimbra Leite Costa. II. Universidade Federal do Rio Grande do Sul. III. Faculdade de Engenharia. IV. Em busca da geração automática da tabela de covariância.

CDU 02:141:005.7

Jonas Kloeckner

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

Esta dissertação foi analisada e julgada adequada para a obtenção do título de Mestre 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, ____ de _____ de 2018:

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

Dr. Áttila Leães Rodrigues
IF - UFRGS

Dr. Marcel Antônio Arcari Bassani
FLE

Porto Alegre
2018

Agradecimentos

“Sou cético com relação ao mito da meritocracia e do protagonismo assim como sou com relação a tantos outros. A sessão de agradecimentos de um trabalho é um espaço de resistência a essa ideia tão percolada na sociedade e da qual a ciência, mais do que qualquer outra instituição humana, tem o dever de se libertar”. Este trecho eu retirei da dissertação de um dos meus melhores amigos, quase mestre em biologia, mas nunca submeteu o artigo necessário, inclusive já feito, para a obtenção do título. No fim das contas, sou o primeiro da minha família e amigos a obter este título e a seguir para o doutorado. E, como ele, também sou cético: não tenho mais mérito que outros e o único protagonismo é ser espectador e personagem principal do meu dia a dia.

Talvez mérito em entender e aceitar os perrengues que envolvem esta escolha, enquanto crio e sigo o próprio caminho. A carreira acadêmica é uma adaptação do mito de Sísifo. Pode ser entendida como o personagem de ficção Bojack Horseman certa vez disse, em um lapso niilista: *“Laura! Clear our my schedule! I have to push a boulder up a hill and then have it roll over me time and time again with no regard for my well-being!”.* Mas ao contrário deste personagem, eu não me isolo e não sou tão autodestrutivo. Estou mais para o personagem babuíno, conhecedor da montanha, que ao ver Bojack nessa situação, explica: *“It gets easier... Every day it gets a little easier... But you gotta do it every day - that’s the hard part. But it does get easier.”.*

Subir e descer a montanha, carregando uma imensa rocha, quase de maneira esquizofrênica, seria tarefa Homérica sem orientação e apoio. Discutir e rever consigo mesmo, ideias e devaneios tão solitários e abstratos, de transformadas de Fourier a tabelas de covariância. Portanto, como nomes em um artigo e agradecimentos no final de cada apresentação, o time que esteve comigo nessa jornada e a estrutura aqui presente são os reais protagonistas e compartilhadores do mérito.

Gostaria de agradecer primeiramente ao meu orientador e professor João Felipe Coimbra Leite Costa. Um exemplo de profissional, extremamente dedicado, acessível e receptivo. Com seu discurso sólido me disse: *“Primeiro as cadeiras e depois teu tema de dissertação. De que adianta escolher um tema se não conheces os possíveis problemas a enfrentar?”.* Como a célebre afirmação “se não sabes para onde vai, qualquer caminho serve”, hoje, sei onde cheguei, orientado pelo caminho correto e no tempo certo. Muito obrigado JF.

Agradeço também aos meus principais companheiros neste tema tão complexo, Áttila Leães Rodrigues e Péricles Lopes Machado. Sem ajuda e paciência de vocês, este trabalho não seria feito.

À minha família, que sempre acreditou no meu potencial, demonstrando através de apoio incondicional. Norma, João Luciano e Francine Kloeckner, vocês são meus os maiores exemplos e estão sempre comigo, seja onde e quando for.

Aos meus amigos de infância: Diogo, Lucas, Marcelo, Paulo e Rafael. É inacreditável a força que vocês me dão e como seguimos tão unidos depois de décadas de amizade. Tenho certeza que isso seguirá até o fim de nossas vidas. Caso tudo não vire um *Black Mirror*.

À minha namorada Tatiana Klaus Sansonowicz. A pessoa mais incrível que já conheci na vida. Mesmo depois de mestre, não sou capaz de calcular o tamanho da minha sorte de te ter ao meu lado. Talvez seja minha limitação cartesiana e eu só poderia fazê-lo com tua ajuda psicanalítica.

Gostaria de agradecer também a todos colegas do Laboratório de Pesquisa Mineral e Planejamento Mineiro (LPM). Vocês fazem desta instituição o que ela grita aos quatro ventos ser: de excelência.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e a Fundação Luis Englert (FLE). Não há pesquisa sem recurso, não há recurso sem pesquisador e não há pesquisa sem pesquisador.

*"Nobody exists on purpose
Nobody belongs anywhere
Everybody's gonna die.
Come watch TV."
(Morty - Rick and Morty)*

Resumo

Os modelos de covariância são ferramentas geoestatísticas essenciais para mapear a continuidade espacial. A abordagem atual busca um modelo de continuidade espacial lícito com mínima ou até mesmo sem nenhuma interferência do usuário. Alinhado a essa visão moderna, é proposto obter uma tabela de covariância que visa substituir na prática o modelo tradicional explicitamente definido de covariância. Essa tabela de covariância é obtida por meio de três etapas: interpolar o conjunto de dados para preencher um grid regular, aplicar a convolução através do algoritmo da transformada rápida de Fourier e, por fim, transformar de volta para o domínio espacial. O modelo base para extrair covariância representa o ponto chave comparando com os métodos anteriores que propuseram o uso da tabela de covariância. Os resultados são satisfatórios, tanto na validação estatística do método, quanto na rapidez de obtenção de uma análise de continuidade espacial. Um estudo de caso tridimensional ilustra a aplicação prática através de krigagem e simulação geoestatística em comparação com a modelagem espacial tradicional.

Palavras-chaves: tabela de covariância, transformada rápida de Fourier, Teorema de convolução, modelo base para extrair covariância.

Abstract

Covariance models are essential geostatistical tools to map spatial continuity. The current approach pursues a licit spatial continuity model with minimum or even no user interference. Aligned with this modern view we propose to obtain a covariance table that aims at replacing in practice traditional covariance explicit defined model. This covariance table is obtained through a three steps work flow: interpolating the dataset to fill up a regular grid, auto convolute via Fast Fourier Transform algorithm and back transform to spacial domain. The base model to extract covariance represents the turning point comparing with previous methods that proposed covariance table usage. The results are satisfactory, both in the statistical validation of the method and in the speed of obtaining a spatial continuity analysis. A three dimensional case study illustrates the practical application for kriging and geostatistical simulation in comparison with traditional spatial modeling.

Keywords: covariance table, fast Fourier transform, convolution Theorem, base models to extract covariance.

Lista de ilustrações

Figura 1 – Comparação entre o método proposto nesta dissertação e o método proposto por Yao. GSCM do inglês <i>Gridded Sample Covariance Map</i>	21
Figura 2 – Fluxograma da metodologia deste estudo. Imagem meramente ilustrativa.	22
Figura 3 – Ilustração de três ondas (2Hz, 3Hz e a soma de ambas) e suas respectivas transformadas de Fourier. Adaptado de (SANDERSON, 2018).	26
Figura 4 – Ilustração (em duas dimensões) da transformada de Fourier. Adaptado de (SANDERSON, 2018).	26
Figura 5 – Ilustração (em duas dimensões) da transformada de Fourier. Adaptado de (SANDERSON, 2018).	27
Figura 6 – Ilustração (em uma dimensão) da retroprojeção como resultado por assumir a natureza periódica de f e g na análise de Fourier. Aumentando o sinal com zeros elimina-se a contribuição de fora da janela espacial observada. Adaptado de (MARCOTTE, 1996).	31
Figura 7 – Modelo exaustivo (a) e MBC (b).	42
Figura 8 – Tabela de covariância do modelo exaustivo (a) e do MBC (b).	43
Figura 9 – OK via tabela de covariância (a) e via variograma (b) em grid 1x1.	44
Figura 10 – OK via tabela de covariância (a) e via variograma (b) em grid 5x5, com elipses vermelhas destacando diferenças em áreas de alto teor e amarelas para baixos teor.	44
Figura 11 – CT de um MBC com uma anisotropia zonal (a), CT de um MBC de efeito pepita puro (b), CT do Walker Lake (c) e CT utilizada no estudo de caso no Capítulo 4 (d). Os elipsoides na região central de cada CT são as elipses de busca para as krigagens e simulações subsequentes.	47
Figura 12 – Casos ilustrativos: (a) um modelo efeito de pepita puro, a realização de SGS como um MBC, a CT extraída dele e uma realização de SGS condicionada pela CT; (b) uma realização SGS de anisotropia zonal como MBC, a CT extraída e uma realização de SGS condicionada pela CT.	48
Figura 13 – Plug-in do software <i>AR2GeMS</i> para geração da CT.	49
Figura 14 – Modelo exaustivo (a) e V 0.1% (b).	52
Figura 15 – Seção em plano horizontal para z igual a 1 do modelo exaustivo (a) e V 0.1% (b).	52
Figura 16 – Histograma do modelo exaustivo e dos dados com 0.1%, 0.5% e 1.0% de V (b).	53

Figura 17 – CT para o modelo exaustivo em seção isométrica com x e z igual zero. O elipsoide que envolve a parte da CT que será incorporada no cálculo, no canto superior direito, representa o elipsoide de busca da krigagem.	54
Figura 18 – CT para o modelo exaustivo em comparação com CT para todos os conjuntos de dados (V01, V05 e V1) em z igual a zero. A linha pontilhada referencia a seção para a direção de maior continuidade espacial.	55
Figura 19 – Variogramas extraídos das seções nas direções de maior e menor continuidade espacial das quatro tabelas de covariância da figura 18. Em verde, o variograma da tabela do modelo exaustivo, em vermelho, rosa e azul, são os variogramas gerados a partir das tabelas de V 0.1%, 0.5% e 1.0%, respectivamente.	56
Figura 20 – Modelo exaustivo em comparação com os MBCs para todos os conjuntos de dados (V01, V05 e V1) para z igual a 1.	57
Figura 21 – OK via CT e variograma para V01, V05 e V1, em seção horizontal em z igual a 1.	58
Figura 22 – CDF do modelo exaustivo em comparação com OK via CT e variograma para V 0.1%.	59
Figura 23 – Análise de deriva do modelo exaustivo em comparação com OK via CT e variograma para V 0.1%.	60
Figura 24 – Seção vertical em x igual a 200 de V 0.1% e de seu respectivo MBC.	60
Figura 25 – Conjunto de realizações via CT e variograma para V 0.1%, V 0.5% e V 1.0%.	61
Figura 26 – CDF para 50 realizações SGS via CT e variograma para V 0.1%, V 0.5% e V 1.0%.	63
Figura 27 – Flutuações ergódicas apresentadas pelo variograma orientado na direção de maior e menor continuidade espacial para SGS via CT e variograma para V 1.0%, 0.5% e 0.1%.	64

Lista de tabelas

Tabela 1 – Complexidade das operações (Adaptado de (MARCOTTE, 1996)). . .	32
Tabela 2 – Estruturas e parâmetros de continuidade espacial usados para gerar a realização via SGS.	51
Tabela 3 – Variância dos modelos de OK via CT e variograma para V 1.0%, V 0.5% e V 0.1%.	59

Sumário

1	INTRODUÇÃO	19
1.1	Meta	21
1.2	Objetivos específicos	21
1.3	Metodologia	22
1.4	Estrutura da dissertação	23
2	REVISÃO BIBLIOGRÁFICA	25
2.1	Transformada de Fourier	25
2.1.1	Transformada Discreta de Fourier	29
2.1.2	Transformada Rápida de Fourier	29
2.1.3	Teorema da Convolução	29
2.2	Computação Rápida de Variogramas via FFT	30
2.3	Modelagem Automática de Tabelas de Covariâncias (Cruzadas) utilizando Transformada Rápida de Fourier	33
3	TABELA DE COVARIÂNCIA: UM MAPEAMENTO RÁPIDO E AUTOMÁTICO DE CONTINUIDADE ESPACIAL	41
3.1	Modelo Base para Extrair Covariância	41
3.2	Gerando a Tabela de Covariância	44
3.3	Algoritmo e Plug-in	49
4	ESTUDO DE CASO	51
4.1	Geração do Depósito Sintético e Banco de Dados	51
4.2	Geração das Tabelas de Covariância	53
4.3	Estimativas	58
4.4	Simulações	61
4.5	Conclusão	65
5	CONSIDERAÇÕES FINAIS	67
5.1	Recomendações para trabalhos futuros	67
	REFERÊNCIAS	69

1 Introdução

Em todas as etapas da mineração, das ciências e engenharias de modo geral, os avanços tecnológicos possibilitam o desenvolvimento de metodologias mais eficientes, automatizadas e que incorporem pensamento computacional. As evidências da influência do pensamento computacional (um tipo de pensamento analítico) em muitos campos são abundantes. A essência do pensamento computacional é abstração. Na computação, abstraímos as noções além das dimensões físicas do tempo e do espaço. Nossas abstrações são extremamente gerais porque são simbólicas, onde as abstrações numéricas são apenas um caso especial. Seu efeito sobre a estatística, através do aprendizado de máquina, automatização de métodos Bayesianos e uso de modelos gráficos probabilísticos, tem tornado possível a identificação de padrões e anomalias em bancos de dados tão volumosos e diversificados quanto mapas astronômicos, além do escaneamento de imagens de ressonância magnética, compras via cartões de crédito e receitas culinárias (WING, 2008). O mesmo vale para a modelagem geológica e de teores. Entretanto, por mais que hoje essa temática esteja em voga (atrelada às técnicas de aprendizado de máquina), as tentativas de automatizar trabalhos repetitivos e maçantes não são novidade. Uma das inspirações e principal referência para esta dissertação é justamente um estudo do final da década de 90, desenvolvido por Tingting Yao (YAO; JOURNEL, 1998), que foca na automatização de um método para a obtenção de um modelo de continuidade espacial, mais precisamente, um modelo de covariância.

Os modelos de covariância são essenciais para a geoestatística como métricas de continuidade espacial. O objetivo, portanto, é a busca pelo melhor modelo de covariância que represente o fenômeno espacial com a mínima - ou até mesmo sem - influência do usuário. O método convencional estabelece um modelo de semivariograma para uma variável de interesse que é válido para todas as distâncias e direções para um certo domínio. Este procedimento implica interferência humana na metodologia, que pode ser contornado com o método aqui proposto, através do uso de uma tabela de covariância como substituto do modelo explícito. Há uma família de métodos focados no ajuste automático do modelo variográfico. Estas técnicas, entretanto, necessitam do usuário para ajustar parâmetros para o variograma experimental. Dessa maneira, portanto, calcular e ajustar um modelo explícito, de semivariograma por exemplo, é um passo crucial para todas as estimativas por krigagem ou simulações baseadas no seu uso (PYRCZ; DEUTSCH, 2006).

O método proposto utiliza a tabela de covariância, que é um conjunto de informações que podem ser usadas para descrever a correlação espacial de um campo aleatório e ser usado em técnicas como krigagem e simulação geoestatística. A tabela de covariância deve conter valores de covariância, aqui chamados de lícitos, para compor uma matriz de

krigagem positiva definida. Portanto, quando é afirmado que um modelo de variograma ou uma tabela de covariância são positivos definidos, isso significa que todos os valores contidos são lícitos. Uma solução foi inicialmente proposta por Deutsch (DEUTSCH; JOURNAL, 1992) e Chu (CHU, 1993) para construir um mapa discreto de covariância diretamente a partir dados, mas somente calculado para as distâncias disponíveis. O mapa é posteriormente interpolado para todas as distâncias. O programa Varmap do GSLIB (DEUTSCH; JOURNAL, 1992) pode ser usado para calcular o mapa de covariância utilizando esta técnica. É uma técnica válida, porém não garante uma tabela de covariância positiva definida, além de ser uma técnica computacionalmente custosa.

Lembrando que os modelos de covariância ou ainda a tabela de covariância fornecem os valores de covariância para a composição matriz de covariância C do sistema de krigagem ordinária 1.1 e que esta matriz C deve ser positiva definida. Cada um dos quatro testes que seguem são necessários e são condição suficiente para uma matriz real e simétrica C ser positiva definida (ISAAKS; SRIVASTAVA, 1989).

$$C \cdot w = D \tag{1.1}$$

$$\begin{bmatrix} C_{11} & \dots & C_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ C_{n1} & \dots & C_{nn} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \vdots \\ w_n \\ \mu \end{bmatrix} = \begin{bmatrix} C_{10} \\ \vdots \\ C_{n0} \\ 1 \end{bmatrix}$$

- $w^t C w > 0$ para todos os vetores não-zeros w ;
- Todos os autovalores de C são maiores que zero;
- Todas as submatrizes de C possuem determinantes positivos;
- Todos os pivôs (sem troca de colunas) são maiores que zero.

Yao et al. (YAO; JOURNAL, 1998) realizaram grandes avanços para a modelagem de tabela de covariância semiautomática (YAO, 2000). Na busca de uma modelagem automática de tabela de covariância, Yao apresentou um método para calcular uma tabela de covariância utilizando dados experimentais com atenção especial a critérios de licitude (positivo definido). Embora fosse um passo em direção à modelagem automática, ainda é necessário inserir parâmetros relacionados ao processo de aquisição de pares e suavização da tabela de covariância no espaço de Fourier. O trabalho de Yao foi um avanço utilizando a transformada rápida de Fourier (FFT) aplicada a análise de continuidade espacial proposta por Marcotte em 1996 (MARCOTTE, 1996). Todavia, todos estes métodos possuem certas limitações em diferentes etapas, ao consumir, por exemplo, muito tempo computacional,

ou ainda, por não serem de fato automáticos, exigindo parâmetros adicionais de suavização que devem ser imputados pelo operador do algoritmo. Dado que a proposta de Yao é a principal referência em todo estudo sobre tabelas de covariância, a comparação entre os métodos é mostrada na figura 1. Pode-se verificar que ambos métodos iniciam e terminam nos mesmos objetos, mas os níveis de complexidade são diferentes, assim como o número de etapas.

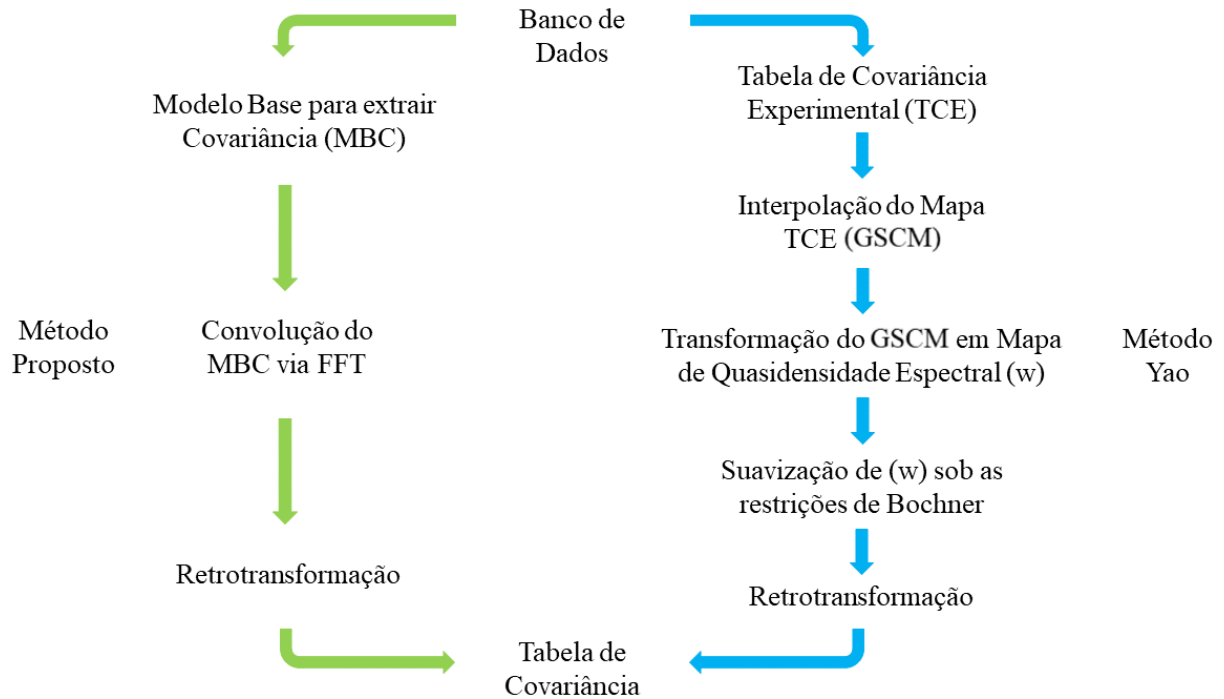


Figura 1 – Comparação entre o método proposto nesta dissertação e o método proposto por Yao. GSCM do inglês *Gridded Sample Covariance Map*.

1.1 Meta

Essa dissertação de mestrado tem como meta o desenvolvimento de uma metodologia automática para a obtenção de uma tabela de covariância como substituto do modelo de covariância explícito, amplamente adotado pela academia e pela indústria. Portanto, contribui na rapidez de aquisição de uma tabela descritiva de um fenômeno espacial.

1.2 Objetivos específicos

A fim de atingir a meta proposta os seguintes objetivos foram delineados:

1. O trabalho visa estabelecer uma metodologia automática para construir a tabela de covariância.

2. Desenvolver o método e implementá-lo no *software* geoestatístico de código aberto *AR2GeMS*, desenvolvendo um *plug-in* funcional em *Python*®;
3. Conduzir um estudo de caso em um banco de dados sintético e verificar a qualidade das estimativas e das realizações por simulação geoestatística gerado pela metodologia proposta, comparando-os a modelos gerados por variografia tradicional.

1.3 Metodologia

O método proposto é dividido em três partes. O primeiro passo é construir um modelo base para extrair covariância considerando os dados experimentais, aqui denominado de modelo base para covariância (MBC). 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 MBC em grid completamente estimado e regular construído a partir dos dados experimentais. O segundo passo é a auto convolução do modelo via FFT. O terceiro passo, por fim, é retrotransformar a tabela de covariância do domínio espectral para o domínio de dados. A figura 2 ilustra o fluxo de trabalho do método proposto, partindo de um conjunto de dados até a obtenção de uma tabela de covariância seguido de aplicações geoestatísticas como estimativa e simulação baseadas em variogramas.

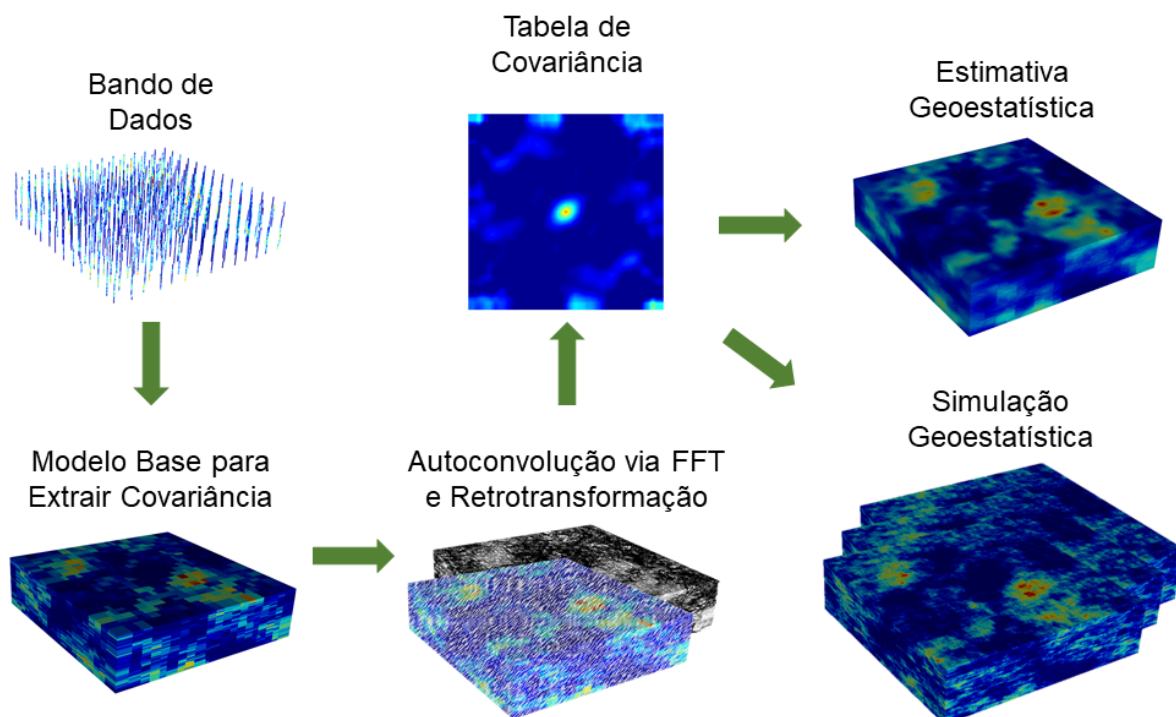


Figura 2 – Fluxograma da metodologia deste estudo. Imagem meramente ilustrativa.

1.4 Estrutura da dissertação

O capítulo 2 é dedicado à revisão de conceitos fundamentais e a apresentar o estado da arte sobre geração de tabelas de covariância para análise de continuidade espacial. Portanto, é feita uma breve revisão de transformada de Fourier, dada sua importância para o assunto, e os dois principais artigos no tema da dissertação.

O capítulo 3 apresenta a metodologia proposta, explorando o denominado modelo base para extrair covariância e o tipo de tabela obtida a partir deste. Apresenta também o *plug-in* desenvolvido para o *software AR2GeMS* e os resultados da validação do algoritmo proposto tendo como referência resultados gerados para a modelagem tradicional via variografia.

O capítulo 4 discute os resultados do estudo de caso conduzido em um banco de dados sintético. Avalia a aplicabilidade do método como substituto ou método auxiliar à metodologia tradicional de modelagem explícita variográfica, comparando estimativa e simulação geoestatística a partir da metodologia proposta à tradicional.

O capítulo 5 encerra o estudo com as conclusões desta dissertação e sugerindo trabalhos futuros relacionados.

2 Revisão bibliográfica

2.1 Transformada de Fourier

Grande parte dos trabalhos e artigos desenvolvidos em processamento de imagens, de sinais, e tantas outras áreas que tratam de transformada de Fourier, o fazem apresentando algoritmos e fórmulas de maneira densa. Assim, é fácil perder-se diante de tanta matemática aparentemente incompreensível. Entretanto, o conceito básico por trás de toda complexidade é, de fato, simples: é possível obter qualquer função $f(x)$ através de uma somatória de séries de senos e cossenos de frequências crescentes (YOO, 2001). Ou seja, qualquer dado que varie no espaço ou tempo pode ser transformado para um domínio denominado de frequência.

O termo frequência pode ser entendido como a descrição de uma característica de comportamento periódico, como cores em uma imagem ou teores em um grid. Já o domínio de frequência coleta e agrupa informações de tal forma que torna-se fácil a utilização de filtros ou medições, que de outra forma seriam de difícil implementação no domínio dos dados. Esta é uma grande vantagem no processamento de imagens ou som (YOO, 2001). Para ilustrar a transformada de Fourier e algumas de suas propriedades, um brilhante matemático da Universidade de Stanford chamado Grant Sanderson produziu uma coleção de vídeos (SANDERSON, 2018). Alguns conceitos e imagens foram adaptados para esta revisão de forma a esclarecer termos e fundamentos na obtenção da tabela de covariância.

O exemplo, que será desenvolvido, está apresentado na figura 3: três ondas (2Hz, 3Hz e a soma de ambas) e suas respectivas transformadas de Fourier. A primeira constatação a ser feita é o agrupamento de informação possibilitado pela transformada, anteriormente comentado. Para a onda de 2Hz, o pico no domínio de frequência acontece em 2.0 ciclos/segundo; para 3Hz, em 3.0 ciclos/segundo; por fim, para a soma de 2Hz e 3Hz o pico acontece em ambos ciclos/segundo. Os filtros e as medições possíveis no domínio de frequência acontecem em função desta característica. Por exemplo, digamos que seja de interesse isolar a onda de 3Hz da onda 2Hz + 3Hz. Basta realizar a transformada para a onda 2Hz + 3Hz, reduzir o pico evidente em 2.0 ciclos/segundo e realizar a transformada inversa, resultando precisamente na onda de 3Hz.

A segunda constatação é a de que a soma de ondas no domínio tempo também é válida no domínio de frequência. Portanto, para chegar na função 2.0 + 3.0 ciclos/segundo no domínio de frequência, dois caminhos são possíveis. O primeiro é transformar individualmente a onda de 2Hz e 3Hz e somá-las no domínio de frequência. Já o segundo caminho é somá-las no domínio tempo e posteriormente transformar a onda resultante.

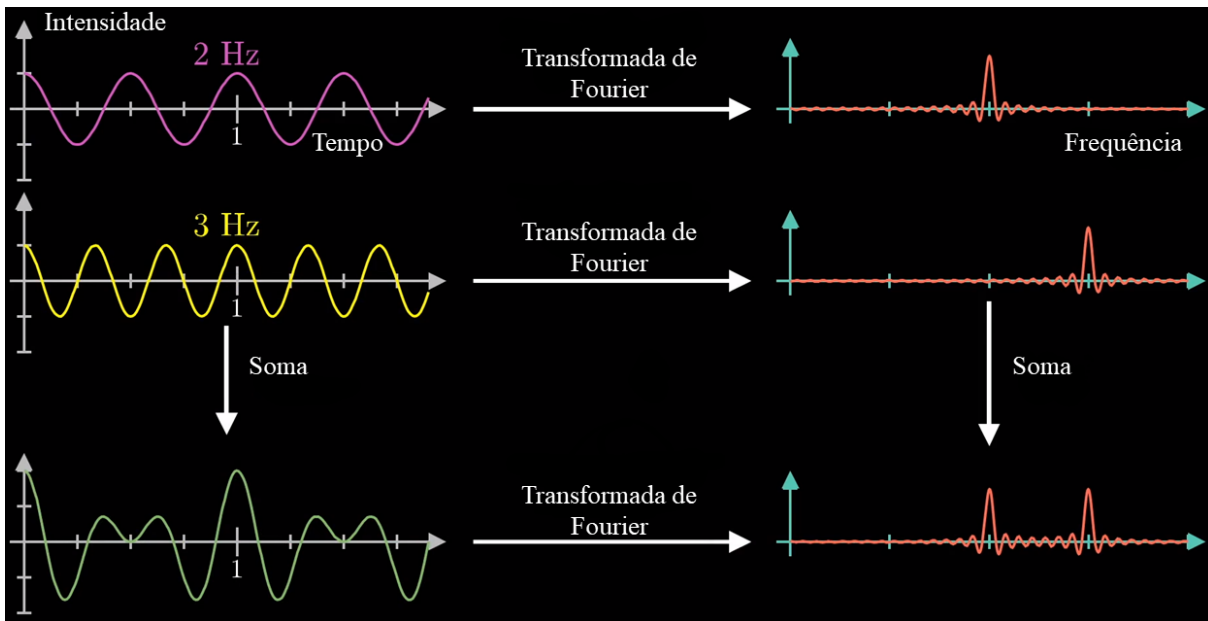


Figura 3 – Ilustração de três ondas (2Hz, 3Hz e a soma de ambas) e suas respectivas transformadas de Fourier. Adaptado de (SANDERSON, 2018).

Sanderson traça um paralelo para explicar o que seria a onda transformada no domínio de frequência. Similar ao centro de massa de um corpo em rotação, haveria um ponto análogo ao “centro de massa” de uma onda em rotação. A figura 4 apresenta três pontos de leitura (0.2, 2.0 e 3.0 ciclos/segundo) do “centro de massa” da onda 2Hz + 3Hz rotando em torno de um eixo imaginário.

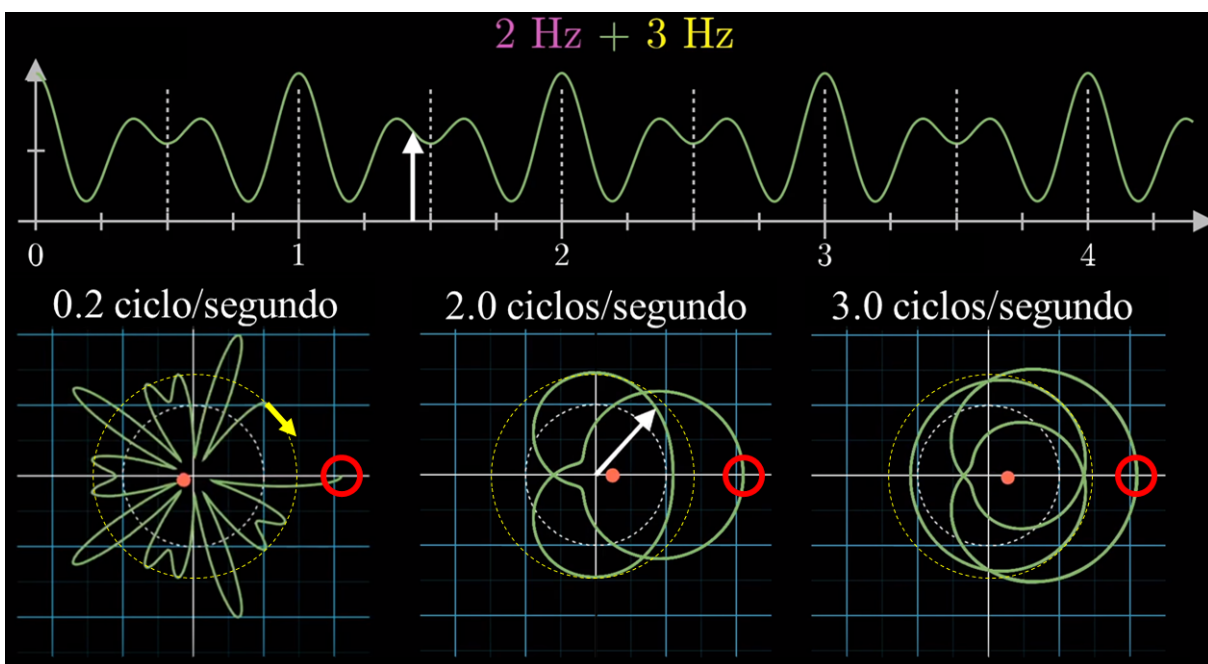


Figura 4 – Ilustração (em duas dimensões) da transformada de Fourier. Adaptado de (SANDERSON, 2018).

Inicia-se a rotação da onda entorno do eixo imaginário fixando o primeiro valor de tempo x intensidade (círculo vermelho). A onda é então estendida (seta amarela) e rotaciona em torno do eixo imaginário. Conforme isto ocorre, é verificado o deslocamento do “centro de massa” em relação ao eixo (ponto rosa). Assim, é possível determinar uma nova função, que nada mais é do que a transformada de Fourier da onda original. A figura 5 apresenta a onda original $2\text{Hz} + 3\text{Hz}$, sua rotação em 2.0 ciclos/segundo e a onda transformada no domínio frequência. E, uma vez que todo o estudo trata de análise de continuidade espacial, pode-se entender a onda original como covariância c em função de um espaçamento h na forma de $c(h)$, e sua transformada de Fourier será a densidade espectral $s(w)$.

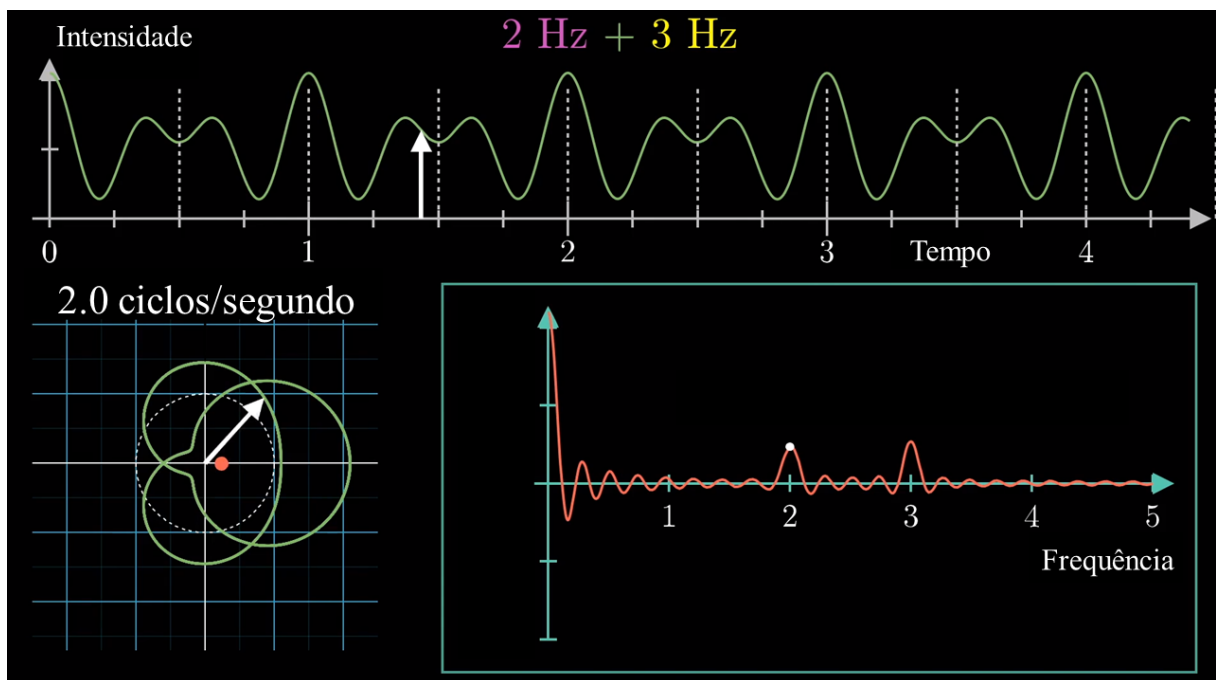


Figura 5 – Ilustração (em duas dimensões) da transformada de Fourier. Adaptado de (SANDERSON, 2018).

Uma vez que estamos trabalhando com transformações e rotações que envolvem parte real e parte imaginária, será necessária a apresentação de algumas definições matemáticas. Um número complexo pode ser escrito na forma:

$$R + iI, \quad (2.1)$$

onde R e I são números reais e i é igual a $\sqrt{-1}$. R denota a parte real e I denota a parte imaginária de um número complexo. Números reais, portanto, podem ser entendidos como um subconjunto dos números complexos, com $I = 0$. Do ponto de vista geométrico, todos os números reais podem ser encaixados uma linha infinitamente longa em um espaço de 1 dimensão. Se adicionarmos uma dimensão, então será possível representar ainda mais

números. Assim, os números complexos podem ser entendidos como um plano ao invés de uma linha. O eixo horizontal desta representação é chamado de *eixo real* e o eixo vertical de *eixo imaginário*. Dessa forma, qualquer número $R + iI$ possui uma coordenada (R, I) no plano (YOO, 2001). A mesma representação de uma coordenada em um plano, pode ser feita em coordenada polar, ou seja, uma distância da origem e um ângulo na forma:

$$r(\cos \theta + i \sin \theta), \quad (2.2)$$

onde $r = \sqrt{R^2 + I^2}$ e $\theta = \tan^{-1}(I/R)$. Já a fórmula de Euler:

$$e^{i\theta} = \cos \theta + i \sin \theta, \quad (2.3)$$

onde $e = 2.71828\dots$ e θ é o ângulo.

Por fim, podemos definir a equação da transformada de Fourier de uma função $z(x)$ na forma:

$$Z(\xi) = \int_{-\infty}^{+\infty} z(x)e^{-i2\pi\xi x} dx, \quad (2.4)$$

onde ξ é a chamada variável frequência e a função $Z(\xi)$ é a transformada de $z(x)$, pertencente ao espaço frequência, portanto. A somatória de senos e cossenos não é óbvia na equação 2.4, mas aplicando a equação de Euler teremos:

$$Z(\xi) = \int_{-\infty}^{+\infty} z(x)(\cos 2\pi\xi x + i \sin 2\pi\xi x) dx, \quad (2.5)$$

Já a transformada inversa de $z(x)$ é dada por:

$$z(x) = \int_{-\infty}^{+\infty} Z(\xi)e^{i2\pi\xi x} dx, \quad (2.6)$$

As equações 2.4 e 2.6 são chamados de *pares da transformada de Fourier* e somente são possíveis caso a função $z(x)$ seja contínua e integrável, e caso a função transformada $Z(\xi)$ seja integrável (YOO, 2001). E, justamente em função desta condição, que a transformada de Fourier da covariância obtida diretamente das amostras não é possível, pois a função $c(h)$ não será contínua. E, o artifício usado para satisfazer essa condição será estimar os dados para todo os nós do grid, o que será explorado no capítulo [Capítulo 3](#). Uma vez que, na prática, dificilmente trabalharemos com funções contínuas de $z(x)$. Por este motivo, trabalha-se na prática com *transformadas discretas de Fourier*.

2.1.1 Transformada Discreta de Fourier

Para aplicar as transformadas de Fourier na prática, é necessário a utilização do artifício matemático chamado *transformada discreta de Fourier*, que basicamente substitui uma integral por uma somatória (YOO, 2001). Dado N amostras discretas de uma função $z(x)$, temos:

$$Z(\xi) = \frac{1}{N} \sum_{x=0}^{N-1} z(x) e^{-i2\pi\xi x/N}, \quad (2.7)$$

para $\xi = 0, 1, 2, \dots, N - 1$, e

$$z(x) = \sum_{\xi=0}^{N-1} Z(\xi) e^{i2\pi\xi x/N}, \quad (2.8)$$

para $x = 0, 1, 2, \dots, N - 1$, e

2.1.2 Transformada Rápida de Fourier

Embora a transformada discreta de Fourier nos permita calcular a transformada de Fourier utilizando programação, esta não é eficiente. O número de multiplicações complexas e adições requeridas nas equações 2.7 e 2.8 é igual a $O(N^2)$, onde N o tamanho do vetor e O o número de operações.

O algoritmo *transformada rápida de Fourier* (FFT) pode acelerar o processo através de uma decomposição apropriada da equação 2.7 para $O(N \log_2 N)$ operações. A diferença entre $N \log_2 N$ e N^2 é imensa. Com $N = 10^6$, por exemplo, a diferença é de aproximadamente 30 segundos de processamento para 2 semanas (PRESS, 1992) Há diversas maneiras de implementar esta decomposição, por exemplo, explorando a simetria da definição discreta de transformação de Fourier. Entretanto, estas maneiras não serão exploradas nesta dissertação.

2.1.3 Teorema da Convolução

Uma das propriedades mais importantes da transformada de Fourier é o teorema da convolução que mapeia a convolução entre duas funções como:

$$\mathcal{F}[z(x) \otimes w(x)] = \mathcal{F}\left[\int_{-\infty}^{\infty} z(x')w(x-x')dx'\right] = Z(\xi)W(\xi), \quad (2.9)$$

onde $z(x)$ e $w(x)$ são funções contínuas periódicas. $Z(\xi)$ é usado para representar a função transformada $z(x)$, $W(\xi)$ para $w(x)$. $\mathcal{F}[\cdot]$ é o operador para a transformada de Fourier e \otimes é o símbolo para convolução.

Até aqui a revisão foi estritamente matemática, de maneira a ambientar e padronizar os termos e símbolos que serão utilizados na metodologia proposta nesta dissertação. Além das duas principais revisões que seguem. A primeira revisão geoestatística, propriamente dita, será feita do artigo de 1996 de Denis Marcotte intitulado *Computação Rápida de Variogramas via FFT*. Marcotte introduziu o algoritmo de transformada rápida de Fourier às funções de continuidade espacial na geoestatística. Preocupou-se com a eficiência computacional ao calcular variogramas experimentais, trazendo a técnica ao campo do conhecimento de seus pares. Por este motivo, seu artigo foi de grande importância à geoestatística.

2.2 Computação Rápida de Variogramas via FFT

Em 1996, Denis Marcotte publicou seu artigo intitulado *Computação Rápida de Variogramas via FFT*, buscando uma maneira moderna para computar funções de continuidade espacial. A abordagem clássica para a computação de variogramas é considerar, de preferência, todos os pares possíveis de dados. Para cada par, é computado uma distância, uma orientação e a diferença quadrática entre os valores. Quando os dados estão em um grid regular, o método mais eficiente é considerar somente algumas direções e um determinado valor de espaçamento. A computação dos dados pode ser feita a um custo menor simplesmente deslocando a matriz de dados. O par é formado pela sobreposição dos pontos originais, e dos mesmos deslocados. Essa abordagem, pressupondo um grid regular ($N \times M$), possui uma complexidade computacional dada por:

$$\sum_k \sum_l (N - k) \cdot (M - l), \quad (2.10)$$

onde k e l representam a diferença entre os deslocamentos na direção x e y .

A seguir, é apresentada a abordagem computacional via domínio espectral, que é a chave do artigo de Marcotte e foi utilizada diretamente nos estudos desenvolvidos nesta dissertação. Considere duas funções randômicas $f(x,y)$ e $g(x,y)$. No caso contínuo, a covariância cruzada não centrada entre $f(x,y)$ e $g(x,y)$ a uma dada distância h_x e h_y é dada pela seguinte expressão:

$$C(h_x, h_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) g(x + h_x, y + h_y) dx dy, \quad (2.11)$$

A partir de uma transformada de Fourier, nós obtemos:

$$\mathcal{F} [C(h_x, h_y)] = \overline{F}[(x, y)] \cdot G[(x, y)], \quad (2.12)$$

onde \mathcal{F} é o operador transformada de Fourier, $\bar{F}[(x, y)]$ é o conjugado complexo de $F(x, y)$, sendo $F(x, y)$ a transformada de Fourier de $f(x, y)$, da mesma forma, $G(x, y)$ para $g(x, y)$.

Uma covariância não centrada pode assim ser obtida computando a transformada de Fourier de $f(x, y)$ e $g(x, y)$, multiplicando ambas, e por fim, fazendo a transformada inversa. O resultado é diretamente transposto para forma discreta, dado que $f(x, y)$ e $g(x, y)$ são preenchidas por zeros até o tamanho de $2N-1$ por $2M-1$. Marcotte afirma que este acréscimo deve ser feito para viabilizar a condição de periodicidade das funções f e g (figura 6), a fim de evitar a retroprojeção (em inglês “wraparound”).

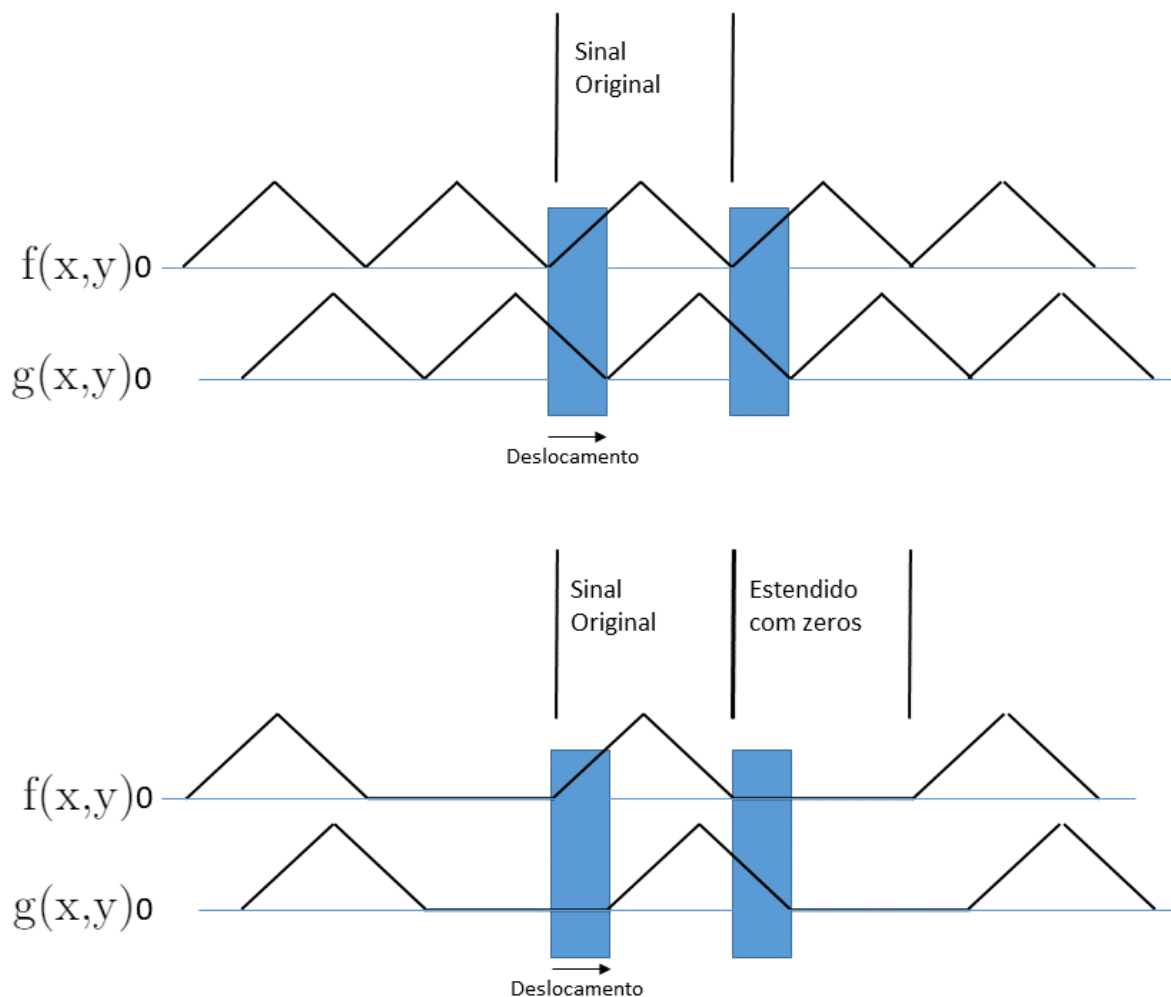


Figura 6 – Ilustração (em uma dimensão) da retroprojeção como resultado por assumir a natureza periódica de f e g na análise de Fourier. Aumentando o sinal com zeros elimina-se a contribuição de fora da janela espacial observada. Adaptado de (MARCOTTE, 1996).

Cabe aqui chamar a atenção para este passo do acréscimo de zeros nas funções discretas $f(x, y)$ e $g(x, y)$. Um dos pontos cruciais no método proposto nesta dissertação é o preenchimento das funções discretas por outros valores que não zero. O preenchimento com outros valores será descrito em maiores detalhes no capítulo no capítulo [Capítulo 3](#).

De toda forma, FFT é um algoritmo eficiente para computar transformadas discretas de Fourier. A partir de uma matriz retangular de tamanho $N \times M$ e acrescentando zeros para $(2N - 1) \times (2M - 1)$ - assumindo que $M < N$ - a complexidade computacional é:

$$(2M - 1) \cdot (2N - 1) \log_2(2N - 1), \quad (2.13)$$

A tabela 1 apresenta os valores obtidos de complexidade de operações para uma matriz quadrada de tamanho $N \times N$ utilizando a abordagem tradicional, a abordagem espacial de grid customizado e FFT. Para a equação 2.10, oito direções foram consideradas e um espaçamento máximo de $N/2$. Verifica-se na tabela 1, que o algoritmo FFT é mais rápido que ambos métodos para tais valores de N . Por exemplo, dado N igual a 128, FFT é 8 vezes mais rápido do que o método espacial (de oito direções). Para uma imagem de 512×512 , FFT pode ser de 30 a 300 vezes mais rápido do que a abordagem espacial dependendo do número de direções e espaçamento escolhido (MARCOTTE, 1996).

Tabela 1 – Complexidade das operações (Adaptado de (MARCOTTE, 1996)).

$N(N \times M)$	Abordagem espacial para todos os pares $(N \times M) \times (N \times M - 1)/2$	Abordagem espacial com oito direções, com espaçamento máximo de $N/2$ lags $\sum_k \sum_l (N-k) \times (N-l)$	Abordagem espectral $(2M-1) \times (2N-1) \log_2(2N-1)$
2	6	6	14
4	120	90	138
8	2,016	868	879
16	32,640	7,560	4,761
32	523,776	62,992	23,724
64	8,386,560	514,080	112,720
128	134,209,536	4,153,408	519,833
256	2,147,450,880	33,390,720	2,349,353
512	34,359,607,296	267,780,352	10,463,815
1024	549,755,289,600	2,144,862,720	46,089,347

Marcotte foi essencial para introduzir o uso da transformada rápida de Fourier no campo da análise de continuidade espacial na geoestatística. Seu objetivo era uma computação eficiente de variogramas, correlogramas, entre outras estatísticas. Entretanto, a referência fundamental desta dissertação é o trabalho desenvolvido por Yao (YAO; JOURNAL, 1998). Este trabalho, por sua vez, foca diretamente na descrição do fenômeno espacial e não somente em uma computação eficiente. Resumidamente, Yao propõem transformar a tabela de covariância (direta ou cruzada) experimental em uma tabela de quasi-densidade espectral utilizando a transformada rápida de Fourier (FFT). Esta tabela de quasi-densidade espectral é então suavizada sob a condição de ser positiva definida e somatória igual a um; o produto é então retro-transformado via FFT para obter a tabela de covariância (YAO, 1998a). A seguir, será feito uma revisão detalhada de seu trabalho,

para comentar as principais contribuições, fragilidades e como o método se compara ao estudo proposto e desenvolvido nesta dissertação.

2.3 Modelagem Automática de Tabelas de Covariâncias (Cruzadas) utilizando Transformada Rápida de Fourier

Em 1998, Tingting Yao propõem 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. O algoritmo requer que o usuário defina um certo número de parâmetros, como diversos tamanhos de janelas, similar aos parâmetros de tolerância para o cálculo do variograma experimental (YAO, 2000). De toda forma, o seu trabalho contribuiu de maneira expressiva para o tema; por este motivo, uma revisão detalhada de seu artigo foi feita e será apresentada a seguir.

Yao argumenta que uma alternativa não paramétrica para modelar analiticamente a covariância poderia ser baseada no Teorema de Bochner (BOCHNER, 1939). A simplificação do Teorema feita por Yao, em seu artigo, afirma que uma função $C(h)$ é positiva definida se e somente se esta pode ser expressa como uma transformada de Fourier de uma medida limitada $S(w)$ positiva não-decrescente:

$$C(h) = \int_{\mathfrak{R}^d} e^{2\pi i/h \cdot w} dS(w), \quad (2.14)$$

onde h é um vetor distância em um espaço \mathfrak{R}^d de d dimensões e w é o vetor frequência no domínio correspondente de frequência. Caso a função estacionária randômica (RF) possua um modelo $Z(u)$ com uma variância unitária, isto é, $C(0) = Var Z(u) = 1$, então $S(w)$ pode ser visto como uma função de distribuição cumulativa (cdf), na forma $dS(w) = s(w)dw$. O teorema de Bochner necessita então que a densidade espectral $s(w) = dS(w)/dw$ seja a função de probabilidade de densidade (pdf), isto é, de forma que:

$$s(w) \geq 0, \forall w \quad (2.15)$$

$$\int_{\mathfrak{R}^d} s(w)dw = 1, \quad (2.16)$$

$s(w)$ é a densidade espectral da RF $Z(u)$. A integral descrita pode ser aproximada em

uma somatória discreta:

$$C(h) = \sum_{k=1}^K e^{2\pi i/h \cdot w} s(w_k), k = 1, \dots, K \quad (2.17)$$

Em (REHMAN, 1995), é proposto uma solução para determinar as séries $s(w_k)$. Entretanto, para casos anisotrópicos, cujo número de casos reais é a grande maioria, o algoritmo proposto por Rehman requer uma discretização na ordem de k^3 termos, o que leva a problemas de convergência. Por este motivo, soluções analíticas são evitadas, conforme Yao nos apresenta em seu trabalho.

A ideia fundamental de Yao é ainda baseada no teorema de Bochner, mas não requer ajustar uma longa série de valores de $s(w_k)$ descrito em 2.17. O resultando final é uma tabela de covariância permissível para a construção de sistema de krigagens a ser usada durante uma estimativa ou simulação. Esta tabela de covariância deve ser positiva definida e a chave para tal ajuste “automático” é a transformada rápida de Fourier (FFT), pois este é feito no espaço de frequência. Uma primeira FFT transforma a pseudo ¹ tabela ou mapa de covariância experimental em uma quasi-densidade espectral, um algoritmo de suavização é então aplicado para se obter uma tabela de densidade espectral lícita. Esta densidade espectral pode ser usada diretamente para simulações espectrais ou pode ser inversamente transformada pela FFT em uma tabela lícita permissível de covariância para estimativas ou simulações no domínio espacial dos dados.

A tabela de covariância experimental (obtida diretamente dos pares de amostras disponíveis) é normalmente não positiva definida, pois os valores foram gerados a partir de um processo que não teve como prioridade a criação de dados em que houvesse uma inversão de matriz sem instabilidade numérica. Caso utilizada, pode acarretar em matrizes singulares de krigagem ou ainda estimar variâncias negativas. Além disso, nem todos os valores de covariância necessários estão disponíveis, por este motivo Yao argumenta que valores amostrais de covariância devem ser interpolados para preencher toda tabela necessária.

Uma importante diferença entre o estudo de Yao e o método proposto nesta dissertação está em calcular a tabela de covariância experimental e interpolar as covariâncias faltantes. O custo computacional para calcular a tabela experimental sem o uso da FFT utilizando somente os dados disponíveis é enorme, conforme visto em (MARCOTTE, 1996) na tabela 2.13. Por isso, é proposto o preenchimento de todo grid com valores estimados do atributo em questão, para ser posteriormente transformado via FFT. Novamente, a interpolação proposta por Yao pode somente prover o que é chamado de pseudo tabela de covariância, pois não necessariamente obedece a condição de ser positiva definida.

¹ Por “pseudo” tabela de covariância, Yao et al. referem-se a uma tabela que talvez não cumpra todas as condições de ser positiva definida. De maneira similar, uma tabela de “quasi” densidade espectral pode não cumprir a condição 2.16.

Através da transformada rápida de Fourier, uma pseudo tabela de covariância $C(h)$ pode ser transformada em uma pseudo tabela espectral $s(w)$. Em 2D esta tabela espectral é definida pela integral:

$$s(\omega_1, \omega_2) = \frac{1}{4\pi^2} \int_{-\infty}^{+\infty} dh_1 \int_{-\infty}^{+\infty} C(h_1, h_2) \exp[-i(\omega_1 h_1 + \omega_2 h_2)] dh_2, \quad (2.18)$$

Em sua versão discreta:

$$s(j_1, j_2) = \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} C(k_1, k_2) \exp[-2\pi i(k_1 j_1 + k_2 j_2)], \quad (2.19)$$

$$j_1 = 0, \dots, N_1 - 1, j_2 = 0, \dots, N_2 - 1$$

De acordo com o teorema de Bochner, a forma geral de uma função real positiva definida de $C(h_1, h_2)$, continua em $h = 0$ é:

$$c(h_1, h_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \cos(\omega_1 h_1 + \omega_2 h_2) \cdot dS(\omega_1, \omega_2) \quad (2.20)$$

onde $dS(\omega_1, \omega_2)$ é uma medida limite positiva.

Em função da tabela de covariância amostral não ser definida positivamente, a tabela de densidade espectral amostral pode não ser lícita e, dessa forma, resultar em valores para s menor do que 0 e a soma destes valores ser diferente de 1. Para assegurar a positividade de uma tabela 2D $s(\omega_1, \omega_2)$ e soma unitária, é possível uma suavização dos valores originais de s sob estas duas condições. E tal suavização também remove valores de flutuações não desejadas. O resultado é agora uma pdf espectral lícita. A retro transformada FFT desta pdf espectral será uma tabela de covariância suavizada e positiva definida.

Yao resume o método de obtenção da tabela de covariância final como uma ida-e-volta via FFT mais um processo de suavização intermediária. Diferente da aproximação por série de Fourier para com os valores originais de covariância descrito na relação 2.17. Yao afirma que o algoritmo proposto é mais rápido e flexível (e mais próximo dos valores de covariância amostral) do que o algoritmo de Rehman, dado o contexto da década de 90. Porém, mesmo hoje com níveis de processamento muito maiores, observa-se que são algoritmos lentos e custosos computacionalmente, acarretando em inviabilidade para a grande maioria dos banco de dados observados nas ciências da Terra.

Após discorrer sobre as preposições iniciais para a utilização do método, o passo inicial é o calculo da tabela covariância experimental. A partir das amostras, em grid ou ponto, a tabela de covariância discreta pode ser calculada para todos os espaçamentos disponíveis (CHU, 1993). Como para qualquer cálculo desse tipo, uma tolerância apropriada deve ser escolhida para um dado vetor h a fim de evitar uma estimativa de valores com

poucos pares. O mapa (ou tabela) é posteriormente interpolado para todas as distâncias. O tamanho e discretização da tabela de covariância depois da interpolação deve ser tal, que seja possível encontrar valores de covariância necessários para os futuros sistemas de krigagem.

A fórmula para o cálculo do mapa de covariância:

$$C_Z(h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} Z(u_\alpha)Z(u_\alpha + h) - m_{Z-h}m_{Z+h}, \quad (2.21)$$

onde m_{z-h} é a média dos pares da cauda do vetor e m_{z+h} da cabeça do vetor. Vale lembrar que a auto-covariância é simétrica em h e $-h$.

Já, a fórmula para o cálculo do mapa de covariância cruzada é:

$$C_{ZY}(h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} Z(u_\alpha)Y(u_\alpha + h) - m_{Z-h}m_{Y+h}, \quad (2.22)$$

$$C_{YZ}(h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} Y(u_\alpha)Z(u_\alpha + h) - m_{Y-h}m_{Z+h}, \quad (2.23)$$

Por sua vez, para a covariância cruzada, $C_{ZY}(h) \neq C_{ZY}(-h)$ e $C_{ZY}(h) = C_{YZ}(-h)$.

Por mais que seja possível gerar as tabelas de covariância cruzada no método proposto nessa dissertação, a possibilidade será meramente comentada. É possível verificar nas equações 2.22 e 2.23, que a tabela gerada não é simétrica e a ordem dos fatores altera produto.

O programa *Varmap* do GSLIB pode ser utilizado para calcular o mapa de covariância (DEUTSCH; JOURNAL, 1997). É uma técnica válida, porém bastante limitada e computacionalmente custosa. Em função do baixo número de dados, o número de pares utilizado para certas células do mapa de covariância pode ser muito pequeno e assim introduzir ruído ao mapa. Por isso, a importância em permitir tolerância sobre a aquisição de dados para o cálculo dos valores de covariância amostral e fornecendo uma suavização preliminar (CHU, 1993).

Em função dos dados serem esparsos, podem haver grandes flutuações e valores faltantes para o cálculo da tabela de covariância original. Yao propõem que a tabela deve ser completada e suavizada antes de ser transformada via FFT. Em seu estudo, Yao utilizou um programa chamado *intpmap* que permite o preenchimento dos valores faltantes ao mesmo tempo que os valores originais do mapa de covariância são suavizados. Para cada entrada, o valor suavizado é uma média ponderada dos valores originais através de uma janela móvel.

O peso associado a cada valor original é proporcional ao número de pares utilizados para originá-lo e decai com a distância em relação ao centro da janela. Para cada covariância de entrada na fileira k_1 e coluna k_2 com coordenadas polares (ρ, θ) , a janela de suavização é um círculo com ângulo $\theta + \Delta \theta$ e seu raio $(1 \pm \omega) \rho$. O valor para a covariância suavizada é:

$$C(k_1, k_2) = \sum_{i=1}^N \lambda_i C_i, \quad (2.24)$$

onde o N é o número total de valores originais de entrada encontrados pela janela de suavização, C_i é o valor para i covariâncias e λ_i os pesos correspondentes definidos por:

$$\lambda_i = np(i)/dist(i), \text{ with } : \sum_{i=1}^N \lambda_i = 1 \quad (2.25)$$

onde $np(i)$ é o número de pares utilizado para o cálculo para o valor de covariância experimental original C_i , $dist(i)$ é a distância entre os i valores de covariância e a entrada $C(k_1, k_2)$ é então suavizada.

O raio da janela $(1 +/ - \omega) \rho$ aumenta com a distância ρ , da mesma forma, o aumento da suavização com a distância lag. De fato, normalmente há uma flutuação indesejada experimental para grandes lags e era necessária sua remoção. Os parâmetros ω e $\Delta \theta$ devem ser escolhidos tais que incluam dados suficientes para cada círculo de suavização. E, uma vez que tabelas de autocovariância são simétricas, ou seja, $C(h) = C(-h)$, é somente necessário calcular metade do mapa de covariância.

Uma vez que a tabela de covariância amostral está completa, a FFT fornece o mapa de pseudo densidade espectral correspondente. A expressão 2.19 considera uma ponderação igual para todos os valores amostrais de covariância $C(k_1, k_2)$. Entretanto, como a confiança dos valores de covariância $C(k_1, k_2)$ decrescem conforme cresce k_1, k_2 , é mais apropriado dar mais peso de forma a favorecer os valores de maior confiança e menor lag k_1, k_2 , ou seja:

$$\hat{s}(j_1, j_2) = \sum_{k_1=0}^{M_1} \sum_{k_2=0}^{M_2} \lambda_{k_1, k_2} C(k_1, k_2) \exp[-2\pi i(k_1 j_1 + k_2 j_2)], \quad (2.26)$$

$$j_1 = 0, \dots, N_1 - 1, j_2 = 0, \dots, N_2 - 1$$

onde M_1, M_2 são os limites truncados definindo a somatória dada pela janela e λ_{k_1, k_2} são pesos a serem aplicados pela janela. Dois tipos de pesos são frequentemente usados (CHATFIELD, 1996) e são eles:

- Janela de Turkey-Hanning:

$$\lambda_{k_1, k_2} = \frac{1}{2} \left(1 + \cos \pi \left(\frac{k_1}{M_1} + \frac{k_2}{M_2} \right) \right), \quad (2.27)$$

$$k_1 = 0, \dots, M_1, k_2 = 0, \dots, M_2$$

- Janela de Parzen:

$$\lambda_{k_1, k_2} = 1 - 6 \left(\frac{k_1}{M_1} + \frac{k_2}{M_2} \right)^2 + 6 \left(\frac{k_1}{M_1} + \frac{k_2}{M_2 - 2} \right)^3 \quad (2.28)$$

$$0 \leq k_1 \leq M_1/2, 0 \leq k_2 \leq M_2/2$$

$$2 \left(1 - \frac{k_1}{M_1} + \frac{k_2}{M_2} \right)^3$$

$$M_1/2 \leq k_2 \leq M_1, M_1/2 \leq k_2 \leq M_2/2$$

Os resultados para dois primeiros métodos estimam valores similares de densidade espectral. De toda forma, o tamanho da janela (M_1 e M_2) é difícil de determinar e Yao admite em seu artigo que a melhor estratégia é por tentativa e erro. Este fato evidencia a importância de uma técnica automatizada; pois, ao errar o valor do tamanho da janela, faz-se necessário uma nova tentativa e o dobro de tempo computacional gasto, sem saber em que valor cessar as tentativas.

De toda forma, uma janela móvel simples também pode ser utilizada:

$$\hat{s}(j_1, j_2) = \frac{1}{m_1 m_2} \sum_{l_1=-m_1/2}^{m_1/2} \sum_{l_2=-m_2/2}^{m_2/2} s(j_1 + l_1, j_2 + l_2), \quad (2.29)$$

onde m_1 e m_2 definem uma janela de suavização simétrica em torno das células (j_1, j_2) .

Resumindo: o algoritmo de suavização processa os dados da seguinte maneira:

- Através da FFT, um mapa amostral de covariância em grid é transformado em um grid de quase-densidade espectral pdf $s(w)$ utilizando a expressão 2.19;
- É feita uma suavização de $s(w)$ através de médias móveis do tipo 2.29 sob as condições de positividade e soma unitária para assegurar uma pdf espectral lícita e suavizada; As condições para o mapa de densidade espectral são:

- Os valores de densidade espectral são reais e não-negativos;
- A integral para toda a densidade espectral em todas as frequências w deve ser igual a $C(0)$, portanto, se o mapa de covariância for estandardizado para uma variância unitária, então em \mathfrak{R}^2 : $\sum_{k_1=1}^{N_1} \sum_{k_2=1}^{N_2} s(k_1, k_2) = 1$;

- Além disso, a densidade espectral deve ser simétrica na frequência zero para assegurar que a transformada reversa de Fourier resulte em valores de covariância reais (não imaginários). A razão pela qual Yao afirma a necessidade de passar para o domínio espectral é a de que se torna mais simples obter uma condição de positividade para a densidade espectral em oposição ao fazê-lo no domínio espacial.
- iii. É possível utilizar o espectro $s(w)$ lícito e suavizado no domínio de frequências para simulações espectrais, seja não condicional (PARDO-IGUZQUIZA; CHICA-OLMO, 1993) ou condicional (YAO, 1998b), ou ainda pode ser retro transformado para o domínio espacial e utilizado como uma tabela de covariância lícita e permissível para sistema de krigagens ou simulações condicionais.

Ao capturar diretamente os valores da tabela de covariância lícita e permissível, é possível construir os sistemas de krigagem sem a necessidade de utilizar qualquer tipo de modelo analítico de covariância. Todos os programas tradicionais de estimativa e simulação podem ser modificados para lerem diretamente desta tabela de covariância. Os parâmetros para os modelos de variogram ou covariância nos arquivos de parâmetros para estimativa e simulação no programa GSLIB, por exemplo, podem ser substituídos por um arquivo impute de dados contendo o mapa de covariância lícito em 2D ou um cubo nos casos 3D. Yao ressalta que, mesmo para os casos 3D, não haveria uma preocupação com o tamanho do cubo de covariância, pois na prática, os dados utilizados pela krigagem estarão limitados pela vizinhança de busca centrada no ponto a ser estimado e a uma máxima distância h .

O método proposto por Yao foi, de fato, inovador e ousado. Sua abordagem foi apresentada como uma modelagem automática da tabela de covariância, de forma a entregar uma tabela definida positiva. Entretanto, conforme foi apresentado na revisão, este método é bastante complexo e não automático. Diversos parâmetros são necessários e ainda há o risco de uma suavização excessiva, de modo a comprometer inclusive uma possível anisotropia dos dados. Por fim, vimos até aqui uma breve revisão sobre as transformadas de Fourier, na teoria e na prática, o método proposto por Marcotte e uma densa revisão bibliográfica do método Yao. Seguimos para o capítulo em que a metodologia proposta é apresentada.

3 Tabela de covariância: um mapeamento rápido e automático de continuidade espacial

Uma vez que todas as técnicas e conceitos fundamentais foram apresentados, é possível descrever o fluxo de trabalho proposto. O primeiro passo é preencher o grid com valores estimados para construir um modelo base para a extrair a covariância (MBC) aceitável para executar o algoritmo de FFT. Este algoritmo precisa de um grid regular sem nós ausentes para gerar a tabela de covariância. O primeiro passo representa a proposta fundamental deste estudo. Ao invés de interpolar os valores de covariância faltantes do conjunto de dados - e ainda por cima suavizá-los - conforme descrito por Yao em (YAO; JOURNAL, 1998), é proposto usar um MBC que reproduza satisfatoriamente as características espaciais dos dados, preenchendo por completo os nós do grid.

O segundo passo consiste em realizar a convolução do MBC via transformada de Fourier através do algoritmo FFT e obter a densidade espectral. Foi mencionado anteriormente, que a densidade espectral pode ser usada para simulação espectral, como por exemplo via método Integral de Fourier (PARDO-IGUZQUIZA; CHICA-OLMO, 1993), ou pode ser retrotransformada de volta para o domínio dos dados como uma tabela de covariância lícita para imputá-la em métodos de krigagem ou simulação geoestatística.

Este método é um mapeamento automático de continuidade espacial e não suprime passos anteriores, como a escolha de domínios estacionários, ou ainda posteriormente na análise quantitativa de vizinhança de krigagem (QKNA) (VANN; JACKSON; BERTOLI, 2003) e na otimização dos parâmetros de krigagem localizada (LKPO) (RUBIO et al., 2018). Será apresentado neste capítulo a importância da vizinhança de krigagem - ou elipsoide de busca - descrito em QKNA e LKPO para o método proposto. A seguir, cada passo descrito acima será explicado e explorado.

Por fim, o uso da tabela de covariância exige que os algoritmos de krigagem e simulação sejam modificados. Além disso, a incorporação da tabela de covariância foi feita migrando os dados para o grid com mesma resolução. Dessa forma, todos os dados estão nas distâncias em que as covariâncias estão tabeladas.

3.1 Modelo Base para Extrair Covariância

É importante, primeiramente, apresentar a definição de modelo base para extrair covariância (MBC) no contexto deste trabalho. A continuidade espacial, ou a falta dela, é uma característica intrínseca de atributos das ciências da Terra. Portanto, um modelo que

carrega uma continuidade espacial semelhante poderia ser um modelo base para extrair uma tabela de covariância. Este trabalho concentra-se na estimativa de um MBC usando uma única variável, mas é possível usar dados correlacionados secundários, exaustivos ou não, e até mesmo uma imagem de treinamento para obter uma tabela de covariância.

Embora questionável quanta similaridade é necessária entre o modelo que desejamos e seu MBC para obter uma boa tabela de covariância, o ponto crucial deste trabalho é entender que mesmo um modelo construído com algum método muito simples pode conter características similares de continuidade espacial em comparação com a tabela de covariância que desejamos. Foi escolhido o método do vizinho mais próximo (NN) após uma extensa investigação sobre diversos possíveis modelos, dentre eles: inverso da distância linear e quadrático, k -vizinhos mais próximos (KNN) com diversos valores de k . Obtêve-se tabelas de covariância via NN, ou KNN para $k = 1$. Além de sua simplicidade e estar contido na grande maioria dos softwares geoestatísticos, seu sumário estatístico tende a ser representativo dos dados desagrupados, preservando sua variância. De toda forma, é importante ressaltar que o MBC via NN pode distorcer a tabela de covariância caso haja grandes extrapolações.

A figura 7 ilustra a comparação de dois modelos muito diferentes em termos de resolução. O primeiro é o modelo exaustivo Walker Lake (a) e o segundo é um modelo estimado via vizinho mais próximo (b), obtido a partir do conjunto de 470 dados (veja (ISAACS; SRIVASTAVA, 1989)). A figura 8 mostra duas tabelas de covariância, a primeira foi obtida utilizando o modelo exaustivo (a) e a segunda através do MBC (b). A diferença entre ambas tabelas de covariância é bastante sutil, especialmente no centro da imagem, que será realmente capturada pelo elipsoide de busca ao executar métodos de krigagem.

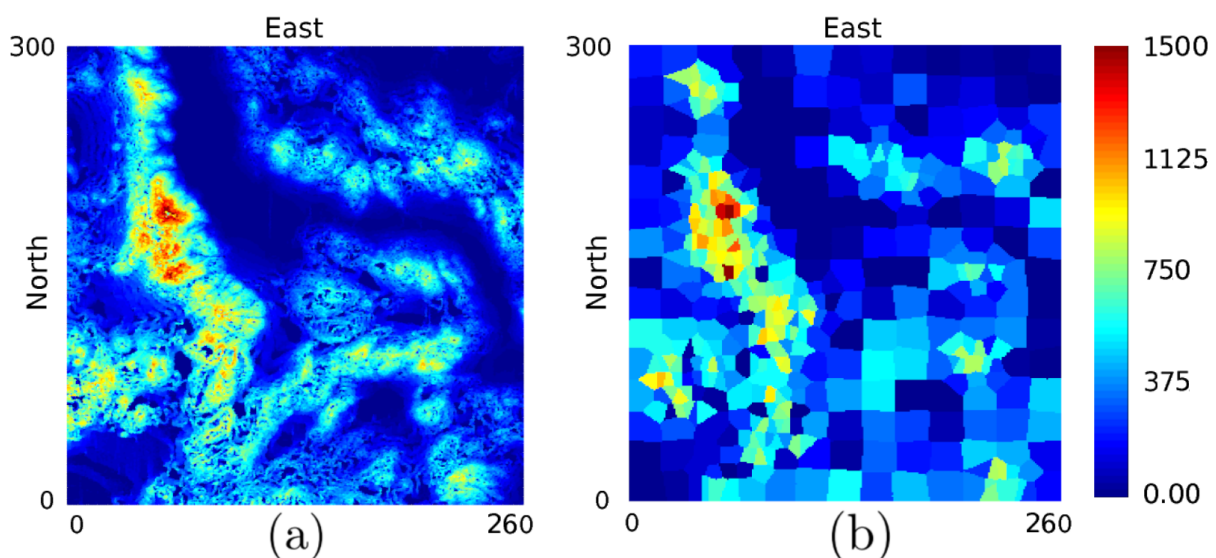


Figura 7 – Modelo exaustivo (a) e MBC (b).

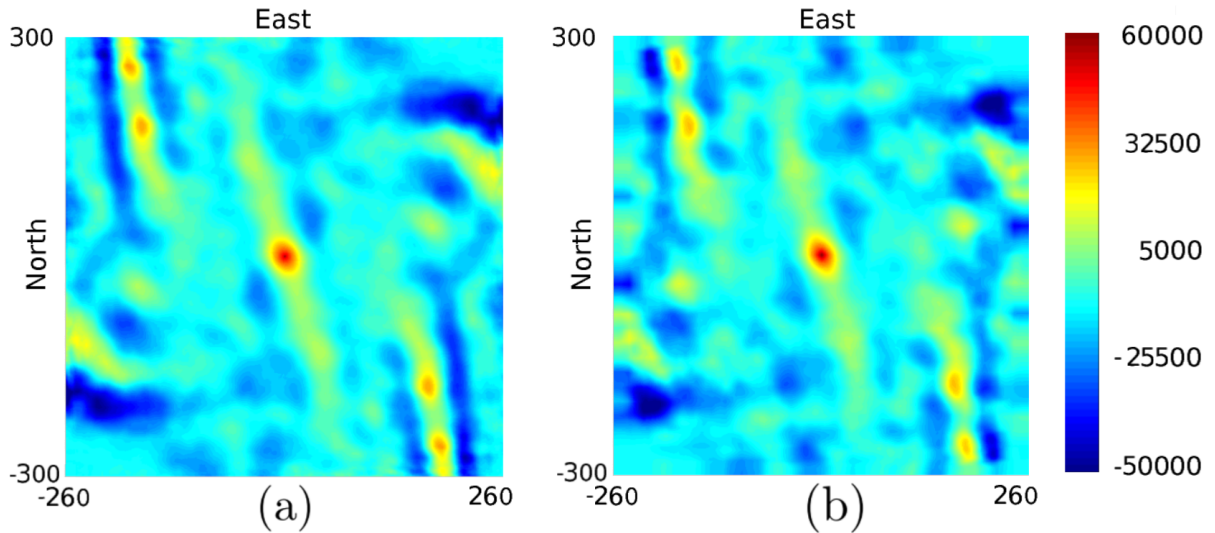


Figura 8 – Tabela de covariância do modelo exaustivo (a) e do MBC (b).

Nesta etapa do estudo, serão apresentadas duas estimativas via krigagem ordinária (OK) para fins ilustrativos, e, no [Capítulo 4](#), o tema será aprofundado. A figura 9 apresenta dois mapas estimados através de krigagem ordinária, ambas com os mesmos parâmetros, com exceção do modelo de continuidade espacial, para um grid de 1×1 - mesma dimensão de bloco da tabela de covariância. O primeiro mapa estimado (a) utilizou a tabela de covariância extraída do MBC via NN, já o segundo mapa utilizou a variografia tradicionalmente calculada e modelada (b) para os 470 dados. As diferenças entre as estimativas feitas podem ser devido a tabela de covariância não considerar o efeito pepita, ou seja, para $h = 0$, a variância é 0 e covariância a máxima. Dessa forma, a tabela atribui maior peso a amostra mais próxima. O modelo variograma, por outro lado, tem o efeito pepita e distribui melhor os pesos entre as amostras, o que pode gerar uma maior suavização.

Para estimar em grids de diferente resolução da tabela de covariância, blocos de 5×5 dada uma tabela de blocos de 1×1 , por exemplo, é possível utilizar a operação de *upsampling*, ou seja, extrair as médias para os blocos pertencentes ao bloco maior. Os mapas presentes na figura 10 exemplificam a operação para um grid com blocos 5×5 . É notável a semelhança do mapa estimado via tabela com o modelo exaustivo e isso se dá, pois, a tabela de covariância pode conter diversas subestruturas extraídas do MBC, que serão reproduzidas em todas as estimativas e simulações decorrentes dessa tabela. Além da questão já comentada dado efeito pepita. Da mesma forma, nota-se significativas diferenças entre os mapas estimados (a) e (b), principalmente nas regiões marcadas de alto teor (elipse vermelha) e de baixo teor (elipses amarelas).

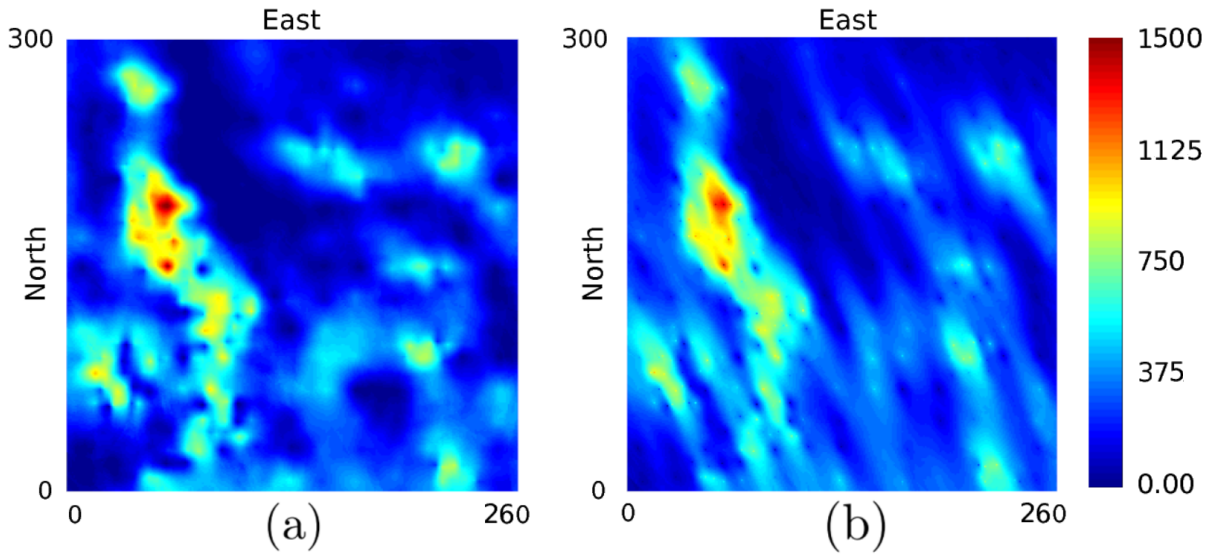


Figura 9 – OK via tabela de covariância (a) e via variograma (b) em grid 1x1.

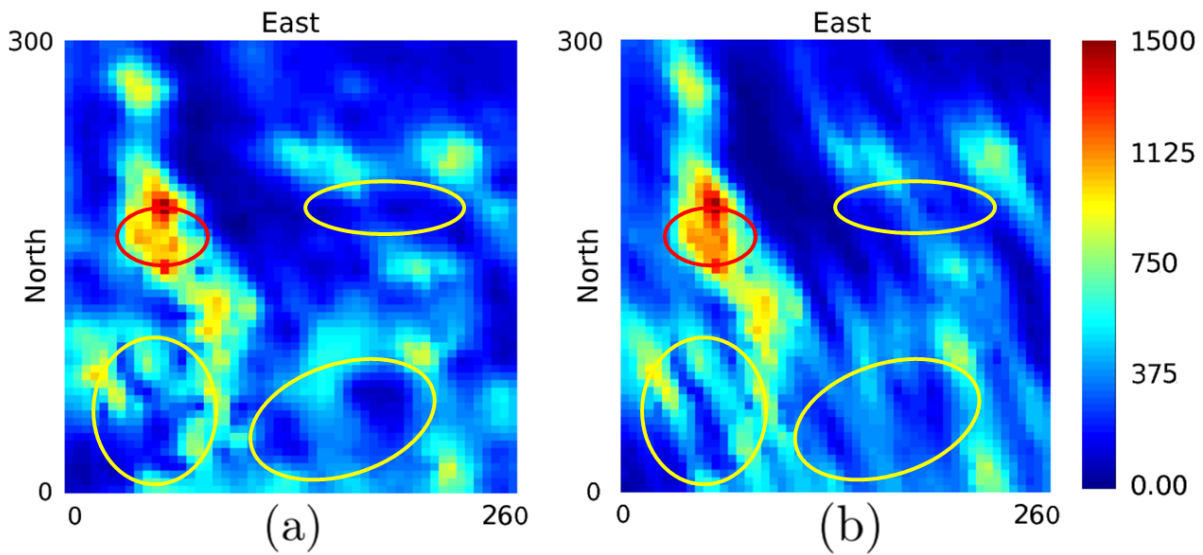


Figura 10 – OK via tabela de covariância (a) e via variograma (b) em grid 5x5, com elipses vermelhas destacando diferenças em áreas de alto teor e amarelas para baixos teor.

3.2 Gerando a Tabela de Covariância

Conforme foi visto no capítulo anterior, a tabela de covariância (CT) consiste em literalmente calcular valores de covariância para todos os espaçamentos hs , definida na forma:

$$C_{ZW}(h) = \frac{1}{N(h)} \sum_{\alpha=1}^{N(h)} Z(u_{\alpha})W(u_{\alpha} + h) - m_{Z-h}m_{W+h}, \quad (3.1)$$

onde m_{Z-h} e m_{W+h} indica respectivamente os valores médios da cauda e cabeça do par de vetores de dados Z e W , $N(h)$ indica o número de pares que podem ser encontrados em um intervalo h . Utilizando a correspondência dada pelo teorema de convolução aplicando uma transformada de Fourier, temos:

$$\mathcal{F}[C_{ZW}(h)] = \mathcal{F}[Z] \cdot \overline{\mathcal{F}[W]}, \quad (3.2)$$

é possível obter a tabela de covariância desejada aplicando o retrotransformação na equação 3.2, resultando em:

$$C_{ZW}(h) = \mathcal{F}^{-1} \left[\mathcal{F}[Z] \cdot \overline{\mathcal{F}[W]} \right]. \quad (3.3)$$

A definição da tabela de covariância na equação 3.3 usa apenas transformações de Fourier e retrotransformação, tornando esta definição adequada para usar as vantagens numéricas da FFT. Embora a transformação de Fourier seja computacionalmente cara, cada transformação requer $O(N^2)$ multiplicações complexas ou operações (N o tamanho do vetor), a Transformada de Fourier Discreta (DFT) pode acelerar o processo para $O(N \log N)$ operações, conforme visto no capítulo anterior. Isto é possível devido ao algoritmo de Transformação de Fourier Rápida (FFT) que, dependendo da forma com que foi implementada, explora a simetria da definição discreta de transformação de Fourier.

A densidade espectral $s_{ZW}(\vec{k})$ é definida como a transformada de Fourier da função de covariância de Z e W :

$$\begin{aligned} s_{ZW}(\vec{k}) &= \mathcal{F}[C_{ZW}(h)], \\ s_{ZW}(\vec{k}) &= \mathcal{F}[Z] \cdot \overline{\mathcal{F}[W]}. \end{aligned} \quad (3.4)$$

A equação 3.4 define, portanto, a densidade espectral $s(\vec{k})$. Esta equação pode ser calculada como $O(n \log(n))$, com $n = n_x, n_y$ e n_z , sendo n o número de células para cada dimensão. A densidade espectral pode ser usada para realizar simulações espectrais, ou pode ser retrotransformada novamente através da FFT.

Como esse trabalho foca em uma continuidade espacial global de uma única variável, a convolução via FFT pode ser entendida como uma autocorrelação descrita na forma:

$$\begin{aligned} \mathcal{F}[C_{Z,Z}(h)] &= \mathcal{F}[Z] * \overline{\mathcal{F}[Z]} = |\mathcal{F}[Z]|^2, \\ A_{Z,Z}(h) &= \mathcal{F}^{-1} \left[|\mathcal{F}[Z]|^2 \right]. \end{aligned} \quad (3.5)$$

onde $A_{Z,Z}(h)$ denota a autocorrelação para a variável Z .

Conforme foi visto em (YAO; JOURNAL, 1998), modelos de covariância devem ser positivos definidos, provendo valores lícitos de covariância para qualquer espaçamento e direção. O método aqui proposto não garante uma tabela de covariância definida positiva. Entretanto, uma vez que todas as tabelas de covariâncias geradas neste estudo vieram de um modelo de base estimado a partir de um conjunto de dados, sempre foi encontrada uma região central lícita. Ainda assim, o método não garante que sempre haverá esta região. Essa área central lícita pode ser identificada e está diretamente relacionada ao fenômeno espacial. Portanto, ela pode ser utilizada uma vez limitada pelo elipsoide de busca em métodos de krigagem.

A figura 11 apresenta quatro tabelas de covariância utilizadas no estudo com suas respectivas áreas centrais lícitas, limitadas pelo elipsoide de busca. A primeira imagem (a) é a CT de um MBC com uma anisotropia zonal, e segunda (b) é uma CT de um MBC de efeito pepita puro. Ambos são casos extremos particulares, afim de ilustrar o método proposto em condições em que poderia haver uma instabilidade quanto a licitude dos valores, ou ainda uma dificuldade na identificação da região a ser utilizada pelo elipsoide de busca. A imagem (c) é a CT do Walker Lake e a última imagem (d) é a CT utilizada no estudo de caso do [Capítulo 4](#).

Ambos casos particulares, anisotropia zonal (a) e efeito pepita puro (b), foram postos a prova e o resultado pode ser visto na figura 12. O fluxo de trabalho começa gerando uma realização não condicionada através da simulação sequencial gaussiana (SGS) com uma estrutura de covariância inserida. O modelo resultante foi um MBC para a CT. Então, uma segunda SGS não condicional foi executada usando o CT para obter o modelo final.

Assim, é possível afirmar que o método de obtenção da tabela de covariância proposto é um mapeamento automático rápido de continuidade espacial. Com, literalmente, um apertado de botão e cerca de 90 segundos de tempo de processamento em um laptop com processador Intel i7, é possível gerar uma tabela de covariância para dois milhões de nós de grid.

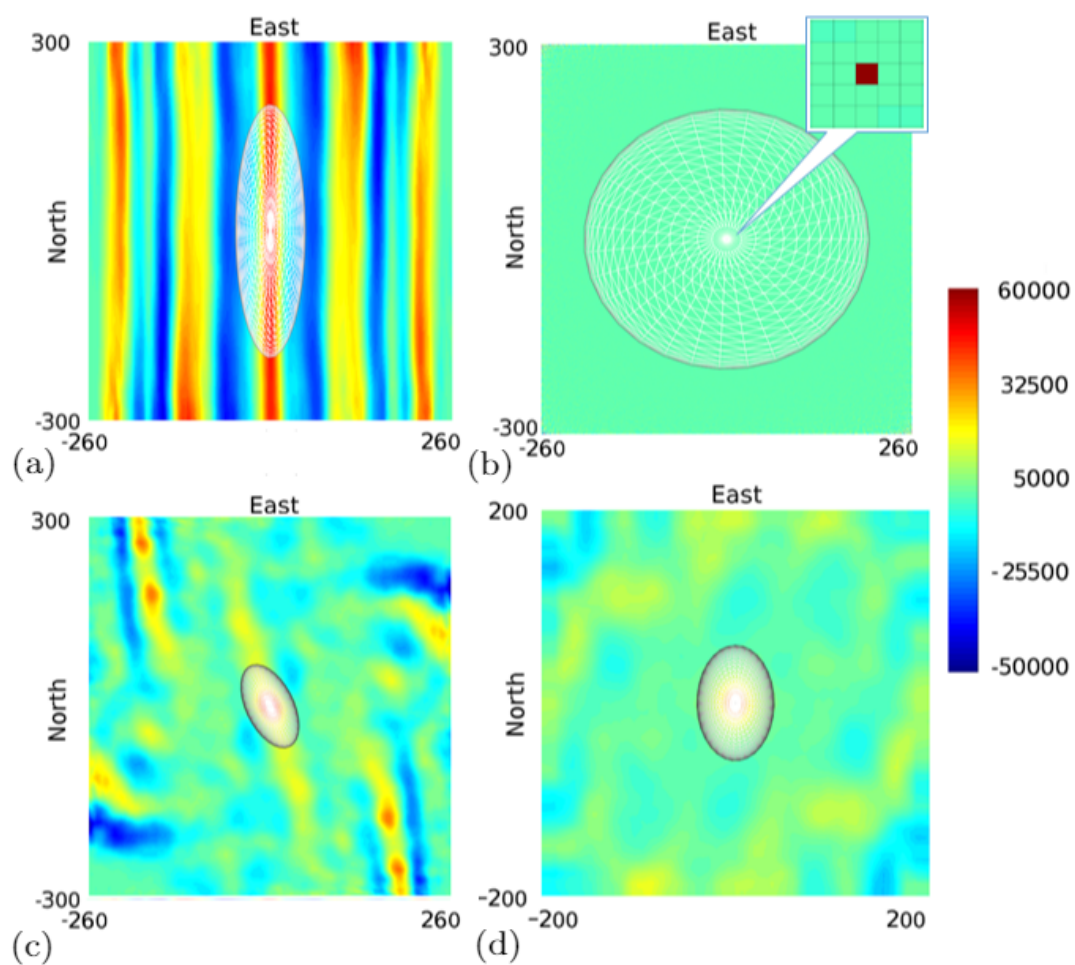


Figura 11 – CT de um MBC com uma anisotropia zonal (a), CT de um MBC de efeito pepita puro (b), CT do Walker Lake (c) e CT utilizada no estudo de caso no [Capítulo 4](#) (d). Os elipsoides na região central de cada CT são as elipses de busca para as krigagens e simulações subsequentes.

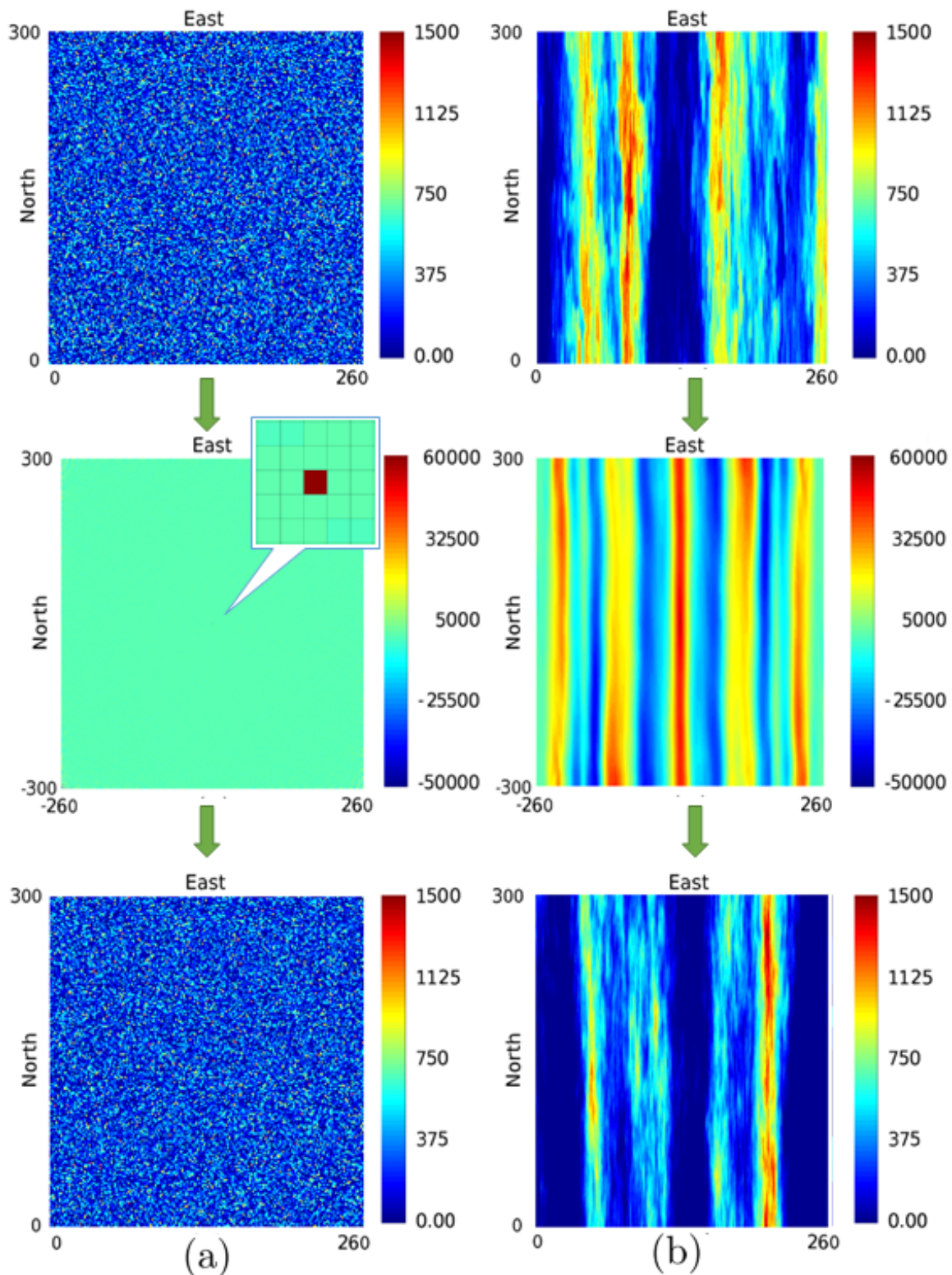


Figura 12 – Casos ilustrativos: (a) um modelo efeito de pepita puro, a realização de SGS como um MBC, a CT extraída dele e uma realização de SGS condicionada pela CT; (b) uma realização SGS de anisotropia zonal como MBC, a CT extraída e uma realização de SGS condicionada pela CT.

3.3 Algoritmo e Plug-in

O código implementado executa a convolução via FFT de um grid MBC, resultando em uma tabela de covariância associada e, portanto, necessita da criação prévia de um MBC (via NN nos casos aqui apresentados). Apesar de ser um assunto denso e complexo, o código é relativamente simples. Em poucas linhas, o pseudocódigo para a implementação do plug-in está descrito abaixo.

Require: \mathbf{d} é o banco de dados. Construir um MBC \mathbf{m} utilizando uma técnica de interpolação a partir do banco de dados \mathbf{d} .

$M \leftarrow \text{FFT} \{ \mathbf{m} \}$

for all (i, j, k) in $\text{index}(M)$

$s(i, j, k) = M(i, j, k)^2$

end for

$C \leftarrow \text{FFT}^{-1} \{ \mathbf{s} \}$

C é a tabela de covariância associada com o banco de dados \mathbf{d} .

O plug-in é de fácil utilização, bastando entender poucos parâmetros é possível gerar uma tabela de covariância. A figura 13 apresenta a janela do plug-in implementado no software *AR2GeMS* para geração da CT, descrevendo os parâmetros utilizados. Há duas entradas para MBC (parâmetro A e B), possibilitando a geração de tabelas de covariâncias cruzadas.

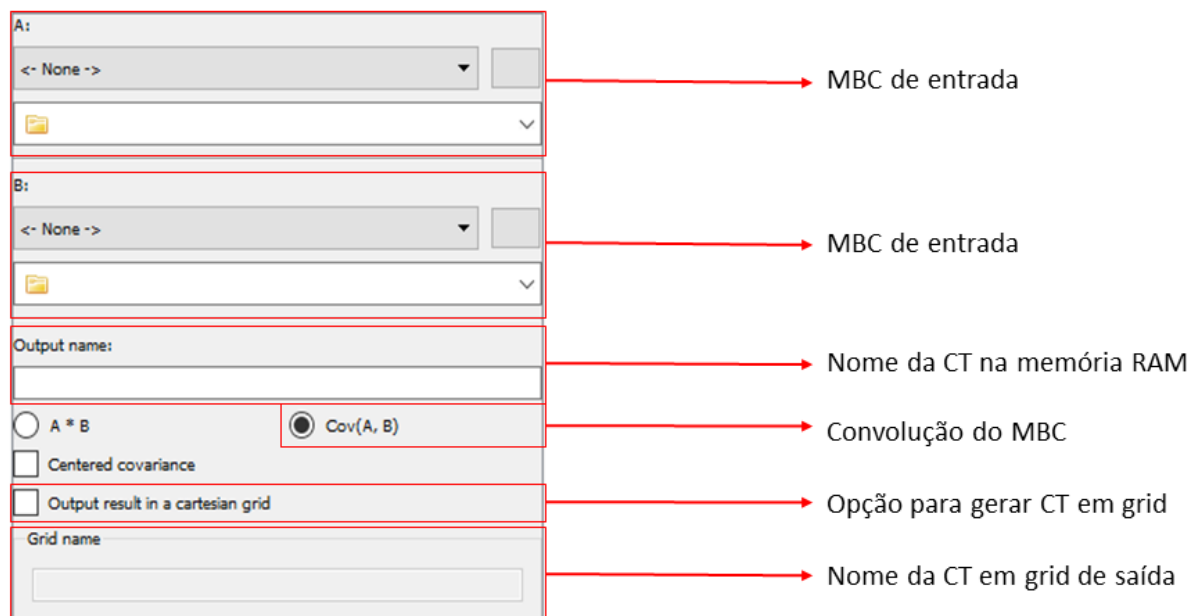


Figura 13 – Plug-in do software *AR2GeMS* para geração da CT.

4 Estudo de caso

Para a discussão de resultados, é proposto um estudo de caso. Foi escolhido demonstrar a técnica sob ambiente controlado, ou seja, gerando um depósito sintético em três dimensões e amostrando este em diferentes proporções. De modo que será possível a comparação dos resultados utilizando a técnica contra o método tradicional e ambos contra o depósito sintético. Também, se estuda como a baixa densidade amostral afeta a geração da tabela de covariância, novamente, em comparação com a variografia.

4.1 Geração do Depósito Sintético e Banco de Dados

Uma realização não condicional foi criada através da técnica de simulação sequencial gaussiana (SGS) que honra a estrutura de covariância de entrada, conforme descrito na tabela 2. A variável V foi criada e preenche por completo um grid de dimensões 200 blocos na direção Norte-Sul, 200 na direção Leste-Oeste e 50 blocos na direção vertical, e cada bloco possui dimensões de $1 \times 1 \times 1$.

Estrutura	Contribuição	Azimute	Maior	Intermediário	Menor
Efeito Pepita	0.10	-	-	-	-
Esférica	0.30	0	30	20	10
Esférica	0.60	0	60	40	25

Tabela 2 – Estruturas e parâmetros de continuidade espacial usados para gerar a realização via SGS.

Foram gerados três conjuntos de dados similares a furos reais com diferentes densidades amostrais: 1,0%, 0,5% e 0,1% do total de dados V no grid exaustivo. Os sumários estatísticos para os três conjuntos de dados representam estatisticamente os dados exaustivo. Isto foi feito para investigar a influência da densidade amostral sobre a técnica proposta em comparação com a modelagem tradicional do variograma. A figura 14 apresenta o modelo exaustivo em vista isométrica (a) e o banco de dados cobrindo 0.1% do total dos nós do grid (b). A figura 15 apresenta uma seção em plano horizontal para z igual a 1 (em 50 seções possíveis) em ambos, modelo exaustivo (a) e V 0.1% (b). Todas as seções para fim de comparação neste estudo de caso serão feitas para este valor de z . Uma vez que todos os furos começam em z igual a zero e terminam em 50, não haverá diferença na escolha da seção para comparação.

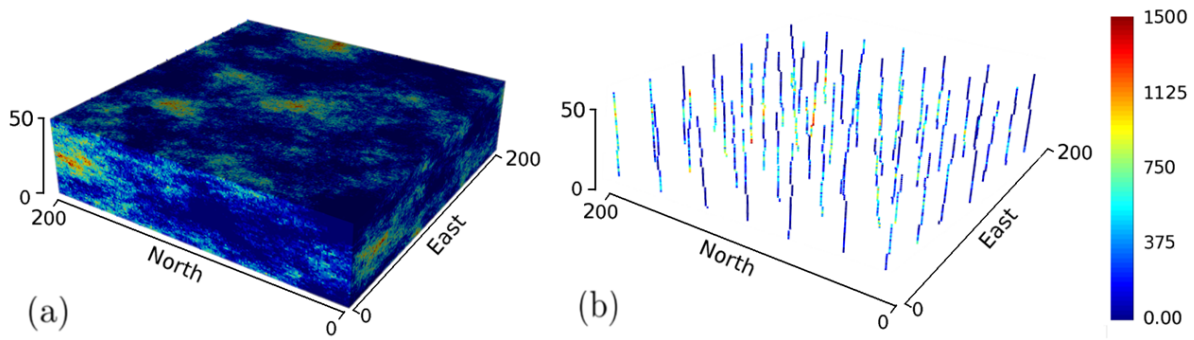


Figura 14 – Modelo exaustivo (a) e V 0.1% (b).

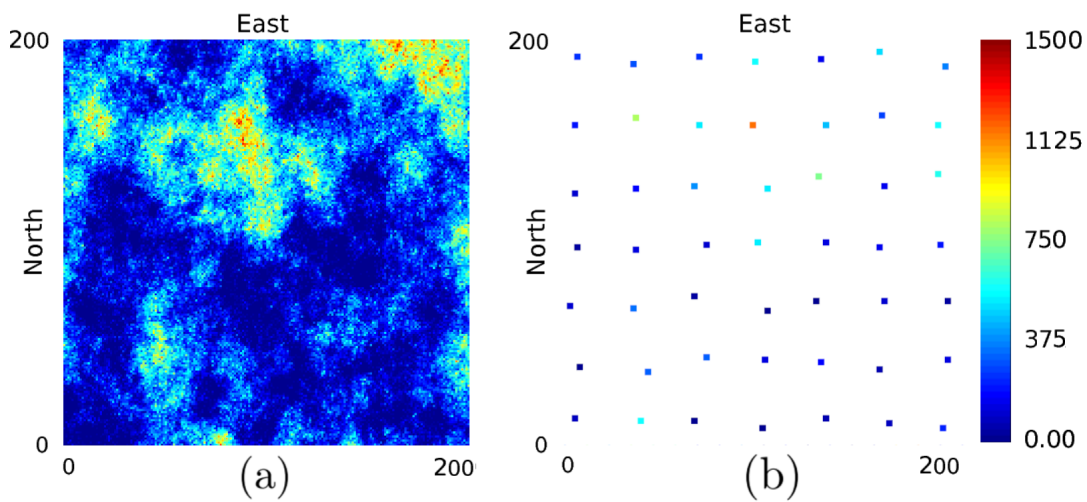


Figura 15 – Seção em plano horizontal para z igual a 1 do modelo exaustivo (a) e V 0.1% (b).

A figura 16 mostra o histograma dos dados exaustivos em comparação com o histograma das amostras. A informação mais relevante para o estudo nos histogramas é a variância total dos dados, que estará representada no centro da tabela de covariância para todos os casos.

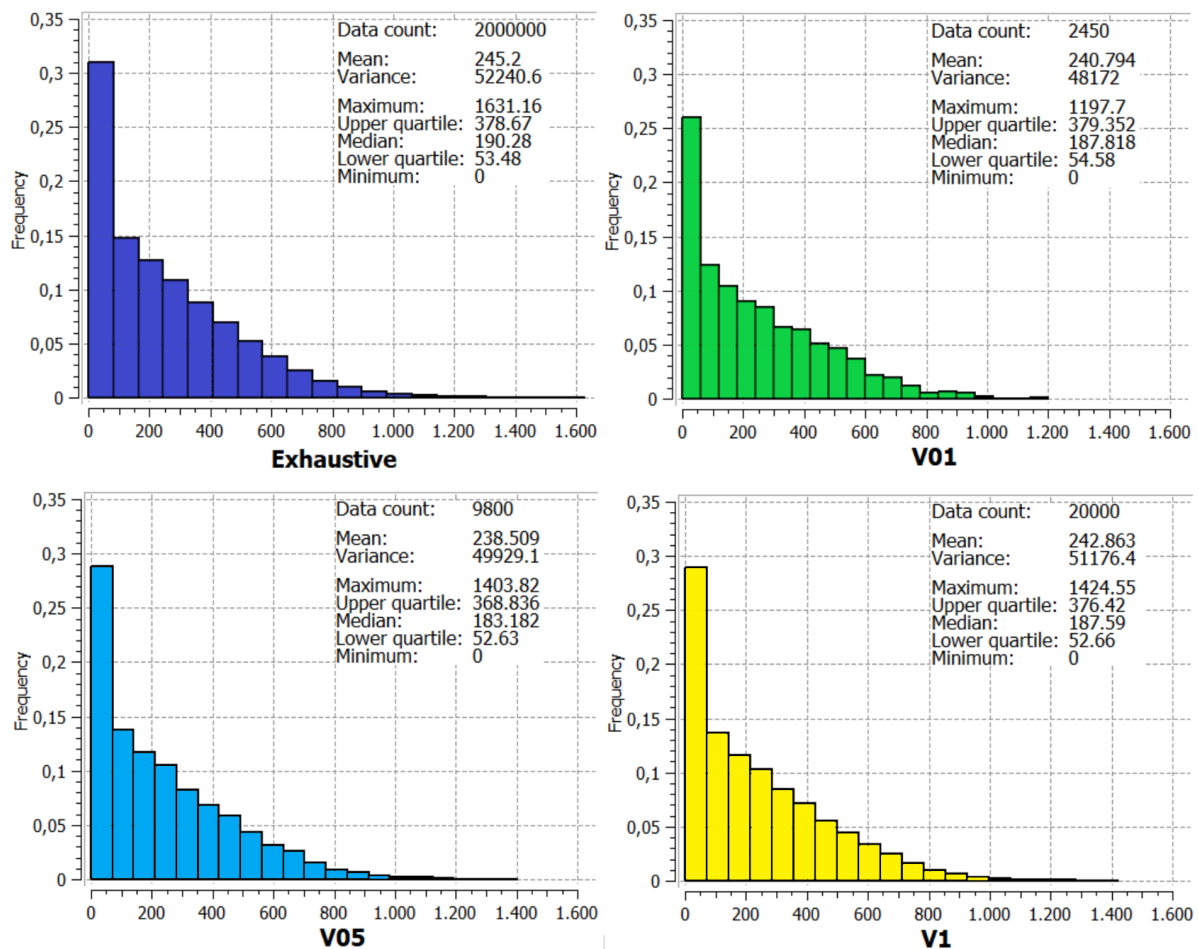


Figura 16 – Histograma do modelo exaustivo e dos dados com 0.1%, 0.5% e 1.0% de V (b).

4.2 Geração das Tabelas de Covariância

Conforme mencionado no capítulo anterior, o tempo para a geração da tabela de covariância para 2 milhões de nós é de aproximadamente 90 segundos em um laptop com processador Intel i7. As tabelas de covariância resultantes para este grid possuem 16 milhões de nós, pois, para descrever fenômenos tridimensionais, é necessário uma tabela 8 vezes maior do que o grid que contém o fenômeno. Isso ocorre em função da simetria necessária para a tabela representar o fenômeno para a mesma direção e em ambos sentidos. Da mesma forma que a variografia ao gerar um modelo simétrico, tal que $C(h) = C(-h)$.

Para fins ilustrativos, a figura 17 apresenta a tabela de covariância do modelo exaustivo com duas seções isométricas que se cruzam em x e z igual zero. O elipsoide que envolve a parte da CT que será incorporada no cálculo de krigagem, no canto superior direito da imagem, representa o elipsoide de busca. As dimensões da região são de aproximadamente 60 no maior alcance, 40 no alcance intermediário e 25 no menor alcance, dada a tabela que descreve o fenômeno 2.

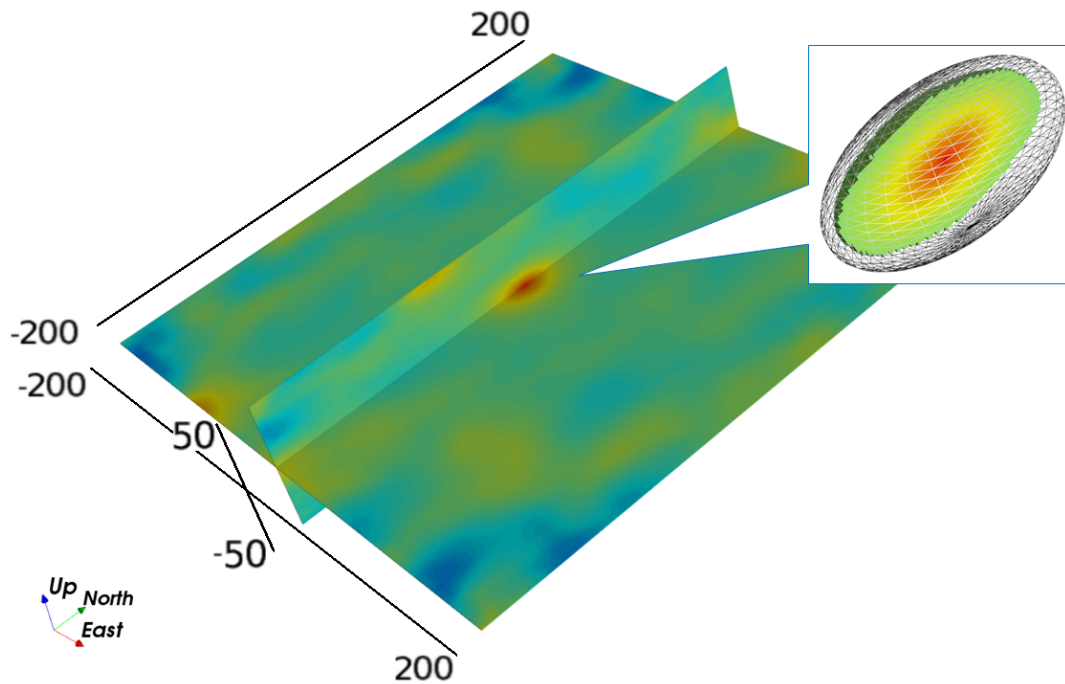


Figura 17 – CT para o modelo exaustivo em seção isométrica com x e z igual zero. O elipsoide que envolve a parte da CT que será incorporada no cálculo, no canto superior direito, representa o elipsoide de busca da krigagem.

A figura 18 mostra a CT para o modelo exaustivo em comparação com CT para todos os conjuntos de dados (V01, V05 e V1) em z igual a zero. Valor escolhido para apresentar a direção de maior continuidade espacial do fenômeno. A diferença entre a CT gerada a partir do modelo exaustivo e do MBC de 1.0% é imperceptível. Para o MBC de 0.5% e 0.1%, uma estrutura em forma de cruz é verificada no centro da CT. Este artefato está provavelmente ligado a forma regular do MBC via NN e da baixa densidade da malha amostral. Entretanto, o alcance para todas as direções de continuidade espacial está corretamente representado em todas as CTs. Ao seccionarmos as quatro tabelas de covariância, é possível extrair graficamente variogramas na orientação desta seção (linha pontilhada da figura 18). A figura 19 apresenta os variogramas destas tabelas, para a direção de maior e menor continuidade espacial.

A figura 20 apresenta o mapa exaustivo e os três MBCs, modelos utilizados para geração das 4 CTs na figura 18 para z igual a 1. Caso a malha amostral fosse um pouco mais desordenada, como é observado no MBC da figura 7, seria possível obter mais pares de diferente valor em todos espaçamentos. O que é normalmente verificado na prática da mineração, devido a sobreposição de diferentes campanhas de sondagem.

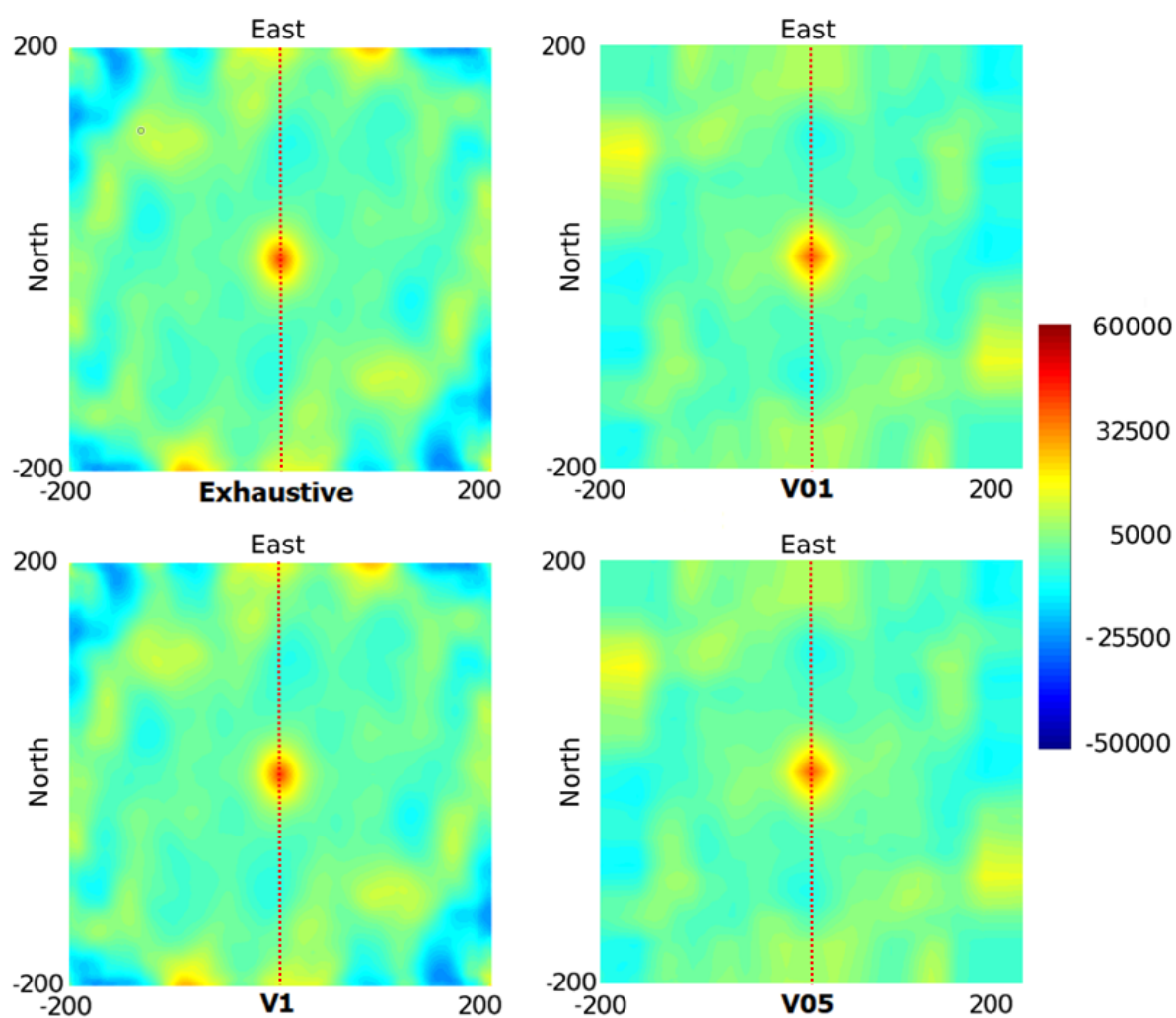


Figura 18 – CT para o modelo exaustivo em comparação com CT para todos os conjuntos de dados (V01, V05 e V1) em z igual a zero. A linha pontilhada referencia a secção para a direção de maior continuidade espacial.

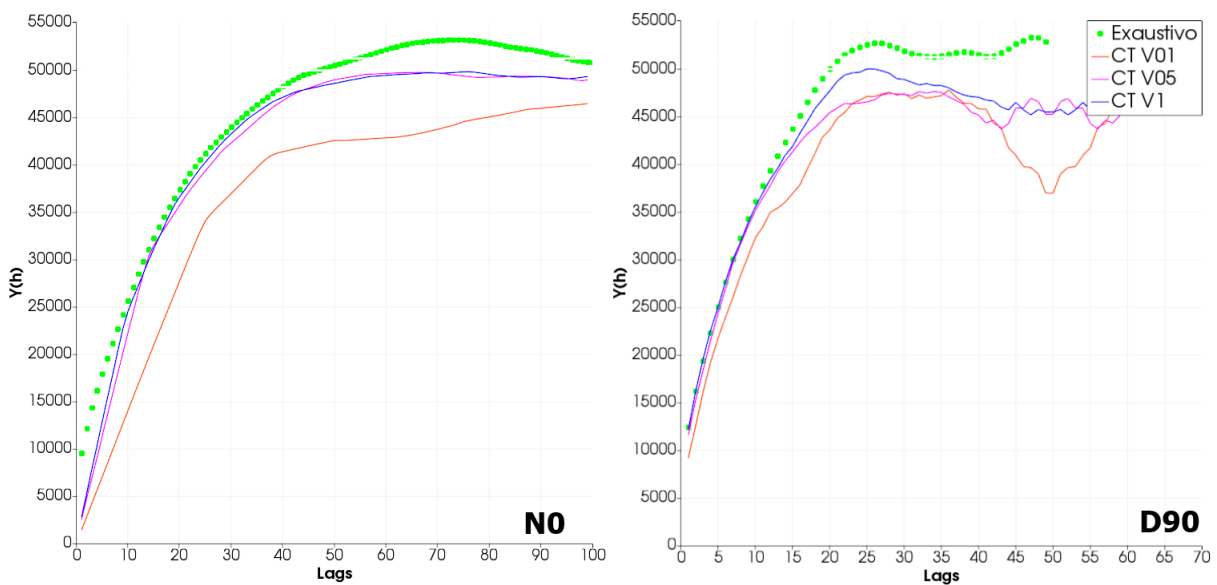


Figura 19 – Variogramas extraídos das secções nas direções de maior e menor continuidade espacial das quatro tabelas de covariância da figura 18. Em verde, o variograma da tabela do modelo exaustivo, em vermelho, rosa e azul, são os variogramas gerados a partir das tabelas de V 0.1%, 0.5% e 1.0%, respectivamente.

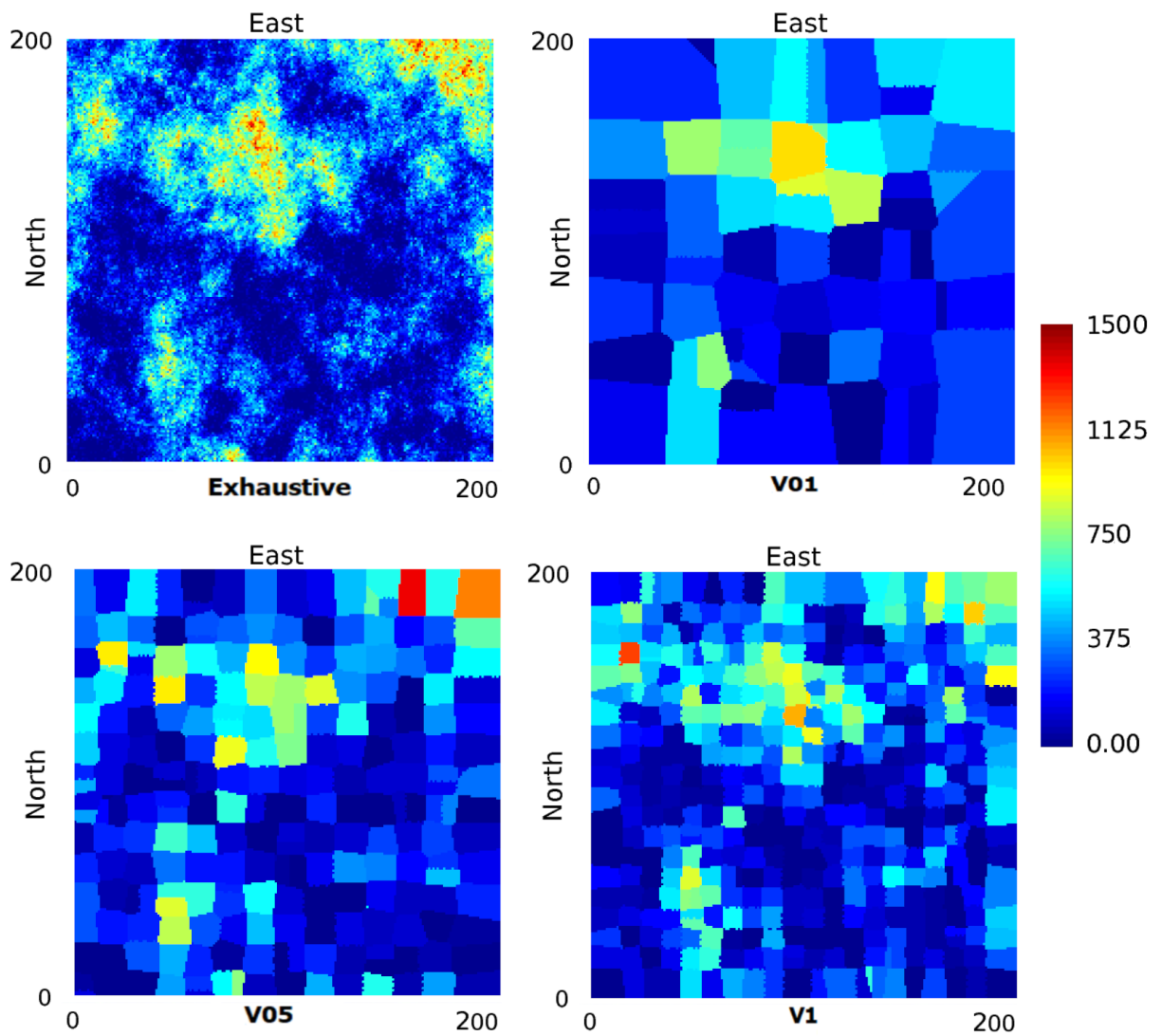


Figura 20 – Modelo exaustivo em comparação com os MBCs para todos os conjuntos de dados (V01, V05 e V1) para z igual a 1.

4.3 Estimativas

A krigagem ordinária (OK) foi realizada para V 1.0%, 0.5% e 0.1% via CT e variograma (Var) utilizando os mesmo parâmetros. A figura 21 abaixo apresenta os mapas resultantes de estimativa em plano horizontal com z igual a 1. O círculo amarelo sobre os mapas CT V01 e CT V05 destaca os artefatos em forma de cruz gerados pelas suas respectivas CTs da figura 18, em contraste aos círculos vermelhos, via variografia.

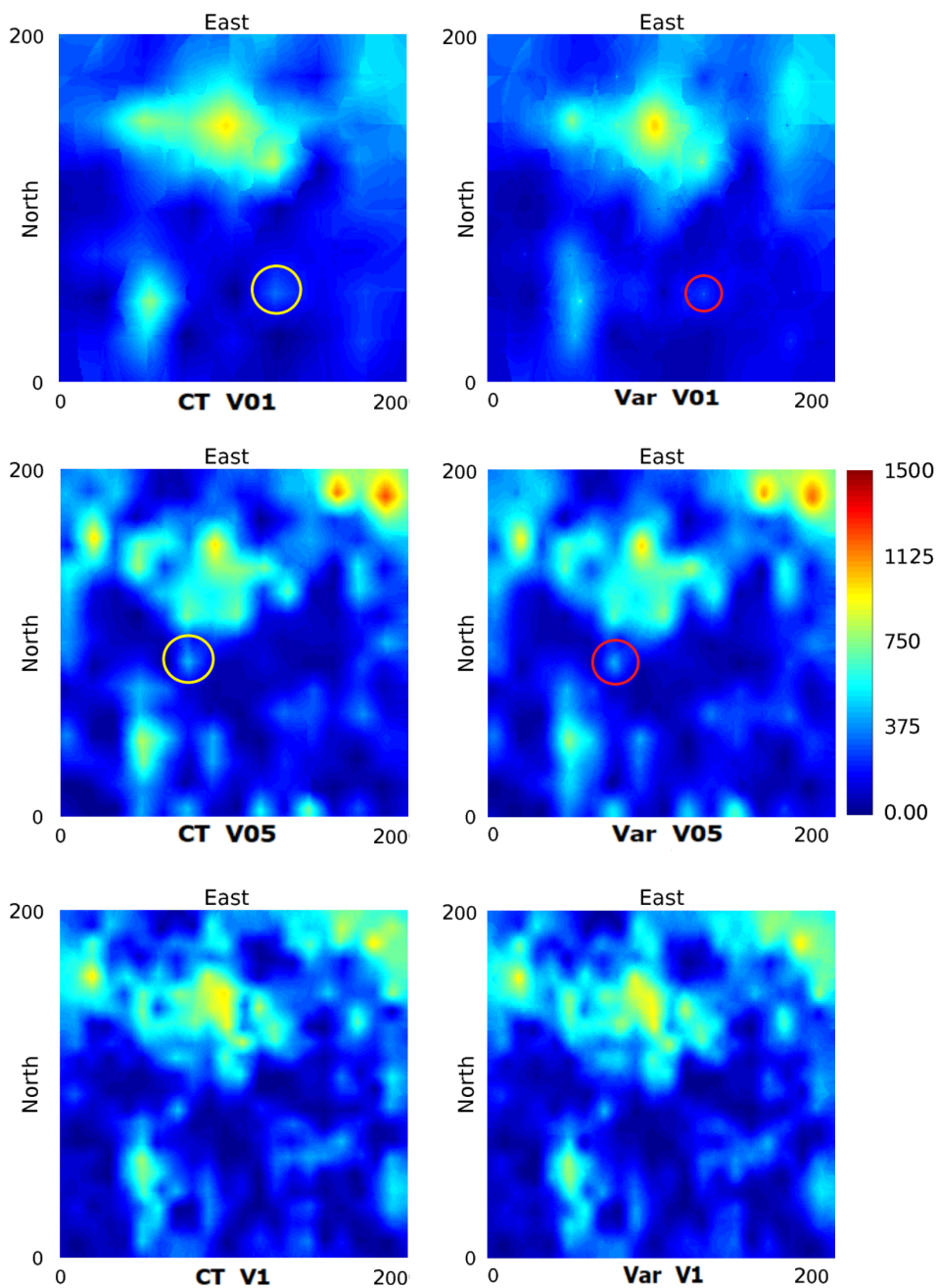


Figura 21 – OK via CT e variograma para V01, V05 e V1, em seção horizontal em z igual a 1.

O coeficiente de correlação para ambas estimativas OK (via CT e variograma) em comparação com o modelo exaustivo foram 0.81 para V 1.0%, 0.76 para 0.5% e 0.62 para 0.1%. Assim, para ambas as técnicas, a baixa relação de densidade de amostragem compromete a qualidade do modelo igualmente, para esta validação. É mostrado na tabela 3, a variância para todos casos. Observe que os modelos estimados via CT tendem a gerar mais variância e menos suavização do que modelos gerados via variograma tradicional.

	V 0.1%	V 0.5%	V 1.0%
OK via CT	27,351	31,491	35,004
OK via Var	22,972	30,310	34,322

Tabela 3 – Variância dos modelos de OK via CT e variograma para V 1.0%, V 0.5% e V 0.1%.

Foi escolhido para ilustrar a diferença de variância encontrada entre os modelos estimados, a função de probabilidade acumulada e a análise de deriva vertical para V 0.1%. A escolha foi feita em função da maior diferença apresentada em ambas análises para esta densidade amostral. As funções de probabilidade acumulada (CDF) dos teores dos modelos podem ser verificadas na figura 22, que apresenta a CDF do modelo exaustivo em comparação com OK via CT e variograma para V 0.1%. Esta diferença foi verificada ao longo de todo estudo feito: CDFs das estimativas via CT mais próximas da CDF do modelo exaustivo para todas densidades amostrais em comparação com as CDFs das estimativas via variografia tradicional.

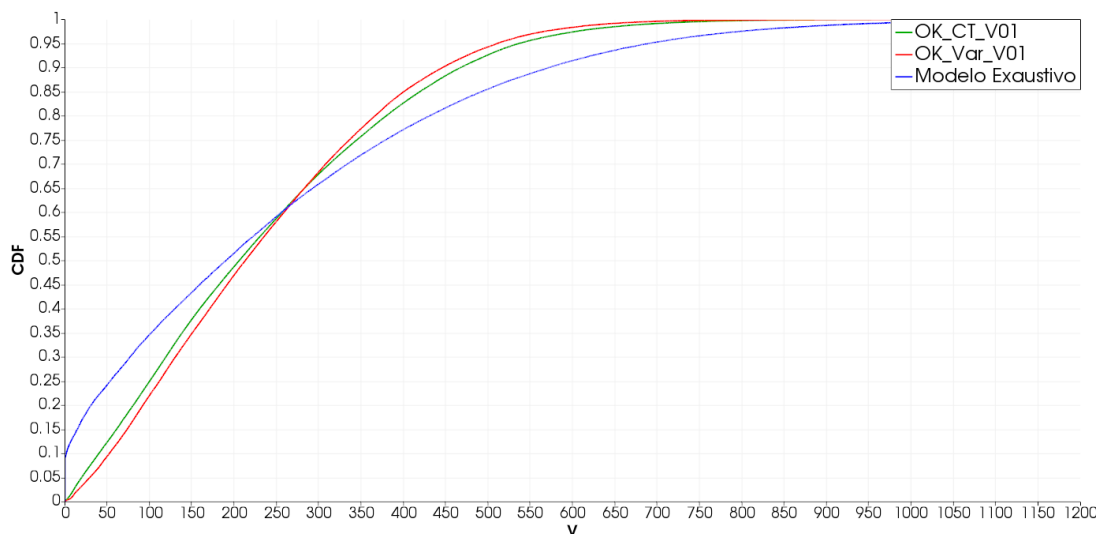


Figura 22 – CDF do modelo exaustivo em comparação com OK via CT e variograma para V 0.1%

A diferença encontrada nas curvas CDFs pode ser atribuída a variação vertical local verificada nos modelos estimados via CT. A figura 23 apresenta a análise de deriva vertical do modelo exaustivo em comparação com OK via CT e variograma para V 0.1%.

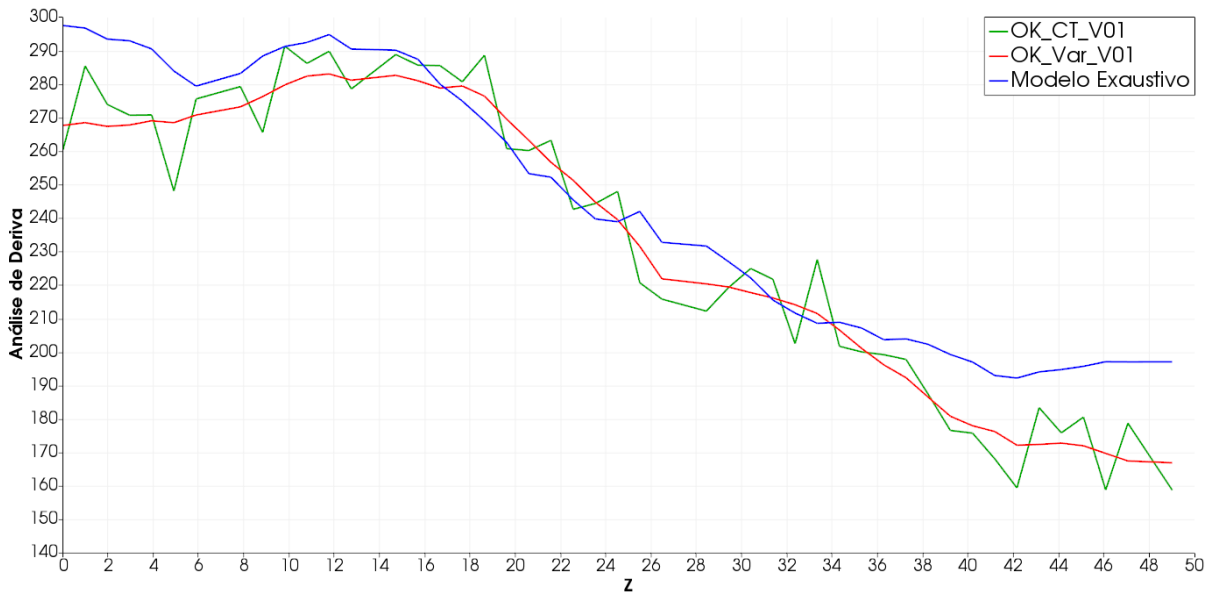


Figura 23 – Análise de deriva do modelo exaustivo em comparação com OK via CT e variograma para V 0.1%.

A figura 24 evidencia, através uma seção vertical dos dados de V 0.1% com x igual a 200 e de seu respectivo MBC, a grande variação nos teores nesta direção. Esta variação não é verificada nas seção horizontais por não haver dados, e, portanto, o MBC possui baixa variabilidade (figura 20).

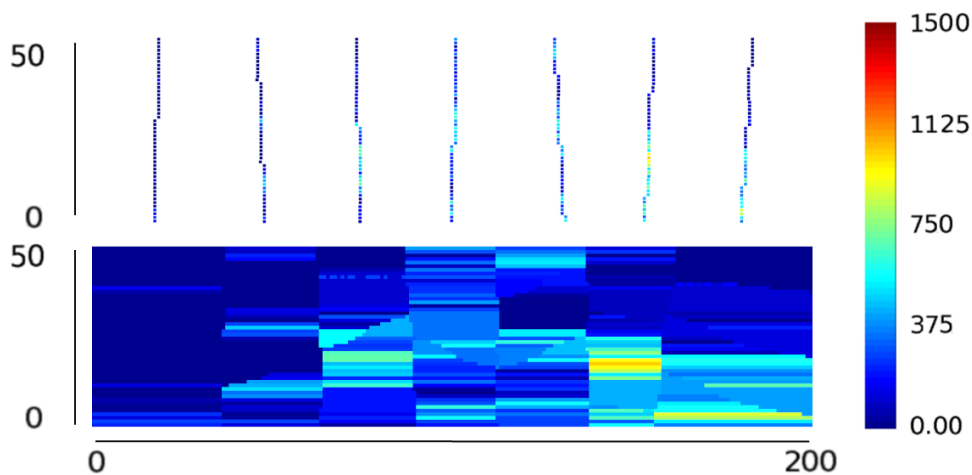


Figura 24 – Seção vertical em x igual a 200 de V 0.1% e de seu respectivo MBC.

4.4 Simulações

Foram também geradas simulações geoestatísticas utilizando a técnica de simulação sequencial gaussiana (SGS). Da mesma forma, os parâmetros foram os mesmos para ambas, com exceção do modelo de covariância utilizado, resultando em 50 realizações. A figura 25 apresenta uma realização utilizando para cada caso, via tabela de covariância e variograma.

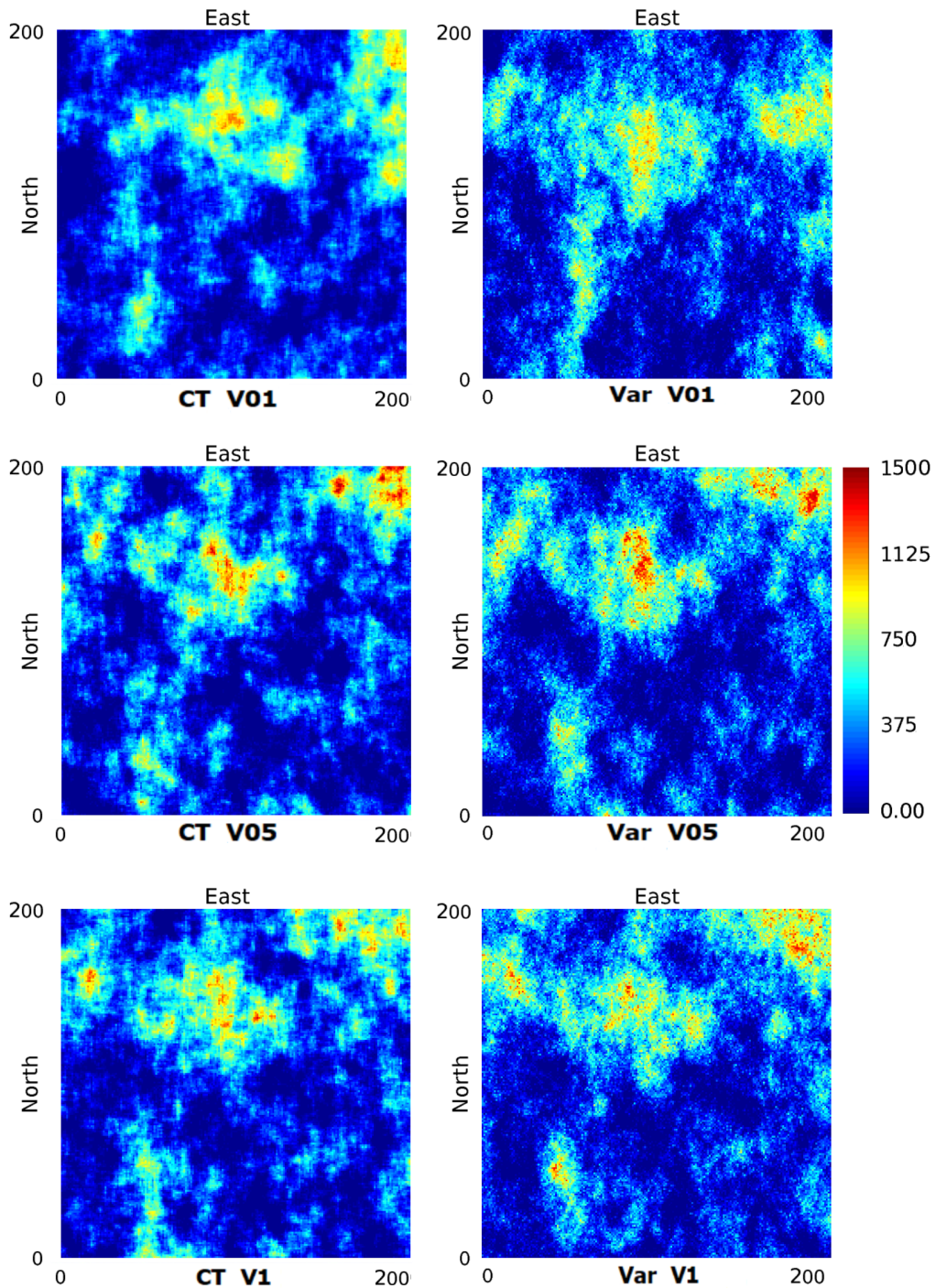


Figura 25 – Conjunto de realizações via CT e variograma para V 0.1%, V 0.5% e V 1.0%.

É possível verificar que a principal diferença nas realizações está no curto alcance. Este é mais contínuo para as SGS via CT. Evidenciando, novamente, que a CT reproduz as características de seu MBC e, apesar da simulação via CT ser mais rápida, a SGS via variograma apresentou melhores resultados. A figura 26 apresenta as funções de probabilidade acumulada para todos os casos em comparação com o modelo exaustivo. Não há significativa diferença entre as técnicas para este resultado.

A figura 27 encerra o estudo de caso, e, talvez, seja a mais importante e representativa do mesmo. Esta apresenta os variogramas para todas as 50 realizações SGS via tabela de covariância e utilizando a variografia clássica nas direções de maior e menor continuidade espacial. A figura também apresenta o variograma experimental para todos os casos de V, os modelos variográficos gerados e a curva variográfica extraída das tabelas de covariância (dada pela relação $\gamma(h) = C(0) - C(h)$).

Conforme discutido ao longo do estudo de caso, a notável diferença nos resultados obtidos pela variografia e pela CT é verificada no curto alcance. Quando o modelo variográfico é obtido a partir do variograma experimental, extrapolamos o curto alcance verificado ao longo do furo para todas as direções. Este é um passo natural, dado que sempre haverá mais dados ao longo do furo em relação aos dados espaçados furo a furo. A tabela de covariância, por sua vez, extrai seus valores do MBC gerado via NN. Portanto, a CT irá replicar as características do MBC. Neste estudo, por haver uma malha regular, o curto alcance sempre apresentou baixa variabilidade no plano horizontal (em destaque no círculo azul), ao contrário do plano vertical.

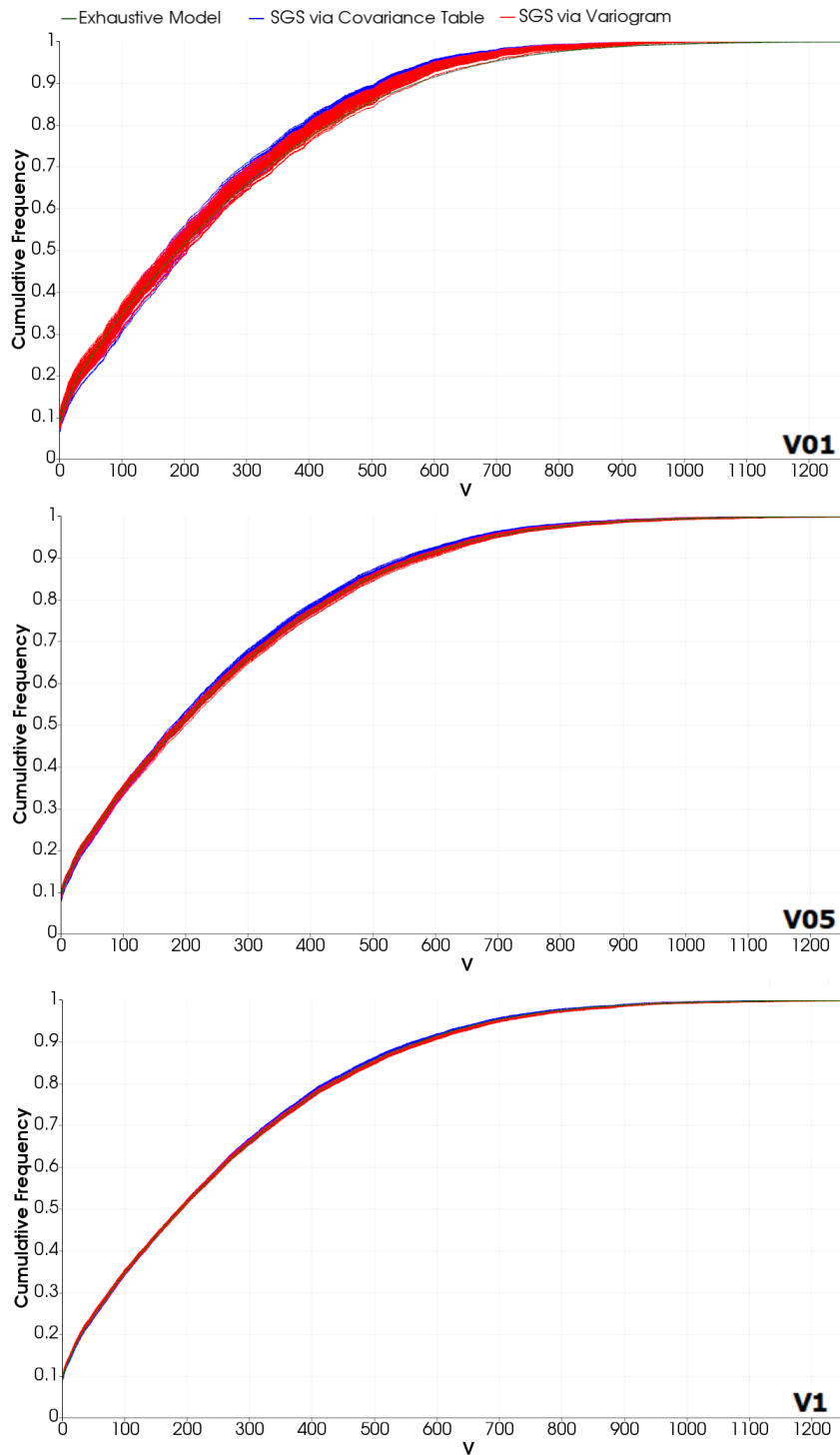


Figura 26 – CDF para 50 realizações SGS via CT e variograma para V 0.1%, V 0.5% e V 1.0%.

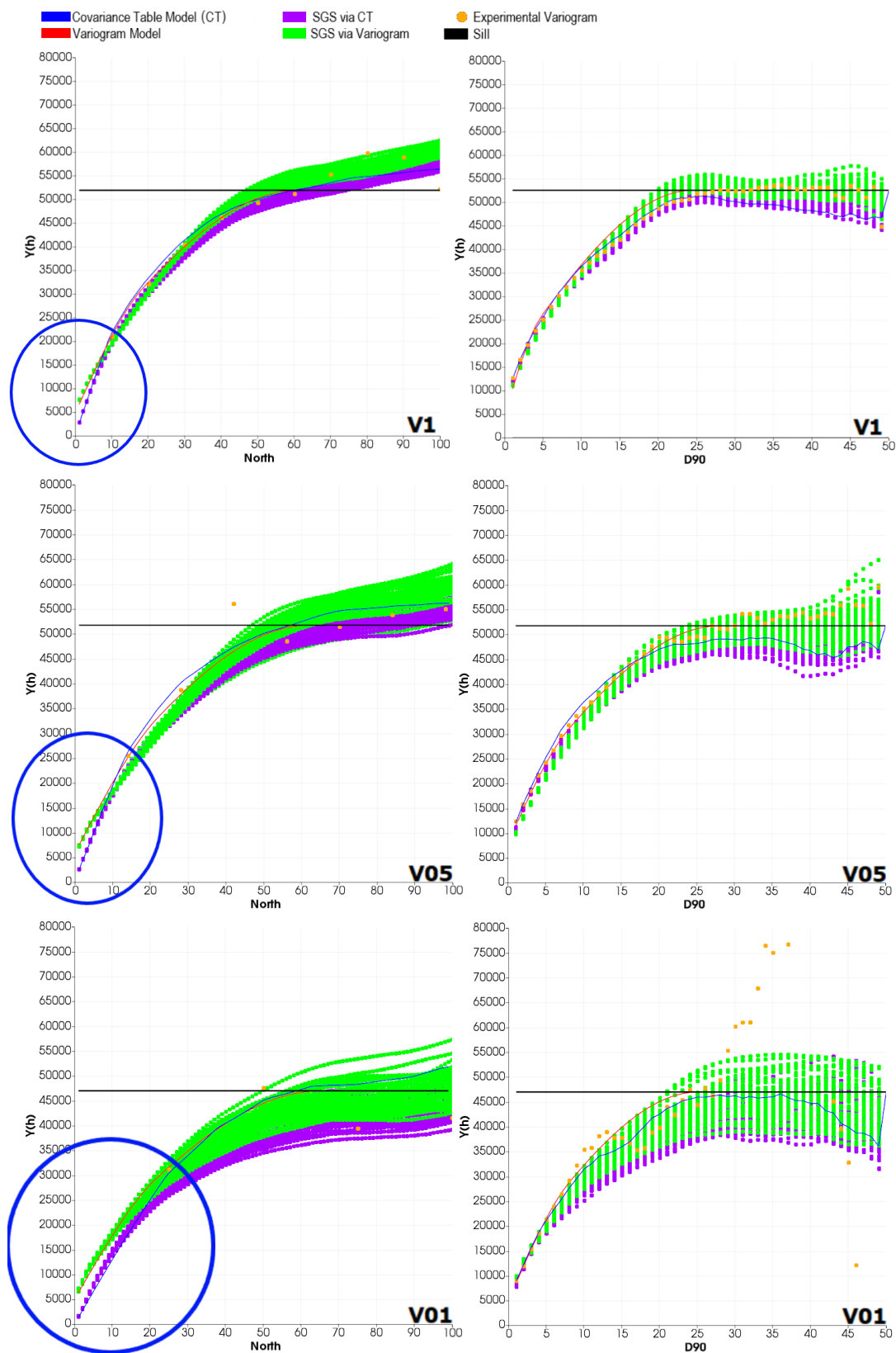


Figura 27 – Flutuações ergódicas apresentadas pelo variograma orientado na direção de maior e menor continuidade espacial para SGS via CT e variograma para V 1.0%, 0.5% e 0.1%.

4.5 Conclusão

O desenvolvimento do presente estudo demonstrou a eficiência da técnica proposta. Esta é baseada na extração de uma tabela de covariância (CT) a partir de um modelo conceitual, denominado neste estudo de modelo base para extrair covariância (MBC). A técnica provou sua eficácia tanto por meio da validação estatística quanto pela velocidade de obtenção de uma análise continuidade espacial. Verifica-se que é um método dependente do MBC e, portanto, os resultados serão tão bons quanto o MBC possibilitar. Esta tabela de covariância é obtida através de um fluxo de trabalho de três etapas: interpolando o banco de dados para preencher todos os nós do grid regular (MBC), aplicar uma auto convolução por meio do algoritmo de transformada rápida de Fourier (FFT) e retrotransformar de volta para o domínio original. O MBC representa uma solução simples para um problema complexo, comparando com métodos anteriores que propuseram o uso da tabela de covariância. Entretanto, a geração do MBC através do método vizinho mais próximo (NN) pode aumentar artificialmente o curto alcance no plano xy , para os casos em que há efeito pepita significativo.

O método proposto é um mapeamento automático rápido de continuidade espacial, mas não suprime passos anteriores de modelagem como escolher domínios estacionários, nem a análise quantitativa da vizinhança da krigagem. Ainda, é importante definir um elipsoide de busca de krigagem como no método tradicional via variografia, pois uma tabela de covariância positiva definitiva não é garantida pelo método. Por outro lado, todas as tabelas de covariâncias geradas através deste estudo apresentaram uma região central que pode ser identificada e está diretamente relacionada com o fenômeno espacial. Krigagem ordinária via tabela de covariância tende a gerar mais variância e menos suavização do que via variografia tradicional. Já a simulação sequencial gaussiana replicou a baixa variabilidade horizontal e alta vertical, verificada na tabela de covariância extraída do MBC via NN para este estudo de caso.

5 Considerações finais

Essa dissertação propôs como meta investigar a aplicabilidade da tabela de covariância como método substituto ou auxiliar aos métodos clássicos de modelagem variográfica. Para tanto, foram apresentados dois objetivos específicos:

1. Operacionalizar o método no software geoestatístico de código aberto *AR2GeMS*, desenvolvendo um *plug-in* funcional.
2. Conduzir um estudo de caso em um banco de dados sintético e verificar a qualidade dos modelos estimados e simulados pela metodologia proposta em comparação com a tradicional.

Ambos objetivos foram cumpridos e validados. Resultando em um *plug-in* intuitivo e de fácil utilização. Uma alternativa mais ágil, e passível de gerar melhores resultados. E evidenciado no [Capítulo 4](#), é possível concluir que a geração da tabela de covariância é um método simples para um mapeamento automático de continuidade espacial. Apesar de não substituir completamente os passos anteriores e posteriores a modelagem, os quais não eram os objetivos, o método gerou modelos semelhante ao modelos tradicionais.

5.1 Recomendações para trabalhos futuros

Como recomendação para trabalhos futuros:

- Verificar se a tabela de covariância gerada é definida positiva e, caso necessário, investigar métodos de correção, como os descritos na tese de Yao ([YAO, 1998a](#));
- Estudar em quais situações o método vizinho mais próximo não seria o MBC recomendado para obter a tabela de covariância, e/ou encontrar formas para incorporar o efeito pepita nesta tabela;
- Repetir o estudo de caso em banco de dados real, com uma malha amostral complexa e não regular;
- Investigar a viabilidade em fenômenos estacionários não claros, com uma continuidade espacial complexa;
- Investigar o uso da densidade espectral gerada para simulações espectrais como o método integral de Fourier ([PARDO-IGUZQUIZA; CHICA-OLMO, 1993](#));
- Estudar casos multivariados, em que será necessário covariâncias cruzadas.

Referências

- BOCHNER, S. Additive set functions on groups. *Annals of Mathematics*, JSTOR, p. 769–799, 1939. Citado na página 33.
- CHATFIELD, C. Model uncertainty and forecast accuracy. *Journal of Forecasting*, Wiley Online Library, v. 15, n. 7, p. 495–508, 1996. Citado na página 37.
- CHU, J. *Conditional fractal simulation, fast sequential indicator simulation, and interactive variogram modeling*. Tese (Doutorado) — Stanford University, 1993. Citado 3 vezes nas páginas 20, 35 e 36.
- DEUTSCH, C.; JOURNEL, A. *Geostatistical software library and user's guide*. [S.l.]: Oxford University Press, New York, 1997. Citado na página 36.
- DEUTSCH, C. V.; JOURNEL, A. G. *GSLIB: Geostatistical Software Library and User's Guide. Hauptbd*. [S.l.]: Oxford university press, 1992. Citado na página 20.
- ISAAKS, E. H.; SRIVASTAVA, R. M. *An Introduction to Applied Geostatistics*. [S.l.]: Oxford University Press, 1989. Citado 2 vezes nas páginas 20 e 42.
- MARCOTTE, D. Fast variogram computation with fft. *Computers & Geosciences*, Elsevier, v. 22, n. 10, p. 1175–1186, 1996. Citado 6 vezes nas páginas 13, 15, 20, 31, 32 e 34.
- 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 3 vezes nas páginas 39, 41 e 67.
- PRESS, W. H. *Numerical Recipes in FORTRAN: The Art of Computing*. [S.l.]: Cambridge University Press, 1992. Citado na página 29.
- PYRCZ, M. J.; DEUTSCH, C. V. Semivariogram models based on geometric offsets. *Mathematical geology*, Springer, v. 38, n. 4, p. 475–488, 2006. Citado na página 19.
- REHMAN, S. U. *Semiparametric modeling of cross-semivariograms*. Tese (Doutorado) — Georgia Institute of Technology, 1995. Citado na página 34.
- RUBIO, R. et al. Localized kriging parameter optimization based on absolute error minimization. Unpublished, 2018. Citado na página 41.
- SANDERSON, G. *But what is the Fourier Transform? A visual introduction*. 2018. Acessado: 10-07-2018. Disponível em: <<https://www.youtube.com/watch?v=spUNpyF58BY>>. Citado 4 vezes nas páginas 13, 25, 26 e 27.
- VANN, J.; JACKSON, S.; BERTOLI, O. Quantitative kriging neighbourhood analysis for the mining geologist—a description of the method with worked case examples. In: BENDIGO, AUSTRALIA. MELBOURNE: AUSTRALIAN INST MINING & METALLURGY. *5th International Mining Geology Conference*. [S.l.], 2003. v. 8, p. 215–223. Citado na página 41.

- WING, J. M. Computational thinking and thinking about computing. *Philosophical transactions of the royal society of London A: mathematical, physical and engineering sciences*, The Royal Society, v. 366, n. 1881, p. 3717–3725, 2008. Citado na página 19.
- YAO, T. *Automatic covariance modeling and conditional spectral simulation with Fast Fourier Transform*. [S.l.: s.n.], 1998. Citado 2 vezes nas páginas 32 e 67.
- YAO, T. Conditional spectral simulation with phase identification. *Mathematical Geology*, Springer, v. 30, n. 3, p. 285–308, 1998. Citado na página 39.
- YAO, T. Note and corrections to “automatic modeling of (cross) covariance tables using fast fourier transform” by t. yao and ag journal. *Mathematical Geology*, Springer, v. 32, n. 1, p. 147–148, 2000. Citado 2 vezes nas páginas 20 e 33.
- YAO, T.; JOURNEL, A. G. Automatic modeling of (cross) covariance tables using fast fourier transform. *Mathematical Geology*, v. 30, n. 6, p. 589–615, 1998. Citado 5 vezes nas páginas 19, 20, 32, 41 e 46.
- YOO, Y. *Tutorial on Fourier theory*. 2001. Acessado: 10-07-2018. Disponível em: http://www.cs.otago.ac.nz/cosc453/student_tutorials/fourier_analysis.pdf. Citado 3 vezes nas páginas 25, 28 e 29.