

Roberto Mentzingen Rolo

**MODELAGEM GEOLÓGICA E AVALIAÇÃO DE
INCERTEZA EM MODELOS GEOLÓGICOS
USANDO FUNÇÕES DISTÂNCIA ASSINALADAS**

Porto Alegre - RS

2021

Roberto Mentzingen Roló

**MODELAGEM GEOLÓGICA E AVALIAÇÃO DE
INCERTEZA EM MODELOS GEOLÓGICOS USANDO
FUNÇÕES DISTÂNCIA ASSINALADAS**

Tese submetida ao Programa de Pós-graduação
em Engenharia de Minas, Metalúrgica e Materiais (PPGE3M) da Universidade Federal do Rio
Grande do Sul como requisito parcial à obtenção
do título de Doutor em Engenharia.

Universidade Federal do Rio Grande do Sul – UFRGS

Escola de Engenharia

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

Orientador: João Felipe Coimbra Leite Costa

Porto Alegre - RS

2021

AGRADECIMENTOS

Agradeço à minha Mãe, às minhas irmãs Todia e Cláudia e meu cunhado Marcelo pelo apoio incondicional. À minha parceira para qualquer tempo Camila. Aos amigos de JF Skeli, Disco, Gunzo, Alves e Dêniel. À república Sparta. Ao meu orientador e grande amigo, Professor João Felipe. Aos colegas de laboratório, principalmente, os parceiros de pesquisa Flávio Amarante e David Drummond. Sem esquecer de outras grandes amizades: Augusto, Marcel, Cristina, Jonas, Luiz, Ryu, Myolo, Ricardinho, Ricardo Colômbia, Kbelo e Anuar. A turma da programação: Péricles e Átila. Viva o software livre! Agradeço também ao LPM, à Fundação Luiz Englert, à CAPES, à AR2Tech, e ao Governo do Estado do Rio Grande do Sul pelas oportunidades e por incentivarem a conclusão desse doutorado.

"All models are wrong, but some are useful"

George Box

RESUMO

Métodos geoestatísticos para estimativa e simulação de teores de minério requerem a delimitação prévia de domínios estacionários. Apesar da ampla adoção pela indústria da mineração da modelagem implícita com funções distância assinaladas como alternativa à modelagem explícita para esse fim, os modelos criados dessa forma não avaliam a incerteza em relação à posição real dos contatos geológicos em subsolo. Em muitos casos, a incerteza do modelo geológico pode ser uma fonte de incerteza importante para o projeto e ignorá-la pode ser devastador para o empreendimento. Esta tese propõe e investiga três diferentes metodologias para avaliar a incerteza em modelos geológicos multi-categóricos usando funções distância assinaladas. Uma das metodologias propostas é baseada em interpolação, enquanto as outras duas são baseadas em simulações Gaussianas não condicionais dentro de uma zona de incerteza. Um estudo de caso foi realizado para cada metodologia. Os resultados foram comparados com a simulação sequencial dos indicadores mostrando que as metodologias propostas podem gerar limites contínuos entre as diferentes litologias do depósito mineral com o realismo geológico esperado. Foram desenvolvidos softwares implementando as metodologias propostas que apresentam um alto nível de automação.

Palavras-chaves: Modelagem geológica. Modelagem geológica implícita. Avaliação de incerteza. Simulação de contatos geológicos.

ABSTRACT

Geostatistical methods for estimation and simulation of ore grades require the previous delimitation of stationary domains. Despite the widespread adoption by the mining industry of implicit modeling with signed distance functions as an alternative to explicit modeling for this purpose, models created in this way do not assess uncertainty related to the real position of the geological contacts in the subsurface. In many cases, the uncertainty of the geological model can be a source of important project uncertainty, and ignoring it can be devastating for the enterprise. This thesis proposes and investigates three different methodologies for assessing uncertainty in multi-categorical geological models using signed distance functions. One of the proposed methodologies is based on interpolation, while the other two are based on unconditional Gaussian simulations inside an uncertainty zone. A case study was conducted for each methodology. The results were compared with sequential indicator simulation showing that the proposed methodologies can generate continuous boundaries between the different lithologies of the mineral deposit with the expected geological realism. Software implementing the proposed methodologies that have a high level of automation were developed.

Key-words: Geological modeling. Implicit geologic modeling. Uncertainty assess. Boundary simulation.

LISTA DE ILUSTRAÇÕES

Figura 1 – Ilustração esquematizando a modelagem geológica explícita usando polilinhas e linhas guias.	15
Figura 2 – Ilustração esquemática destacando a incerteza na localização do limite entre os domínios azul e vermelho, entre os dados amostrais.	16
Figura 3 – Ilustração esquemática da modelagem geológica implícita.	20
Figura 4 – Mapa de localização das 259 amostras do <i>dataset Swiss Jura</i>	21
Figura 5 – Mapa de localização das 259 amostras dos indicadores para a categoria Quaternary do <i>dataset Swiss Jura</i>	22
Figura 6 – Ilustração esquemática mostrando o cálculo das distâncias assinaladas. . . .	22
Figura 7 – Mapa de localização das distâncias assinaladas para a categoria Quaternary do <i>dataset Swiss Jura</i>	23
Figura 8 – Variogramas omnidirecionais dos indicadores.	25
Figura 9 – Efeito da resolução do <i>grid</i> na reprodução de estruturas geológicas.	27
Figura 10 – Categoria Quaternary do <i>Swiss Jura</i> interpolada por krigagem dual em um grid 3x3 metros.	29
Figura 11 – Modelo geológico para a categoria Quaternary do <i>Swiss Jura</i>	32
Figura 12 – Exemplo mostrando a extração da isosuperfície zero e visualização em três dimensões.	33
Figura 13 – Figura esquemática mostrando os passos para a criação do modelo geológico multi-categórico no <i>Swiss Jura</i>	34
Figura 14 – Particionamento do domínio.	36
Figura 15 – Esquema mostrando o funcionamento do algoritmo cubos marchantes. . . .	37
Figura 16 – Esquema mostrando o refinamento dos contatos em três iterações reduzindo as dimensões do <i>grid</i> pela metade para o <i>Swiss Jura</i>	38
Figura 17 – Esquema mostrando os vetores de anisotropia local e as amostras representadas pelos círculos pretos para cada nó do <i>grid</i>	39
Figura 18 – Esquema mostrando os vetores de anisotropia local e as amostras, representadas pelos círculos pretos, para cada centro de partição representadas pelas linhas tracejadas vermelhas.	40
Figura 19 – Esquema mostrando o refinamento iterativo para funções de bases radiais. . . .	41
Figura 20 – Cenário unidimensional simples ilustrando o viés devido à assimetria nos modelos geológicos.	42

Figura 21 – Esquema mostrando a extração da iso-superfície usando indicadores e triangulação de Voronoi. À esquerda a figura mostra o CDF dos indicadores interpolados e em vermelho a proporção (1-0.2) ocupada pelos triângulos que representam amostras no interior do domínio que equivale a um valor de indicador de 0.313. À direita um <i>cutoff</i> feito em 0.313 nos indicadores interpolados para definir os limites do domínio.	43
Figura 22 – Distâncias estimadas e transformadas em probabilidades para um mesmo bloco do <i>Swiss Jura</i>	45
Figura 23 – Relação entre o parâmetro ω e a entropia de Shannon.	46
Figura 24 – Distâncias transformadas em probabilidades para a categoria Quaternary usando valores diferentes para ω	47
Figura 25 – Parametrização das amostras.	49
Figura 26 – Calibração do parâmetro C para um esquema em duas dimensões.	53
Figura 27 – Figura esquemática da curva de calibração.	54
Figura 28 – Simulação das distâncias dentro da zona de incerteza para um esquema em duas dimensões.	55
Figura 29 – Esquema mostrando como os contatos são simulados dentro da zona de incerteza.	56
Figura 30 – Duas realizações da simulação dos contatos para um esquema em duas dimensões.	57
Figura 31 – Amostras em posição 1, 10, 70 e 90. Amostras em azul não pertencem ao domínio enquanto amostras vermelhas pertencem.	58
Figura 32 – Amostras em uma dimensão no gráfico na posição inferior e distâncias assinaladas calculadas para cada uma das amostras no eixo y do gráfico na posição superior. As curvas coloridas representam <i>kernels</i> associados a cada amostra. A reta vertical na figura da direita na posição 40 representa o contato entre as categorias.	59
Figura 33 – Amostras em uma dimensão no gráfico na posição inferior e distâncias assinaladas calculadas para cada uma das amostras no eixo y do gráfico na posição superior. As curvas coloridas representam os <i>kernels</i> associados à cada amostra.	60
Figura 34 – Caso base e cenários otimista e pessimista para o exemplo unidimensional proposto. As retas verticais representam os pontos de contato entre as categorias.	60
Figura 35 – Distâncias assinaladas interpoladas para a categoria Quaternary do banco de dados <i>Jura</i> para o caso base, e cenários pessimista e otimista.	61
Figura 36 – Modelos geológicos para o <i>Swiss Jura</i> criados a partir da parametrização do suporte do <i>kernel</i> . As cores representam cada uma das cinco categorias enquanto a legenda mostra a qual cenário aquela categoria pertence.	62

Figura 37 – Modelo conceitual composto por rocha encaixante, mostrada em vermelho, um dique, mostrado em verde, e estruturas lenticulares, mostradas em azul.	63
Figura 38 – Realizações para o modelo geológico sintético geradas a partir dos furos verticais. As cores representam cada uma das três categorias enquanto a legenda mostra a qual cenário aquela categoria pertence.	64
Figura 39 – Da esquerda para a direita: (1) zona de incerteza de 12 metros; (2) distâncias estimadas para cada uma das três categorias um bloco específico; (3) probabilidades transformadas.	66
Figura 40 – Um valor de probabilidade simulado de 0,51 é usado para amostrar a categoria vermelha de uma distribuição condicional em um bloco específico.	67
Figura 41 – Realizações Gaussianas e modelos geológicos correspondentes.	68
Figura 42 – Vista em perspectiva das amostras do depósito de ouro. As cores representam as cinco diferentes litologias do depósito mineral.	69
Figura 43 – Variogramas dos indicadores para todas as cinco categorias do depósito mineral modelados com estruturas esféricas.	70
Figura 44 – Visão em perspectiva de seções verticais e horizontais de duas realizações do modelo geológico juntamente com as amostras em suporte pontual.	72
Figura 45 – Comparativo de seções verticais xz entre a metodologia proposta, a SISIM e um modelo explícito.	73
Figura 46 – Uma realização para o <i>Swiss Jura</i> criada usando um variograma esférico para a simulação não condicional.	74
Figura 47 – Entropia calculada a partir das dez realizações no <i>Swiss Jura</i>	75
Figura 48 – Variação de área para todas as categorias do <i>Swiss Jura</i> para 10 realizações com larguras de banda de 12 e 20 metros.	75
Figura 49 – Matriz de confusão para o depósito de ouro mostrando proporções das categoria preditas nas colunas e das categorias das amostras (real) nas linhas.	76
Figura 50 – Proporção de cada uma das categorias nos dados.	77
Figura 51 – Proto-modelo e blocos sinalizados como contatos pelo algoritmo cubos marchantes para o <i>Swiss Jura</i>	78
Figura 52 – Regra hierárquica definida pelo algoritmo proposto.	80
Figura 53 – Mapa de localização das amostras dos quatro grupos.	81
Figura 54 – Distâncias modificadas e interpoladas para cada um dos grupos.	82
Figura 55 – Uma realização dos valores Gaussianos simulados, dentro da zona de incerteza, para cada grupo.	83
Figura 56 – Uma realização da simulação de limites para todos os grupos.	84
Figura 57 – Duas realizações para o modelo geológico para o <i>Swiss Jura</i>	85
Figura 58 – Vista em perspectiva para o conjunto de dados no suporte pontual que descreve as cinco categorias.	85

Figura 59 – Regra de hierarquia para o conjunto de dados do estudo de caso definido pelo algoritmo proposto.	86
Figura 60 – Visão em perspectiva para os grupos de amostras em suporte pontual.	86
Figura 61 – Variogramas dos indicadores para cada grupo.	87
Figura 62 – Calibração do parâmetro C por <i>jackknife</i>	89
Figura 63 – Visão em perspectiva de seções verticais e horizontais de duas realizações do modelo geológico juntamente com as amostras em suporte pontual.	90
Figura 64 – Seções equidistantes em XZ comparando uma realização do modelo geológico do SISIM com uma realização da metodologia proposta.	91
Figura 65 – Seções equidistantes em XZ comparando a entropia do SISIM com uma realização da metodologia proposta.	92
Figura 66 – Matriz de confusão para o depósito de cobre mostrando proporções das categoria preditas nas colunas e das categorias das amostras (real) nas linhas.	93
Figura 67 – Efeito do parâmetro C.	94
Figura 68 – Variável Cr do <i>Swiss Jura</i>	97
Figura 69 – Análise de risco para o conteúdo metálico de Cr no domínio Argovian.	97
Figura 70 – LPM-Geomod_Suite no painel de algoritmos do <i>SGeMS/AR2GeMS</i>	109
Figura 71 – Interface do <i>plugin point_transform</i>	110
Figura 72 – Interface do <i>plugin block_transform</i>	111
Figura 73 – Interface do <i>plugin auto_grid</i>	111
Figura 74 – Interface do <i>plugin vtk_export</i>	112
Figura 75 – Interface do <i>plugin deterministic</i>	113
Figura 76 – Arquivo de texto com os modelos variográficos.	114
Figura 77 – Interface do <i>plugin svm</i>	115
Figura 78 – Modelo geológico para o dataset <i>Swiss Jura</i> criado com máquina de vetores de suporte.	115
Figura 79 – Interface do <i>plugin kernel_factor_uncertainty</i>	117
Figura 80 – Interface do <i>plugin boundary_simulation_pfields</i>	118

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Contextualização	13
1.2	Motivação	16
1.3	Declaração de tese	17
1.4	Objetivos	17
1.4.1	Desenvolvimento de software	17
1.4.2	Avaliação de incerteza pela parametrização do suporte do <i>kernel</i>	17
1.4.3	Avaliação de incerteza usando funções distância assinaladas e campos de probabilidade	18
1.4.4	Simulação de contatos hierárquica para múltiplas categorias	18
1.5	Estrutura da tese	18
2	REVISÃO BIBLIOGRÁFICA	19
2.1	Modelagem geológica determinística	19
2.1.1	Modelagem geológica com funções distância assinaladas	21
2.1.1.1	Cálculo das distâncias assinaladas	21
2.1.1.2	Variografia	24
2.1.1.3	Interpolação	26
2.1.1.3.1	Krigagem global	29
2.1.1.3.2	Krigagem dual	30
2.1.1.3.3	Funções de bases radiais (RBF)	30
2.1.1.4	Visualização do modelo geológico	31
2.1.1.5	Adaptação para múltiplas categorias	33
2.1.1.6	Acelerando o processo	35
2.1.1.6.1	Decomposição do domínio	35
2.1.1.6.2	Refinamento dos contatos	36
2.1.1.7	Incorporação de estacionariedade de segunda ordem	38
2.1.1.7.1	Krigagem com anisotropia local variável	38
2.1.1.7.2	Funções de bases radiais com anisotropia local variável	39
2.1.1.8	Desvantagens	41
2.2	Modelagem geológica estocástica	43
2.2.1	Avaliação heurística da incerteza	45
2.2.2	<i>Boundsim</i>	47
2.2.3	Condicionamento das amostras	48
2.2.4	Simulação direta das distâncias assinaladas	49

2.2.5	Avaliação de incerteza usando imagens de treinamento para simulação multi-ponto geradas a partir dos dados	50
2.2.6	Modelagem estocástica de estruturas curvilineares	51
2.2.7	Avaliando e visualizando a incerteza com movimento estocástico	51
2.2.8	Simulação dos contatos	51
3	METODOLOGIAS PROPOSTAS	58
3.1	Parametrização do suporte do <i>kernel</i>	58
3.1.1	Estudo de caso	62
3.1.2	Discussão	64
3.2	Avaliação de incerteza usando funções distância assinaladas e campos de probabilidades	65
3.2.1	Estudo de caso	68
3.2.2	Discussão	72
3.3	Simulação hierárquica dos contatos	77
3.3.1	Estudo de caso	85
3.3.2	Discussão	90
3.4	Conclusão	94
4	OBSERVAÇÕES FINAIS	99
4.1	Sumário das contribuições	99
4.2	Sugestões para trabalhos futuros	100
	REFERÊNCIAS	102
	APÊNDICES	108
	APÊNDICE A – SOFTWARE DESENVOLVIDO	109
A.1	Transformações	110
A.1.1	Em pontos	110
A.1.2	Em blocos	110
A.2	Auxiliares	111
A.2.1	Criação automática do <i>grid</i>	111
A.2.2	Exportação em formato VTK	112
A.3	Modelagem determinística	112
A.3.1	Máquina de vetores de suporte (SVM)	114
A.4	Modelagem estocástica	116
A.4.1	Parametrização do suporte <i>kernel</i>	116

A.4.2	Avaliação de incerteza usando funções distâncias assinaladas e campos potenciais	117
A.4.3	Simulação de contatos hierárquica	118

1 INTRODUÇÃO

Este capítulo traz uma breve contextualização com o objetivo de justificar e construir um caminho lógico para a motivação da tese. Além disso, a declaração de tese é apresentada e a estrutura do documento é delineada.

1.1 Contextualização

Estimativas de recursos e reservas minerais demandam a construção de modelos numéricos de teores de longo prazo, que abrangem toda a extensão do depósito mineral e compreendem todo o tempo de vida da mina. Esses modelos são atualizados em intervalos de um a três anos de operação. Modelos de médio prazo podem ser construídos para planejar de um a seis meses no futuro do empreendimento. Enquanto modelos de curto prazo, são construídos com o objetivo de balizar, semanalmente ou diariamente, as decisões relativas a controle de teores e planejamento detalhado da mina.

Construir modelos numéricos de longo, médio e curto prazo para avaliação de recursos/reservas e planejamento de mina exige quatro grandes atividades ([ROSSI; DEUTSCH, 2013](#)):

- i Coleta e gerenciamento de dados;
- ii Interpretação e modelagem geológica;
- iii Atribuição de teores;
- iv Avaliação e gerenciamento da incerteza geológica e de teores.

Após a coleta, gerenciamento e checagem dos dados, é necessário identificar diferentes domínios. A determinação dos domínios deve ser baseada no conhecimento geológico, i.e. zonas de oxidação, litologias, alteração ou limites estruturais e deve ser suportada por extensiva análise estatística (análise exploratória dos dados), incluindo análises variográficas. É importante que a definição dos domínios faça sentido espacial e geológico. Os domínios são, algumas vezes, baseados na combinação de duas ou mais variáveis para as quais a correlação entre os teores pode ser demonstrada. O procedimento pode ser bastante demorado, principalmente em depósitos multivariados, entretanto, as estimativas são melhoradas significativamente quando cuidadosamente limitadas por variáveis geológicas ([ROSSI; DEUTSCH, 2013](#)).

A definição de diferentes domínios de estimativas, ou zonas estacionárias no interior do depósito, é necessária para a inferência geoestatística. Geralmente, o método usado na

construção dos modelos numéricos de teores (MCLENNAN, 2007). Isto é, os teores em cada domínio estacionário pertencem a populações estatísticas distintas, são caracterizados através de modelos de distribuição e de covariância específicos e devem ser modelados como funções aleatórias estacionárias diferentes. (JOURNEL; HUIJBREGTS, 1978).

Uma função aleatória estacionária é uma representação probabilística de uma propriedade petrofísica com valor esperado e covariância constantes independente da localização (MCLENNAN, 2007). Estacionariedade, definida por Matheron (1962), é uma propriedade do modelo de função aleatória, não é uma característica intrínseca da variável. É uma decisão tomada pelo geomodelador e necessária para a inferência.

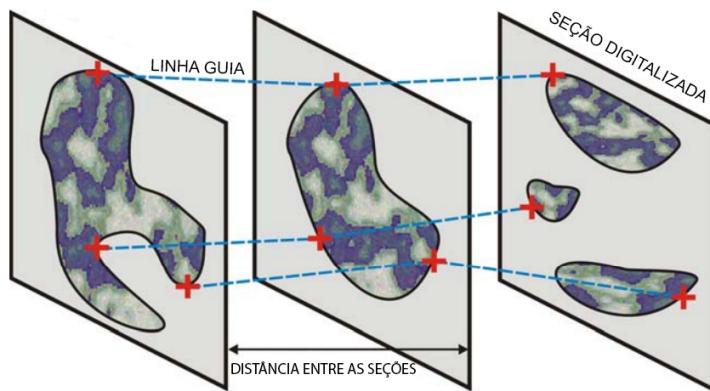
Os domínios estacionários não devem ser muito pequenos, dessa forma não incluiriam dados suficientes para descrição e inferência estatística confiável, e nem muito grandes, já que os dados, provavelmente, poderiam ser subdivididos em grupos geologicamente mais homogêneos. A definição adequada dos domínios estacionários é uma etapa importante na avaliação de recursos/reservas, a mistura de populações produz estimativas sub ótimas que superestimam ou subestimam teores e tonelagens. Além disso, nenhuma técnica geoestatística pode compensar uma definição de estacionariedade inadequada (ROSSI; DEUTSCH, 2013).

Uma vez que os domínios estacionários tenham sido definidos, um modelo tridimensional com os limites de cada função aleatória estacionária deve ser construído: o modelo geológico. É necessário identificar a "jurisdição espacial" de cada função aleatória estacionária. A etapa de interpretação e modelagem geológica compreende a definição dos domínios estacionários e a criação do modelo geológico tridimensional que forma o alicerce para todo o trabalho de estimativa subsequente. Muitas vezes, o modelo geológico é o fator de maior importância na estimativas das tonelagens mineralizadas (ROSSI; DEUTSCH, 2013).

A abordagem tradicional para a criação de modelos geológicos tridimensionais é através da triangulação de polilinhas¹ (*strings*) gerando sólidos que representam os corpos minerais. As polilinhas são digitalizadas manualmente em seções deslocadas, a partir dos dados de sondagem e conhecimento geológico prévio, por um geomodelador. As seções são unidas por linhas guias, que demarcam a conectividade entre as seções digitalizadas para triangulação adequada dos sólidos. Esse procedimento é referido como modelagem geológica explícita, já que os limites entre os diferentes domínios são definidos explicitamente por coordenadas (x,y,z) que posicionam a "colcha de retalhos" de triângulos (MCLENNAN; DEUTSCH, 2006; COWAN et al., 2003). A Figura 1 ilustra a metodologia explícita para um típico depósito de veio.

¹ Em computação gráfica: uma linha composta por vários segmentos, usada para compor imagens na tela (OXFORD, 2018).

Figura 1 – Ilustração esquematizando a modelagem geológica explícita usando polilinhas e linhas guias.



Modificado de [McLennan e Deutsch \(2006\)](#)

De acordo com [McLennan e Deutsch \(2006\)](#), embora a metodologia seja direta e simples, principais motivos da sua ampla implementação na prática, e que softwares modernos de mineração forneçam ferramentas computacionais para visualizar os dados de sondagem e agilizar o processo de digitalização ([SILVA, 2015](#)), existem uma série de limitações e desvantagens. O processo é tedioso e demorado, digitalizar manualmente as polilinhas e uni-las por linhas guia, exige muito tempo de um profissional experiente. Em depósitos de alta complexidade, não é raro o geomodelador trabalhar até três meses no modelo geológico explícito. A geometria dos corpos geológicos muitas vezes precisa ser simplificadas para que o modelo seja concebido em tempo hábil. Por esse motivo, para a maioria das minas, apenas um modelo é construído e mantido, raramente há oportunidade de modelar interpretações geológicas alternativas e comparar estimativas de recursos baseadas em diferentes modelos ([COWAN et al., 2003](#)).

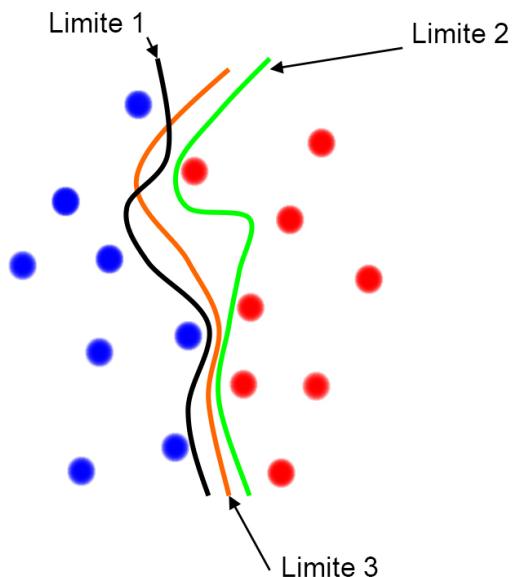
Os modelos criados explicitamente são subjetivos e não replicáveis. O volume mineralizado é essencialmente composto por uma série de pequenas decisões subjetivas, já que cada ponto da polilinha em cada seção é escolhido por um geomodelador. Inevitavelmente, a "assinatura" do profissional é impressa nos limites dos domínios geológicos. Geomodeladores diferentes criam modelos diferentes a partir do mesmo banco de dados. Ademais, modelos explícitos são inflexíveis, é laborioso e demorado atualizar modelos explícitos a medida que novos dados são obtidos já que uma nova digitalização manual é necessária.

Embora a metodologia explícita seja demorada, laboriosa, subjetiva, não replicável, inflexível e incapaz de avaliar incertezas, o limites entre domínios resultantes, geralmente, são realistas. Realismo geológico é um dos principais objetivos da modelagem e pode ser diretamente controlado durante o processo de digitalização.

1.2 Motivação

Em virtude da onerosidade em construir múltiplos modelos geológicos explícitos, é custoso avaliar a incerteza na localização dos limites dos diferentes domínios entre os dados amostrais, como mostrado na ilustração esquemática na [Figura 2](#). Além da tediosa redigitalização manual de polilinhas e retriangulação, não existe forma direta de incorporar múltiplas realizações possíveis para a localização dos limites que representem a incerteza associada.

Figura 2 – Ilustração esquemática destacando a incerteza na localização do limite entre os domínios azul e vermelho, entre os dados amostrais.



Modificado de [Cáceres et al. \(2011\)](#)

Em muitos casos, a incerteza do modelo geológico pode ser uma fonte de incerteza crucial. Em depósitos de veio de ouro, por exemplo, o volume mineralizado é um indicador econômico vital para o gerenciamento do projeto. Ignorar a incerteza volumétrica, considerando apenas um único modelo geológico explícito, pode ser devastador para o empreendimento. A incerteza associada ao modelo geológico deve ser avaliada.

Dadas as desvantagens do método explícito de modelagem geológica, pesquisas vêm sendo realizadas com o objetivo de automatizar, ou pelo menos, semi automatizar a modelagem e avaliar incerteza dos modelos geológicos.

Alguns dos métodos de simulação para variáveis categóricas, amplamente utilizados pela indústria da mineração, geram realizações que apresentam excessivo ruído. Além da necessidade de avaliar a incerteza dos modelos geológicos, existe também a preocupação de que os modelos finais apresentem limites contínuos e realistas entre as diferentes litologias do depósito. Não só para que os modelos tenham uma aparência mais agradável aos olhos, mas

também para que a tarefa de definir *diglines*, entre minério e estéril, seja facilitada e menos subjetiva.

1.3 Declaração de tese

A modelagem geológica implícita com funções distância assinaladas vem sendo utilizada há mais de uma década com sucesso como uma alternativa matemática à modelagem explícita na criação de modelos geológicos determinísticos. A metodologia está presente em diversos softwares comerciais de ampla utilização na indústria. Embora exista a necessidade de avaliar a incerteza associada ao modelo geológico, não há na literatura diferentes opções de métodos de avaliação de incerteza baseadas em funções distância assinaladas, especialmente para modelos multi-categóricos.

1.4 Objetivos

O objetivo dessa tese é propor e investigar três diferentes metodologias para avaliação de incerteza em modelos geológicos multi-categóricos usando funções distância assinaladas. As realizações obtidas a partir das metodologias propostas devem apresentar feições geológicas realistas e contínuas. Os métodos foram desenvolvidos e implementados em forma de *plug-ins* para o software *SGeMS/AR2GeMS* ([REMY; BOUCHER; WU, 2009](#)) ou *jupyter notebooks* baseados em softwares da biblioteca *GSLib* ([DEUTSCH; JOURNEL, 1992](#)).

1.4.1 Desenvolvimento de software

Foi desenvolvida uma *suite* de *plug-ins* de modelagem geológica para o software geostatístico *SGeMS/AR2GeMS*: [LPM-Geomod-Suite](#). Os *plug-ins* abrangem todas as etapas da modelagem geológica com funções distância assinaladas e indicadores desde transformações, interpolação, e avaliação da incerteza associada ao modelo geológico.

Um manual de instalação e uso dos *plug-ins* é apresentado no [Apêndice A](#).

1.4.2 Avaliação de incerteza pela parametrização do suporte do *kernel*

É proposto um método para avaliação de incerteza baseado na parametrização do suporte do *kernel*, que é a base para a interpolação das distâncias assinaladas utilizando funções de bases radiais (RBF). O *kernel* é análogo ao variograma da krigagem. Para cada uma das variáveis, são geradas interpolações: para um cenário base, sem parametrização, um cenário otimista e um cenário pessimista. As diferentes propriedades interpoladas são combinadas para gerar diferentes realizações para o modelo geológico.

1.4.3 Avaliação de incerteza usando funções distância assinaladas e campos de probabilidade

É proposto um método de avaliação de incerteza utilizando funções distância assinaladas e campos de probabilidade. Uma zona de incerteza é definida ao redor dos contatos, então, as distâncias assinaladas interpoladas no interior da zona de incerteza são transformadas em probabilidades. Em seguida, os campos de probabilidades gerados a partir de simulações Gaussianas não condicionais são utilizados para simular as diferentes realizações para o modelo geológico a partir do método de Monte Carlo.

1.4.4 Simulação de contatos hierárquica para múltiplas categorias

Uma adaptação do método proposto por [Wilde e Deutsch \(2012\)](#) para múltiplas categorias simultâneas é proposto. Um algoritmo que identifica e define a frequência de ocorrência dos diferentes contatos é usado para agrupá-los de forma hierárquica. Os contatos representados por cada grupo são simulados de forma independente e a regra hierárquica é usada para construir as diferentes realizações do modelo geológico multi-categórico.

1.5 Estrutura da tese

O capítulo 1 dessa tese traz uma introdução ao assunto modelagem geológica e avaliação de incerteza e apresenta: a motivação, declaração e objetivos da tese.

O capítulo 2 é uma revisão extensiva da literatura em modelagem geológica e avaliação de incerteza em modelos geológicos usando funções distância assinaladas.

O capítulo 3 apresenta as metodologias propostas utilizando o conhecido *dataset Swiss Jura* ([GOOVAERTS, 1997](#)) como prova de conceito em 2D. Um estudo de caso é conduzido para cada uma das metodologias propostas mostrando sua eficiência. As características e particularidades de cada uma delas e os resultados obtidos são discutidos.

Finalmente, o capítulo 4 apresenta as observações finais, um sumário das contribuições da tese e sugere trabalhos futuros.

2 REVISÃO BIBLIOGRÁFICA

Esse capítulo é uma revisão da literatura em modelagem geológica automática e avaliação de incerteza em modelos geológicos com ênfase no uso de funções distância assinaladas.

2.1 Modelagem geológica determinística

A dispendiosidade na construção e a inflexibilidade dos modelos explícitos impulsionaram o desenvolvimento de técnicas de modelagem geológica automáticas. As técnicas determinísticas geram apenas um único modelo, consequentemente, não possibilitando a avaliação da incerteza associada ao modelo geológico.

Métodos matemáticos menos sofisticados, como o vizinho mais próximo podem ser utilizados na criação automática de modelos geológicos determinísticos. Esse algoritmo constrói ao redor de cada amostra sua área de influência, ou volume de influência em três dimensões. A toda a área de influência é atribuída a categoria, ou litologia, correspondente àquela amostra.

Métodos baseados em aprendizado de máquina também podem ser utilizados. Smirnoff, Boisvert e Paradis (2008) propuseram o uso de máquina de vetores de suporte (*support vector machine*) para a criação de modelos geológicos determinísticos multi-categóricos a partir de amostras esparsas de diferentes origens.

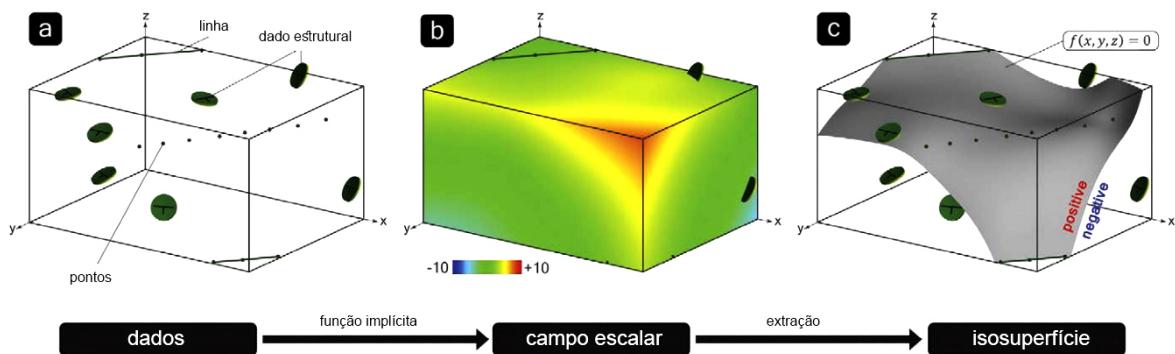
Métodos geoestatísticos também têm aplicação na criação de modelos geológicos de forma automática, a krigagem dos indicadores (ALABERT, 1987) consiste na geração de uma distribuição acumulada de probabilidades a partir da transformação não linear dos dados. Para variáveis categóricas, é possível estimar diretamente a probabilidade de cada local não amostrado pertencer a cada uma das diferentes litologias do depósito mineral. A partir das probabilidades um modelo geológico pode ser criado.

A família de métodos ditos métodos implícitos compreendem técnicas importadas da computação gráfica. O uso da modelagem implícita foi introduzida no campo da computação gráfica para a geração de objetos de diferentes geometrias e complexidades por Bloomenthal e Bajaj (1997). A ideia geral é usar uma função implícita para demarcar regiões de diferentes formas e extensões no espaço. Cowan et al. (2002) e Cowan et al. (2003) popularizaram a modelagem implícita no contexto geológico, baseados no trabalho de Savchenko et al. (1995), que modela objetos tridimensionais a partir de dados esparsos interpolando uma função volume.

Modelos geológicos implícitos são criados a partir de dados esparsos, geralmente, dados categóricos (Figura 3 (a)). Os pontos amostrais podem ser usados para derivar uma função implícita que fornece uma representação matemática contínua de um atributo através de um

volume, um campo escalar ou modelo implícito (Figura 3 (b)). Modelos implícitos contém um número infinito de isosuperfícies. Para visualizar o modelo geológico uma isosuperfície específica deve ser extraída do modelo e mostrada no espaço tridimensional (geralmente a isosuperfície zero), $f(x, y, z) = s$, onde s é o valor de interesse no campo escalar (Figura 3 (c)). A localização dessa superfície é conhecida implicitamente como função da localização no domínio (MARTIN; BOISVERT, 2017a).

Figura 3 – Ilustração esquemática da modelagem geológica implícita.



Modificado de (AILLERES et al., 2015)

Diferentes tipos de dados amostrais podem ser usados para derivar diferentes funções volume na modelagem implícita. Mallet (2004) propõe uma função volumétrica cronológica, levando em consideração a posição estratigráfica das diferentes unidades geológicas, enquanto Lajaunie, Courrioux e Manuel (1997), Chilès et al. (2004), Calcagno et al. (2008) usam cokrigagem de incrementos em um campo potencial, omitindo a função volume. O procedimento não é simples já que a covariância do campo potencial é particularmente difícil de inferir uma vez que não há *hard data* de campo potencial disponível.

A função volume mais comumente utilizada é a função distância assinalada (OSHER; FEDKIW, 2003). A aplicação dessa metodologia é encontrada por toda a literatura de interpolação de dados esparsos, uma das aplicações é a reconstrução de superfícies a partir de escaneamento. Na modelagem geológica, a metodologia é competente em capturar a geometria e extensão de corpos geológicos e tem sido aplicada com sucesso há mais de uma década na exploração mineral tendo ganhado espaço em diferentes *softwares* comerciais. Para bancos de dados pontuais, que representam a posição de diferentes litologias no espaço, o uso das funções distância assinaladas torna os processos de interpretação, cálculo, modelagem, modificação do modelo de acordo com a interpretação dos geomodeladores e avaliação da incerteza diretos e rápidos. Além disso, informações a respeito da orientação das unidades geológicas (dados estruturais) podem ser imputada como restrição na criação do modelo geológico (MARTIN; BOISVERT, 2017a).

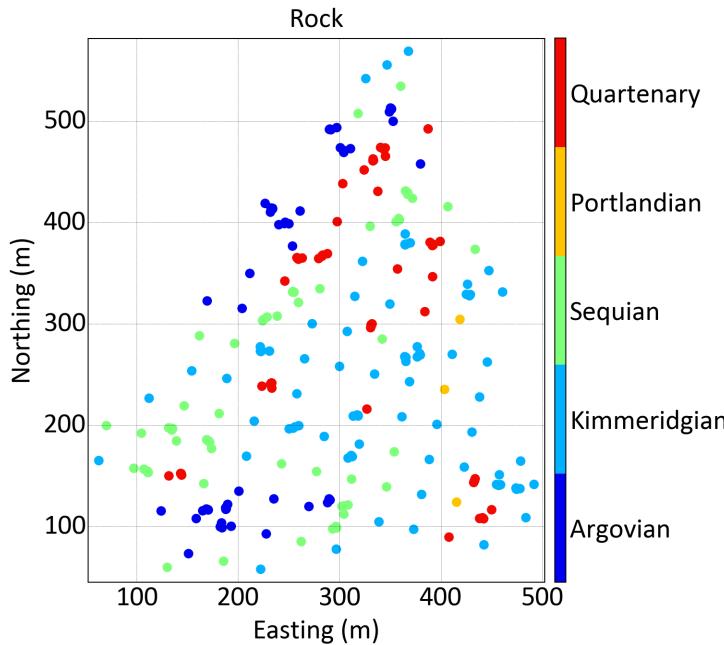
As próximas seções trazem uma revisão bibliográfica extensiva da modelagem geológica com funções distância assinaladas.

2.1.1 Modelagem geológica com funções distância assinaladas

Para ilustrar a metodologia será utilizado o *dataset Swiss Jura* (GOOVAERTS, 1997). O *dataset* compreende 259 amostras da propriedade *rock type* (RT). Um mapa com a localização das amostras é mostrado na Figura 4.

A variável RT possui cinco categorias: *Argovian*; *Kimmeridgian*; *Sequian*; *Portlandian*; e *Quartenary*. A área é coberta parcialmente com amostras uniformemente espaçadas, algumas áreas são densamente amostradas com agrupamento adicional de amostras.

Figura 4 – Mapa de localização das 259 amostras do *dataset Swiss Jura*.



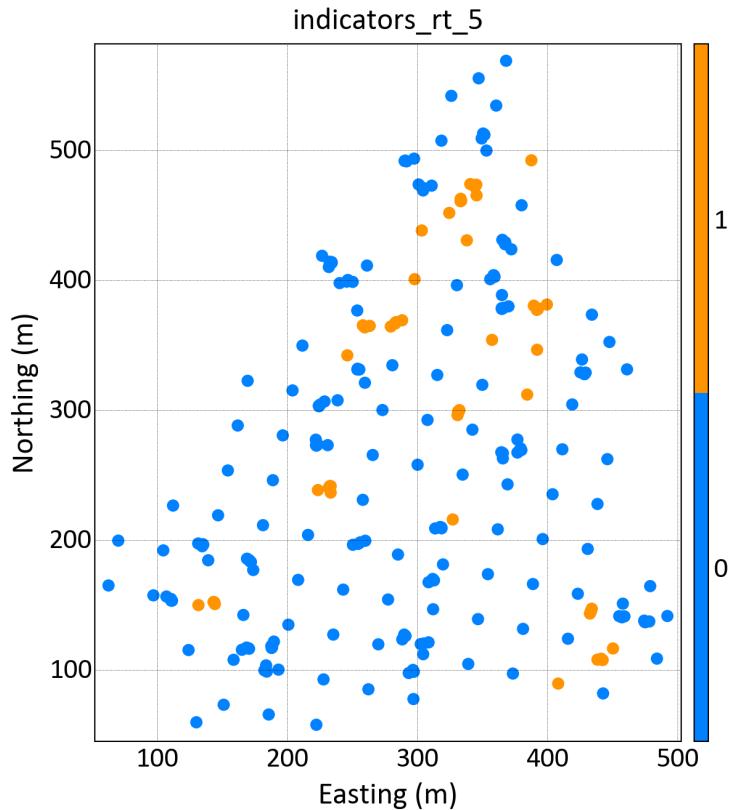
2.1.1.1 Cálculo das distâncias assinaladas

O primeiro passo no cálculo das distâncias assinaladas é transformar as amostras $z(u_\alpha)$, $\alpha = 1, \dots, n$ em indicadores binários de acordo com a Equação 2.1, especificando se a amostra pertence ou não ao domínio k que está sendo modelado.

$$i_k(u_\alpha) = \begin{cases} 1, & \text{se } z(u_\alpha) \text{ se pertence ao domínio } k \\ 0, & \text{se } z(u_\alpha) \text{ caso contrário} \end{cases} \quad (2.1)$$

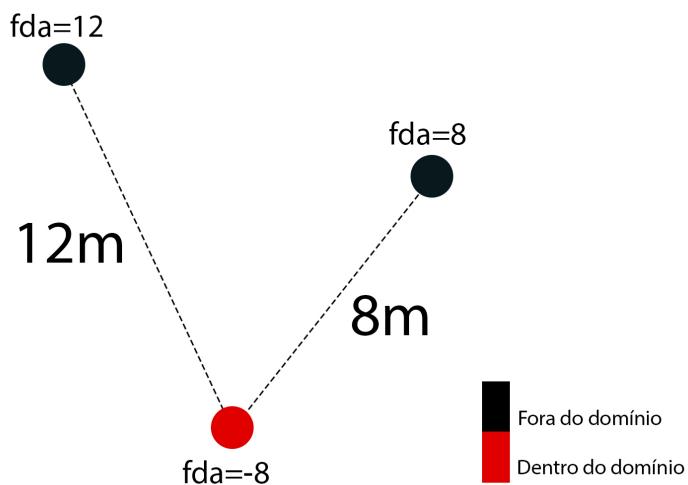
Como exemplo será utilizada a categoria Quaternary do *Swiss Jura*. A Figura 5 mostra as amostras do *Swiss Jura* transformadas em indicadores para essa categoria.

Figura 5 – Mapa de localização das 259 amostras dos indicadores para a categoria Quaternary do dataset Swiss Jura.



O segundo passo é o cálculo das distâncias assinaladas. Para cada ponto amostral, a menor distância euclideana até um outro ponto amostral que pertença a um indicador oposto é computada, e esse valor atribuído àquele ponto, com o sinal negativo caso pertença ao domínio modelado e com o sinal positivo, caso contrário, como ilustrado na [Figura 6](#).

Figura 6 – Ilustração esquemática mostrando o cálculo das distâncias assinaladas.



Desse modo, o valor da função distância assinalada d_k para a o domínio k modelado no

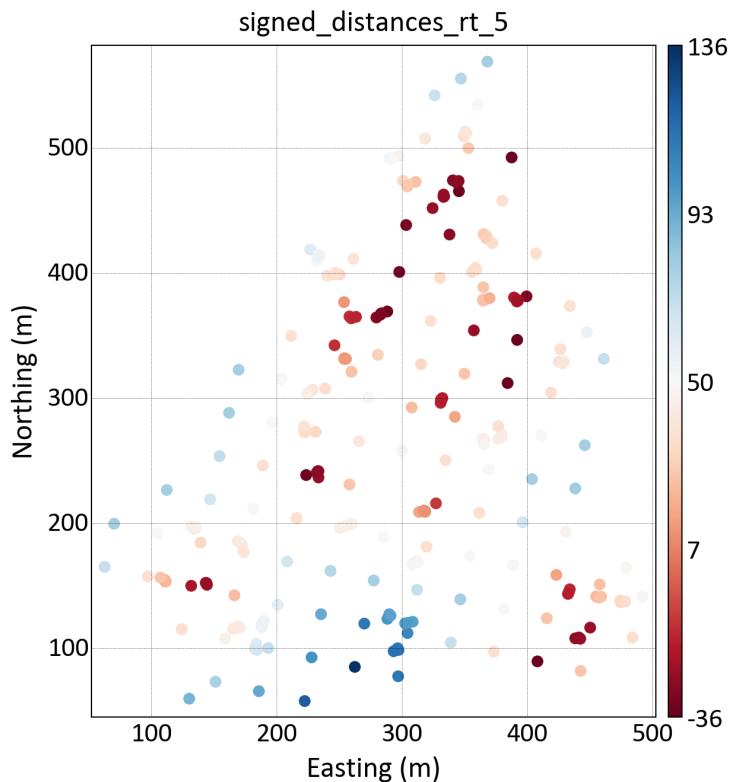
local u_α é:

$$d_k(u_\alpha) = \begin{cases} -\|u_\alpha - u_\beta\|, & \text{se } u_\alpha \text{ pertence ao domínio} \\ +\|u_\alpha - u_\beta\|, & \text{se } u_\alpha \text{ não pertence ao domínio} \end{cases} \quad (2.2)$$

O local u_β corresponde à amostra mais próxima codificada com um indicador diferente de u_α .

A Figura 7 mostra as distâncias assinaladas calculadas para a categoria Quaternary do dataset Swiss Jura.

Figura 7 – Mapa de localização das distâncias assinaladas para a categoria Quaternary do dataset Swiss Jura.



Para corpos geológicos que se estendem muito mais em uma direção em relação às demais, como os corpos tabulares por exemplo, as distâncias calculadas podem ser anisotrópicas. Sendo assim, as coordenadas originais x , y e z dos dados devem ser rotacionadas e/ou contraídas/dilatadas, a partir da transformação da Equação 2.3. Então, as distâncias euclidianas (Equação 2.17) são calculadas normalmente para as novas coordenadas x'' , y'' , z'' .

$$\begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} = \begin{bmatrix} \frac{1}{a_{max}} & 0 & 0 \\ 0 & \frac{1}{a_{min}} & 0 \\ 0 & 0 & \frac{1}{a_{vert}} \end{bmatrix} \begin{bmatrix} \cos \alpha \cos \phi - \sin \alpha \sin \beta \sin \phi & -\sin \alpha \cos \phi - \cos \alpha \sin \beta \sin \phi & \cos \beta \sin \phi \\ \sin \alpha \cos \beta & \cos \alpha \cos \beta & \sin \beta \\ -\cos \alpha \sin \phi - \sin \alpha \sin \beta \cos \phi & \sin \alpha \sin \phi - \cos \alpha \sin \beta \cos \phi & \cos \beta \cos \phi \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (2.3)$$

Onde a_{max} , a_{min} e a_{vert} são as dimensões dos eixos máximo, médio e mínimo, e α , β e ϕ , os ângulos de azimute, mergulho e *rake* que caracterizam os eixos de anisotropia.

2.1.1.2 Variografia

No caso das distâncias assinaladas, obter um modelo de covariância a partir dos variogramas experimentais não é trivial. Distâncias assinaladas não são estacionárias, por isso, o variograma não se estabiliza em um patamar. Além disso, o caráter extremamente contínuo das distâncias torna a identificação analítica das direções principais um processo embaraçoso.

[Manchuck e Deutsch \(2015\)](#) sugerem alternativas para modelagem do variograma das distâncias:

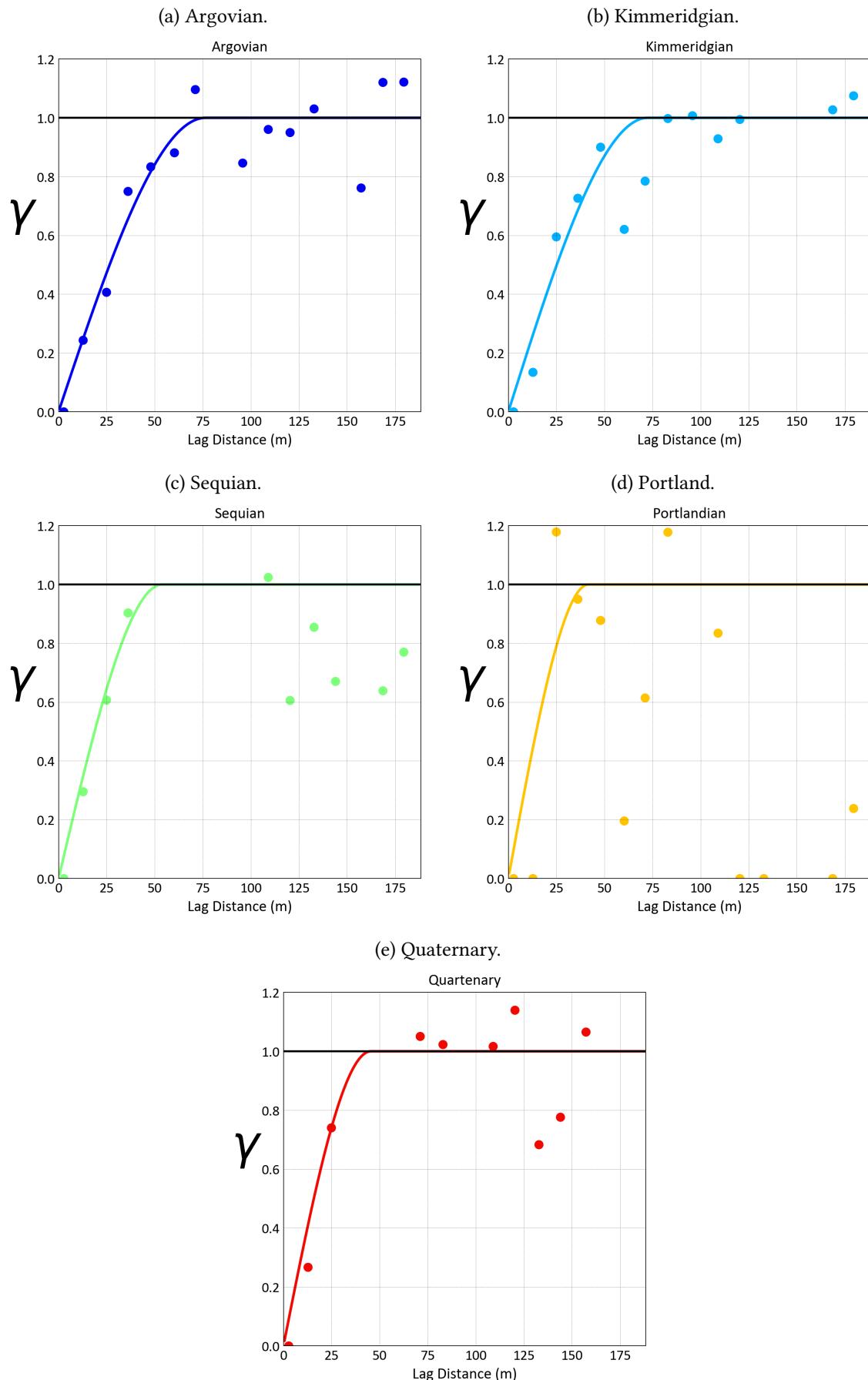
- Treinar o variograma usando validação cruzada;
- Tentar modelar interativamente os variogramas experimentais;
- Calcular e modelar os variogramas para as propriedades de indicadores e transformá-los em um equivalente Gaussiano para as distâncias assinaladas baseado na similaridade entre a simulação Gaussiana truncada e a modelagem geológica com funções distância assinaladas. Um campo Gaussiano diferenciável e contínuo é gerado e truncado em um limiar para que a proporção correta de um indicador particular seja o resultado. Entretanto, na prática, o variograma dos indicadores é usado para parametrizar o variograma das distâncias assinaladas de forma direta ([MARTIN; BOISVERT, 2017a](#));
- Inferir um modelo de covariância plausível a partir de mapas delineados a mão em softwares de iso-contorno.

Uma outra alternativa ao cálculo e modelagem de variogramas é o uso de tabelas de covariância proposto por [Kloeckner et al. \(2018\)](#), esse estudo propõe a substituição da modelagem tradicional dos variogramas por tabelas de covariância, tornando o processo de modelagem geológica completamente automático e livre da subjetividade do geomodelador.

O alcance do variograma, juntamente com a anisotropia, controlam a extensão e forma dos domínios modelados. Os variogramas experimentais das distâncias assinaladas não exibem efeito pepita. Porém, pode ser adicionado arbitrariamente pelo usuário para controlar a interconectividade dos domínios, adicionar ruído ao modelo quando desejável e controlar a extração que pode ser excessiva em alguns casos devido à alta continuidade dos modelos Gaussianos.

A [Figura 8](#) mostra o variograma omnidirecional dos indicadores para as categorias do Swiss Jura. Os variograma foram modelados com uma estrutura esférica e efeito pepita zero.

Figura 8 – Variogramas omnidirecionais dos indicadores.



As equações: Equação 2.4, Equação 2.5, Equação 2.6, Equação 2.7, Equação 2.8 representam os variogramas das categorias *Argovian*; *Kimmeridgian*; *Sequian*; *Portlandian*; e *Quartenary*, respectivamente.

$$\gamma(h) = 1.0Sph_1 \left[\frac{OMNI}{76m} \right] \quad (2.4)$$

$$\gamma(h) = 1.0Sph_1 \left[\frac{OMNI}{72m} \right] \quad (2.5)$$

$$\gamma(h) = 1.0Sph_1 \left[\frac{OMNI}{53m} \right] \quad (2.6)$$

$$\gamma(h) = 1.0Sph_1 \left[\frac{OMNI}{41m} \right] \quad (2.7)$$

$$\gamma(h) = 1.0Sph_1 \left[\frac{OMNI}{45m} \right] \quad (2.8)$$

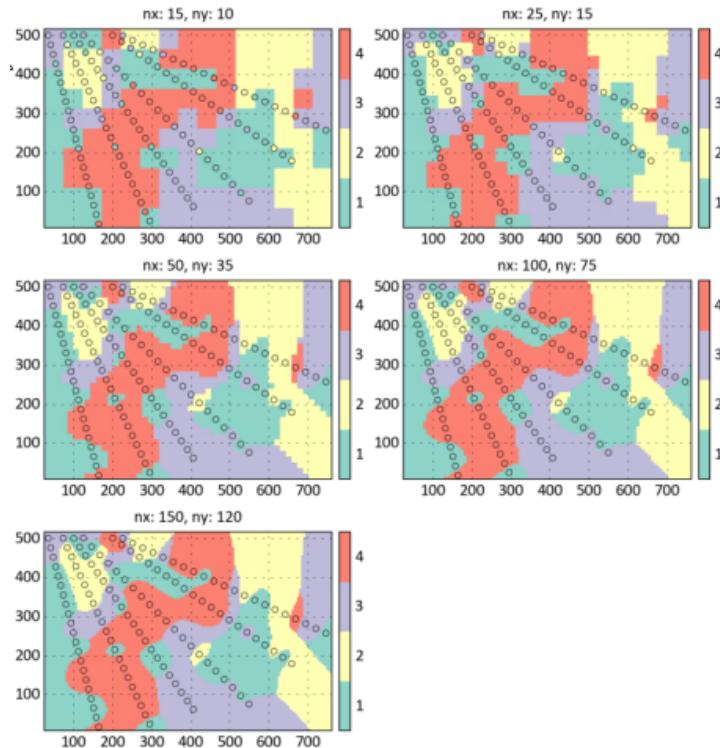
2.1.1.3 Interpolação

Uma vez que as distâncias tenham sido calculadas a partir da localização das amostras, o próximo passo é interpolar as propriedades distância assinaladas para o *grid* definido para a modelagem geológica. Embora métodos implícitos não demandem um *grid* - são métodos conhecidos como *gridless* - o objetivo final é preencher o *grid* de estimativas com as categorias correspondentes ou visualizar o modelo através de isosuperfícies. Técnicas de extração de isosuperfícies exigem a função volumétrica definida exaustivamente em um *grid* retilíneo (**MARTIN; BOISVERT, 2017a**).

Para um número constante de amostras, o tempo necessário para interpolação da função volume para todos os nós do *grid* aumenta linearmente com o número de nós. Os parâmetros do *grid* de estimativas, muitas vezes, são determinado por requisitos de engenharia do projeto. Todavia, o *grid* para definição do modelo geológico pode ser diferente do *grid* definido para os modelos numéricos de teores. A consideração mais importante é a capacidade de reproduzir a menor estrutura geológica de interesse.

Considere as seções de um modelo geológico na [Figura 9](#) gerados em *grids* de diferentes resoluções.

Figura 9 – Efeito da resolução do *grid* na reprodução de estruturas geológicas.



Fonte: [Martin e Boisvert \(2017a\)](#)

A resolução adequada para esse modelo é entre $(nx, ny) = (50, 35)$ e $(nx, ny) = (100, 75)$. Para os dois primeiros *grids* mais grosseiros há classificação errônea dos nós colocados com amostras, além de $(nx, ny) = (100, 75)$ há pouco benefício na reprodução das estruturas levando em consideração o aumento do tempo de interpolação.

É preciso encontrar um balanço entre resolução e tempo de interpolação. A resolução do *grid* pode ser aumentada a medida que o projeto avança e mais informação é adquirida ([MARTIN; BOISVERT, 2017a](#)).

Qualquer método de interpolação pode ser utilizado, mesmo os métodos do inverso da distância produzem resultados realistas. A krigagem e as funções de bases radiais permitem incorporar informação adicional através dos modelos de covariância parametrizados a partir das amostras.

[Hosseini e Deutsch \(2007\)](#) utilizaram inverso da distância, [Silva \(2015\)](#) utilizou krigagem global, [Rolo \(2017\)](#) utilizou krigagem ordinária, [Silva e Deutsch \(2015\)](#) aplicaram krigagem dual. [Boisvert \(2013\)](#) gerou modelos implícitos através de distâncias assinaladas com anisotropia variável local (*Locally varying anisotropy kriging - LVA*) e [Cowan et al. \(2002\)](#) propuseram a utilização de interpoladores baseados em funções de bases radiais (*radial basis functions - RBF*), [Manchuck e Deutsch \(2018\)](#) ainda propuseram a utilização de mínimos quadrados móveis para incorporar interpretação manual em seções e avaliar incerteza. A escolha do interpolador deve

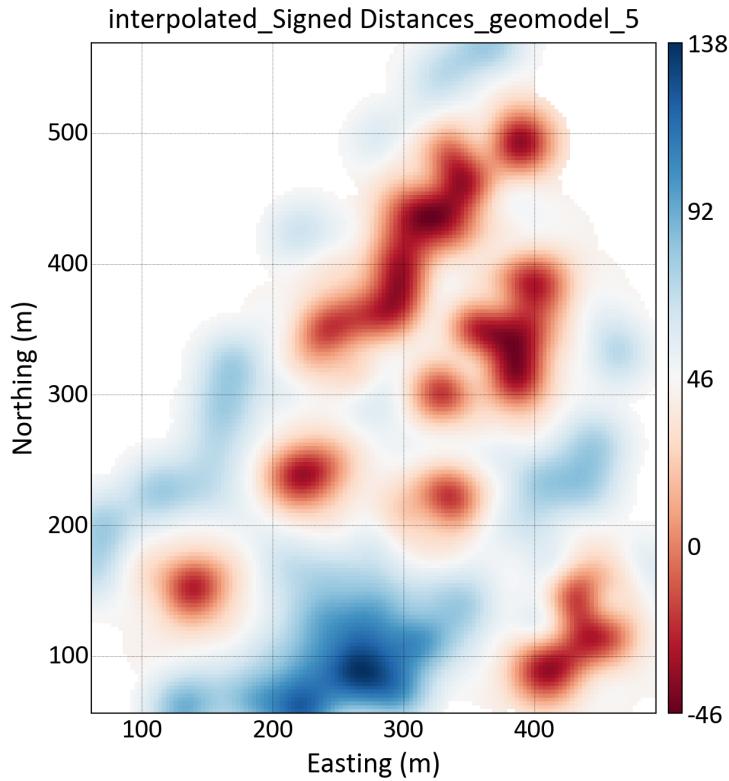
ser baseada em suas propriedades, na informação exigida para parametrizá-lo (o variograma, por exemplo) e nas sua capacidade de se ajustar a formas geológicas.

Especialmente para as funções distância assinaladas, existe um problema adicional relacionado à estacionariedade de primeira e segunda ordem. Existe uma forte tendência (*trend*) que restringe os estimadores baseados em krigagem para que a estacionariedade seja mantida. Além disso, formas geológicas são curvilíneas e não são bem descritas por funções de covariância "tradicionalis" onde a distância euclideana entre os pares de pontos são medidas em linha reta. Por esses motivos muitos autores preferem as funções de bases radiais já que ela não se baseia em estacionariedade de primeira ordem e honram localmente formas geológicas sem a necessidade de parametrização especial (MARTIN; BOISVERT, 2017a).

Independente do método de interpolação, os modelos gerados devem ser livres de artefatos para que os artefatos do modelo implícito não sejam transferidos para os modelos geológicos. Por esse motivo, interpoladores globais devem ser usados por utilizarem todas as amostras em cada estimativa, já para os métodos baseados em inverso da distância e krigagem (não global) os parâmetros de busca são uma escolha crítica (MARTIN; BOISVERT, 2017a). Em contrapartida, métodos globais possuem limitações em bancos de dados volumosos por dependerem de muito processamento e memória RAM.

A Figura 10 mostra as distâncias assinaladas para a categoria Quaternary do Swiss Jura interpolada por krigagem dual (subseção 2.1.1.3.2) em um grid 3x3 metros. O variograma dos indicadores da Figura 8e foi usado para para parametrizar o variograma Gaussiano - mesmas estruturas, mesmos ranges, mesmas relações e ângulos de anisotropia.

Figura 10 – Categoria Quaternary do *Swiss Jura* interpolada por krigagem dual em um grid 3x3 metros.



Os principais interpoladores globais são listados a seguir.

2.1.1.3.1 Krigagem global

A krigagem global (NEUFELD; WILDE, 2005) não produz nenhum artefato relacionado à estratégia de busca, porque não há estratégia de busca. O algoritmo lê todos os dados e usa todos os valores dos dados para estimar cada local não amostrado. Um benefício adicional da krigagem global é a redução do tempo de CPU para o cálculo das estimativas já que não é necessário buscar amostras no espaço e a matriz de covariâncias entre as amostras é calculada e invertida uma única vez. O valor estimado da função distância assinalada para um local não amostrado é dado pela [Equação 2.9](#):

$$d^*(u) = \sum_{i=1}^n \lambda_i(u) d(u_i) \quad (2.9)$$

Onde: d^* é a distância estimada no local não amostrado u , $\lambda_i(u)$ é o peso dado para a distância assinalada d observada no local u_i .

Os pesos podem ser calculados resolvendo as equações de krigagem. Neste caso, as equações de krigagem ordinária:

$$\begin{bmatrix} C_{11} & \dots & C_{1n} & 1 \\ \vdots & \ddots & \vdots & 1 \\ C_{n1} & \dots & C_{nn} & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} C_{10} \\ \vdots \\ C_{n0} \\ 1 \end{bmatrix} \quad (2.10)$$

Onde C_{ij} é a covariância entre as amostras i e j , C_{i0} é a covariância entre a amostra i e o local a ser estimado, n é o número de amostras no dataset e μ é o operador lagrangeano necessário para a krigagem ordinária (ISAAKS; SRIVASTAVA R, 1989). A condição de não enviesamento é $\sum_{i=1}^n \lambda_i = 1$.

2.1.1.3.2 Krigagem dual

A krigagem dual (ROYER; VIEIRA, 1984) é uma representação alternativa da krigagem na qual as estimativas são expressas como uma combinação linear das covariâncias ao invés das amostras. Dessa forma, o estimador da Equação 2.11 torna-se:

$$d^*(u) = \sum_{i=1}^n \omega_i C(u - u_i) + m_{OK}^* \quad (2.11)$$

Onde ω_i é o peso dual associado à covariância $C(u - u_i)$ e m_{OK}^* é uma estimativa da tendência (*trend*). A condição de não enviesamento é $\sum_{i=1}^n d_i = 0$

Ao contrário da krigagem simples e ordinária o sistema de equações não resulta da minimização da variância do erro, mas sim da propriedade de exatidão do estimador de krigagem.

Os pesos são calculados resolvendo a Equação 2.12:

$$\begin{bmatrix} C_{11} & \dots & C_{1n} & 1 \\ \vdots & \ddots & \vdots & 1 \\ C_{n1} & \dots & C_{nn} & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_n \\ m_{OK}^* \end{bmatrix} = \begin{bmatrix} d_1 \\ \vdots \\ d_n \\ 0 \end{bmatrix} \quad (2.12)$$

Onde d_n é o valor da amostra da distância assinalada n. m_{OK}^* e ω_i precisam ser calculados apenas uma única vez e usados para todas as estimativas.

A krigagem dual não necessariamente deve ser global, todavia, sem nenhuma adaptação, seus benefícios só podem ser capitalizados se uma vizinhança global for usada (AUÑÓN; GÓMEZ-HERNÁNDEZ, 2000).

2.1.1.3.3 Funções de bases radiais (RBF)

A interpolação com funções de bases radiais (FASSHauer, 2007) é uma técnica global semelhante à krigagem dual. As RBFs são baseadas no posicionamento de um *kernel* radial em

cada ponto amostral. A superfície interpolada é uma combinação linear desses *kernels*. O *kernel* é análogo ao variograma da krigagem. O estimador é descrito pela Equação 2.13:

$$d^*(u) = \sum_{i=1}^n \lambda_i \theta(u - u_i) \quad (2.13)$$

Onde θ é o *kernel*, que pode ser parametrizado a partir do variograma dos indicadores para uma determinada categoria, e λ_i são os pesos calculados a partir da solução da Equação 2.14:

$$\begin{bmatrix} \theta_{11} & \dots & \theta_{1n} \\ \vdots & \ddots & \vdots \\ \theta_{n1} & \dots & \theta_{nn} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} \quad (2.14)$$

θ_{ij} é o valor do *kernel* entre as amostras i e j .

Fasshauer (2007) advoga que o *kernel* pode ser parametrizado pela distância que representa a maior esfera que pode ser colocada entre amostras no domínio. Para modelagem geológica, essa técnica tende a super estimar o suporte em relação à parametrização pelo variograma calculado a partir das amostras (MARTIN; BOISVERT, 2017a).

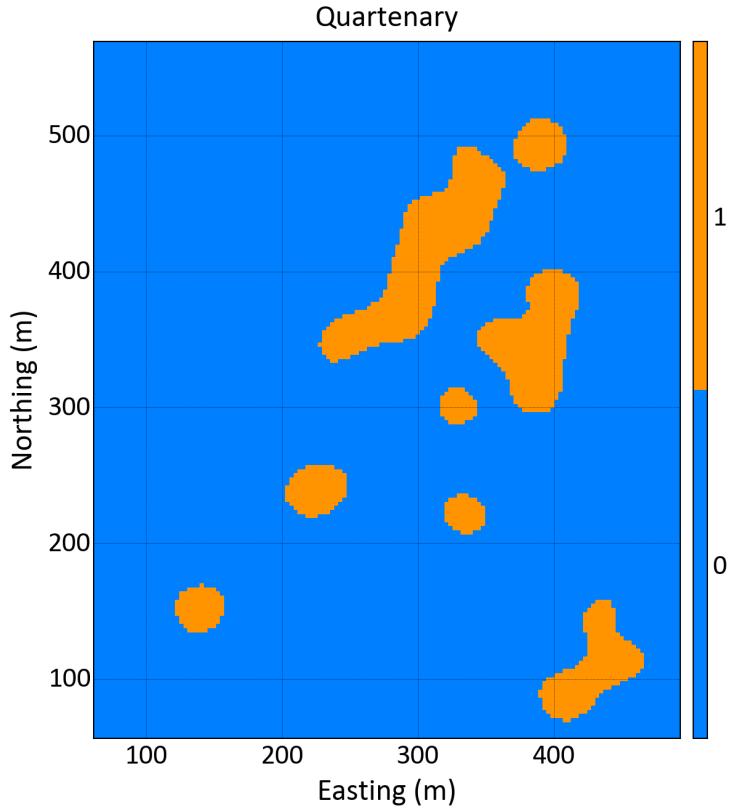
2.1.1.4 Visualização do modelo geológico

Como as funções distância são negativas no interior do domínio e positivas no exterior, um bom palpite inicial para a interface que separa os domínios no espaço, seria a linha (em duas dimensões) ou superfície (em três dimensões) que corresponda ao valor zero da função distância assinalada (WILDE; DEUTSCH, 2011). Dessa forma, blocos em que a distância estimada tem valor negativo, são classificados como pertencentes ao domínio. Blocos em que a distância estimada tem valor positivo, classificados como não pertencentes ao domínio de acordo com a Equação 2.15:

$$i^*(u) = \begin{cases} 1, & \text{se } d^*(u) \leq 0 \\ 0, & \text{caso contrário} \end{cases} \quad (2.15)$$

A Figura 11 mostra o modelo geológico para a categoria Quaternary do Swiss Jura. Uma animação que mostra as coordenadas das distâncias assinaladas estimadas nos eixos x e y e o valor da distância assinalada no eixo z pode ser vista [aqui](#). O plano corta as distâncias assinaladas na isolinha zero.

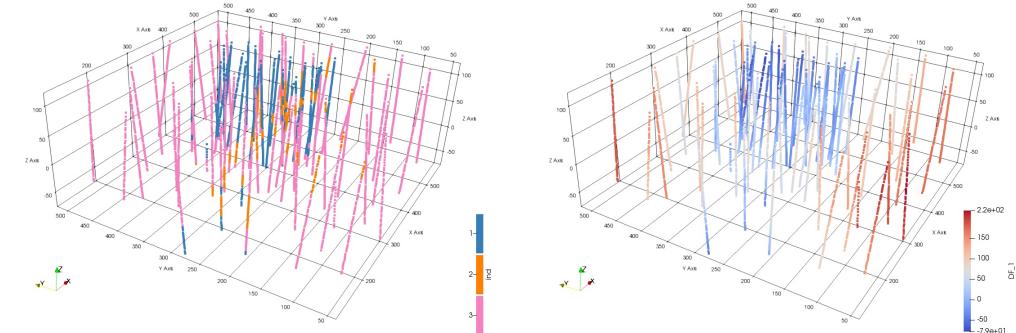
Figura 11 – Modelo geológico para a categoria Quaternary do *Swiss Jura*.



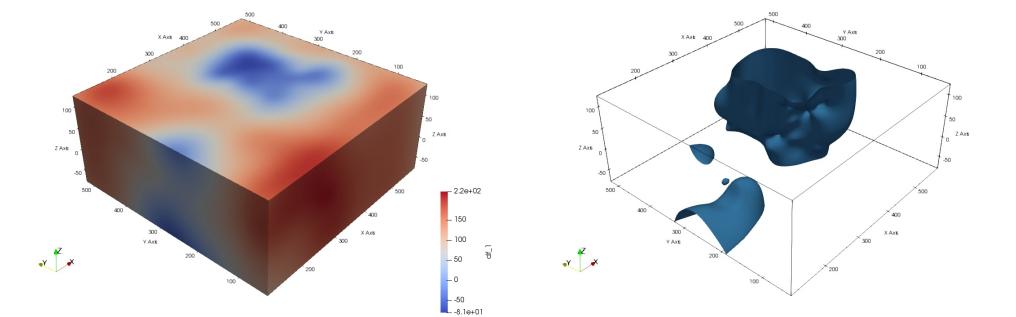
Em três dimensões um algoritmo de extração de isosuperfícies pode ser aplicado ao modelo implícito para extrair a isosuperfície zero que pode ser posteriormente visualizada como mostrado na [Figura 12](#). Esse método cria superfícies suaves já que não são delimitadas pelos blocos do *grid* como na [Figura 11](#).

Figura 12 – Exemplo mostrando a extração da isosuperfície zero e visualização em três dimensões.

- (a) Amostras categóricas do banco dados tridimensional.
 (b) Amostras das distâncias assinaladas calculadas para a categoria sendo modelada



- (c) Modelo implícito interpolado por RBF para a categoria sendo modelada.
 (d) Isosuperfície zero extraída a partir do modelo implícito.



Na presença de múltiplos domínios, as isosuperfícies ou isolinhas, devem ser extraídas uma a uma, e algum tipo de regra hierárquica baseada na idade de cada litologia deve ser aplicada para a criação dos modelos multi-categóricos.

2.1.1.5 Adaptação para múltiplas categorias

Silva e Deutsch (2012) propuseram, uma adaptação para o método com o objetivo de modelar múltiplos domínios geológicos simultaneamente, de forma similar ao caso binário.

Se existirem K múltiplos domínios no depósito mineral, para cada ponto amostral $z(u_\alpha)$, $\alpha = 1, \dots, n$, um vetor de indicadores de K elementos é codificado de acordo com a Equação 2.16.

$$i_k(u_\alpha) = \begin{cases} 1, & \text{se } z(u_\alpha) = k \\ 0, & \text{se } z(u_\alpha) \text{ caso contrário} \end{cases} \quad k = 1, \dots, K \quad (2.16)$$

A função distância é calculada, individualmente para cada elemento k do vetor, de acordo com a Equação 2.17.

$$d_k(u_\alpha) = \begin{cases} -\|u_\alpha - u_\beta\|, & \text{se } i_k(u_\alpha) = 1 \\ +\|u_\alpha - u_\beta\|, & \text{se } i_k(u_\alpha) = 0 \end{cases} \quad k = 1, \dots, K \quad (2.17)$$

As distâncias calculadas são então interpoladas pelo método escolhido, individualmente para todos os nós do *grid* de acordo com a Equação 2.18.

$$d_k^*(u) = \sum_{\alpha=1}^n \lambda_\alpha(u) d_k(u_\alpha) \quad k = 1, \dots, K \quad (2.18)$$

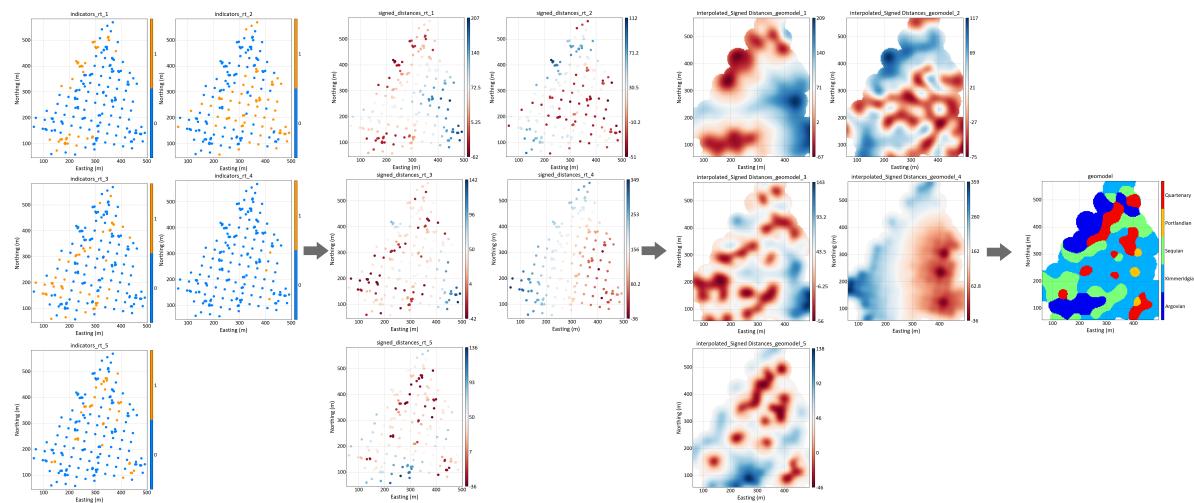
Por fim, cada bloco é classificado pela Equação 2.19

$$i^*(u) = k' \text{ de modo que } d_{k'}^* = \min\{d_k^*(u)\}_{k=1}^K \quad (2.19)$$

As distâncias estimadas fornecem uma medida de proximidade ao domínio oposto mais próximo. Sendo assim, a mínima distância assinalada estimada é tida como o domínio mais provável de ser encontrado numa região não amostrada. A Equação 2.19 sumariza essa ideia (SILVA, 2015). A categoria associada com a menor distância estimada é retida em cada bloco.

A Figura 13 mostra, da esquerda para a direita, as amostras do *Swiss Jura* transformadas em indicadores para cada uma das cinco categorias do depósito, as distâncias assinaladas calculadas para cada uma das propriedades de indicadores, as distâncias estimadas para um grid 3x3 metros por krigagem dual usando equivalentes Gaussianos dos variogramas da Figura 8 e finalmente, o modelo geológico multi-categórico criado a partir da retenção da categoria responsável pela menor distância assinalada estimada em cada local não amostrado. Uma animação que mostra as coordenadas das distâncias assinaladas estimadas nos eixos x e y e o valor da distância assinalada no eixo z para cada uma das cinco categorias pode ser vista [aqui](#).

Figura 13 – Figura esquemática mostrando os passos para a criação do modelo geológico multi-categórico no *Swiss Jura*.



2.1.1.6 Acelerando o processo

Embora o uso de métodos globais permita a criação de modelos realistas e livres de artefatos, existem limitações em bancos de dados volumosos já que métodos globais demandam muito processamento e memória RAM. Existem algumas técnicas para superar essa limitação como a solução iterativa (BEATSON; CHERIE; MOUAT, 1999) por exemplo. Dois métodos que apresentam bons resultados no contexto da modelagem geológica com distâncias assinaladas serão apresentados.

2.1.1.6.1 Decomposição do domínio

A decomposição do domínio (*Partition of unity - POU*) (WENDLAND, 2004), transforma um problema volumoso e que demanda muito esforço computacional em diversos problemas menores e eficientes que são, ao final, unidos.

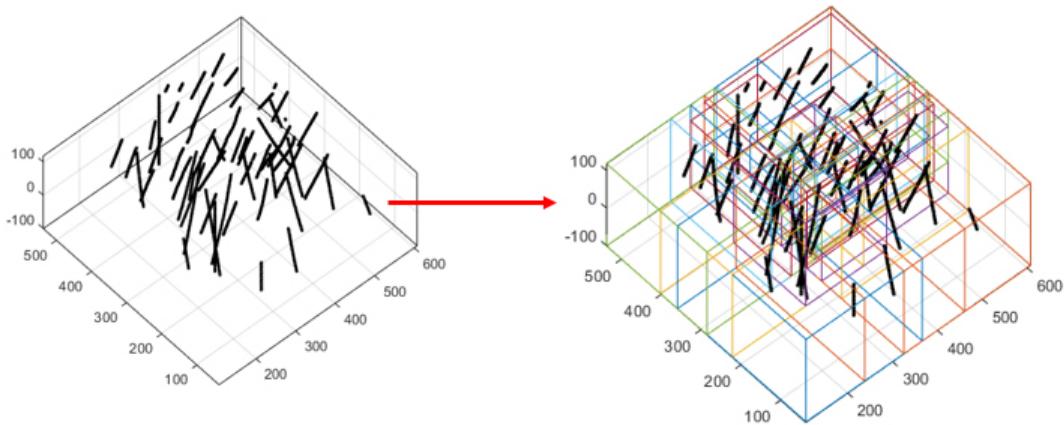
A demonstração das equações pode ser encontrada em Wendland (2004) e Martin e Boisvert (2015). Um domínio A é subdividido em uma série de partições sobrepostas β , como mostrado na Figura 14 à direita, de modo que a união de todas as k partições em A , $\{\beta\}_{j=1}^K$ compreende o domínio. Para cada partição, os dados correspondentes são utilizados para interpolar a função escalar $s_j(x)$.

A função peso para cada partição deve ser igual a 1 no centro e 0 na fronteira, a contribuição de cada partição nos locais de sobreposição é dado pela função peso.

As partições são definidas encontrando uma série de coordenadas para seus centros. Diversos métodos podem ser utilizados: um grid regular, k-means, árvore binária, *oct tree*.

A Figura 14 mostra um domínio particionado: o particionamento começa considerando o domínio completo como sendo uma partição, então a partição "total" é subdividida recursivamente em duas, mantendo um nível de sobreposição. O particionamento termina quando cada partição individual tem menos amostras que o máximo especificado pelo usuário (MARTIN; BOISVERT, 2017a; MARTIN; BOISVERT, 2017c). Dessa forma, áreas com alta densidade amostral ficam em partições pequenas; enquanto, áreas esparsas ocupam partições maiores.

Figura 14 – Particionamento do domínio.



Fonte: [Martin e Boisvert \(2017c\)](#)

Deve haver um número de amostras em cada partição e sobreposição das partições suficientes para evitar o surgimento de artefatos quando as partições independentes forem unidas.

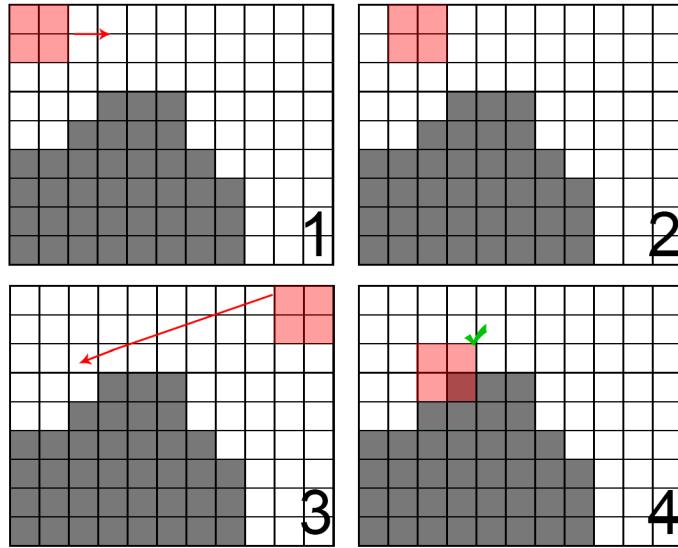
O método deve ser utilizado para um único domínio por vez. Na presença de múltiplos domínios uma regra hierárquica deve ser definida ou a categoria responsável pela menor distância estimada deve ser retida.

2.1.1.6.2 Refinamento dos contatos

[Silva e Deutsch \(2015\)](#) propuseram uma técnica baseada no algoritmo dos cubos marchantes ([LORENSEN; CLINE, 1987](#)) para acelerar o processo de modelagem em modelos multi-categóricos. O primeiro passo é criar o modelo geológico usando distâncias assinaladas em um *grid* grosso, como mostra o mapa a esquerda na [Figura 16](#).

Depois disso, o algoritmo dos cubos marchantes é aplicado ao modelo geológico com o objetivo de identificar a região ao redor dos contatos. Neste algoritmo, um cubo composto por 8 células em três dimensões ou 4 células em duas dimensões percorre o *grid*, conforme mostrado na [Figura 15](#), da esquerda para a direita e de cima para baixo visitando todo o *grid*. O algoritmo verifica se pelo menos um dos nós que constituem o cubo pertence a uma categoria diferente dos demais. Nesse caso, esses blocos são marcados como região de contato.

Figura 15 – Esquema mostrando o funcionamento do algoritmo cubos marchantes.

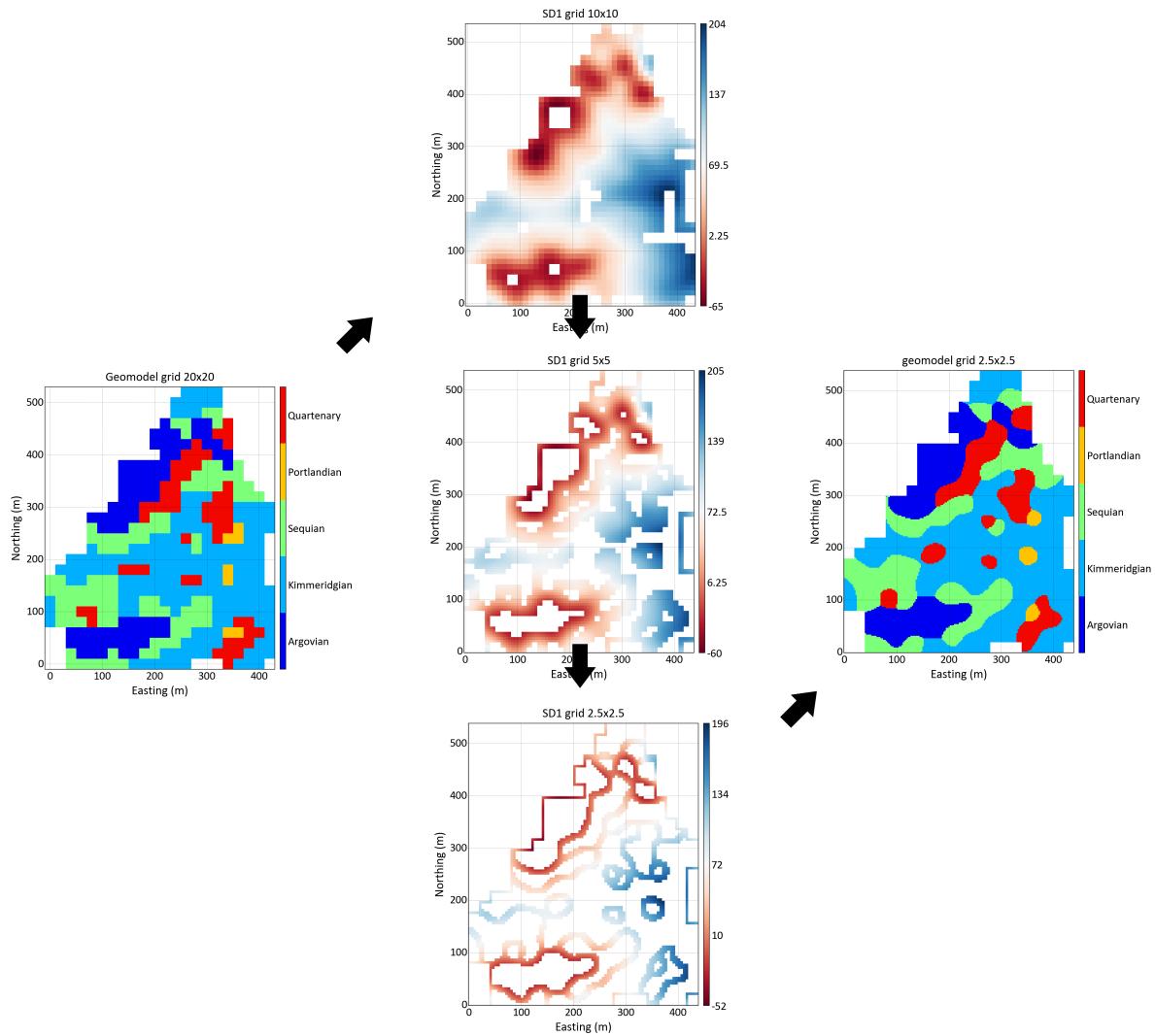


Por fim, o *grid* é reduzido (*downscaled*) e uma nova interpolação das distâncias e posterior classificação é realizada apenas nas regiões marcadas como contato. Os demais blocos são classificados com a mesma categoria que pertenciam no modelo no *grid* grosso. A posição real do contato necessariamente está localizada na região demarcada pelo algoritmo.

Os passos são repetidos de forma automática, um número de vezes definido pelo usuário, até que a resolução necessária seja atingida.

A [Figura 16](#) mostra, seguindo a direção das flechas, o modelo inicial para o *Swiss Jura* criado em um *grid* de dimensões 20x20 m. O algoritmo cubos marchantes então identifica a região dos contatos. Na primeira iteração, as dimensões do *grid* foram reduzidas pela metade, esse parâmetro é escolhido pelo usuário. As distâncias são estimadas no *grid* 10x10 m somente nos blocos marcados como contato pelos cubos marchantes, essa região fica mais estreita a cada iteração. Um novo modelo geológico refinado é criado nesse *grid*. O processo é repetido até que a dimensão dos blocos seja 2.5x2.5 m, como mostrado na imagem da direita na [Figura 16](#).

Figura 16 – Esquema mostrando o refinamento dos contatos em três iterações reduzindo as dimensões do *grid* pela metade para o Swiss Jura.



2.1.1.7 Incorporação de estacionariedade de segunda ordem

Corpos geológicos são complexos da escala macro à escala micro. Essa característica pode tornar difícil a captura de suas feições com funções de covariância. Segundo [Martin e Boisvert \(2017a\)](#), a não estacionariedade de segunda ordem deve ser incorporada quando estruturas complexas estão sendo modeladas.

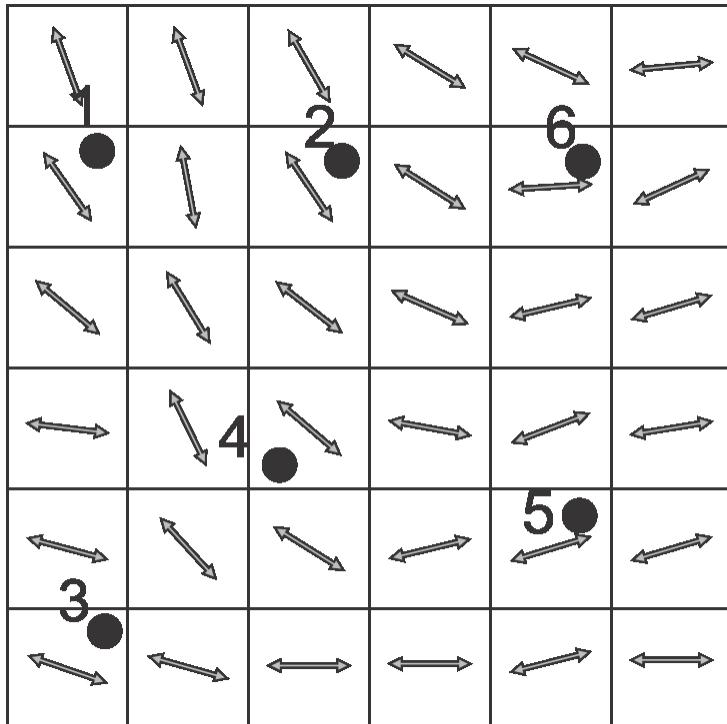
Anisotropia é definida como o conjunto de rotações e relações anisotrópicas: azimute, mergulho, *rake*, $r1 = \frac{a_{hmin}}{a_{hmax}}$, $r2 = \frac{a_{vert}}{a_{hmax}}$.

2.1.1.7.1 Krigagem com anisotropia local variável

A krigagem com anisotropia variável exige que os parâmetros locais de anisotropia sejam definidos em todos os nós do grid ([Figura 17](#)), e variem de forma suave pelo domínio. Isso permite que estruturas curvilíneas em escala menor que o espaçamento entre as amostras

sejam capturadas ([MARTIN; BOISVERT, 2017a](#)). Porém, torna o método computacionalmente exigente.

Figura 17 – Esquema mostrando os vetores de anisotropia local e as amostras representadas pelos círculos pretos para cada nó do grid.



Fonte: [Martin e Boisvert \(2017a\)](#)

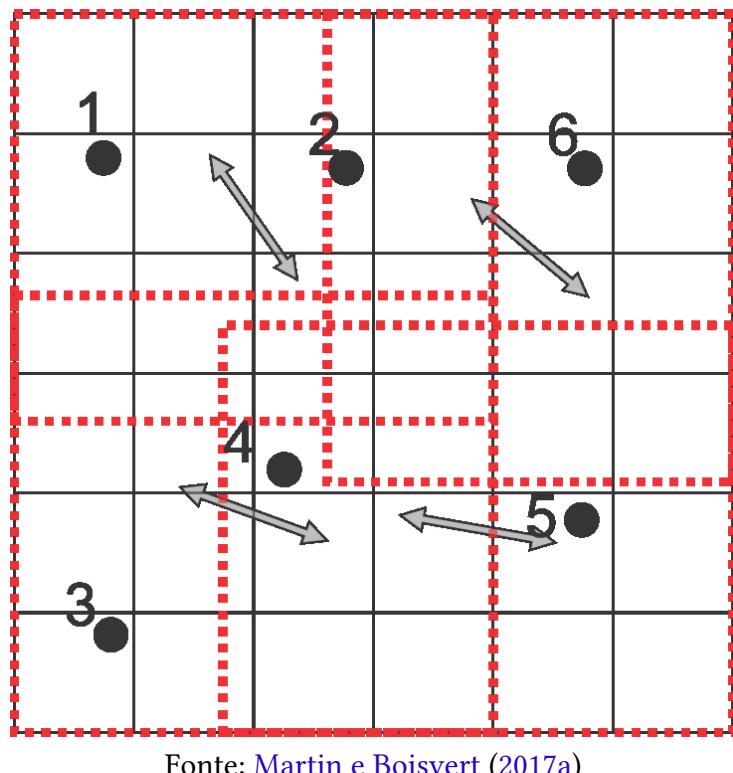
Existem diversos métodos para definição da anisotropia local ([LILLAH; BOISVERT, 2015](#)): pode ser inferida por um geomodelador, estimada diretamente a partir dos dados usando técnicas automáticas, interpolada de uma série de pontos onde a orientação foi medida ou pode ser extraída de dados em um grid que representam a variabilidade local (teores krigados, por exemplo). A biblioteca *GSLib* tem diversos softwares para definição de anisotropia local.

2.1.1.7.2 Funções de bases radiais com anisotropia local variável

Interpolação com funções de bases radiais com anisotropia local variável é baseada na decomposição de domínios ([subseção 2.1.1.6.1](#)), um kernel anisotrópico pode ser usado para melhor suportar os dados em cada partição, já que as partições são independentes, uma matriz de rotação é usada para definir *kernels* anisotrópicos locais que melhor se ajustam às propriedades espaciais de cada partição. O interpolador de bases radiais é obtido independentemente para cada partição e a solução final é obtida ponderando as soluções independentes nos locais de sobreposição ([MARTIN; BOISVERT, 2017a](#)).

Ao contrário da krigagem com anisotropia local variável, que requer os vetores de anisotropia em todas os nós do *grid*, a implementação para funções de bases radiais requer os vetores somente no centro de cada partição (Figura 17) variando suavemente pelo domínio.

Figura 18 – Esquema mostrando os vetores de anisotropia local e as amostras, representadas pelos círculos pretos, para cada centro de partição representadas pelas linhas tracejadas vermelhas.



Fonte: [Martin e Boisvert \(2017a\)](#)

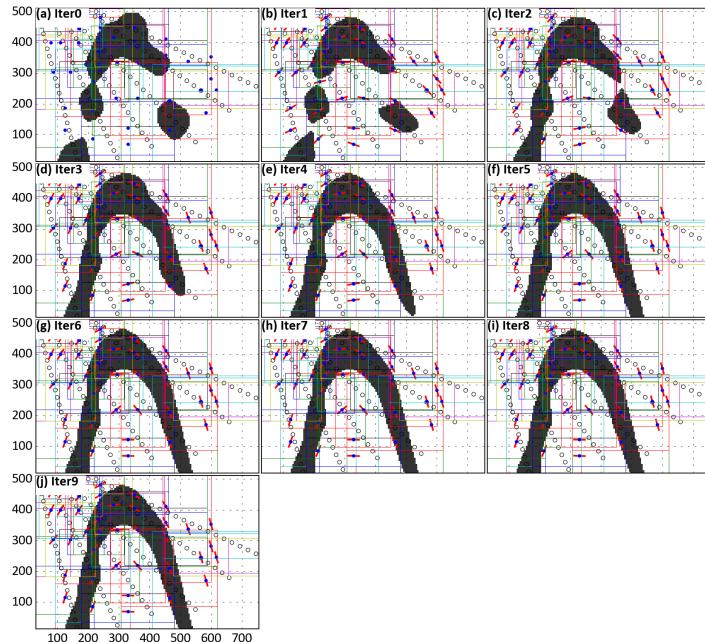
Extrair orientações locais de um modelo criado com anisotropia global e utilizar essas orientações em uma nova interpolação não estacionária resulta em um modelo geológico mais refinado considerando informação local. Esse processo pode ser repetido continuamente refinando o campo de anisotropias locais.

O refinamento iterativo (MARTIN; BOISVERT, 2017c) com funções de bases radiais tem um benefício extra (em relação à krigagem com variação local de anisotropia), a possibilidade de incluir algum tipo de critério de parada, quando o modelo for considerado refinado. Como as partições são independentes, se o refinamento de uma partição em particular não resultar em mudanças significativas, poderá ser omitido na solução final. Essa característica torna o processo significativamente mais rápido (MARTIN; BOISVERT, 2017a).

A Figura 19 mostra um processo de refinamento iterativo onde a semente é um modelo isotrópico. O algoritmo começa sua execução inferindo orientações anisotrópicas em cada flanco da dobra, que são rasos demais para que sejam definidos adequadamente. Porém, nas próximas iterações do refinamento, o mergulho se torna mais profundo e os flancos são definidos

implicitamente.

Figura 19 – Esquema mostrando o refinamento iterativo para funções de bases radiais.



Fonte: [Martin e Boisvert \(2017c\)](#)

2.1.1.8 Desvantagens

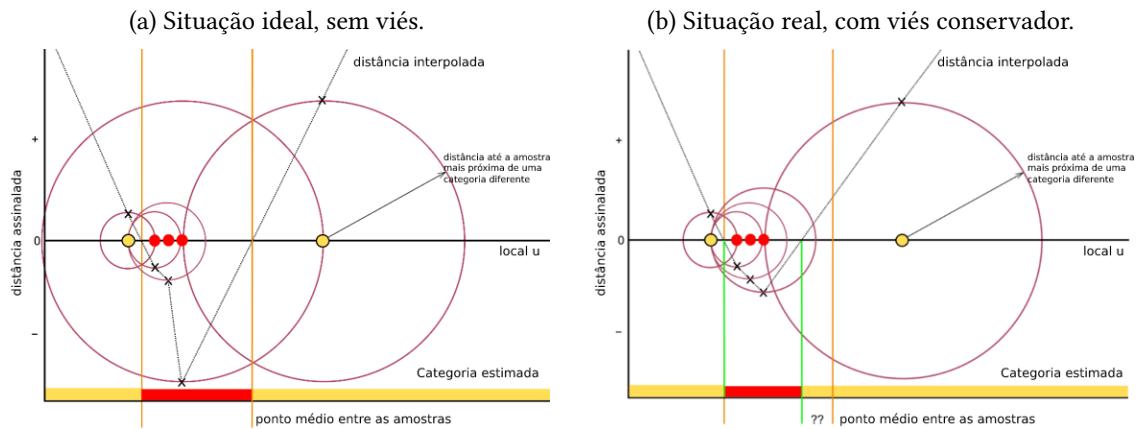
[Mancell e Deutsch \(2019\)](#) apontam que na presença de dados esparsos e amostras com espaçamento variável, a incerteza em relação à posição dos contatos de um domínio geológico é significativa. Conforme amostras são coletadas e inseridas no modelo, a incerteza em relação à localização diminui e o risco é reduzido. A assimetria espacial na amostragem e a dispersão das amostras são características inerentes às amostragens para avaliação de recursos e podem levar a situações únicas em que as distâncias assinaladas se mostram incapazes de gerar modelos apropriados.

O problema reside no fato do algoritmo para o cálculo da função distância assinalada buscar a amostra mais próxima codificada com um indicador oposto sem levar em consideração a estrutura e o espaçamento dos dados circundantes. Um cenário unidimensional simples ilustra esse viés devido à assimetria na [Figura 20](#).

A Figura [20a](#) mostra a situação ideal, onde a amostra amarela da esquerda aponta para a primeira amostra vermelha, da esquerda para a direita. A primeira amostra vermelha aponta para a amostra amarela da esquerda, assim como a segunda amostra vermelha. Já a terceira amostra vermelha aponta para a amostra amarela da direita, dessa forma, o melhor palpite para o limite (ou a isosuperfície zero) da direita entre os domínios amarelo e vermelho se localiza no ponto médio entre as amostras.

A Figura 20b mostra a situação que acontece de fato, todas as amostras vermelhas apontam para a amostra amarela da esquerda puxando o limite da direita entre os domínios para perto das amostras vermelhas, diminuindo o tamanho do domínio, ou seja, há um viés conservador no volume dos domínios na presença de forte assimetria nas amostras.

Figura 20 – Cenário unidimensional simples ilustrando o viés devido à assimetria nos modelos geológicos.

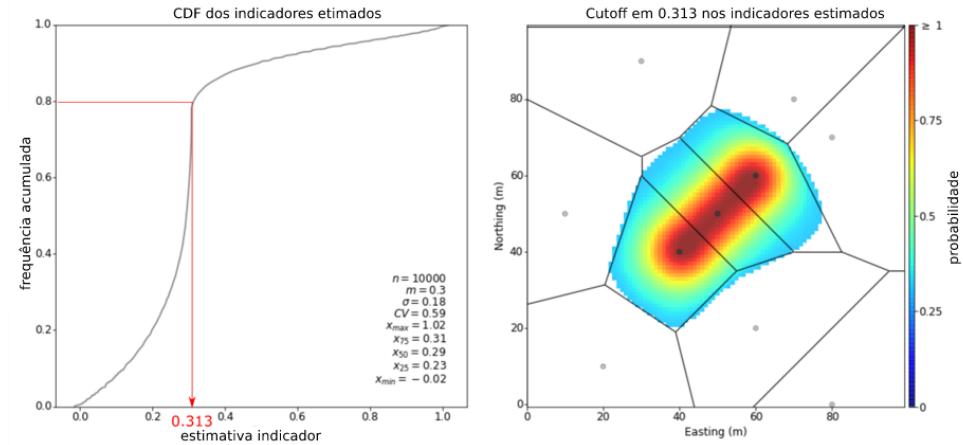


Modificado de [Mancell e Deutsch \(2019\)](#)

Os autores propõe uma solução usando indicadores, nesse caso, a isosuperfície definida pela probabilidade de ocorrência 0,5 não é o melhor palpite para a localização do contato. A solução consiste em:

- i Usar a triangulação de Voronoi para definir a área de influência de cada amostra de indicador;
- ii Calcular a proporção da área total que é ocupada por polígonos que representem amostras do domínio a ser modelado;
- iii Usar a proporção da área ($1-p$) para definir, a partir da distribuição acumulada da interpolação dos indicadores, o limiar no qual a isosuperfície que represente o contato será extraída. Como mostrado no exemplo da [Figura 21](#)

Figura 21 – Esquema mostrando a extração da iso-superfície usando indicadores e triangulação de Voronoi. À esquerda a figura mostra o CDF dos indicadores interpolados e em vermelho a proporção (1-0.2) ocupada pelos triângulos que representam amostras no interior do domínio que equivale a um valor de indicador de 0.313. À direita um *cutoff* feito em 0.313 nos indicadores interpolados para definir os limites do domínio.



Modificado de [Mancell e Deutsch \(2019\)](#)

2.2 Modelagem geológica estocástica

A necessidade de avaliação da incerteza associada ao modelo geológico impulsionou o desenvolvimento de métodos estocásticos baseados em múltiplas realizações equiprováveis de um atributo. Esses métodos podem ser divididos em dois grupos: métodos baseados em objetos e métodos baseados em células.

Os métodos baseados em objetos ([PYRCZ; DEUTSCH, 2014](#)) têm como objetivo a reprodução da morfologia do depósito, como canais meandrantes em um meio fluvial por exemplo. Geralmente os modelos não são condicionados às amostras. Métodos baseados em objetos ainda não têm muita aplicação na mineração encontrando uso na modelagem de depósitos de óleo e gás.

Os métodos baseados em células, em contrapartida, são condicionados pelas amostras e são utilizados amplamente pela indústria da mineração. Um dos principais métodos desse grupo é a simulação sequencial dos indicadores ([ALABERT, 1987](#)). O algoritmo é robusto e proporciona uma maneira simples e direta de incorporar a incerteza em modelos geológicos multi-categóricos. Entretanto, os modelos podem apresentar uma aparência ruidosa e não realista do ponto de vista geológico ([DEUTSCH, 2006](#)).

Na simulação sequencial dos indicadores, os nós do *grid* são visitados segundo um caminho aleatório. Uma distribuição condicional de probabilidades é construída em cada nó do *grid*, a partir das amostras e dos nós previamente simulados dentro de um raio de busca. Então, uma categoria é amostrada aleatoriamente da distribuição e atribuída àquele nó.

Um outro método que se baseia no mesmo princípio é a simulação geostatística multi-ponto (MPS). Esse método utiliza imagens de treinamento para construir a distribuição condicional de probabilidades em cada nó do *grid*. Cada realização deve reproduzir as características geológicas da imagem de treinamento. O método foi inicialmente proposto por [Guardiano e Srivastava \(1993\)](#), porém, diferentes variantes foram desenvolvidas ao longo do tempo ([STREBELLE, 2002](#); [BOUCHER, 2009](#)).

Há ainda a família dos métodos Gaussianos truncados. Esses métodos são aplicados, geralmente, quando as litologias ocorrem em uma ordem fixa, como em sequências estratigráficas por exemplo. A ideia básica por trás da simulação Gaussiana truncada ([MATHERON et al., 1987](#)) é simular uma variável Gaussiana em cada célula do *grid* e então aplicar uma regra de truncagem para converter os valores Gaussianos contínuos em valores categóricos.

Em muitos casos, a abordagem Gaussiana truncada se mostra bastante restritiva. Por esse motivo, a técnica pode ser estendida para duas ou mais variáveis Gaussianas. Essa adaptação é chamada de simulação pluri-Gaussiana truncada ([ARMSTRONG et al., 2011](#)).

O conceito de usar a regra de truncagem em mais de uma dimensão para transformar múltiplas variáveis Gaussianas em uma única variável categórica é limitada, já que a interpretação geológica se degrada quando um número maior de categorias e/ou variáveis gaussianas são consideradas devido à complexidade do modelo.

Por isso, uma nova adaptação foi desenvolvida: [Silva e Deutsch \(2019\)](#) propuseram a simulação pluri-Gaussiana truncada hierárquica. Esse método utiliza uma estrutura em árvore para definir diferentes regras de truncagem em uma única dimensão. Isso facilita a definição do mapeamento entre espaço contínuo (Gaussiano) e categórico, tornando possível a utilização eficiente de qualquer número de variáveis Gaussianas para a modelagem de uma variável categórica.

No braço dos métodos implícitos, alguns autores desenvolveram técnicas para avaliar a incerteza dos modelos geológicos. Avaliar a incerteza em relação à posição dos contatos a partir dos campos potenciais é uma processo direto. [Gonçalves, Kumaira e Guadagnin \(2017\)](#) deram uma abordagem de aprendizagem de máquinas para o método possibilitando a avaliação de incerteza. [Varga, Schaaf e Wellmann \(2019\)](#) desenvolveram uma ferramenta de modelagem geológica estocástica de código aberto que avalia a incerteza usando campos potenciais combinados com uma abordagem Bayesiana.

Outros autores construíram modelos implícitos e avaliaram a incerteza do modelo geológico usando dados estruturais. [Lindsay et al. \(2012\)](#) examinaram a incerteza introduzida pelos dados de orientação geológica, produzindo um conjunto de modelos 3D implícitos gerados a partir de medições de orientação sujeitas a simulações de incerteza. [Pakyuz-Charrier et al. \(2018\)](#) estudaram o efeito da propagação da incerteza dos dados estruturais de entrada na modelagem geológica 3D implícita. [Caumon, Tertois e Zhang \(2007\)](#) avaliaram a incerteza do

modelo geológico ao perturbar os parâmetros geológicos, como superfícies de falha e horizonte de alteração.

Também existem na literatura métodos de avaliação de incerteza baseados em distâncias assinaladas, as próximas seções apresentam os principais. Alguns desses, que são base para alguns dos métodos propostos no [Capítulo 3](#), serão abordados com mais profundidade.

2.2.1 Avaliação heurística da incerteza

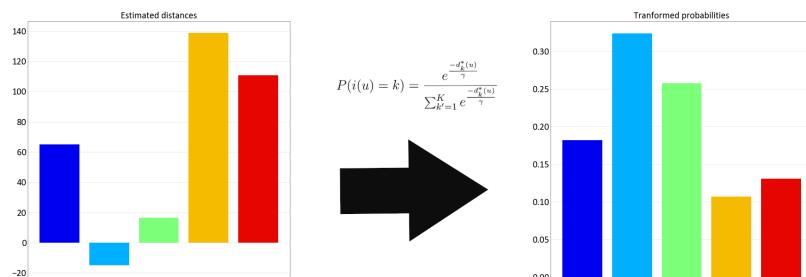
O método mais simples de avaliação de incerteza em modelos geológicos implícitos, proposto por [Silva e Deutsch \(2012\)](#), consiste em transformar distâncias estimadas em probabilidades via *softmax transformation*. Técnica amplamente usada em métodos de classificação para múltiplas classes ([MCCULLAGH; NELDER, 1989](#)). A ideia é transformar as distâncias estimadas em probabilidades a partir da [Equação 2.20](#). Os valores transformados encontram-se entre zero e um e sua soma deve ser igual a um, para cada bloco em que as distâncias assinaladas forem estimadas.

$$P(i(u) = k) = \frac{e^{\frac{-d_k^*(u)}{\omega}}}{\sum_{k'=1}^K e^{\frac{-d_{k'}^*(u)}{\omega}}} \quad (2.20)$$

$P(i(u) = k)$ representa a probabilidade de um local u pertencer à categoria k , $d_k^*(u)$ é a distância estimada para a categoria k e ω é um parâmetro que regula a inter-relação entre as probabilidades das K diferentes categorias. Quanto maior ω , maior as diferenças entre as probabilidades (maior a banda de incerteza). [Silva e Deutsch \(2013\)](#) advogam que o parâmetro pode ser o maior módulo entre as distâncias estimadas em um nó.

A [Figura 22](#) mostra, à esquerda, as distâncias estimadas em uma determinada célula do *grid* criado para a modelagem do *Swiss Jura* e a direita as distâncias transformadas em probabilidades usando a [Equação 2.20](#). O parâmetro ω é o maior módulo entre as distâncias estimadas naquele bloco.

Figura 22 – Distâncias estimadas e transformadas em probabilidades para um mesmo bloco do *Swiss Jura*

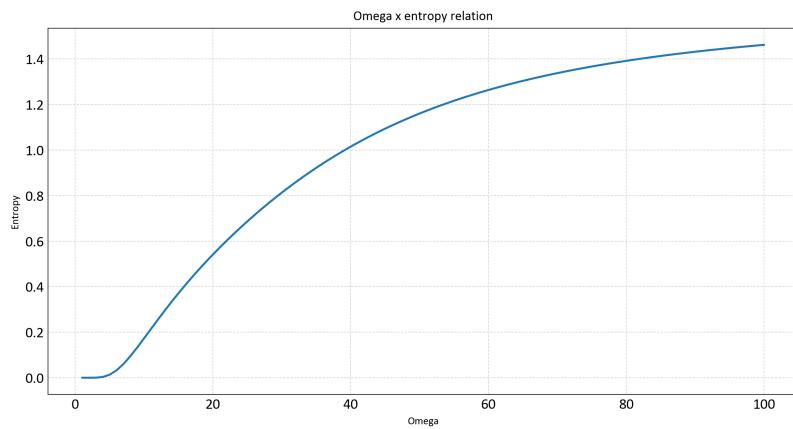


A [Figura 23](#) mostra a relação entre ω e a entropia de Shannon ([SHANNON, 1948](#))

calculada pela [Equação 2.21](#) para o bloco da [Figura 22](#). $P(X_k)$ é a probabilidade para a categoria k . Para pequenos valores de ω , uma das categorias será transformada em uma probabilidade muito alta, enquanto as outras serão transformadas em probabilidades muito baixas, nesse caso a entropia é mínima. À medida que o parâmetro ω aumenta, a diferença entre as probabilidades torna-se menos pronunciada até que a probabilidade para todas as categorias seja igual, caso em que a entropia é maximizada.

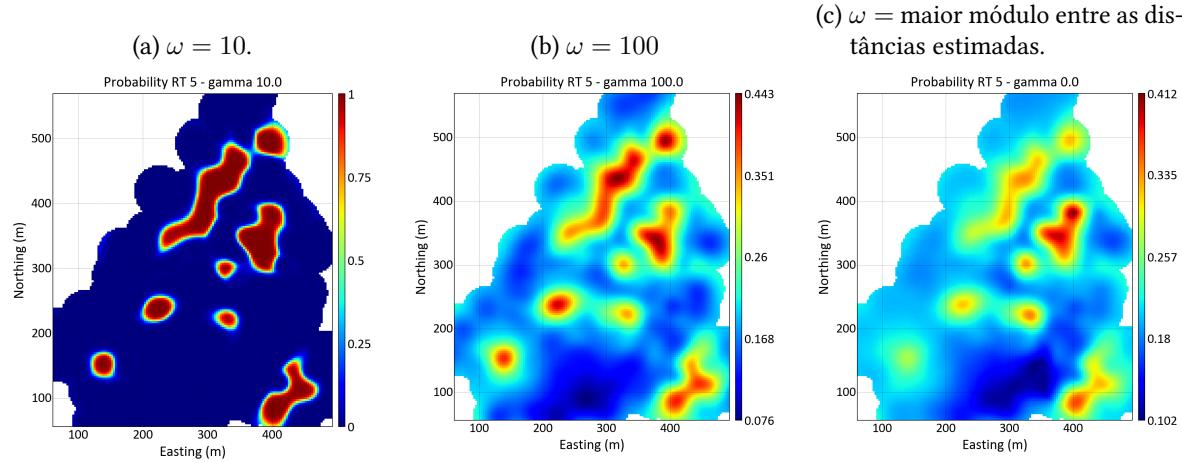
$$H(X) = - \sum_{k=1}^K P(X_k) \log(P(X_k)) \quad (2.21)$$

Figura 23 – Relação entre o parâmetro ω e a entropia de Shannon.



A [Figura 24](#) mostra, à esquerda, as probabilidades para a categoria Quaternary obtidas transformando as distâncias assinaladas com o valor de ω igual a 10. Para esse valor do parâmetro há muitas regiões com probabilidade zero ou um de ocorrência e algumas regiões de transição próximas aos contatos. Ao centro, as probabilidades calculadas a partir das distâncias para ω igual a 100, nesse caso, há uma menor diferença entre as probabilidades. Por fim, à direita, as probabilidades foram calculadas usando como parâmetro ω o maior módulo entre as distâncias estimadas.

Figura 24 – Distâncias transformadas em probabilidades para a categoria Quaternary usando valores diferentes para ω .



Apesar do método ser simples e direto, suportar múltiplas categorias simultaneamente, seu resultado não pode ser utilizado em etapas subsequentes do processo de avaliação de recursos, por não gerar múltiplas realizações do modelo geológico.

2.2.2 *Boundsim*

Uma outra metodologia proposta por [McLennan \(2007\)](#), [McLennan e Deutsch \(2006\)](#) consiste em realizar um *bootstrap* espacial ([DEUTSCH, 2004](#)) da média da distância assinalada calculada para uma dada categoria. O procedimento de *bootstrap* é amplamente utilizado para quantificar incerteza em parâmetros estatísticos. As duas mais importantes suposições são: os dados são representativos e os dados são independentes. Por esse motivo, a técnica de amostragem deve ser modificada para variáveis que apresentam continuidade espacial.

Diferentes valores para a média são tomados do histograma da média da distância assinalada, gerado pelo *bootstrap* espacial. Esses valores podem ser correspondentes ao p10, p50 e p90 por exemplo. Então, as distâncias assinaladas são interpolados para os nós do *grid* por krigagem simples, que tem como hipótese a estacionariedade da média por todo o domínio. A média estacionária para a krigagem simples são os valores tomados do histograma.

O último passo consiste em extrair a isosuperfície zero dos modelos implícitos gerados por krigagem simples com os diferentes valores da média estacionária tomadas do histograma da média. Em tese, quanto maior o percentil menor o risco associado, o volume do sólido modelado expande ou contrai de acordo com o valor escolhido para a média estacionária.

A metodologia é simples e direta, porém, deve ser aplicada à uma categoria por vez, na presença de múltiplos domínios alguma abordagem hierárquica deve ser estabelecida. A sensibilidade da krigagem simples à média depende da configuração espacial das amostras, muitas vezes a diferença entre os volumes obtidos é irrelevante.

O produto final do método pode ser utilizado nas etapas subsequentes do processo de avaliação, todavia, o método não avalia incerteza de forma adequada. Os autores do estudo original aplicaram a metodologia apenas em bancos de dados sintéticos e de geometria regular.

2.2.3 Condicionamento das amostras

[McLennan \(2007\)](#) também propôs uma metodologia baseada na parametrização das amostras e posterior estimativas. Os pesos de krigagem e da interpolação pelo inverso da distância dependem do arranjo geométrico das amostras mas não dos valores das amostras. A proposta é alterar os pesos dados às amostras a partir de um fator de condicionamento $f_i^{DC}(u)$, como apresentado na [Equação 2.22](#)

$$d^*(u) = \sum_{i=1}^n f_i^{DC}(u) \lambda_i(u) d(u_i) \quad (2.22)$$

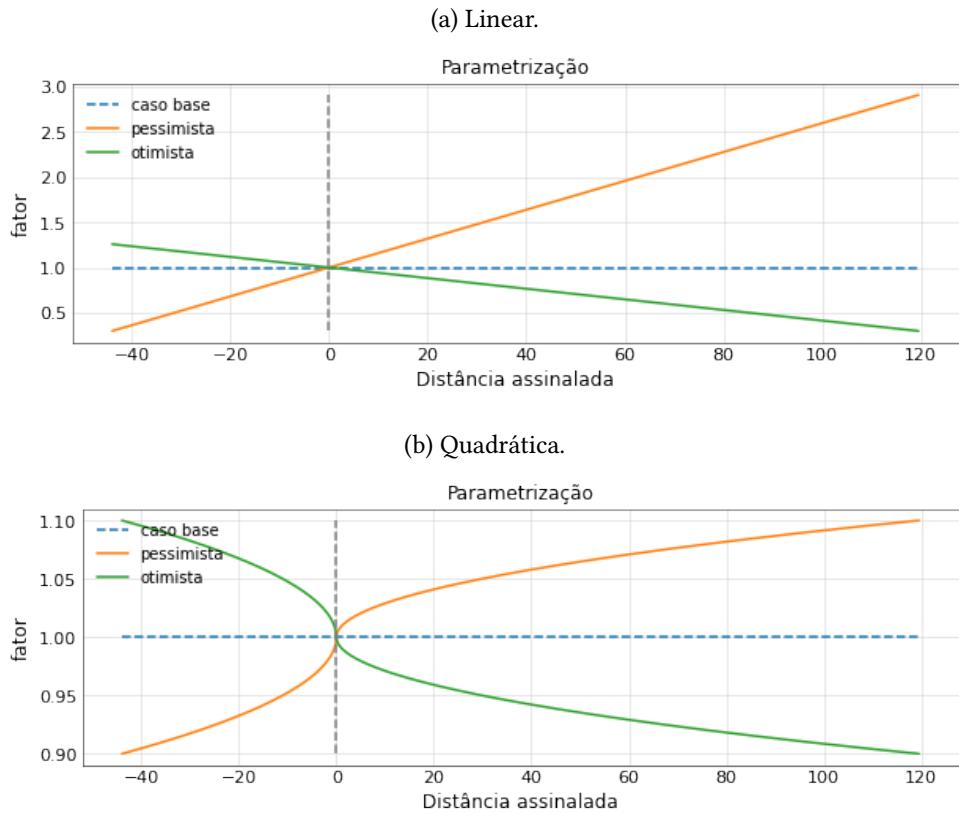
Onde $d^*(u)$ é a distância interpolada no local não amostrado u , $d(u_i)$ são as distâncias assinaladas calculadas para as amostras, $\lambda_i(u)$ são os pesos dados pela krigagem ou inverso da distância e $f_i^{DC}(u)$ são os fatores de condicionamento que dependem do valor numérico das amostras. A krigagem ou inverso da distância clássicos apresentam um fator de condicionamento igual a um para todas as amostras.

Limites pessimistas ou otimistas podem ser gerados a partir das parametrizações mostradas na [Figura 25](#). A reta de inclinação negativa da Figura [25a](#) dá cada vez mais peso para amostras de distância mais negativas, e cada vez menos peso para amostras mais positivas produzindo uma isosuperfície zero otimista. O domínio tem um volume maior em relação ao caso base onde $f_i^{DC}(u)$ é igual a um para todas as amostras independente do seu valor. O contrário acontece para a reta com inclinação positiva.

A magnitude da diferença entre os limites otimistas e pessimistas é dada pelo parâmetro fator mínimo, ou f_{min} , escolhido pelo usuário. Um f mínimo igual a um, gera um fator igual a um para todas as amostras (estimativa clássica). Quanto maior o fator, maior a inclinação da reta, e consequentemente, maior a diferença entre os cenários. O valor de f mínimo pode ser diferente para os cenários otimista e pessimista.

A magnitude da diferença pode ser acentuada usando uma função quadrática ([Figura 25b](#)) ao invés de uma linear.

Figura 25 – Parametrização das amostras.



McLennan (2007) sugere o uso da interpolação pelo inverso da distância ao invés da krigagem porque os pesos negativos, que podem ser produzidos pela krigagem, interferem com o fator de parametrização.

2.2.4 Simulação direta das distâncias assinaladas

Cáceres et al. (2011) e Radtke (2017) propuseram que as distâncias assinaladas sejam simuladas de forma direta: a partir de um modelo geológico implícito multi categórico, como o mostrado na Figura 13, um coeficiente U de incerteza deve ser calculado de acordo com Equação 2.23:

$$U(u) = \frac{\max\{D_{min}\} - \min\{d_k^*(u)\}_{k=1}^K}{\max\{D_{min}\} - \min\{D_{min}\}} \quad (2.23)$$

Onde:

$$D_{min} = \{\min\{d_k^*(u_1)\}, \dots, \{\min\{d_k^*(u_n)\}\}_{k=1}^K\} \quad (2.24)$$

E $d_k^*(u)$ é a distância estimada no local u para a categoria k.

Os valores de U variam entre 0 e 1, um *cutoff* em U determina a extensão da zona de incerteza ao redor dos contatos: quanto mais próximo de 1, mais estreita a zona. As distâncias assinaladas devem ser simuladas no interior da zona de incerteza e a categoria correspondente ao menor valor simulado é retida em cada bloco para cada realização. Os blocos simulados são concatenados com os blocos estáticos, atribuídos pela interpolação.

O método é baseado em múltiplas realizações do modelo geológico, então, seu produto final pode ser usado nas etapas posteriores do processo de avaliação. Entretanto, a metodologia tem muitos passos ([RADTKE, 2017](#))

- i Cálculo das distâncias assinaladas para todas as amostras e categorias;
- ii Variografia das distâncias no espaço original para todas as categorias;
- iii Interpolação das distâncias;
- iv criação do modelo com base na menor distância interpolada;
- v Criação da zona de incerteza;
- vi Transformação Gaussiana das distâncias;
- vii Variografia das distâncias no espaço gaussiano para todas as categorias;
- viii Geração de múltiplos modelos baseados na menor distância simulada;
- ix Validação e pós processamento das realizações.

Além disso, a simulação é muito sensível aos parâmetros, o que muitas vezes, gera modelos com excessivo ruído, que não são geologicamente realistas, e não honram as amostras. O resultado final não justifica o número demasiado de passos e parâmetros. O coeficiente U não controla a magnitude da incerteza, apenas diminui o tempo das simulações, a magnitude da incerteza é controlada pelos parâmetros da simulação.

2.2.5 Avaliação de incerteza usando imagens de treinamento para simulação multi-ponto geradas a partir dos dados

[Silva \(2015\)](#) propôs uma metodologia que integra modelagem geológica implícita com funções distâncias assinaladas e simulação geostatística multi ponto. A geostatística multi-ponto utiliza imagens de treinamento para extrair e reproduzir estruturas geológicas complexas. Geralmente, imagens de treinamento são baseadas em modelos conceituais do fenômeno geológico. Em seu trabalho, [Silva \(2015\)](#) cria imagens de treinamento a partir de modelos geológicos implícitos determinísticos criados com funções distância assinaladas e modelos estocásticos criados por simulação sequencial dos indicadores. No trabalho, é introduzida uma

metodologia para integrar múltiplas imagens de treinamento, uma metodologia para calibrar a contribuição de cada imagem de treinamento e uma medida de entropia multi-ponto ao longo dos furos de sondagem.

[Silva \(2015\)](#) concluiu que imagens de treinamento geradas a partir das amostras produzem modelos geológicos estocásticos, por MPS, melhores.

2.2.6 Modelagem estocástica de estruturas curvilineares

[Lillah e Boisvert \(2013\)](#) propuseram a modelagem de estruturas curvilineares e a avaliação de incerteza desses modelos usando funções distância assinaladas. A metodologia tradicional é modificada para considerar não estacionariedade na forma de anisotropia variável localmente (LVA), que é incorporada na krigagem e na simulação sequencial Gaussiana das distâncias assinaladas. O campo de anisotropia variável captura a direção local, na qual a unidade geológica é mais contínua; enquanto, a simulação das distâncias e posterior extração da isosuperfície avalia a incerteza volumétrica.

2.2.7 Avaliando e visualizando a incerteza com movimento estocástico

[Yang et al. \(2019\)](#) propuseram uma técnica para avaliar e visualizar a incerteza em modelos geológicos multi-categóricos a partir do movimento estocástico. As unidades litológicas são representadas como uma soma de um modelo implícito estocástico conceitual com uma função residual condicionada às amostras e às relações cronológicas das unidades litológicas. Duas técnicas de amostragem para criação do movimento estocástico são propostas. Uma delas baseada na simulação de Monte Carlo e a outra em cadeias de Markov.

A incerteza é visualizada a partir de um filme no qual a transição entre as realizações é suave.

2.2.8 Simulação dos contatos

A metodologia proposta por [Wilde e Deutsch \(2012\)](#) consiste em gerar realizações dos contatos geológicos comparando distâncias modificadas interpoladas com simulações não condicionais.

É necessário definir, antes de qualquer coisa, uma zona de incerteza em torno dos contatos. Blocos localizados fora da zona são considerados como pertencentes a um determinado indicador, enquanto blocos dentro da zona terão sua incerteza avaliada. A zona de incerteza é obtida a partir de um parâmetro C que controla sua espessura.

A escolha do parâmetro C depende do nível de precisão requerido versus o tempo necessário para calibrar o parâmetro. Assim, o parâmetro pode ser determinado de diferentes maneiras: a determinação empírica é baseada em valores pré-determinados e/ou no conhecimento.

mento do fenômeno considerando a geometria da amostragem e a experiência do geomodelador (KARPEKOV; DEUTSCH, 2016). A calibração parcial compara modelos interpolados com um modelo de referência. O valor de C é determinado a partir da interpretação de seções geológicas representativas e da opinião de especialistas. A calibração total é demorada e trabalhosa: são necessários vários modelos de referência para determinar o parâmetro C (KARPEKOV; DEUTSCH, 2016). Para cada modelo de referência, um algoritmo de otimização é usado para calcular o parâmetro C, e para uma única iteração é necessário interpolar todos os modelos de referência com o mesmo parâmetro C. Para alcançar a calibração completa, várias iterações são necessárias usando valores diferentes para o parâmetro C, que demanda um grande esforço computacional e aumenta o tamanho e o número dos modelos de referência usados (MUNROE; DEUTSCH, 2008).

A calibração também pode ser realizada por um método semelhante à metodologia de validação cruzada por *jackknife* (WILDE; DEUTSCH, 2012), utilizando apenas os dados para calibrar o parâmetro C com baixa demanda computacional e alta automatização.

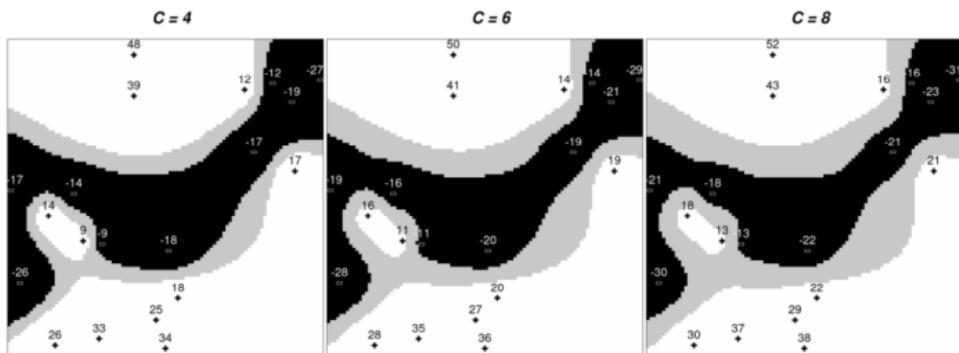
O primeiro passo na calibração de C por *jackknife* é remover um subconjunto dos dados. Isso pode ser feito escolhendo aleatoriamente os furos de sondagem a serem excluídos. Um estudo de calibração adequado envolve a geração de vários conjuntos, com diferentes proporções (50% a 10%) e diferentes configurações de dados removidos (MANCHUCK; DEUTSCH, 2015). Os valores da função distância são então calculados para os dados restantes com um valor inicial para C igual a zero. Essas amostras da função distância assinaladas são usadas para condicionar a estimativa da função distância em cada um dos locais de dados removidos. Existem quatro resultados possíveis desta estimativa. A localização pode ser: corretamente estimada como fora do domínio, corretamente estimada como dentro do domínio, incorretamente estimada como fora do domínio e incorretamente estimada como dentro do domínio. O principal interesse está no número de dados que caem do lado errado da fronteira, ou seja, o número de vezes que a estimativa é positiva, mas os dados são codificados como dentro do domínio e o número de vezes que a estimativa é negativa, mas os dados são codificados como fora do domínio. O número de vezes que um dado cai do lado errado do limite para C igual a zero é o caso base.

O valor de C é então aumentado e os valores da função distância das amostras não removidas são modificados de acordo com a Equação 2.25. Essa função distância modificada é estimada nos locais das amostras removidas. O limite agora é considerado como estando entre -C e C. Os dados caem do lado errado do limite quando um local estimado é codificado como interno, mas tem uma estimativa para a função distância modificada maior que C ou um local estimado é codificado como externo, mas tem uma estimativa da função distância modificada menor que -C. O número de dados que caem do lado errado do limite diminui à medida que C aumenta. C é aumentado até que o número de dados que caem no lado errado do limite seja aceitável. O valor C onde isso ocorre é o valor C calibrado que quantifica a incerteza do limite.

$$d_k(u_\alpha) = \begin{cases} -\| u_\alpha - u_\beta \| -C, & \text{se } u_\alpha \text{ pertence ao domínio} \\ +\| u_\alpha - u_\beta \| +C, & \text{se } u_\alpha \text{ não pertence ao domínio} \end{cases} \quad (2.25)$$

A Figura 26 mostra, para um esquema sintético em duas dimensões, em branco a zona considerada fora domínio, em preto dentro do domínio e em cinza a zona de contato para diferentes valores de C. À medida que C aumenta, o número de classificações errôneas diminui já que qualquer ponto estimado na zona cinza não é uma classificação errônea. Os sinais (+ e -) representam as amostras e sua distância assinalada calculada.

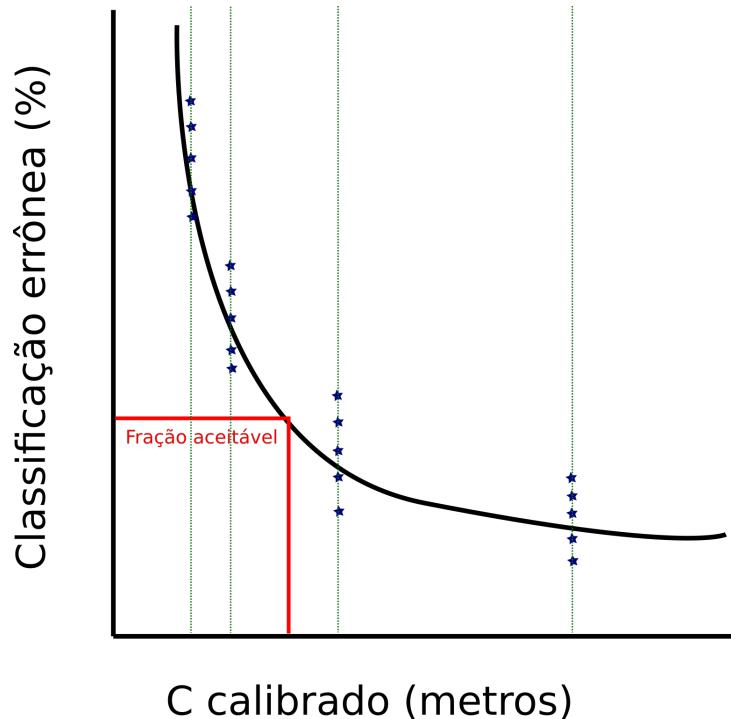
Figura 26 – Calibração do parâmetro C para um esquema em duas dimensões.



Fonte: [Wilde e Deutsch \(2012\)](#)

Uma curva como a da Figura 27 é construída. São escolhidos N valores para C entre zero e um valor definido pelo usuário (linhas tracejadas em verde). Para cada um desses valores de C, são removidos X% dos furos (ou amostras em duas dimensões) Z vezes (estrelas azuis). É indicado que a classificação errônea seja calculada diversas vezes para que a calibração de C seja robusta ([WILDE; DEUTSCH, 2012](#)). O índice de classificação errônea é registrado no eixo Y do gráfico (estrelas azuis) e o valor de C no eixo X (linhas tracejadas em verde). Então traça-se a classificação errônea média (curva em preto).

Figura 27 – Figura esquemática da curva de calibração.



[Manchuck e Deutsch \(2015\)](#) sugerem 2,5% de classificação errônea um nível aceitável para depósitos tabulares, ([MARTIN; BOISVERT, 2017a](#)) definiram como 3.9% para um dos domínios de um depósito de cobre. Uma outra forma de escolha é o método do cotovelo, o valor tomado deve ser o ponto de inflexão da curva de calibração ([MARTIN; BOISVERT, 2017a](#)). C depende do tipo de depósito e da configuração geométrica das amostras. Para depósitos com menor variabilidade, como os de bauxita por exemplo, C deve ser pequeno. Já para depósitos com maior variação, como os de ouro, o valor de C deve ser maior. Para amostragens densas o valor de C deve ser menor em relação à amostragens esparsas. C não deve ser maior do que a malha de sondagem.

Após o valor de C ser selecionado, as distâncias assinaladas calculadas para cada amostra devem ser modificadas a partir da equação [Equação 2.25](#). As distâncias modificadas devem então ser interpoladas, utilizando o mesmo variograma das distâncias originais, para todos os nós do grid, e a truncagem das distâncias modificadas estimadas entre -C e +C define a zona de incerteza.

Valores para a função distância são simulados uniformemente entre -C e +C. Para isso, é necessário realizar uma simulação Gaussiana não condicional. Um dos algoritmos comumente usados é a simulação sequencial Gaussiana ([algoritmo 1](#)). O variograma utilizado na simulação não condicional pode ser o mesmo das distâncias assinaladas. O alcance do variograma determina a natureza do contato entre os domínios, menores alcances geram contatos mais rugosos enquanto maiores alcances, contatos suaves. O efeito pepita controla a inter conectividade entre o domínio.

Algoritmo 1: Simulação sequencial Gaussiana não condicional

```

1. Defina um caminho aleatório através dos nós do grid;
for cada nó do
    1. Busque no espaço por nós previamente simulados;
    if número de nós previamente simulados != 0 then
        1. Construa e resolva o sistema de krigagem;
        2. Amostre a distribuição condicional construída a partir da média e da
           variância;
    else
        1. Amostre a partir de uma distribuição Gaussiana padrão;
    end
    2. Atribua o valor amostrado ao nó do modelo e inclua o nó simulado no conjunto
       de dados simulado anteriormente;
end

```

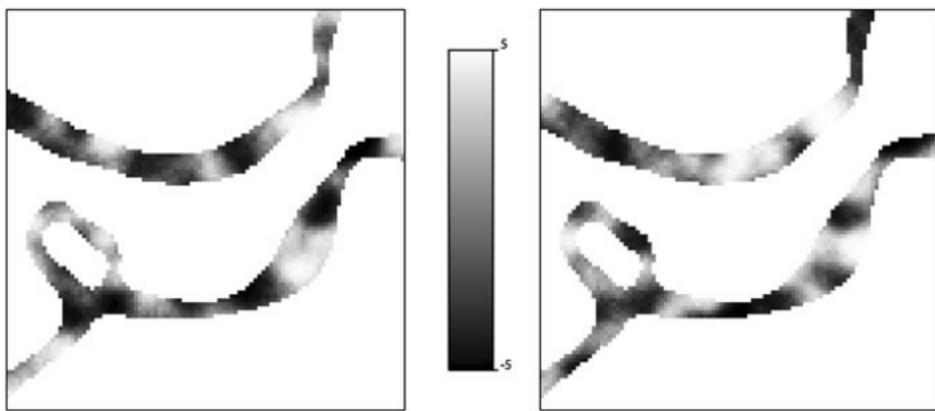
Para que os valores simulados se distribuam de forma uniforme entre $-C$ e $+C$, o valor Gaussiano, deve ser transformado pela relação:

$$df'(u) = 2 * C * G^{-1}(y'(u)) - C \quad (2.26)$$

Onde: $df'(u)$ é o valor da função distância simulada, $y'(u)$ o valor normal padrão da simulação não condicional, e G^{-1} representa a determinação do valor da distribuição acumulada padrão normal correspondente a $y'(u)$. Para garantir que os valores pertençam a região estabelecida os valores são multiplicados por $2C$ e subtraídos de C .

A [Figura 28](#) mostra duas realizações da simulação das distâncias para uma zona de incerteza igual a 5 metros para o esquema em duas dimensões da [Figura 26](#).

Figura 28 – Simulação das distâncias dentro da zona de incerteza para um esquema em duas dimensões.



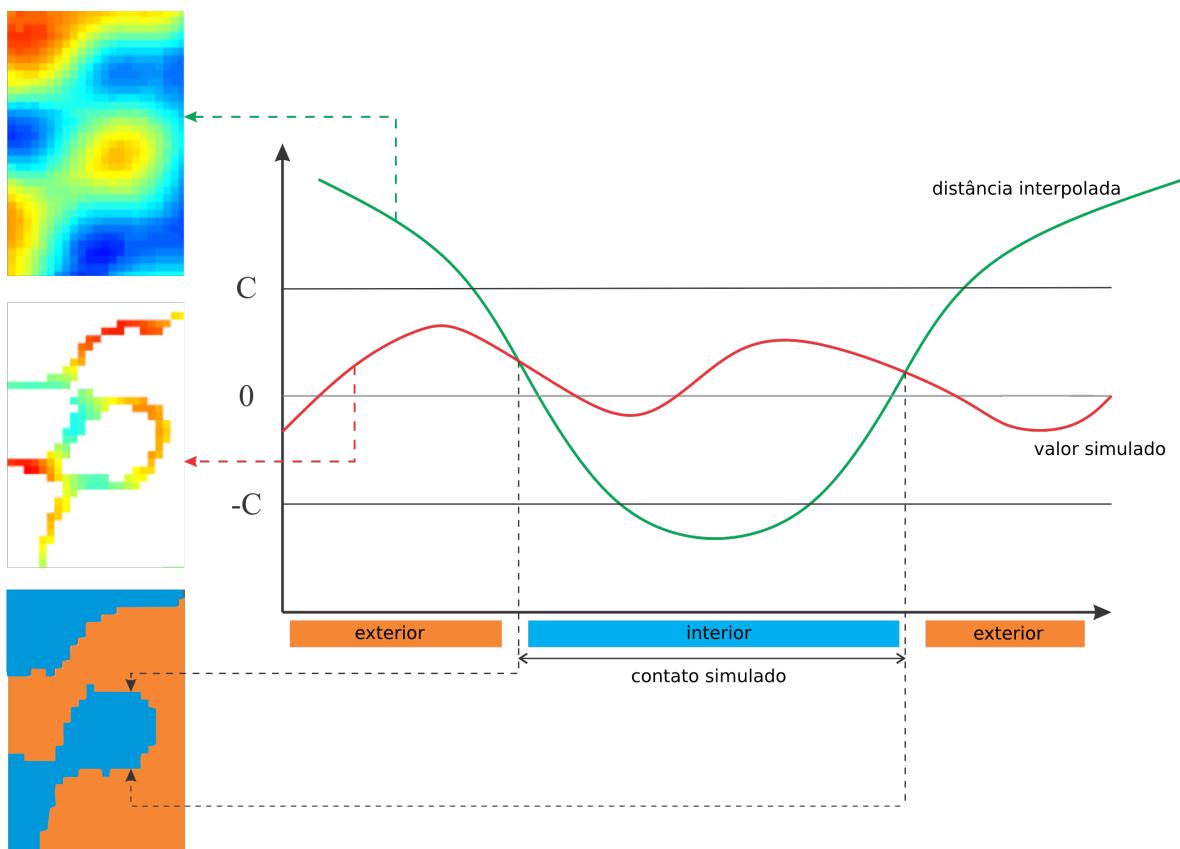
Fonte: [Wilde e Deutsch \(2012\)](#)

Os contatos podem agora ser simulados no interior da zona de incerteza. Caso o valor

interpolado $df^*(u)$ seja menor que o valor simulado $df'(u)$ para a função distância, o local é considerado pertencente ao domínio (indicador 1), caso o valor interpolado seja maior que o simulado, o local é considerado externo ao domínio (indicador 0), como evidenciado pela Equação 2.27. Nos locais onde a interpolação apresenta o mesmo valor da simulação, são estabelecidos os contatos dos domínios como esquematizado na Figura 29.

$$i^*(u) = \begin{cases} \text{indicador 1, se } df'(u) > df^*(u) \\ \text{indicador 0, se } df'(u) < df^*(u) \end{cases} \quad (2.27)$$

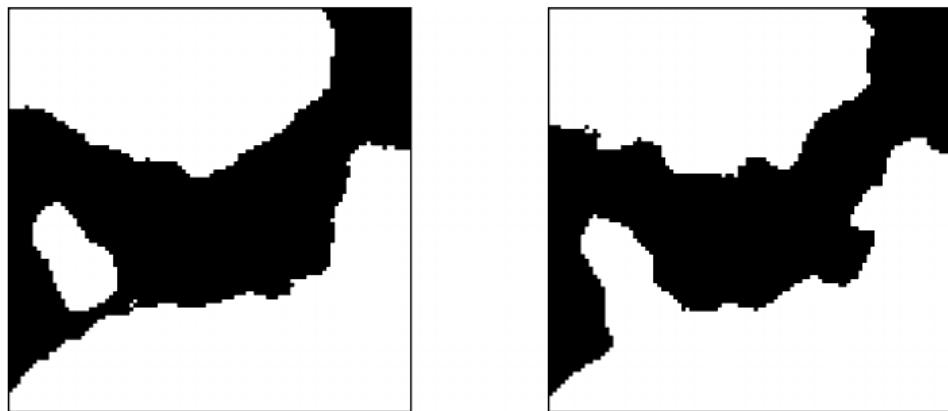
Figura 29 – Esquema mostrando como os contatos são simulados dentro da zona de incerteza.



Modificado de [Amarante, Rolo e Costa \(2021\)](#)

A Figura 30 mostra duas diferentes realizações da simulação dos contatos para o esquema em duas dimensões mostrado na Figura 26.

Figura 30 – Duas realizações da simulação dos contatos para um esquema em duas dimensões.



Fonte: [Wilde e Deutsch \(2012\)](#)

O método é rápido, já que é baseado em simulações não condicionais, gera modelos realistas sem ruído e com fronteiras contínuas. Tanto a magnitude da incerteza, quanto a natureza da fronteira podem ser controlados. Entretanto, o método funciona apenas para uma categoria por vez e mesmo que simplificações tenham sido desenvolvidas, a calibração do parâmetro de incerteza ainda pode ser laboriosa e subjetiva.

3 METODOLOGIAS PROPOSTAS

Nesse capítulo, as três metodologias para avaliação de incerteza em modelos geológicos multi-categóricos usando funções distâncias assinaladas propostas nessa tese são apresentadas. Todos os métodos são acompanhados de uma prova de conceito no banco de dados *Swiss Jura* e um estudo de caso.

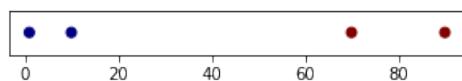
3.1 Parametrização do suporte do *kernel*

Esse método é uma adaptação do condicionamento das amostras apresentado na [subseção 2.2.3](#), para que tenha sua aplicação estendida para múltiplos domínios. No método proposto por [McLennan \(2007\)](#), o interpolador é o inverso da distância já que outros interpoladores, como a krigagem, podem gerar pesos negativos que interferem na ponderação gerada pelo condicionamento das amostras.

Essa adaptação propõe, no lugar de modificar o peso dado às amostras, modificar o suporte dos *kernels* usados na interpolação por funções de bases radiais multiplicando-os por um fator menor ou maior que um. Para cada propriedade de função distância assinalada que representa cada categoria do depósito mineral, os cenários otimista, pessimista e caso base são gerados e combinados para criar múltiplos cenários para o modelo geológico multi-categórico.

Considere o exemplo em uma dimensão da [Figura 31](#). A figura mostra quatro amostras, duas azuis pertencem ao domínio sendo modelado, e duas vermelhas que não pertencem ao domínio sendo modelado. As amostras estão, da esquerda para a direita, nas posições 1, 10, 70 e 90.

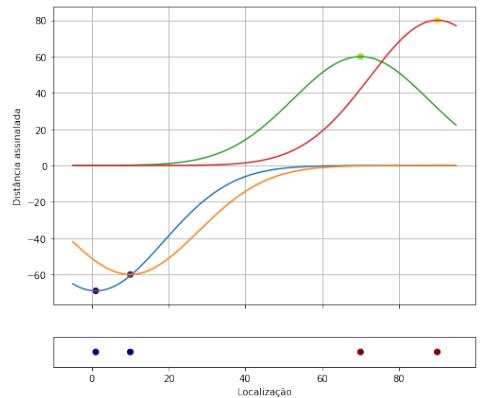
Figura 31 – Amostras em posição 1, 10, 70 e 90. Amostras em azul não pertencem ao domínio enquanto amostras vermelhas pertencem.



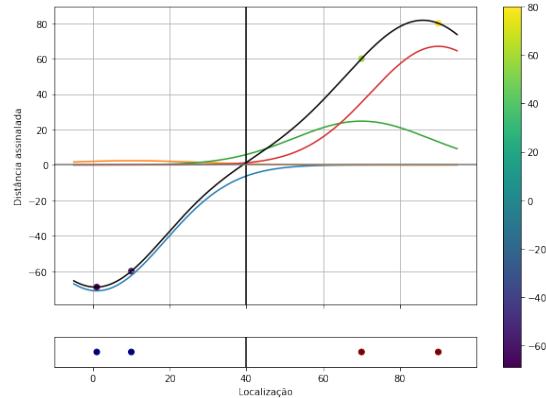
A [Figura 32](#) mostra, no gráfico na posição inferior, as amostras da [Figura 31](#) e no eixo y do gráfico, na posição superior, o valor da função distância assinalada calculada para cada uma das quatro amostras. A [Figura 32a](#) mostra uma função RBF com suporte igual a 25 centrada em cada uma das amostras. A [Figura 32b](#) mostra, em preto, a superfície interpolada a partir da combinação linear das quatro RBFs após o cálculo dos pesos a partir do sistema de equações apresentado na [Equação 2.14](#). O limite entre os domínios é onde a distância assinalada interpolada é igual a zero.

Figura 32 – Amostras em uma dimensão no gráfico na posição inferior e distâncias assinaladas calculadas para cada uma das amostras no eixo y do gráfico na posição superior. As curvas coloridas representam *kernels* associados a cada amostra. A reta vertical na figura da direita na posição 40 representa o contato entre as categorias.

(a) Uma RBF com suporte igual a 25 e centrada em cada uma das amostras.



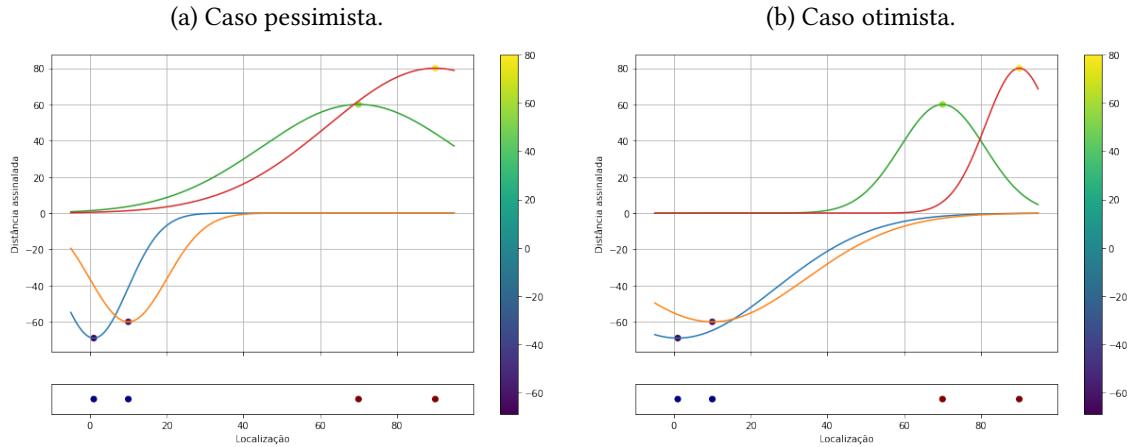
(b) Uma RBF com suporte igual a 25 e centrada em cada uma das amostras após o cálculo dos pesos e sua combinação linear em preto.



O método proposto modifica o suporte de cada uma das RBFs centradas nas amostras de acordo com o valor da distância assinalada calculada para aquela amostra a partir de curvas de parametrização como as mostradas na Figura 25. Dessa forma, cada RBF terá um suporte diferente. O volume do sólido fruto da extração da iso-superfície zero, gerado pelo cenário otimista é maior do que o gerado pelo caso base, ambos são maiores do que o volume do sólido gerado pelo caso pessimista.

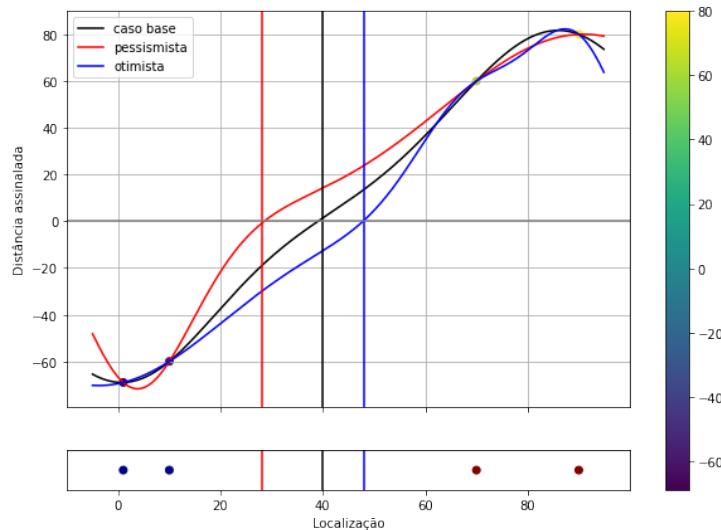
A Figura 33 mostra as RBFs, que na Figura 32 apresentam suporte igual a 25, com seus suportes modificados a partir de curvas de parametrização lineares. Os pesos são então calculados para cada um dos casos (otimista e pessimista) e uma superfície interpolada pode ser gerada a partir da combinação linear das RBFs.

Figura 33 – Amostras em uma dimensão no gráfico na posição inferior e distâncias assinaladas calculadas para cada uma das amostras no eixo y do gráfico na posição superior. As curvas coloridas representam os *kernels* associados à cada amostra.



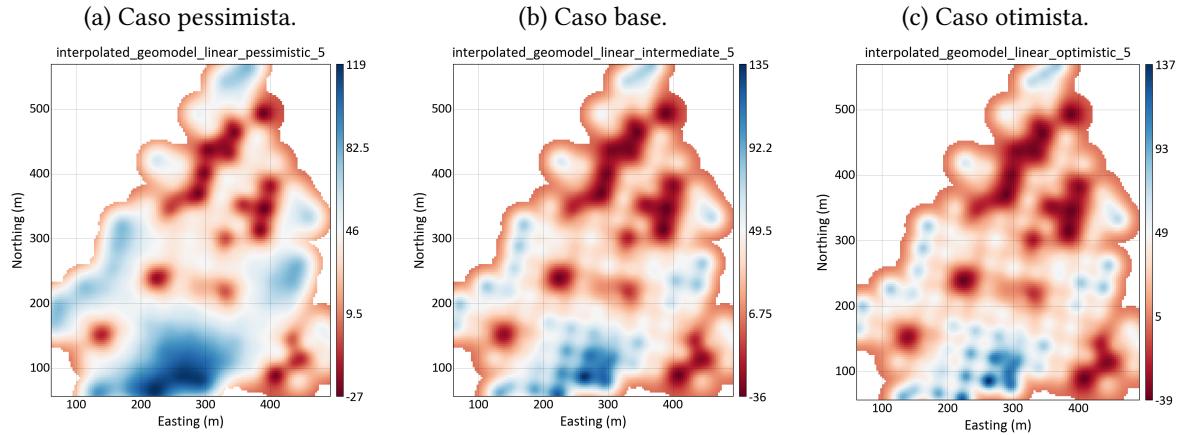
A Figura 34 mostra, em preto, a superfície interpolada para o caso base, em vermelho, para o cenário pessimista e em azul para o cenário otimista.

Figura 34 – Caso base e cenários otimista e pessimista para o exemplo unidimensional proposto. As retas verticais representam os pontos de contato entre as categorias.



A aplicação do método, no banco de dados *Swiss Jura*, consiste em calcular as distâncias assinaladas para cada uma das cinco categorias do banco de dados. Os parâmetros f_{min} para os cenários otimista e pessimista usados na modificação do suporte devem ser escolhidos. Nesse caso, usou-se um modelo linear com valor de f mínimo igual a 0.95 para todas as categorias. A Figura 35 mostra as distâncias assinaladas para a categoria Quaternary interpoladas para o cenário pessimista, caso base e cenário otimista. Os mesmos cenários foram interpolados para as outras quatro categorias presentes no banco de dados.

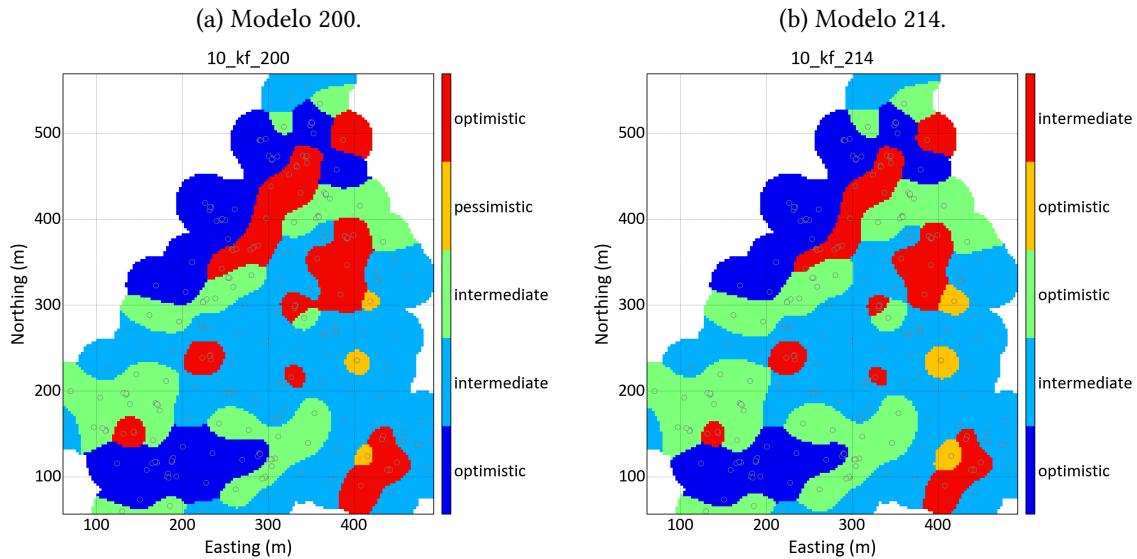
Figura 35 – Distâncias assinaladas interpoladas para a categoria Quaternary do banco de dados *Jura* para o caso base, e cenários pessimista e otimista.



Para cada categoria, há três diferentes cenários que devem ser combinados para criar os modelos geológicos multi-categóricos. Em alguns casos, a combinação do cenário otimista de uma categoria com grande área ou volume com o cenário pessimista de uma categoria de pequena área ou volume pode fazer com que nenhum bloco seja atribuído à categoria de menor proporção em alguns dos 3^K modelos gerados, onde K é o número de categorias do banco de dados. Para contornar esse problema, os modelos devem passar por um teste de validação. O algoritmo checa se o bloco com o centroide mais próximo a cada amostra foi classificado com a mesma categoria daquela amostra. O usuário seleciona um nível mínimo de reprodução, nesse caso 0.9. Desse modo, modelos que não reproduzem, pelo menos, 90% dos dados amostrais, são descartados. Dos 243 modelos gerados para o *Swiss Jura* combinando os três cenários para as cinco diferentes categorias, 24 não foram descartados já que reproduzem, pelo menos, 90% das amostras.

A [Figura 36](#) mostra duas realizações escolhidas aleatoriamente entre as 24 selecionadas, juntamente com as amostras representadas pelos círculos. Uma animação mostrando todas as realizações pode ser vista [aqui](#).

Figura 36 – Modelos geológicos para o *Swiss Jura* criados a partir da parametrização do suporte do *kernel*. As cores representam cada uma das cinco categorias enquanto a legenda mostra a qual cenário aquela categoria pertence.

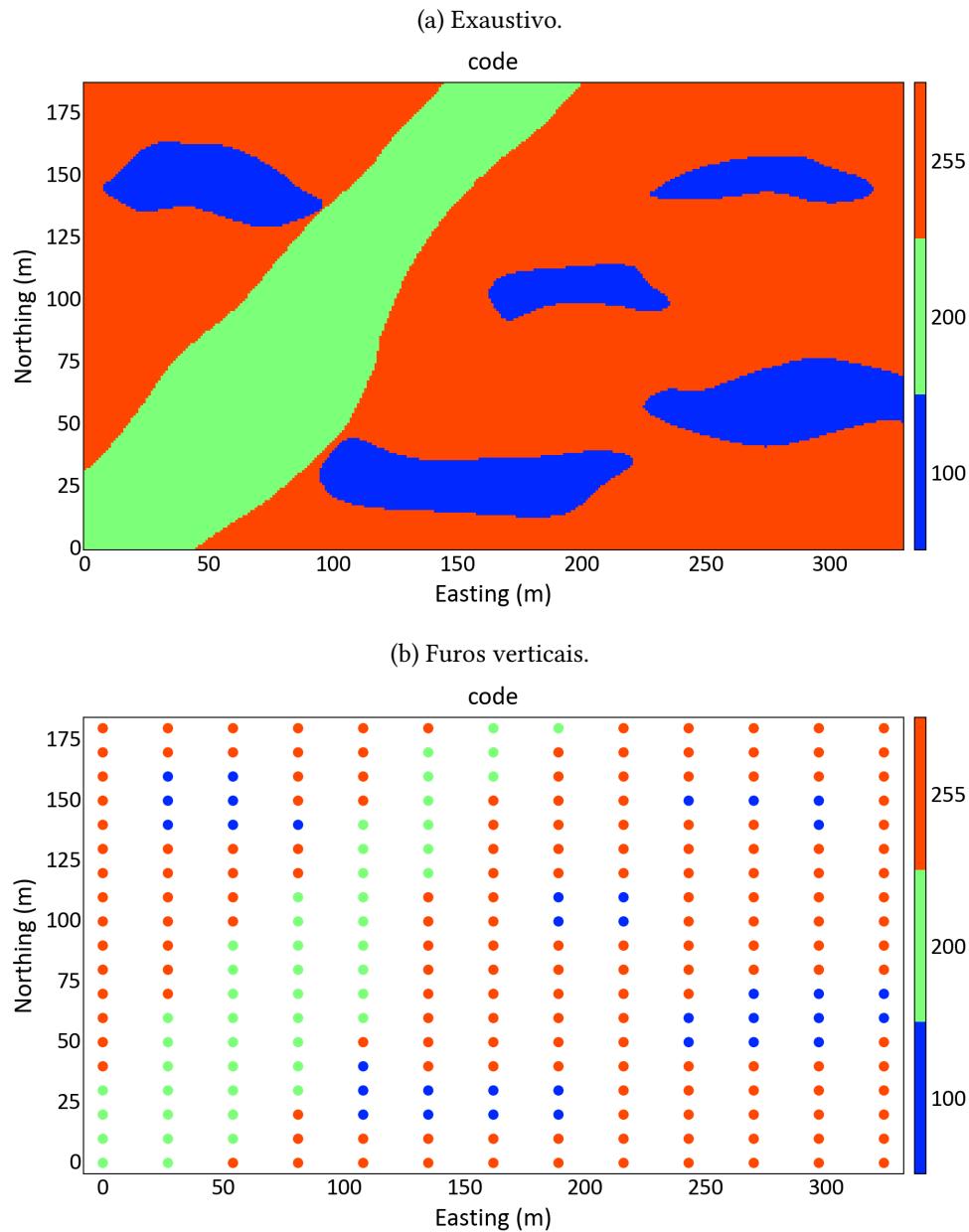


3.1.1 Estudo de caso

Para ilustrar de forma prática a metodologia, uma seção de um modelo conceitual foi produzida e é mostrada na Figura 37a. O modelo é composto por rocha encaixante, mostrada em vermelho, um dique, mostrado em verde, e estruturas lenticulares, mostradas em azul.

O modelo conceitual exaustivo foi amostrado por 12 furos verticais igualmente espaçados. O espaçamento vertical das compostas é de 10 metros. Os furos verticais são mostrados na Figura 37b.

Figura 37 – Modelo conceitual composto por rocha encaixante, mostrada em vermelho, um dique, mostrado em verde, e estruturas lenticulares, mostradas em azul.



O método foi aplicado no banco de dados da Figura 37b. O *grid* de interpolação tem dimensões 1x1 metros. O modelo dos *kernels* é Gaussiano com suportes iguais a 20, 30 e 20 metros para as categorias 255, 200 e 100, respectivamente. Não foi aplicada anisotropia e um efeito pepita de 0.01% foi utilizado para evitar instabilidades nas matrizes.

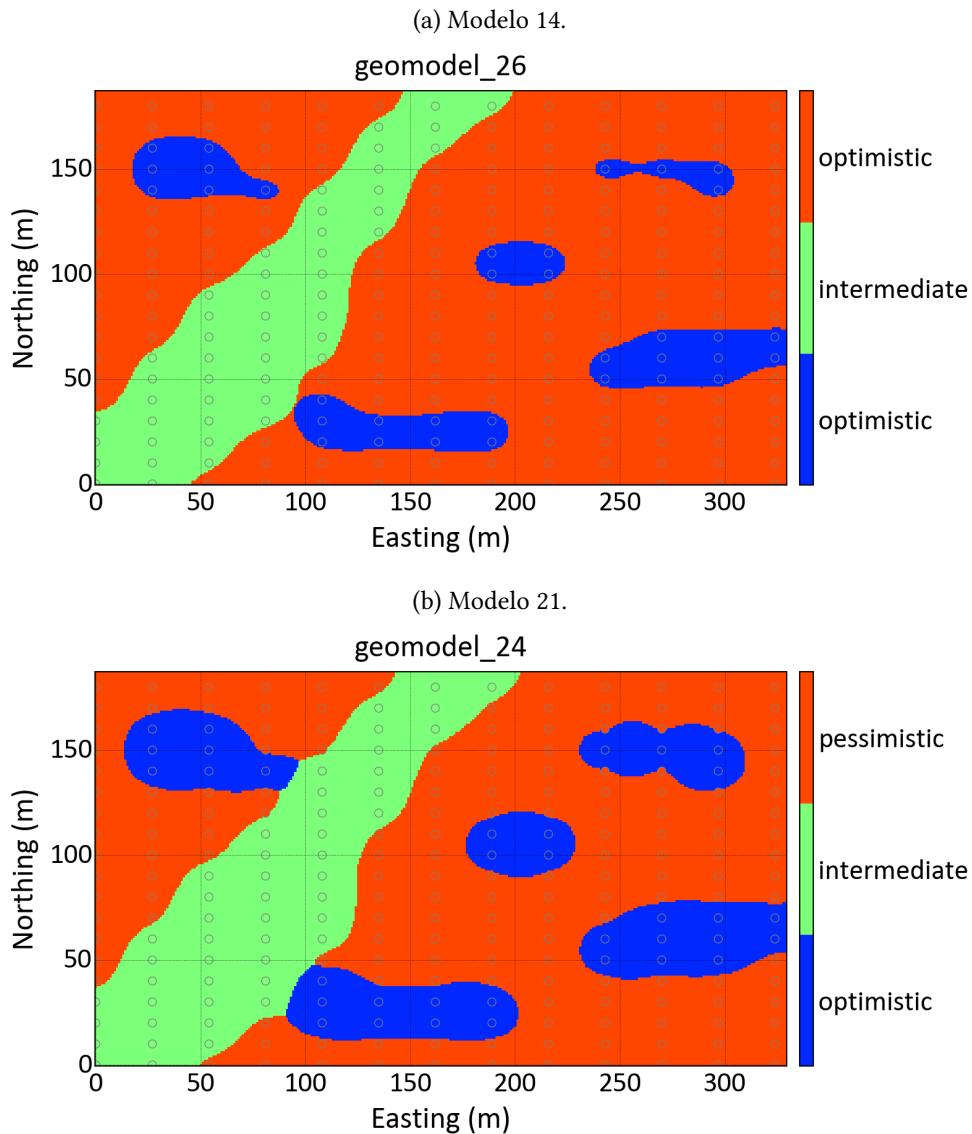
Foi utilizada a parametrização linear com $fmin\ optimista=fmin\ pessimista=0.8$ e critério de aceitação de reprodução das amostras igual a 0.85.

A [Figura 38](#) mostra duas realizações aprovadas no critério de aceitação de reprodução dos dados amostrais escolhidas aleatoriamente.

Uma animação mostrando todas as 17 realizações (aprovadas no teste de reprodução)

pode ser vista [aqui](#).

Figura 38 – Realizações para o modelo geológico sintético geradas a partir dos furos verticais. As cores representam cada uma das três categorias enquanto a legenda mostra a qual cenário aquela categoria pertence.



3.1.2 Discussão

O método proposto por [McLennan \(2007\)](#) é baseado em interpoladores pelo inverso da distância. Por esse motivo, não é possível considerar anisotropia, de forma direta, nos modelos. O autor põe o método em prática apenas em modelos sintéticos de geometria simples, quando aplicado a modelos com geologias complexas, a multiplicação dos pesos diretamente pelo fator calculado para cada amostra a partir da parametrização produziu, muitas vezes, volumes (ou áreas) muito grandes ou muito pequenas que não honram os dados amostrais. Aplicar o fator no suporte do *kernel* torna o método menos sensível aos fatores e dá ao usuário um controle

maior sobre os modelos gerados, já que é possível diminuir artificialmente o suporte do *kernel* para reduzir sua influência em detrimento da influência da parametrização.

O método proposto, por ser baseado em interpolação global, gera modelos sem ruídos, apresentando o realismo geológico esperado. O método simula diferentes interpretações para cada litologia do modelo geológico contraindo ou expandindo os diferentes domínios.

O usuário pode controlar o quanto os modelos honram os dados amostrais; entretanto, uma maior restrição nesse sentido gera menos cenários diferentes.

Embora a adaptação seja uma melhoria em relação ao método inicialmente proposto ainda existem pontos que precisam de ajustes: as propriedades distância assinaladas interpoladas ainda são, embora menos em relação à implementação original, sensíveis à parametrização. Uma pequena variação no parâmetro *fmin* gera uma variação muito grande nas propriedades distâncias interpoladas. De modo geral, é necessário reduzir artificialmente o alcance dos variogramas obtidos a partir das amostras. Caso o alcance não seja reduzido, a área ou volume dos sólidos será muito grande no caso otimista ou muito pequena no caso pessimista, prejudicando a reprodução dos dados amostrais.

A implementação computacional é em Python, o que torna a execução lenta e inviável para estudos de caso tridimensionais com um grande número de amostras e/ou nós de *grid*.

3.2 Avaliação de incerteza usando funções distância assinaladas e campos de probabilidades

[Froidevaux \(1993\)](#) propôs uma abordagem para a simulação condicional de variáveis contínuas. Este método dissocia a tarefa de estimar as funções de distribuição de probabilidade local (PDFs) da produção de realizações equiprováveis. A simulação a partir de campos de probabilidade parte da premissa de que as distribuições condicionais locais são conhecidas. As simulações condicionais são então obtidas extraiendo realizações desses PDFs locais.

A metodologia proposta se baseia nesse método para simular contatos entre as diferentes litologias de um depósito mineral. A primeira etapa define uma largura de banda de incerteza em torno dos contatos entre as diferentes litologias. Como não há incerteza no interior dos domínios, os blocos fora da zona de incerteza são congelados com o valor da categoria responsável pela distância assinalada estimada mais negativa. Simultaneamente, os blocos dentro da zona de incerteza serão classificados como diferentes categorias em diferentes realizações.

A largura da zona de incerteza pode ser definida por um geomodelador com base no tipo de depósito em estudo e na configuração amostral: em depósitos onde a geologia apresenta variabilidade mais significativa, como depósitos com múltiplas rochas intrusivas por exemplo, a zona de incerteza deve ser mais ampla. Em depósitos menos erráticos, como os estratificados, deve ser mais estreita. Em malhas de amostragem com espaçamento próximo, zonas de incerteza

menores são necessárias, enquanto em malhas de amostragem esparsas, zonas mais amplas são necessárias.

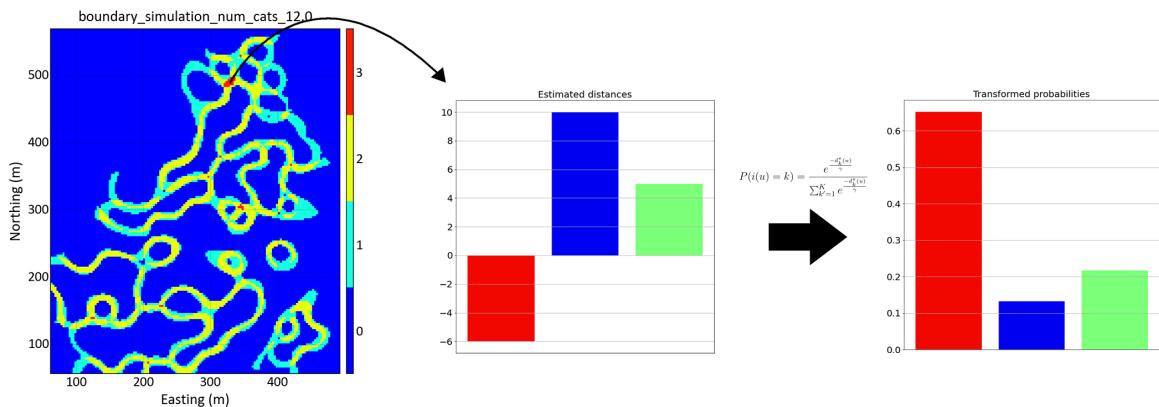
A distância assinalada estimada em um nó do *grid* é a distância de um bloco ao domínio oposto mais próximo. A zona de incerteza é definida pela retenção de todos os blocos onde o valor absoluto da distância estimada, de pelo menos uma das categorias do banco de dados, é menor ou igual ao valor definido pelo geomodelador para a largura de banda de incerteza.

A estimativa da PDF local é feita transformando as distâncias estimadas em probabilidades em cada um dos blocos da zona de incerteza como mostrado na [subseção 2.2.1](#).

A partir da modelagem multi-categórica no banco de dados *Swiss Jura* mostrada na [Figura 13](#), foi determinada uma zona de incerteza de 12 metros ao redor dos contatos. Isto é, para cada uma das distâncias assinaladas interpoladas que representam cada uma das categorias do banco de dados, qualquer bloco em que a distância interpolada pertença ao intervalo [-12,12] é classificado como zona de incerteza.

A [Figura 39](#) mostra, do lado esquerdo, blocos definidos como pertencentes à banda de incerteza de 12 metros para o *Swiss Jura*. As cores indicam quantas categorias têm sua distância assinalada interpolada absoluta menor ou igual ao valor da largura de banda (12 metros).

Figura 39 – Da esquerda para a direita: (1) zona de incerteza de 12 metros; (2) distâncias estimadas para cada uma das três categorias um bloco específico; (3) probabilidades transformadas.



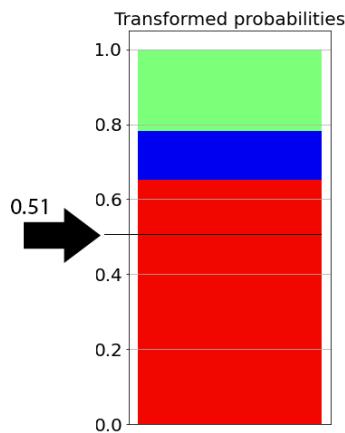
Observe a partir da [Figura 39](#) que, embora haja cinco categorias diferentes no banco de dados, apenas três farão parte da distribuição de probabilidade local em um bloco vermelho; apenas duas em um bloco amarelo e apenas uma em um bloco azul claro. A imagem central mostra as distâncias estimadas para um bloco em específico. Finalmente, no lado direito estão as probabilidades transformadas pela [Equação 2.20](#) para aquele bloco usando o valor de distância absoluta máxima como o parâmetro ω .

A produção de realizações equiprováveis para o modelo geológico é feita simulando primeiro um campo gaussiano não condicional nos blocos pertencentes à zona de incerteza por

simulação sequencial Gaussiana ([algoritmo 1](#)). Os valores Gaussianos devem ser transformados no campo de probabilidade estandardizando-os para que variem entre 0 e 1 em uma distribuição uniforme. Isso é feito calculando os valores de distribuição Gaussiana cumulativos.

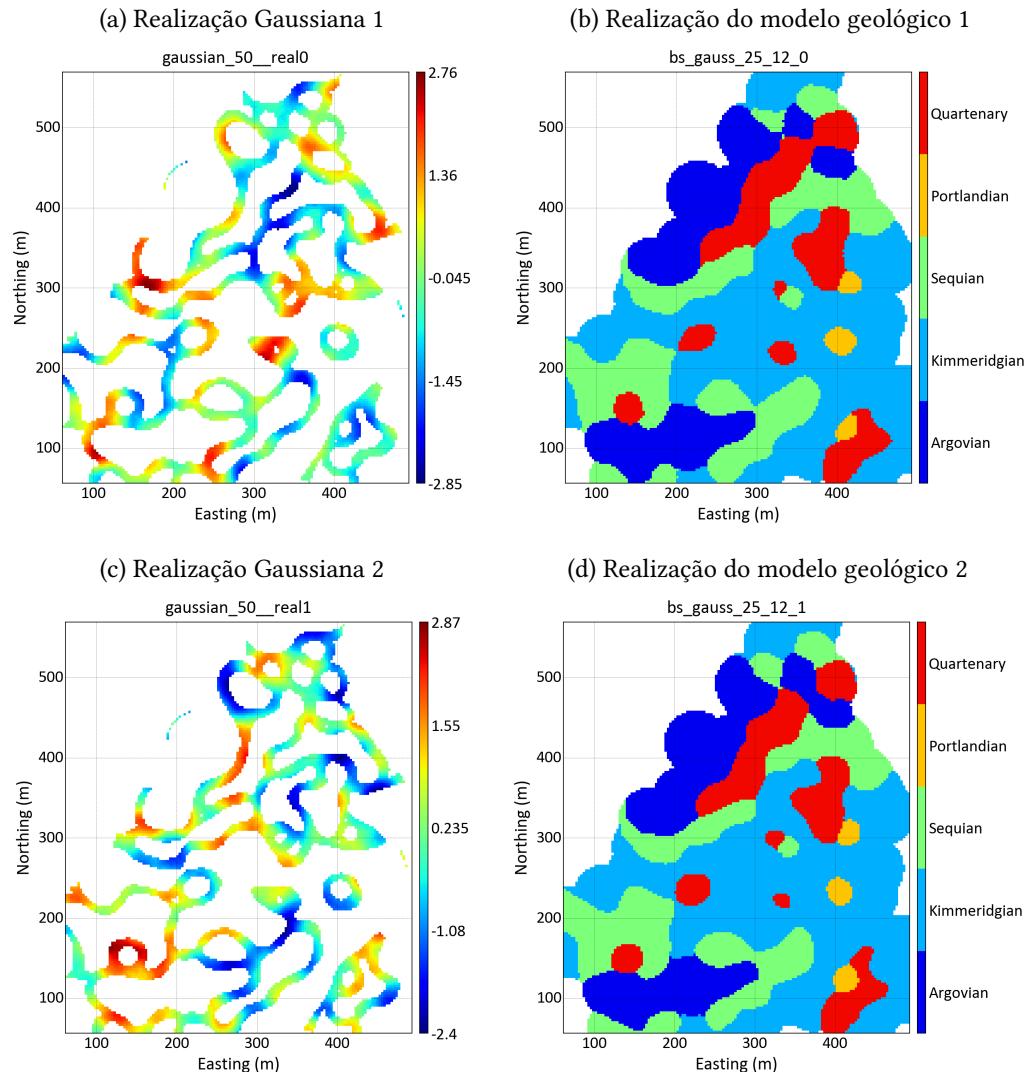
Finalmente, para gerar várias realizações do modelo geológico, uma categoria deve ser amostrada das PDFs locais usando o campo de probabilidade simulado para cada bloco dentro da zona de incerteza em cada realização, como mostrado na [Figura 40](#).

Figura 40 – Um valor de probabilidade simulado de 0,51 é usado para amostrar a categoria vermelha de uma distribuição condicional em um bloco específico.



A [Figura 41](#) mostra, do lado esquerdo, duas realizações diferentes de uma simulação Gaussiana não condicional dentro da zona de incerteza. Ao lado direito, mostra duas realizações do modelo geológico criado pela amostragem de uma categoria extraída das distribuições condicionais locais para o banco de dados *Swiss Jura*. A transição entre as diferentes categorias é suave, pois as simulações Gaussianas têm continuidade espacial. Uma animação mostrando todas as 10 realizações realizadas pode ser vista [aqui](#).

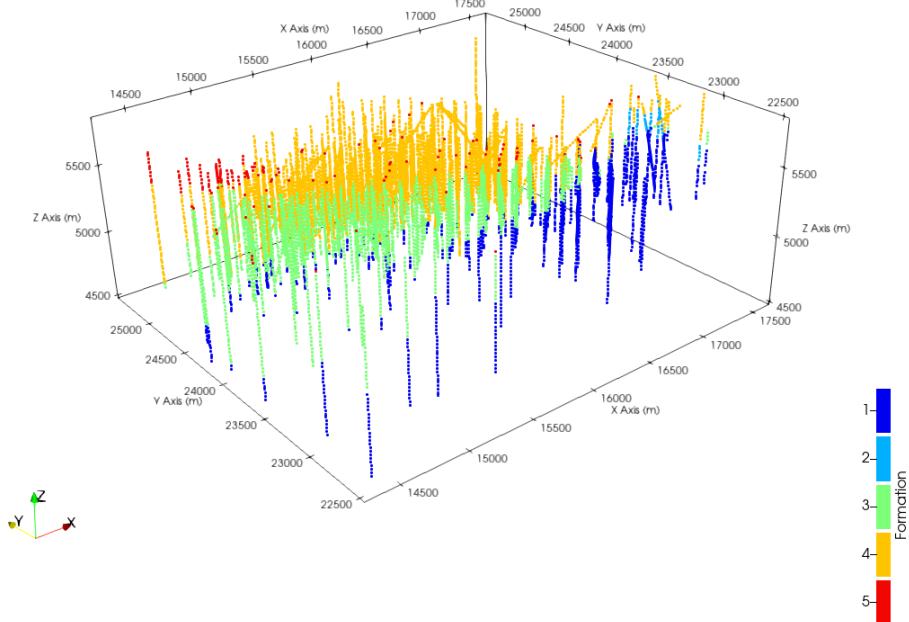
Figura 41 – Realizações Gaussianas e modelos geológicos correspondentes.



3.2.1 Estudo de caso

Esse estudo de caso foi conduzido em um banco de dados real fornecido por uma grande operação de mineração de ouro. O conjunto de dados apresenta 9140 amostras em suporte pontual representando cinco litologias diferentes localizadas em uma área de aproximadamente 10 km² com 1300 metros de profundidade. Uma vista em perspectiva das amostras é apresentada na [Figura 42](#).

Figura 42 – Vista em perspectiva das amostras do depósito de ouro. As cores representam as cinco diferentes litologias do depósito mineral.

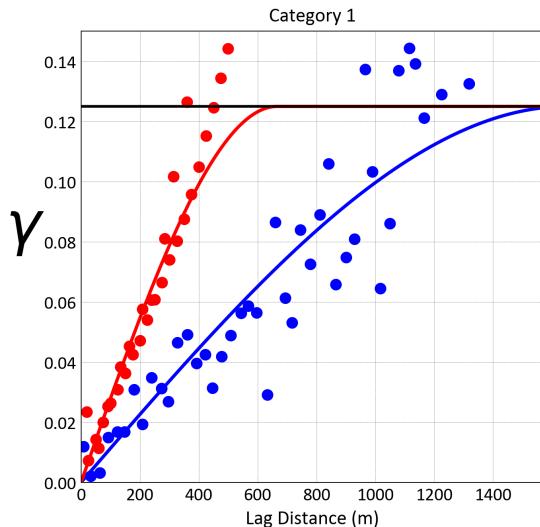


[Rolo \(2017\)](#) e [Rolo, Radtke e Costa \(2017\)](#) realizaram um estudo de caso no mesmo banco de dados; no entanto, os autores utilizaram distâncias assinaladas para criar um modelo geológico multi-categórico determinístico. O presente trabalho é uma extensão natural do trabalho acima mencionado, uma vez que é simples e direto avaliar a incerteza do modelo geológico usando campos de probabilidade a partir das distâncias estimadas usadas para modelagem geológica determinística.

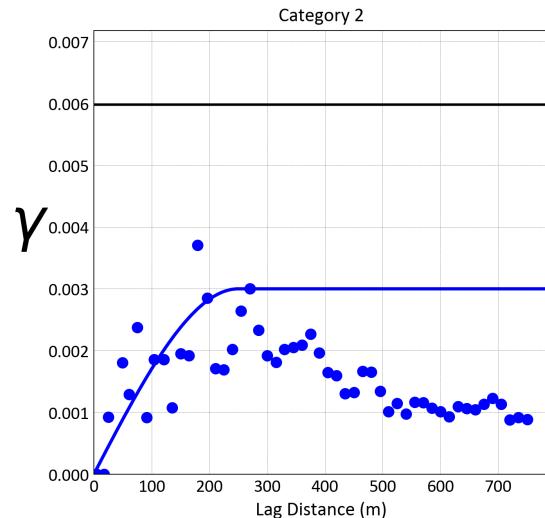
Os variogramas dos indicadores para as cinco categorias do depósito de ouro foram calculados e modelados com estruturas esféricas, conforme mostrado na [Figura 43](#). Para os grupos 2, 3, 4 e 5, um variograma omnidirecional foi calculado e modelado. Já para a categoria 1, foram calculados e modelados variogramas omni-horizontal (em azul) e vertical (em vermelho).

Figura 43 – Variogramas dos indicadores para todas as cinco categorias do depósito mineral modelados com estruturas esféricas.

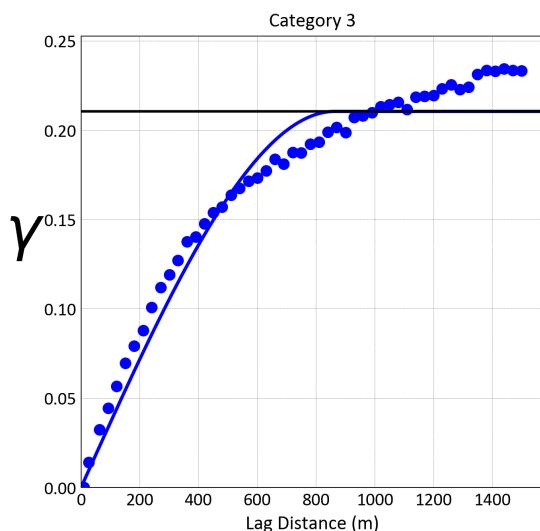
(a) Variograma omni-horizontal em azul e vertical em vermelho para a categoria 1.



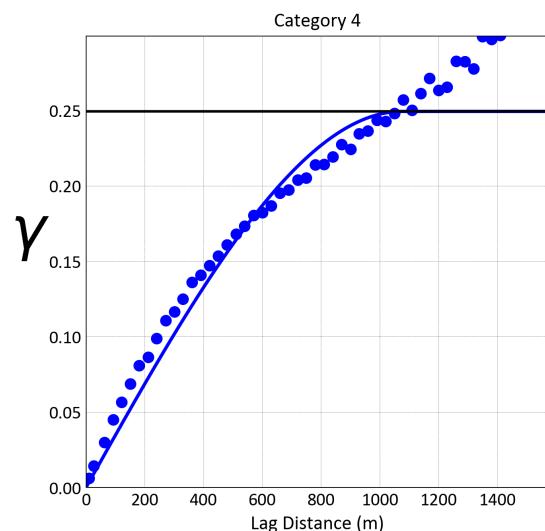
(b) Variograma omnidirecional para a categoria 2.



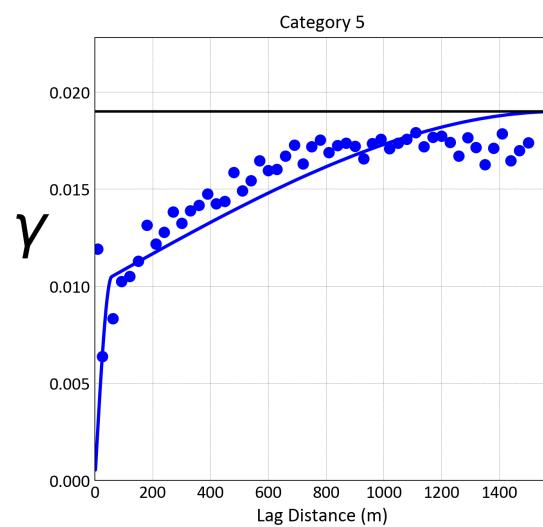
(c) Variograma omnidirecional para a categoria 3.



(d) Variograma omnidirecional para a categoria 4.



(e) Variograma omnidirecional para a categoria 5.



As equações: [Equação 3.1](#), [Equação 3.2](#), [Equação 3.3](#), [Equação 3.4](#) e [Equação 3.5](#) representam os variogramas para as categorias 1, 2, 3, 4 e 5 respectivamente.

$$\gamma(h) = 0.12504 Sph_1 \left[\frac{EW/NS}{1651.53m}, \frac{VERT}{665.98m} \right] \quad (3.1)$$

$$\gamma(h) = 0.003 Sph_1 \left[\frac{OMNI}{250m} \right] \quad (3.2)$$

$$\gamma(h) = 0.21044 Sph_1 \left[\frac{OMNI}{863.81m} \right] \quad (3.3)$$

$$\gamma(h) = 0.24927 Sph_1 \left[\frac{OMNI}{1076.01m} \right] \quad (3.4)$$

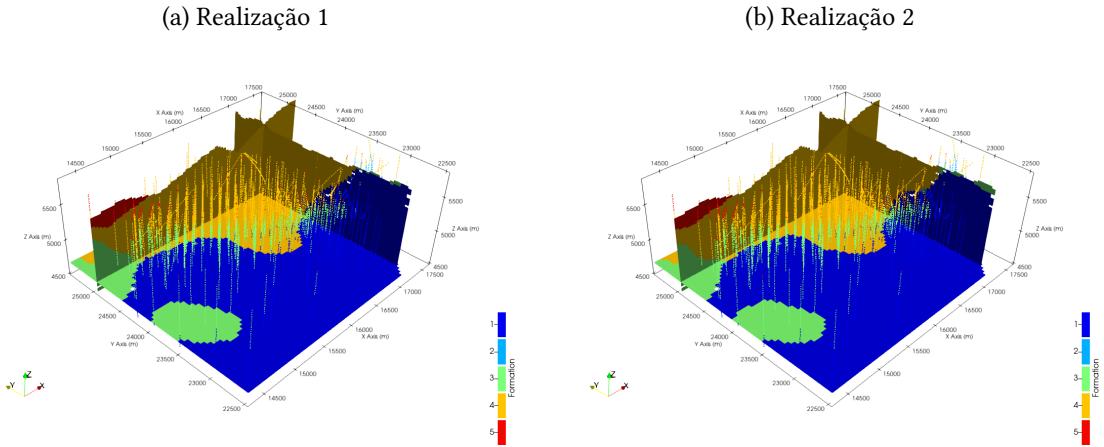
$$\gamma(h) = 0.01 Sph_1 \left[\frac{OMNI}{58.2m} \right] + 0.00899 Sph_2 \left[\frac{OMNI}{1612.38m} \right] \quad (3.5)$$

As distâncias assinaladas calculadas foram interpoladas, por krigagem dual, para um *grid* de 50x50x25 metros usando um equivalente Gaussiano para os variogramas dos indicadores ajustados com modelos esféricos (mesmas contribuições, alcances e razões de anisotropia). Essas dimensões foram escolhidas para o *grid*, porque a interpolação dos teores de ouro e posterior planejamento mineiro são feitos, também, nesse *grid*.

Uma largura de 80 metros foi escolhida para a zona de incerteza com base no tipo de depósito e na configuração geométrica amostral. Blocos fora da zona de incerteza foram classificados com a categoria responsável pela distância estimada mais negativas, enquanto blocos dentro da zona de incerteza foram simulados.

Dez realizações de uma simulação não condicional Gaussiana foram realizadas, dentro da zona de incerteza, usando um variograma Gaussiano com um alcance de 600 metros e um pequeno (0,01%) efeito pepita para evitar instabilidades matemáticas. As realizações foram transformadas em campos de probabilidade e usadas para amostrar uma categoria das distribuições de probabilidade locais. As probabilidades locais foram calculadas transformando as distâncias estimadas, dentro da zona de incerteza, em probabilidades. A distância máxima estimada absoluta foi usada como o parâmetro ω para cada bloco, dessa forma, a magnitude da incerteza foi controlada apenas pela largura de zona de incerteza. A [Figura 44](#) mostra uma vista em perspectiva de seções verticais e horizontais de duas das 10 realizações para o modelo geológico juntamente com as amostras em suporte pontual.

Figura 44 – Visão em perspectiva de seções verticais e horizontais de duas realizações do modelo geológico juntamente com as amostras em suporte pontual.



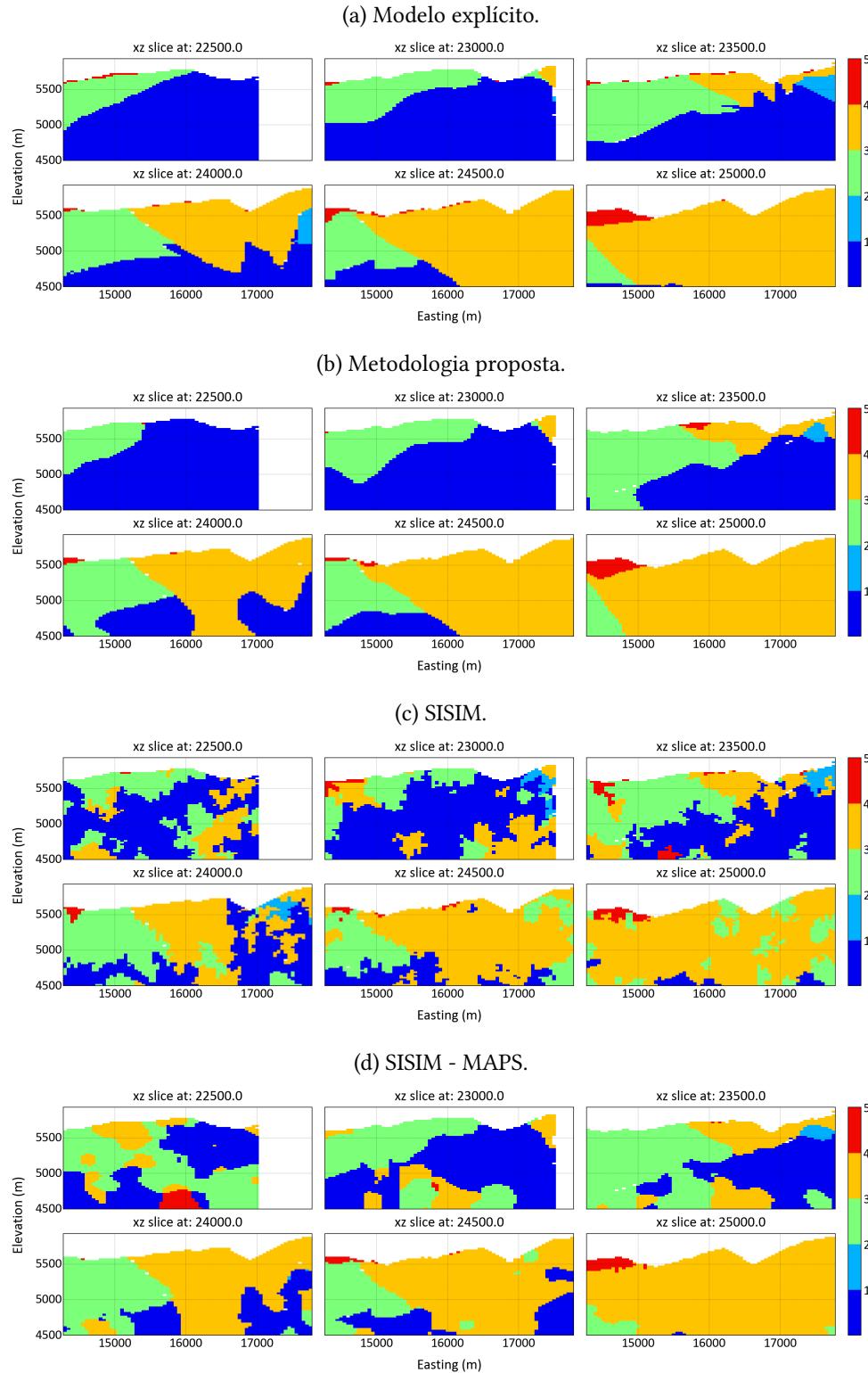
3.2.2 Discussão

Uma das técnicas tradicionalmente usadas para modelar geocorpos e avaliar sua incerteza na presença de várias categorias é a simulação sequencial dos indicadores (SISIM). A SISIM foi aplicada ao mesmo banco de dados a fim de comparar os resultados com a metodologia proposta, uma vez que também depende apenas das coordenadas x, y e z dos compósitos e da propriedade categórica.

A Figura 45 mostra um comparativo de seções verticais XZ entre a metodologia proposta (45b), a SISIM (45c) utilizando o variograma dos indicadores mostrados na Figura 43, e um modelo explícito (45a) elaborado por um geomodelador experiente. Uma animação mostrando as mesmas seções para todas as dez realizações pode ser vista [aqui](#). A SISIM gera realizações com uma textura típica "salt and pepper" e padrões ruidosos, o que não é geologicamente realista. Foram propostas técnicas para limpar as realizações da SISIM (DEUTSCH, 1998; YAMAMOTO et al., 2015). Uma dessas técnicas é chamada MAPS (*maximum a-posteriori selection*), e é comumente utilizada para remover os artefatos ruidosos da SISIM e melhorar a reprodução das proporções das categorias. A Figura 45d mostra seções verticais de uma realização da SISIM pós processada.

A metodologia proposta cria limites contínuos que são mais semelhantes à interpretação do geomodelador, mesmo quando comparada à SISIM pós processada. A ausência de dados de sondagem é responsável pelas regiões onde a metodologia proposta e o SISIM divergem da interpretação do geomodelador. Como o interpolador é global, não há necessidade configurar estratégia de busca na metodologia proposta ao contrário da SISIM, entretanto, é necessário definir um variograma adicional para a simulação não condicional, em relação aos variogramas para interpolar as distâncias assinaladas ou simular os indicadores na SISIM. Ao contrário da SISIM, o algoritmo não possui mecanismo para respeitar as proporções e reproduzir a continuidade espacial de cada categoria, embora isso possa ser contornado alterando iterativamente os parâmetros e verificando os resultados.

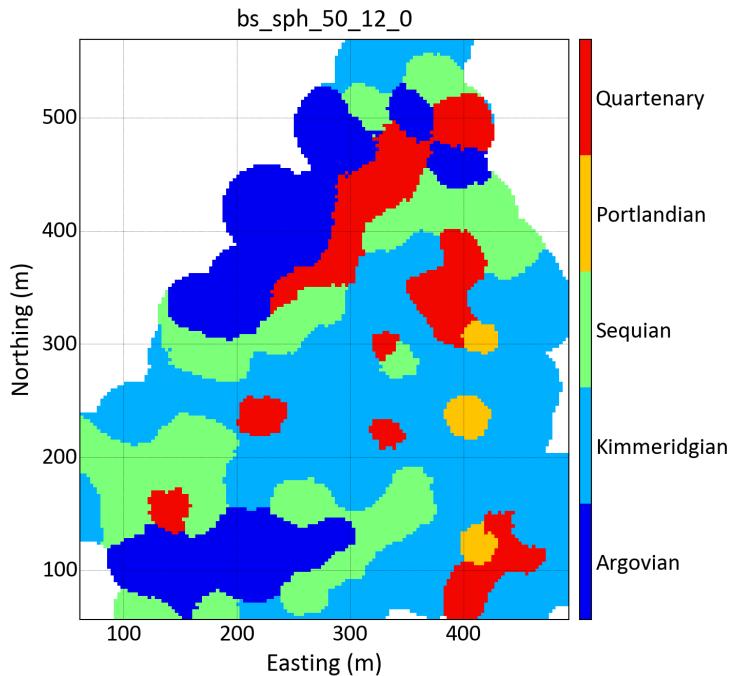
Figura 45 – Comparativo de seções verticais xz entre a metodologia proposta, a SISIM e um modelo explícito.



O comportamento dos contatos é controlado pelo variograma usado na simulação não condicional; alcances menores e efeitos pepita maiores geram limites irregulares e descontínuos. Por outro lado, longos alcances e pequenos efeitos pepita geram limites suaves e contínuos. A

[Figura 46](#) mostra uma realização para o *Swiss Jura* criada usando um variograma esférico com um alcance de 50 metros. Os contatos são mais rugosos do que aqueles vistos na [Figura 41](#), que foram gerados por um variograma Gaussiano.

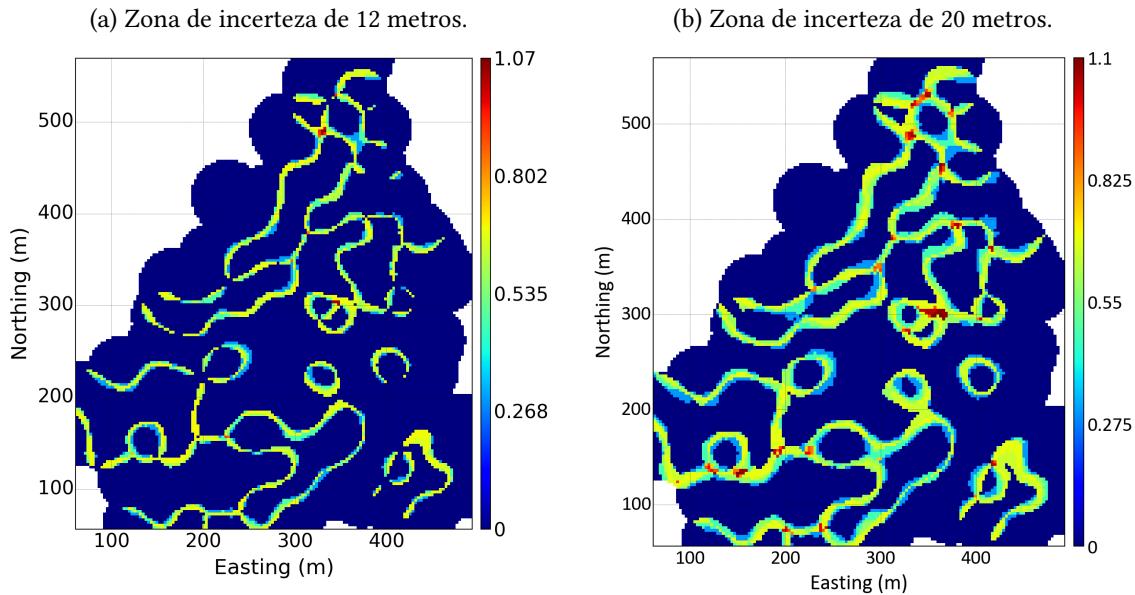
Figura 46 – Uma realização para o *Swiss Jura* criada usando um variograma esférico para a simulação não condicional.



A magnitude da incerteza é controlada tanto pela largura de zona de incerteza quanto pelo parâmetro ω . Um valor para o parâmetro ω pode ser escolhido para todos os blocos ou o maior valor absoluto entre as distâncias assinaladas pode ser utilizado como parâmetro ω bloco a bloco. Nesse caso, somente a largura da banda vai controlar a magnitude da incerteza. Uma animação mostrando 10 realizações feitas para o *Swiss Jura* com o mesmo variograma Gaussiano do exemplo anterior, entretanto, para uma largura da zona de incerteza de 20 metros pode ser vista [aqui](#). Como há mais área para os contatos mudarem de posição, a variação de área é maior do que a observada para a largura de banda de 12 metros.

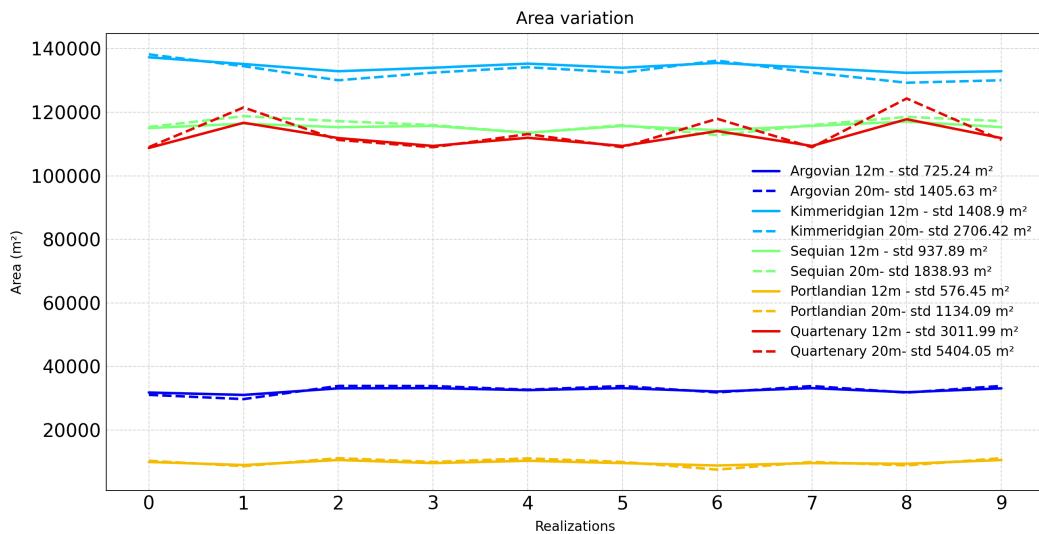
O mesmo comportamento é evidenciado pelos mapa de entropia calculados a partir da [Equação 2.21](#) (e estandardizadas para que os valores estejam entre 0 e 1) para as bandas de incerteza de 12 e 20 metros. Blocos que a entropia tem um valor alto, variam mais entre diferentes categorias nas diferentes realizações.

Figura 47 – Entropia calculada a partir das dez realizações no Swiss Jura.



A Figura 48 mostra a área para as categorias Swiss Jura para 10 realizações com larguras de banda de 12 e 20 metros. Os desvios padrão das áreas também são mostrados. O gráfico mostra que as áreas têm maior variação para todas as categorias quando a largura de banda da incerteza é de 20 metros.

Figura 48 – Variação de área para todas as categorias do Swiss Jura para 10 realizações com larguras de banda de 12 e 20 metros.



A matriz de confusão mostrada na Figura 49 é calculada checando, para cada realização do modelo geológico para o depósito de ouro, a categoria simulada ou estimada nos blocos com centroide mais próximo a cada uma das amostras. A média das proporções verificadas é tomada e os valores registrados na tabela: categoria predita nas colunas e categoria da amostra (real)

nas linhas. Sendo assim, as proporções da diagonal principal são de blocos que reproduzem as amostras, enquanto as proporções fora da diagonal principal são de blocos que não reproduzem as amostras. Considerando as 10 realizações, 91% dos blocos com centroide mais próximos de amostras da categoria 1 foram classificados como categoria 1, ao mesmo tempo, 13% dos blocos com centroide mais próximo às amostras de categoria 1 foram classificados como categoria 2. A reprodução é alta para categorias de alto volume com abundância de amostras. As categorias 2 e 5 têm menos volume e menos amostras em comparação com as categorias 1, 3 e 4, como evidenciado pela Figura 50, o que explica a menor reprodução. Para a categoria 5, o problema é acentuado, uma vez que existem apenas algumas amostras esparsas na superfície do depósito. A distância estimada das categorias de volume maior sempre será mais negativa nessas áreas para casos como esse. Uma maneira de contornar esse problema é pré-processar os modelos atribuindo as categorias das amostra mais próximas aos blocos correspondentes. Esta solução congelaria alguns blocos com base no valor da amostra mais próxima, minimizando o desvanecimento dessas categorias em relação à se fossem interpoladas/simuladas com dados circundantes.

Figura 49 – Matriz de confusão para o depósito de ouro mostrando proporções das categoria preditas nas colunas e das categorias das amostras (real) nas linhas.

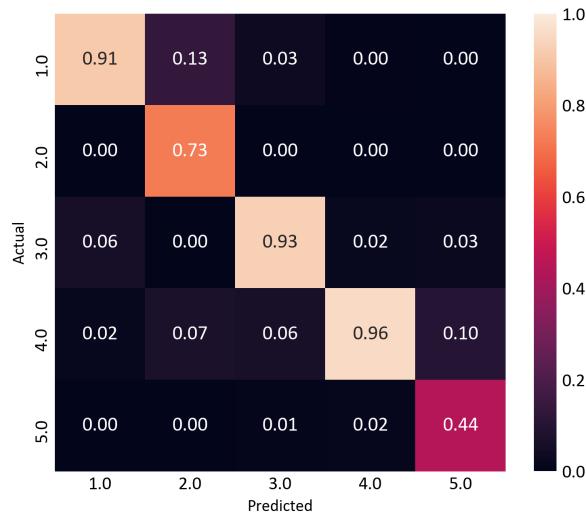
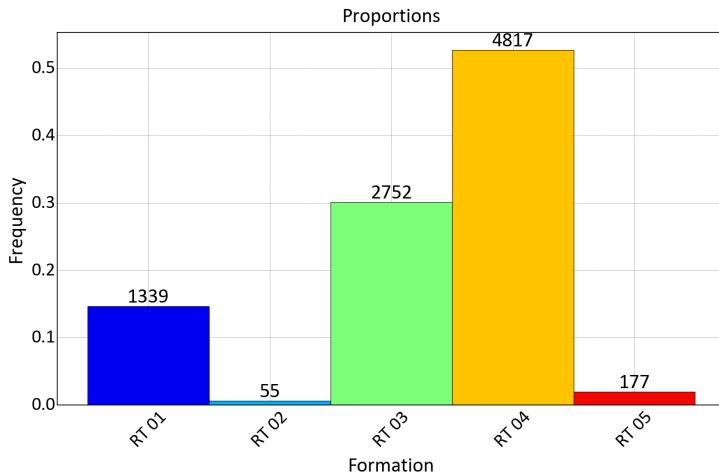


Figura 50 – Proporção de cada uma das categorias nos dados.



O método é bastante simples e conveniente já que dissocia a simulação dos contatos da obtenção de probabilidades locais. Geralmente, as probabilidades foram calculadas previamente para a construção de um modelo geológico determinístico. Assim, para avaliação da incerteza é necessário apenas realizações de uma simulação não condicional. Entretanto, não há uma forma objetiva para definição do variograma das simulações. Além disso, uma mesma simulação com o mesmo variograma, para representar o contato entre todas as diferentes categorias do depósito mineral pode ser restritivo. O contato entre horizontes sulfetados e oxidados podem apresentar um contato suave, enquanto o contato entre rochas intrusivas podem ser mais ásperos em depósitos de cobre pórfiro por exemplo.

A zona de incerteza simétrica ao redor dos contatos também pode ser um fator de restrição de aplicabilidade do método. Considerar uma zona simétrica, embora simplifique o método, não leva em consideração a possibilidade de que amostras podem estar no interior da zona de incerteza. Além disso, pode haver uma maior incerteza em relação a alguns contatos em relação aos demais, nesse sentido, existe a necessidade de zonas de incerteza com espessuras diferentes para diferentes contatos.

3.3 Simulação hierárquica dos contatos

A proposta dessa metodologia é avaliar a incerteza do modelo geológico multi-categórico simulando os contatos entre os diferentes domínios com base no método proposto por [Wilde e Deutsch \(2012\)](#). Os resultados são obtidos a partir dos furos de sondagem compostados, apresentando coordenadas X, Y e Z e propriedade categórica que representa as diferentes litologias do depósito. O método de [Wilde e Deutsch \(2012\)](#) pode ser aplicado apenas a casos binários e fazer sua aplicação ingenuamente em casos de múltiplos domínios é demorado, subjetivo e pode causar sobreposição de blocos atribuídos à diferentes categorias em diferentes realizações ou produzir blocos não atribuídos à nenhuma categoria em algumas das realizações.

Por esse motivo, essa tese propõe um algoritmo para agrupar automaticamente as diferentes litologias e um fluxo de trabalho para simular cada grupo individualmente e, em seguida, reconstruir o modelo geológico multi-categórico, a partir do agrupamento, de forma hierárquica.

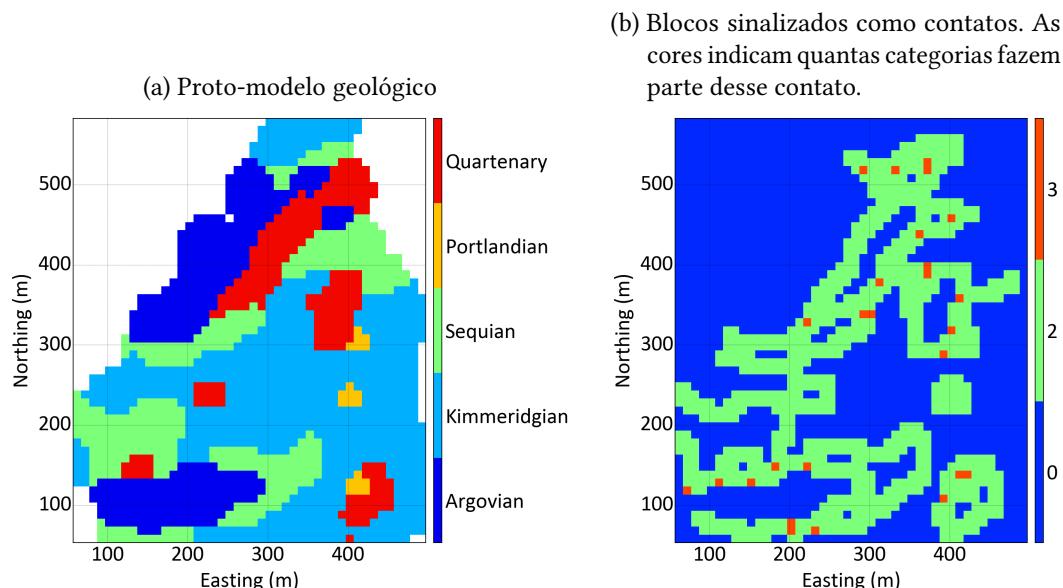
A metodologia foi inicialmente proposta por [Amarante, Rolo e Costa \(2018\)](#) e foi posteriormente aprimorada tendo sua subjetividade reduzida e automatização aumentada ([AMARANTE; ROLO; COSTA, 2021](#)). Um outro estudo de caso foi conduzido em um banco de dados de ferro por [Amarante, Rolo e Costa \(2019\)](#) e o método se mostrou competente, apresentando resultados melhores em relação a outros métodos concorrentes.

A primeira etapa do fluxo de trabalho é definir grupos de amostras que representem os contatos entre as diferentes litologias do depósito mineral. O agrupamento pode ser definido explicitamente por um geomodelador ou pode ser feito automaticamente pelo algoritmo proposto.

O algoritmo de cubos marchantes, apresentado na [subseção 2.1.1.6.2](#), é aplicado a um proto-modelo geológico para identificar e contar os contatos entre as diferentes litologias. O proto-modelo pode ser criado pelo vizinho mais próximo, máquina de vetores de suporte, krigagem dos indicadores ou usando funções distâncias assinaladas em um *grid* mais grosso do que o *grid* de simulação final para acelerar o processo.

A [Figura 51](#) mostra um proto-modelo geológico criado a partir do banco de dados *Swiss Jura* usando distâncias assinaladas em um *grid* de 10x10 metros. A resolução do *grid* pode alterar o agrupamento final.

Figura 51 – Proto-modelo e blocos sinalizados como contatos pelo algoritmo cubos marchantes para o *Swiss Jura*.



O algoritmo dos cubos marchantes, ao identificar um contato, faz sua contagem e o

registra. O número de contatos entre as diferentes litologias, contados pelo algoritmo dos cubos marchantes no proto-modelo da Figura 51a, é mostrado na Tabela 1.

Tabela 1 – Contagem de contatos pelo algoritmo dos cubos marchantes.

Kimmeridgian; Sequian	134
Argovian; Sequian	74
Kimmeridgian; Quartenary	66
Argovian; Quartenary	45
Sequian; Quartenary	40
Kimmeridgian; Portlandian	23
Portlandian; Quartenary	9
Argovian; Kimmeridgian	8
Argovian; Kimmeridgian; Sequian	7
Argovian; Sequian; Quartenary	6
Kimmeridgian; Portlandian; Quartenary	4
Kimmeridgian; Sequian; Quartenary	4

Para definir os ($K-1$) grupos, onde K é o número de categorias do conjunto de dados, o Algoritmo 2 é aplicado. Os primeiros grupos representam os contatos mais frequentes. Na iteração inicial, a variável i é o número par 0, o algoritmo irá criar um grupo onde as amostras das categorias Kimmeridgian e Sequian são codificadas como 1 e as amostras de outras categorias são codificadas como 0. Na segunda iteração, a variável i recebe o número ímpar 1, portanto o algoritmo irá criar um grupo onde as amostras da categoria Kimmeridgian são codificadas como 1 e Sequian são codificadas como 0. Além disso, o algoritmo irá remover da tabela de contagem de contatos quaisquer contatos em que Kimmeridgian ou Sequian façam parte, a saber, linhas 1, 2, 3, 5, 6, 8, 9, 10, 11 e 12. A variável i apresenta o valor par igual a 2 na terceira iteração, assim um grupo é criado onde as amostras das categorias Argovian e Quartenary são codificadas como 1 e as outras categorias da tabela de contagem de contatos, após a remoção das linhas, são codificadas como 0, pois este é o contato mais frequente após a remoção das linhas. O algoritmo assume que os primeiros ($K-1$) contatos mais frequentes serão entre duas, não mais, categorias, o que é razoável em um contexto geológico.

Algoritmo 2: Define os grupos a partir da tabela de contagem dos contatos.

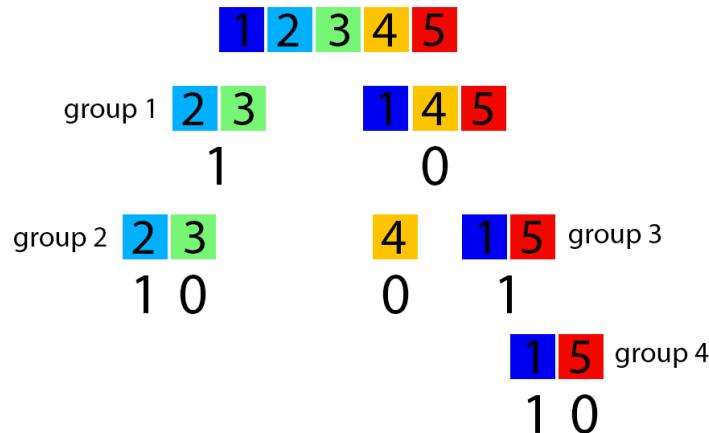
```

for  $i$  in  $(K-1)$  do
    if  $i \% 2 != 0$  then
        1. Cria um grupo: categorias do contato mais frequente são codificadas como 1,
           outras categorias são codificadas como 0;
    else
        1. Cria um grupo: uma categoria do contato mais frequente é codificada como 1,
           a outra como 0;
        2. Remove da lista de contagem de contatos todos os contatos que contêm pelo
           menos uma das categorias do contato mais frequente;
    end
end

```

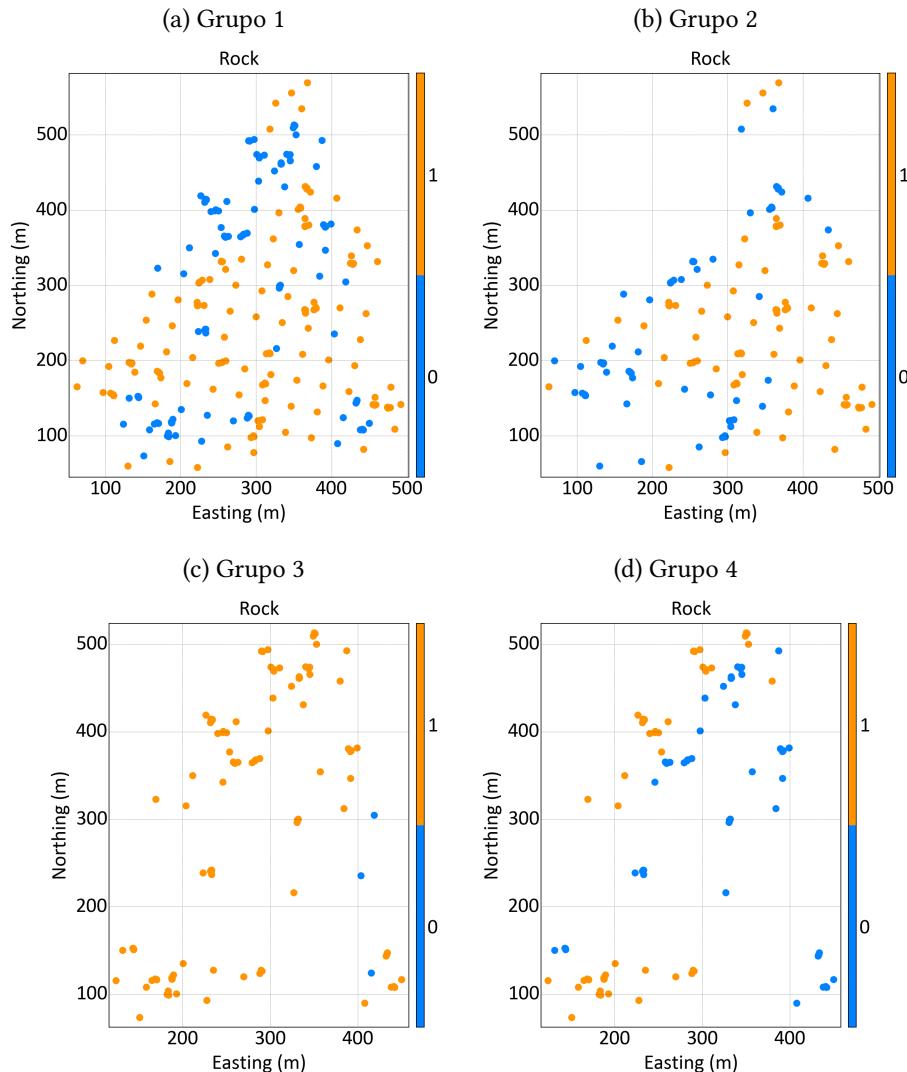
A figura 52 mostra a regra de hierarquia definida pelo algoritmo proposto para o proto-modelo do *Swiss Jura*.

Figura 52 – Regra hierárquica definida pelo algoritmo proposto.



Os quatro grupos de amostras gerados a partir da regra de hierarquia são apresentados na Figura 53. O primeiro grupo contém todas as amostras do conjunto de dados, enquanto os outros são subconjuntos.

Figura 53 – Mapa de localização das amostras dos quatro grupos.



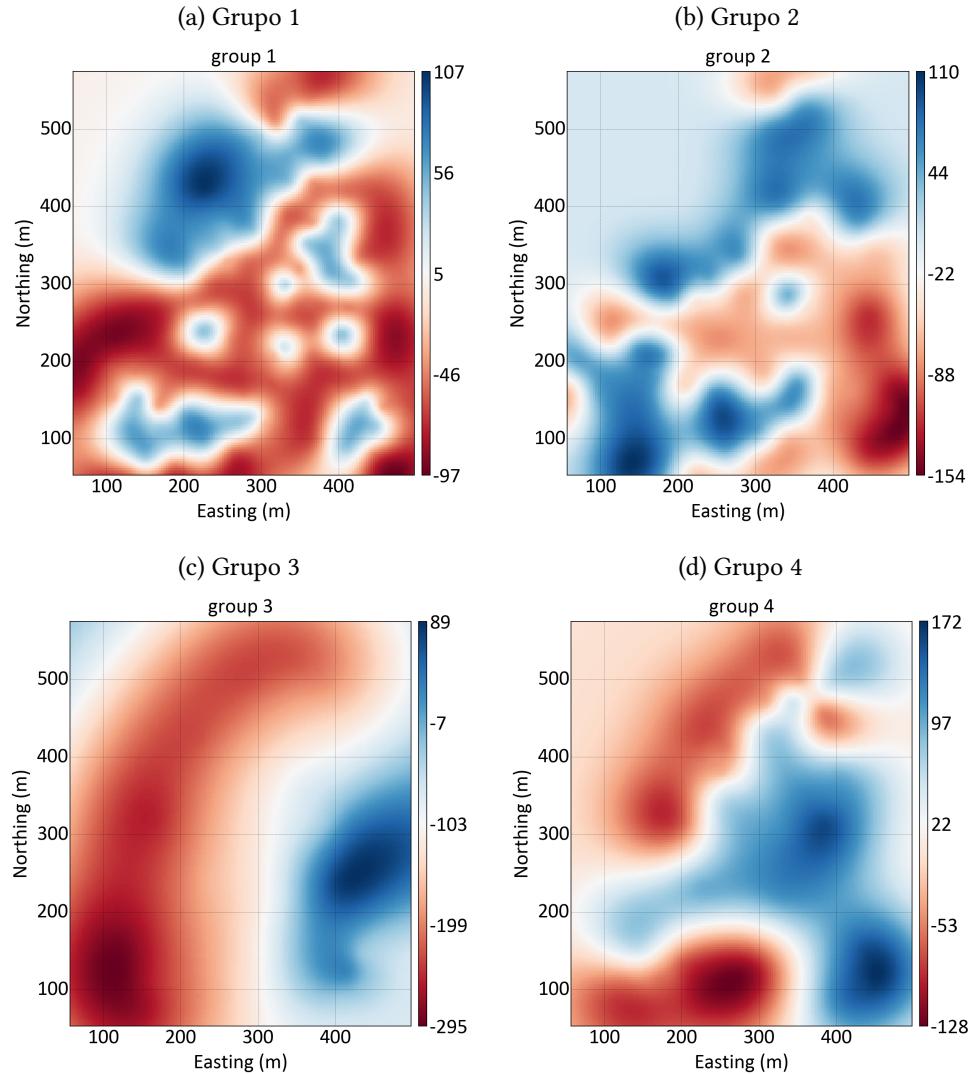
É necessário definir uma zona de incerteza em torno dos contatos entre os indicadores de cada grupo. Blocos localizados fora da zona são considerados como pertencentes a um determinado indicador, enquanto blocos dentro da zona terão sua incerteza avaliada. A zona de incerteza é obtida usando o mesmo parâmetro C apresentado na [subseção 2.2.8](#) para criar e controlar a largura da banda. O parâmetro C deve ser calibrado ou determinado empiricamente para cada grupo. Para o exemplo do *Swiss Jura*, um valor C de 12 metros foi escolhido para todos os quatro grupos.

As propriedades funções distância assinaladas devem ser calculadas, modificadas pelo parâmetro C de acordo com a [Equação 2.25](#), e interpoladas para cada grupo. A interpolação estima, para cada nó do *grid*, o valor da função de distância modificada. A zona de incerteza, para cada grupo, é determinada pelos blocos com valores estimados entre $-C$ e $+C$.

A figura [Figura 54](#) mostra as distâncias assinaladas modificadas interpoladas para cada grupo. A interpolação foi realizada em um *grid* de 3x3 metros por RBF usando *kernels*

Gaussianos equivalentes aos variogramas dos indicadores de cada grupo.

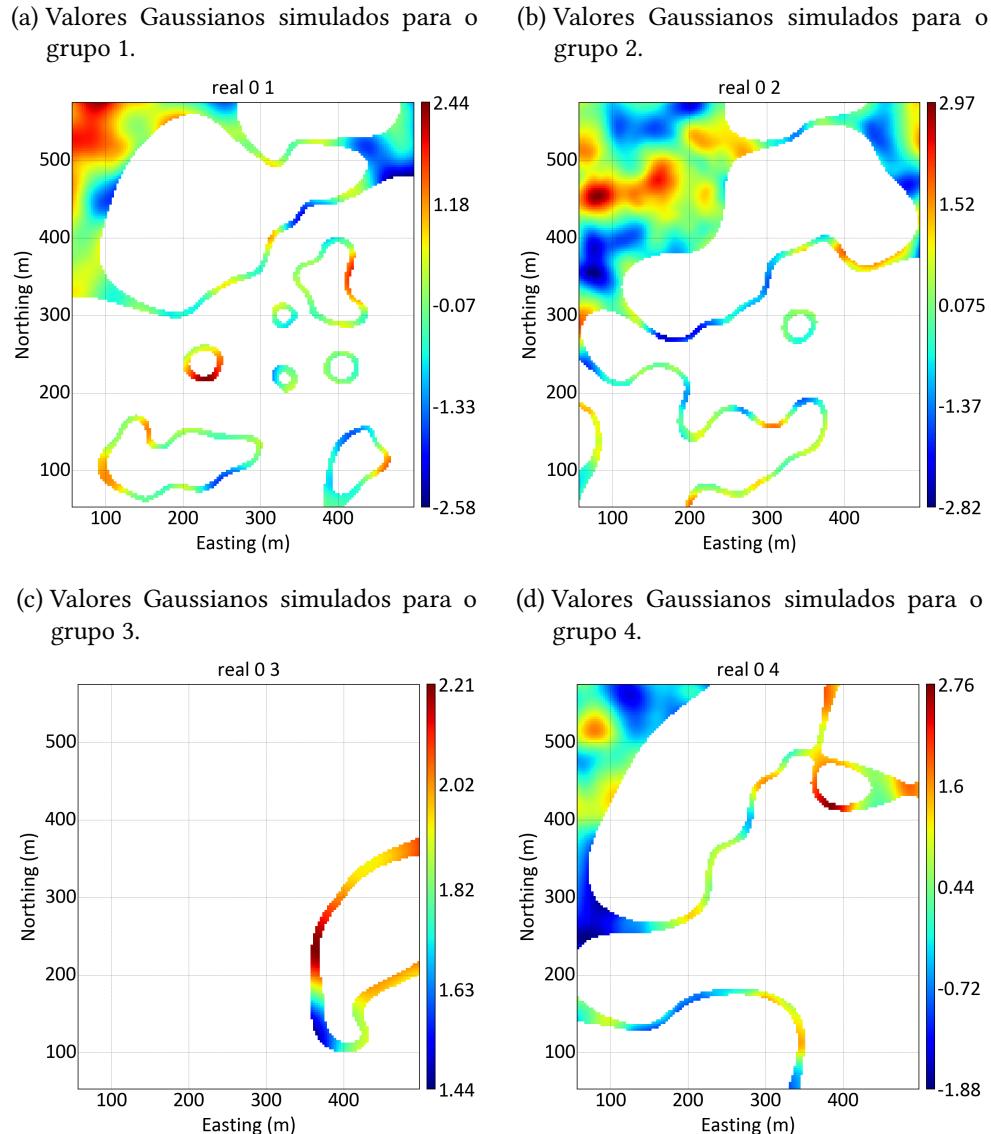
Figura 54 – Distâncias modificadas e interpoladas para cada um dos grupos.



A figura [Figura 55](#) apresenta uma realização, de um total de 10, dos valores Gaussianos simulados dentro da zona de incerteza para cada grupo. A zona de incerteza foi definida truncando os valores interpolados da [Figura 54](#) entre $-C$ e $+C$, nesse caso 12 metros.

Como um contato está sendo simulado, é recomendado o uso de um variograma Gausiano para a simulação não condicional, pois permite que a continuidade de curto alcance seja reproduzida ([WILDE; DEUTSCH, 2012](#)). [Martin e Boisvert \(2017a\)](#) advogam que o mesmo variograma usado para interpolar as distâncias assinaladas deve ser usado para simular os valores Gaussianos que serão usados para simular os contatos.

Figura 55 – Uma realização dos valores Gaussianos simulados, dentro da zona de incerteza, para cada grupo.

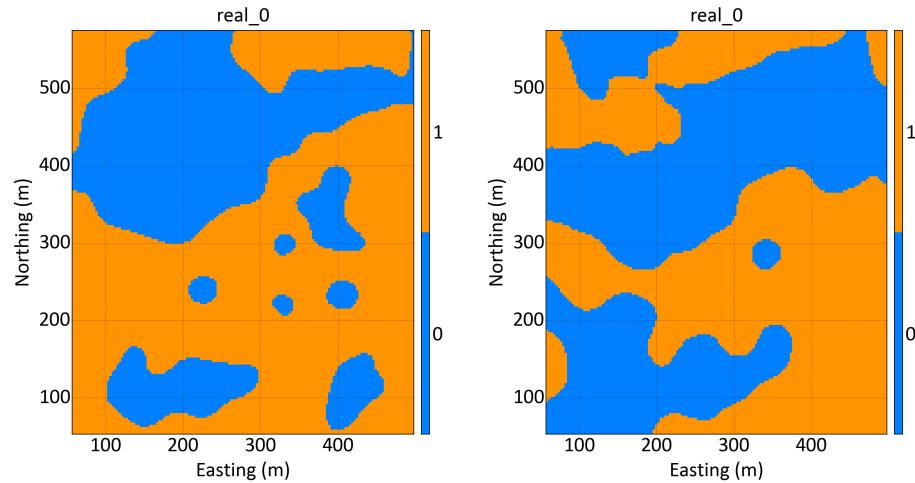


A classificação da localização não amostrada u como indicador 1 ou 0 é realizada comparando os valores interpolados $df^*(u)$ com os valores simulados $df'(u)$ de acordo com a Equação 2.27.

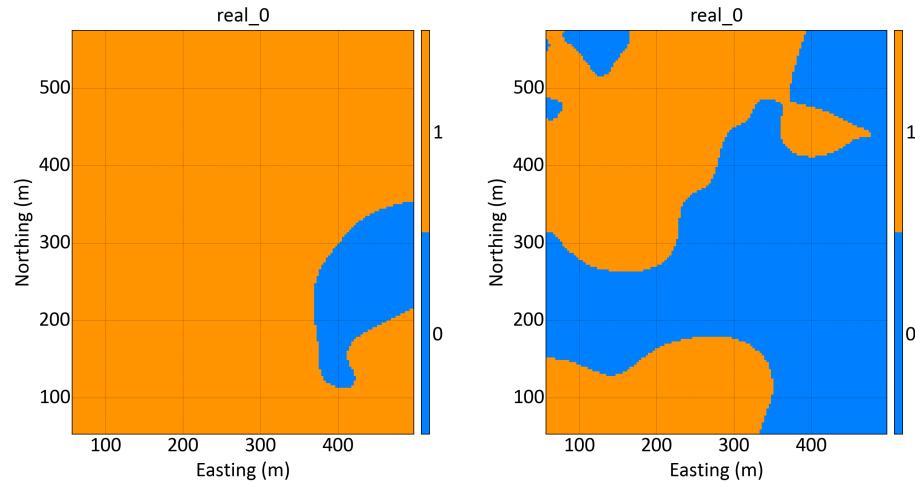
A figura Figura 56 mostra uma realização dos limites simulados para cada grupo.

Figura 56 – Uma realização da simulação de limites para todos os grupos.

(a) Simulação dos limites para o grupo 1. (b) Simulação dos limites para o grupo 2.



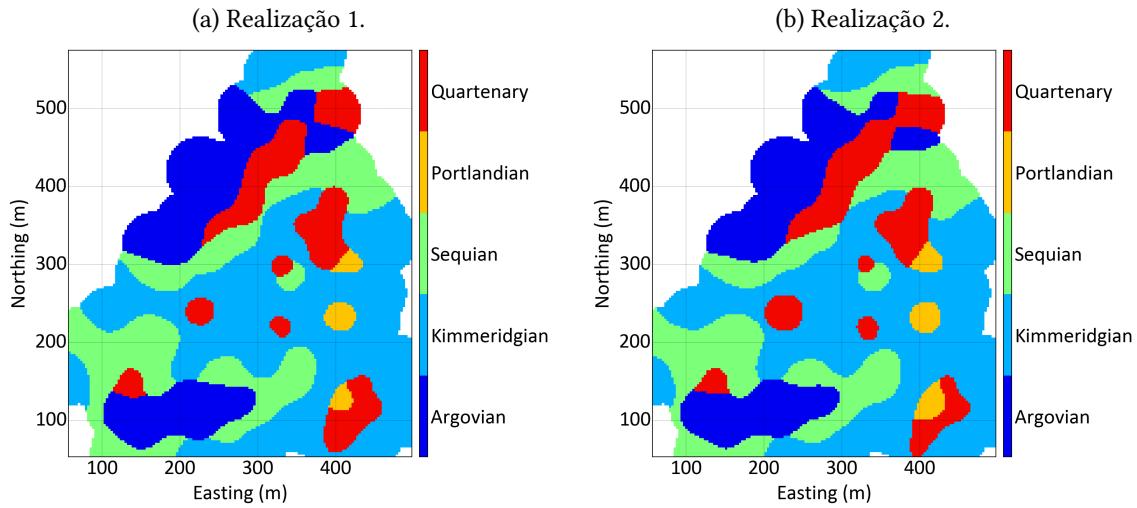
(c) Simulação dos limites para o grupo 3. (d) Simulação dos limites para o grupo 4.



As realizações de cada grupo devem ser sobrepostas para gerar uma realização para o modelo geológico. A categoria correspondente deve ser atribuída a cada nó do *grid* de acordo com as regras de hierarquização da Figura 52. Por exemplo, Kimmeridgian é atribuído a um nó do *grid* se duas condições forem satisfeitas: o bloco pertence à região do grupo 1, classificada com o indicador 1, e pertence à região classificada com o indicador 1 no grupo 2. Da mesma forma, Argovian é atribuído a um nó do *grid* se três condições forem satisfeitas: o bloco pertence à região do grupo 1, classificada com o indicador 0, o bloco pertence à região classificada com o indicador 1 no grupo 3, e o bloco pertence à região classificada com o indicador 1 em grupo 4.

A Figura 57 mostra duas, de 10 realizações, para o modelo geológico. Uma animação das 10 realizações pode ser vista [aqui](#).

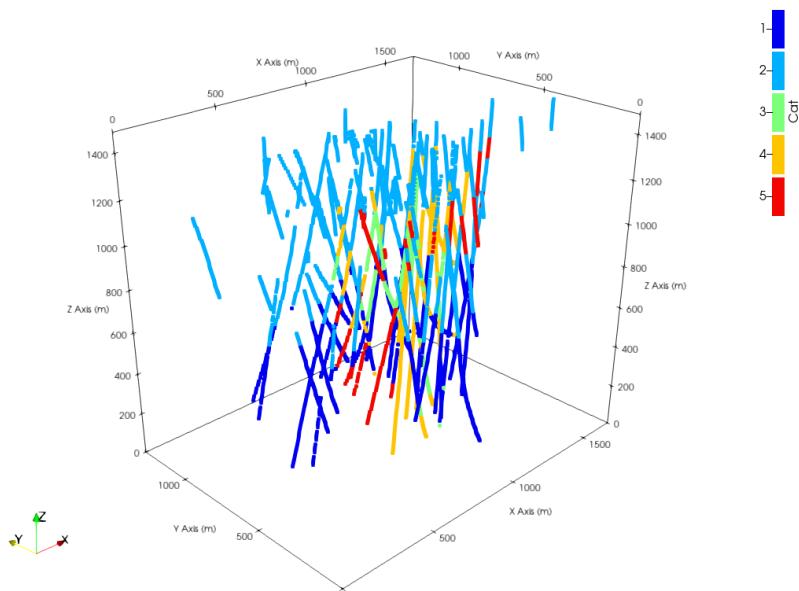
Figura 57 – Duas realizações para o modelo geológico para o Swiss Jura.



3.3.1 Estudo de caso

Esse estudo de caso foi realizado em um depósito de cobre pórfiro para demonstrar a eficiência da metodologia proposta. O depósito mineral consiste em cinco domínios geológicos mapeados com dados de 91 furos de sondagem, que foram compostados em intervalos de 15 metros, o que leva a 3.276 amostras de suporte pontual. Três domínios são rochas intrusivas, um domínio é de rochas oxidadas e o último é de rochas sulfetadas. A Figura 58 mostra uma vista em perspectiva dos furos compostados. As amostras em azul escuro representam rochas sulfetadas da categoria 1; o azul claro representa a categoria 2, rochas oxidadas; e verde, amarelo e vermelho representam as categorias 3, 4 e 5 de rochas intrusivas, respectivamente.

Figura 58 – Vista em perspectiva para o conjunto de dados no suporte pontual que descreve as cinco categorias.



O algoritmo de agrupamento proposto foi aplicado a um proto-modelo criado usando distâncias assinaladas a partir das amostras em suporte pontual do depósito de cobre pôrfiro em um *grid* de 50x50x30 metros. Como resultado, a regra de hierarquia apresentada na [Figura 59](#) é gerada. As amostras agrupadas são mostradas na [Figura 60](#).

Figura 59 – Regra de hierarquia para o conjunto de dados do estudo de caso definido pelo algoritmo proposto.

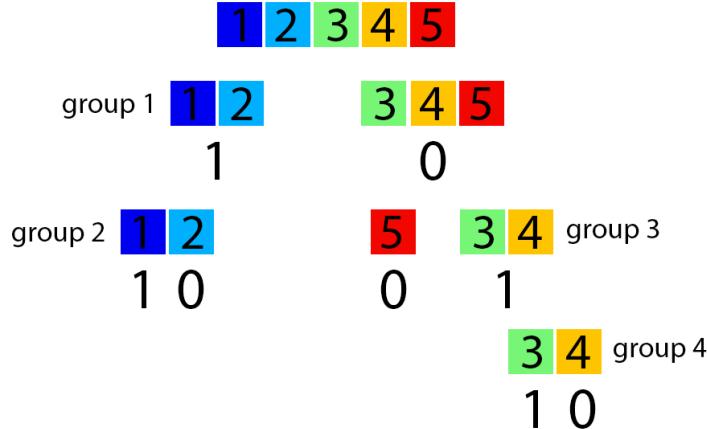
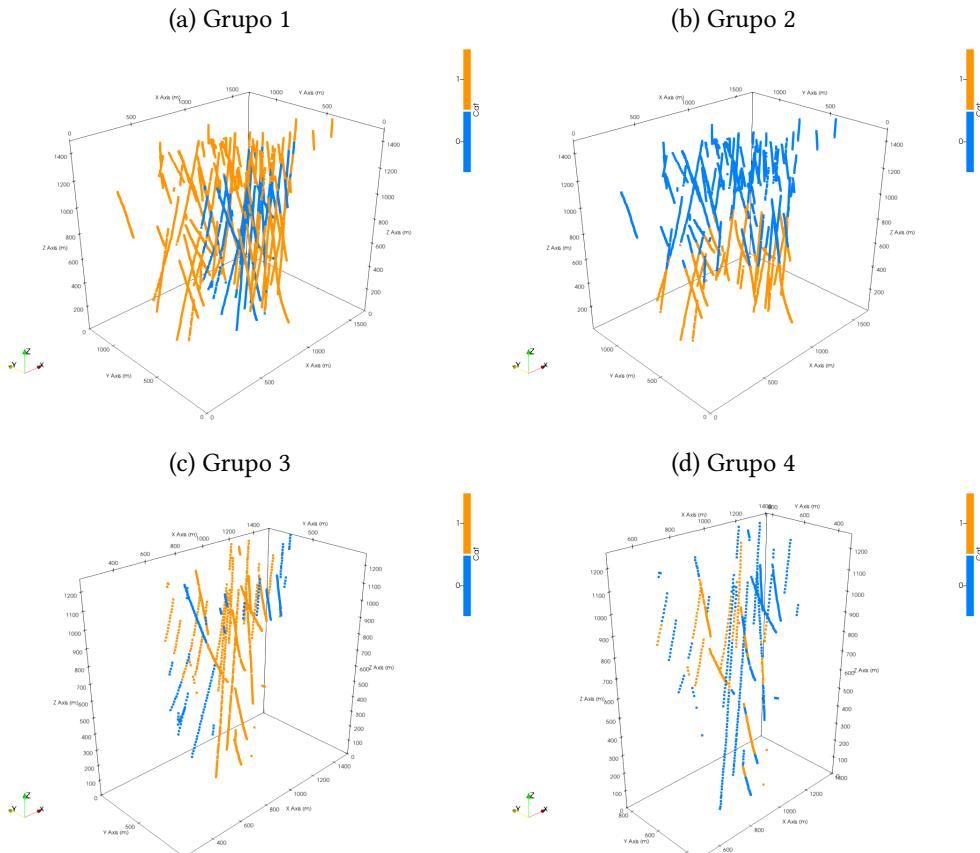


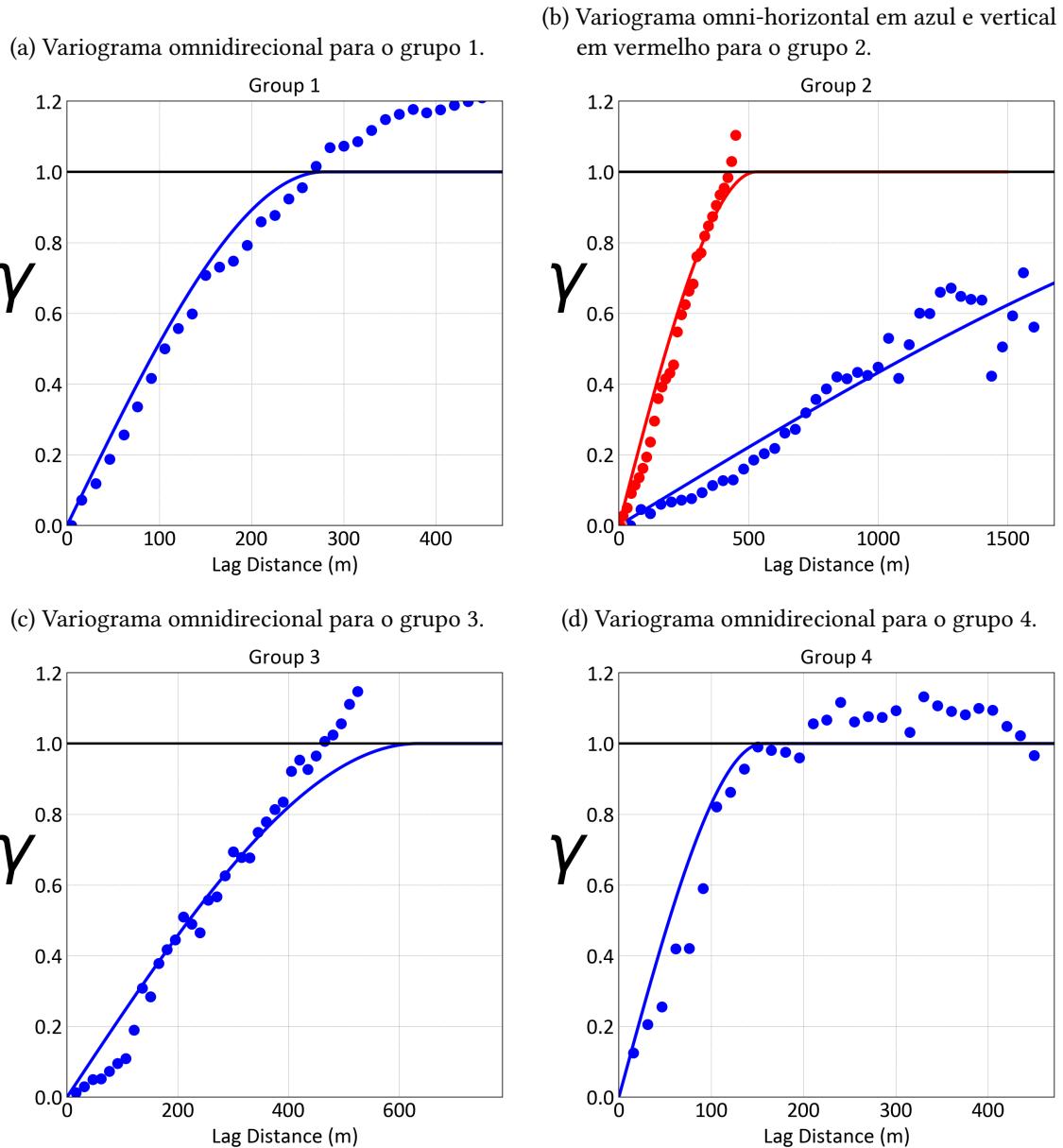
Figura 60 – Visão em perspectiva para os grupos de amostras em suporte pontual.



A [Figura 61](#) mostra os variogramas dos indicadores para cada grupo. Para os grupos

1, 3 e 4, um variograma omnidirecional foi calculado e modelado. Já para o grupo 2, foram calculados e modelados variogramas omni-horizontal (em azul) e vertical (em vermelho).

Figura 61 – Variogramas dos indicadores para cada grupo.



As equações: [Equação 3.6](#), [Equação 3.7](#), [Equação 3.8](#) e [Equação 3.9](#) representam os variogramas para os grupos 1, 2, 3 e 4 respectivamente.

$$\gamma(h) = 1.0Sph_1 \left[\frac{OMNI}{278.77m} \right] \quad (3.6)$$

$$\gamma(h) = 1.0Sph_1 \left[\frac{EW/NS}{3370.22m}, \frac{VERT}{531.09m} \right] \quad (3.7)$$

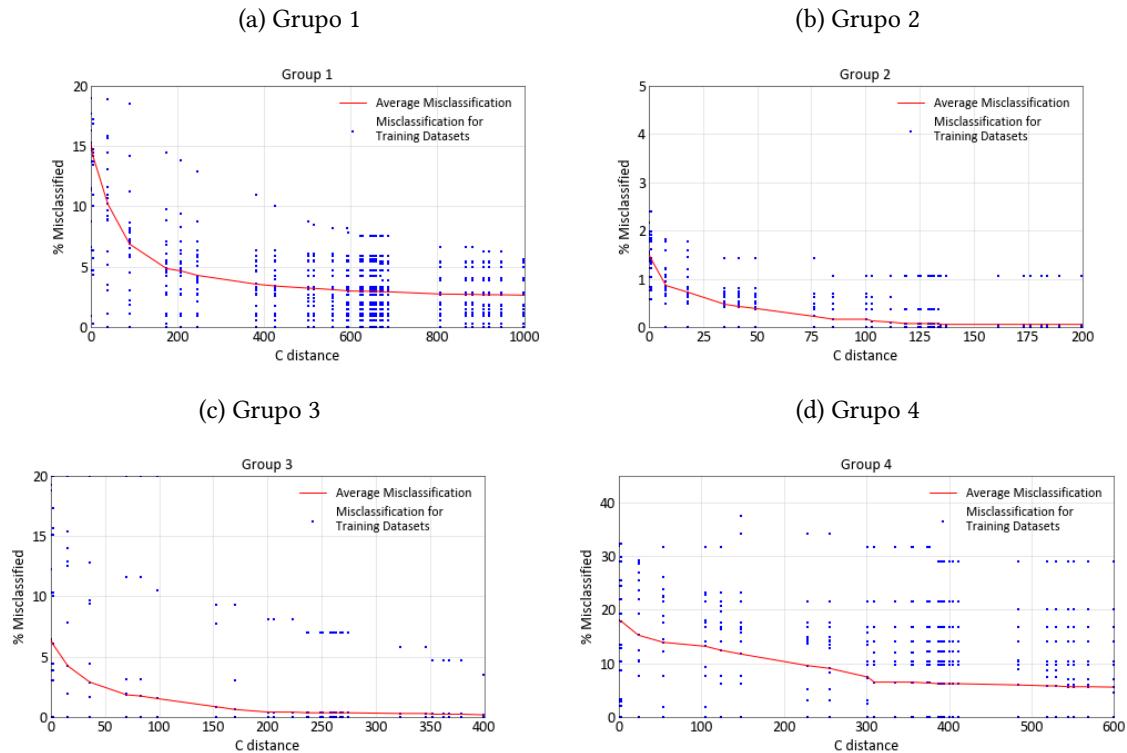
$$\gamma(h) = 1.0Sph_1 \left[\frac{OMNI}{634.57m} \right] \quad (3.8)$$

$$\gamma(h) = 1.0Sph_1 \left[\frac{OMNI}{156.34m} \right] \quad (3.9)$$

A calibração do parâmetro C foi realizada com 20 conjuntos de dados *jackknife* aleatórios para cada grupo, removendo 10% dos furos em cada execução, e com 30 parâmetros C aleatórios indo de zero a um valor em metros determinado pelo geomodelador para cada um dos quatro grupos. A calibração de C deve ser realizada para uma variedade de subconjuntos *jackknife* para garantir que a calibração seja robusta (WILDE; DEUTSCH, 2012). Os resultados são mostrados na Figura 62.

A diretriz para escolher um parâmetro C para cada grupo, a partir do índice de classificação incorreta nas calibrações mostradas na Figura 62, é considerar uma taxa de classificação incorreta aceitável. Valores razoáveis para os parâmetros C foram escolhidos para cada grupo com base nos resultados da calibração, o tipo de depósito e a configuração dos dados.

Os valores escolhidos foram 5%, 0,5%, 2,5% e 10% de taxas de classificação incorreta aceitáveis para os grupos 1, 2, 3 e 4, respectivamente, que correspondem aos valores dos parâmetros C de 170m, 30m, 50m e 200m. A escolha do parâmetro C da calibração *jackknife* é subjetiva, no entanto, a calibração auxilia o geomodelador na escolha de um valor adequado para o parâmetro. Um valor de C não deve ser escolhido na região onde a curva se torna horizontal. Além disso, a largura de banda da incerteza não pode ser maior que o espaçamento das sondagens (ROSSI; DEUTSCH, 2013).

Figura 62 – Calibração do parâmetro C por *jackknife*

As funções distância calculadas para cada grupo foram modificadas com o parâmetro C escolhido, adicionando C às amostras codificadas como 0 e subtraindo C das amostras codificadas como 1. As distâncias modificadas para cada grupo foram então interpoladas para todos os nós do *grid*. As funções de base radial foram escolhidas para interpolar a função distância para um *grid* com dimensões de bloco de 25x25x15 metros. O *kernel* RBF é Gaussiano com efeito pepita de 0,001% para evitar instabilidades nas matrizes. O suporte dos *kernels* é o alcance dos variogramas dos indicadores de cada grupo (Figura 61). As razões de anisotropia de cada grupo também foram consideradas.

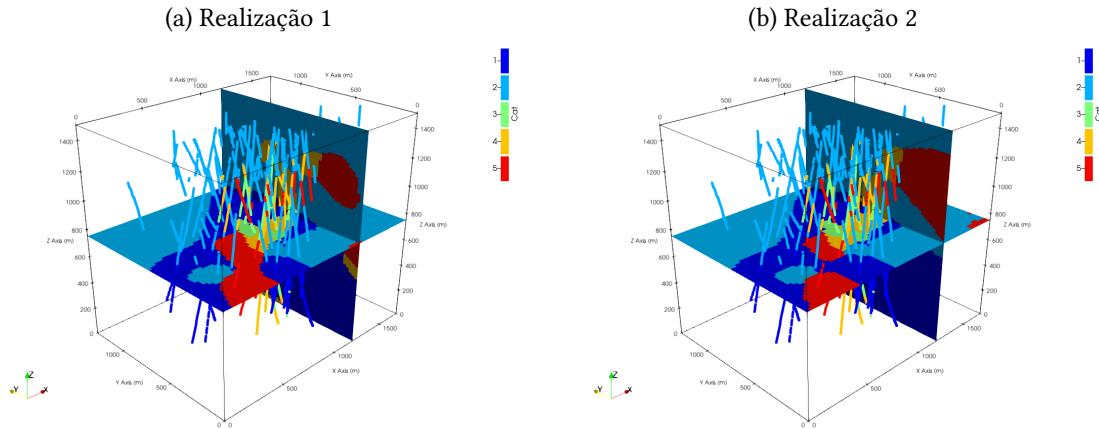
A zona de incerteza é obtida a partir dos modelos implícitos interpolados, truncando-os entre $-C$ e $+C$. Simulações não condicionais foram realizadas nesta região, e cada realização das simulações é transformada para que seus valores estejam entre $-C$ e $+C$ a partir da Equação 2.26. Equivalentes Gaussianos aos variogramas dos indicadores foram utilizados para simular cada um dos grupos. Dez realizações foram geradas para cada grupo comparando o valor simulado com os valores interpolados dentro da zona de incerteza.

As realizações para cada grupo devem ser sobrepostas, respeitando a regra de hierarquização da Figura 59, para gerar uma realização para o modelo geológico. Além disso, as realizações são mescladas respeitando seu índice. A realização 2 obtida na simulação de cada grupo é usada para construir a realização 2 do modelo geológico hierárquico.

A Figura 63 mostra uma seção vertical e uma horizontal para duas realizações do modelo geológico juntamente com os dados amostrais em suporte pontual. Uma animação mostrando

seções verticais em XZ das 10 realizações executadas pode ser vista [aqui](#).

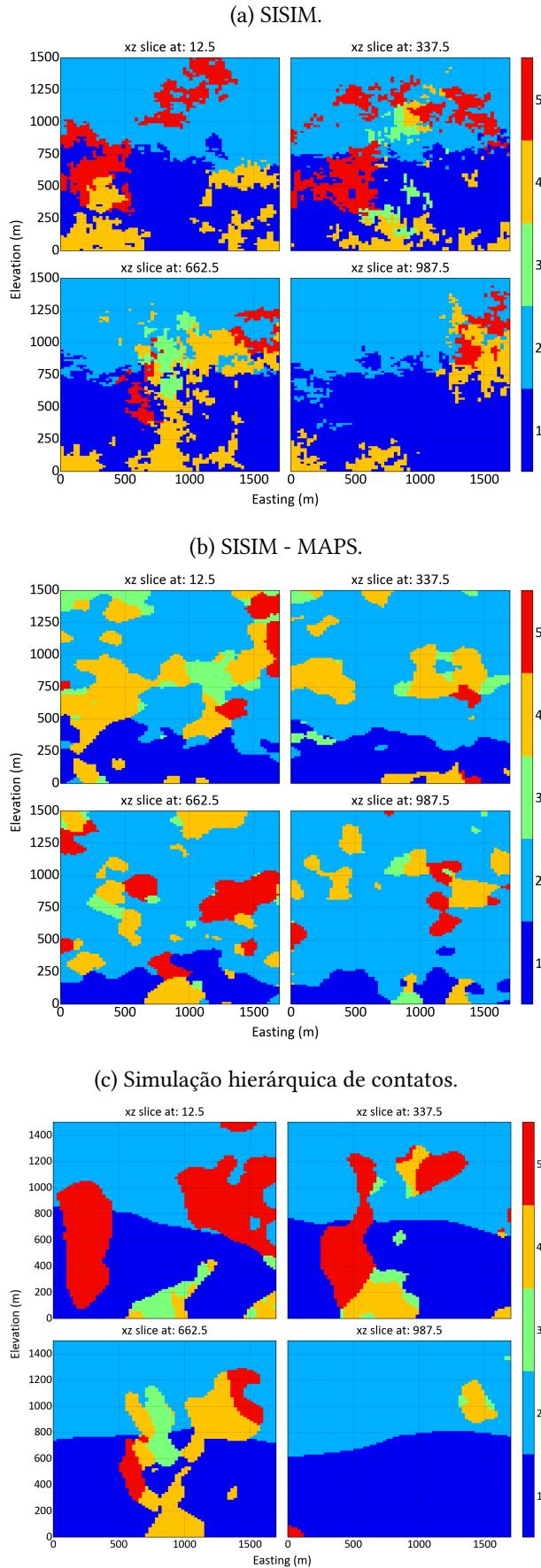
Figura 63 – Visão em perspectiva de seções verticais e horizontais de duas realizações do modelo geológico juntamente com as amostras em suporte pontual.



3.3.2 Discussão

A SISIM foi aplicada ao mesmo banco de dados, assim como na [seção 3.2](#), a fim de comparar os resultados com a metodologia proposta. Os variogramas dos indicadores para as cinco categorias foram calculados e modelados, os indicadores foram simulados para o mesmo *grid* em que a simulação de contatos hierárquica foi realizada. A Figura [64a](#) mostra as mesmas quatro seções verticais ao longo de XZ comparando uma realização da SISIM com a metodologia proposta. A SISIM gera um modelo com uma textura “salt and pepper” e padrões ruidosos também para esse estudo de caso, gerando estruturas que não são geologicamente realistas. A Figura [64b](#) mostra a SISIM pós processada com MAPS, quando a amostragem não é representativa e a proporção de cada categoria nas amostras não reflete as proporções do depósito. como é o caso do depósito de cobre em estudo, o MAPS, ao tentar reproduzir a proporção dos dados no modelo pode criar estruturas não realistas. A simulação hierárquica de contatos (Figura [64c](#)) gera limites contínuos, geologicamente realistas, que honram os dados, como mostra a matriz de confusão na [Figura 66](#). Além disso, apenas quatro variogramas dos indicadores, um para cada grupo, devem ser calculados e modelados, contra cinco variogramas para a SISIM. Como o RBF é um interpolador global, os parâmetros da vizinhança de busca não devem ser configurados, ao contrário da SISIM. Como desvantagens, assim como o algoritmo apresentado na [seção 3.2](#), esse algoritmo também não possui mecanismo para respeitar as proporções e reproduzir a continuidade espacial de cada categoria, embora isso possa ser contornado alterando interativamente os parâmetros e verificando os resultados.

Figura 64 – Seções equidistantes em XZ comparando uma realização do modelo geológico do SISIM com uma realização da metodologia proposta.



Uma forma quantitativa de avaliar a incerteza de modelos geológicos é usando a entropia de informação, essa técnica tem sido aplicada em geociências como uma medida de qualidade de modelos geológicos (YANG et al., 2019). A Figura 65 mostra a entropia da informação calculada para a SISIM e para a metodologia proposta. Como esperado, na metodologia proposta os maiores valores de entropia da informação estão em torno dos contatos, ao contrário da SISIM, onde altos valores de entropia estão espalhados no espaço. A maior parte da incerteza reside em áreas onde várias categorias têm uma alta probabilidade de ocorrência. Essa análise é necessária porque ao contrário do método apresentado na [seção 3.2](#), onde é criada uma zona de incerteza ao redor dos contatos entre as litologias. Nesse caso, a zona de incerteza é criada ao redor dos contatos dos grupos definidos, portanto, é necessário verificar se no modelo geológico a incerteza reside, de fato, ao redor dos contatos entre as litologias.

Figura 65 – Seções equidistantes em XZ comparando a entropia do SISIM com uma realização da metodologia proposta.

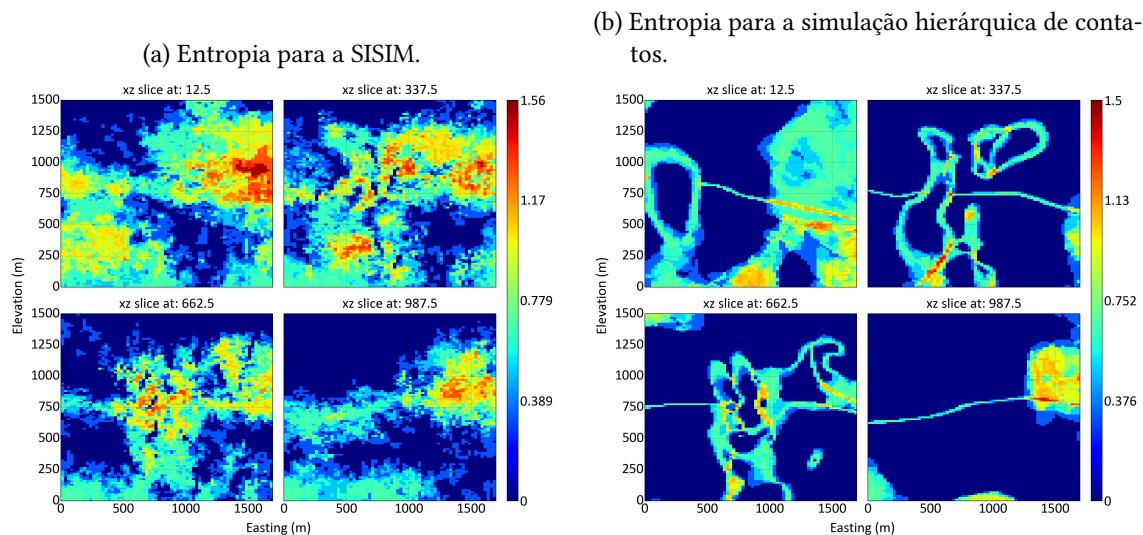
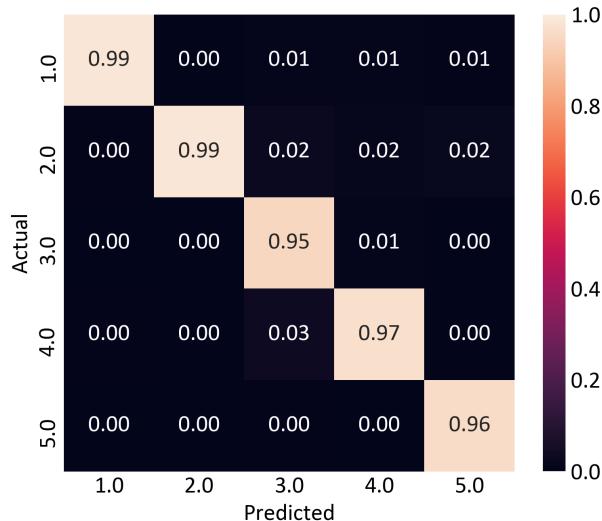


Figura 66 – Matriz de confusão para o depósito de cobre mostrando proporções das categoria preditas nas colunas e das categorias das amostras (real) nas linhas.



Os parâmetros do *grid* de simulação podem ser predeterminados pelos requisitos de engenharia para um determinado projeto. No entanto, as definições do *grid* para o modelo geológico não precisam corresponder às da modelagem numérica; a consideração mais importante é a capacidade de reproduzir as menores feições geológicas de interesse (MARTIN; BOISVERT, 2017b). O parâmetro C e a resolução do *grid* devem ser compatíveis, se o parâmetro C for maior que as dimensões da célula, a zona de incerteza pode ter poucos blocos ou até mesmo nenhum.

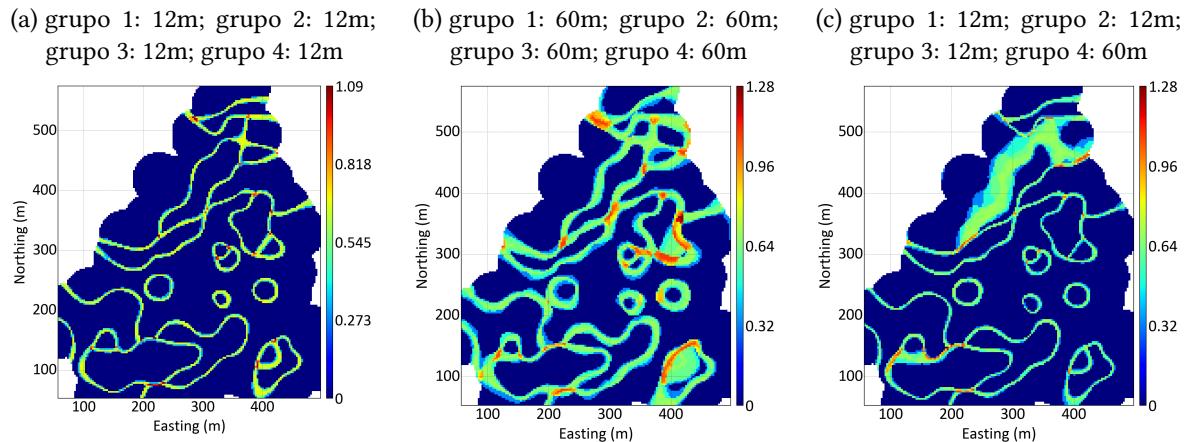
Os parâmetros do variograma de interpolação, a saber, alcance, anisotropia e efeito pepita, controlam a forma e o volume das diferentes categorias. Existe uma estreita relação entre o variograma utilizado para a interpolação de cada grupo e o parâmetros C. O mesmo valor para C pode levar a uma zona de incerteza maior se o alcance do variograma aumentar.

O comportamento dos contatos é controlado pelo variograma usado na simulação não condicional, alcances menores e efeitos pepita maiores geram limites ásperos e descontínuos, por outro lado, alcances longos e efeitos pepita pequenos geram limites suaves e contínuos.

A magnitude da incerteza pode ser controlada pelo parâmetro C, que define a largura da banda da incerteza. O parâmetro C pode ser calibrado a partir dos dados existentes ou determinado por um especialista, com base no conhecimento sobre o depósito. A largura da banda de incerteza não deve ser nem muito grande nem muito pequena, o verdadeiro limite entre domínios deve existir em algum local no interior da zona de incerteza. A variação do volume de cada litologia depende diretamente do parâmetro C definido para cada grupo. A Figura 67 mostra a entropia da informação calculada para o exemplo do *Swiss Jura* com diferentes valores para C. A Figura 67a mostra o caso base, onde os valores de C para todos os grupos são iguais a 12 metros, a Figura 67b mostra a entropia da informação para valores de C igual a 60 metros

para todos os grupos, a largura da banda da incerteza é maior do que o caso base. Finalmente, a Figura 67c mostra um caso onde o parâmetro C é igual a 12 metros para os grupos 1, 2 e 3 e 60 metros para o grupo 4. Como o grupo 4 representa o contato entre as categorias Argovian e Quaternary, a zona de incerteza é maior em torno desses contatos do que nos demais.

Figura 67 – Efeito do parâmetro C.



Vale ressaltar que a principal desvantagem do método consiste em determinar as regras de hierarquia. O algoritmo proposto automatiza o processo, entretanto, a resolução do *grid* do proto-modelo e o método utilizado para criar o proto-modelo podem produzir diferentes agrupamentos. A hierarquia entre as categorias pode ser determinada cronologicamente por um geomodelador. No entanto, essa escolha depende do conhecimento sobre o depósito mineral e da experiência do geomodelador. Também pode ser determinado de forma arbitrária, o que pode levar a modelos geológicos subjetivos, irreproduzíveis e não auditáveis.

3.4 Conclusão

Dos três métodos propostos e investigados nessa tese para avaliação de incerteza em modelos geológicos usando funções distância assinaladas, dois são baseados em simulações gaussianas não condicionais e um deles não envolve simulação, apenas interpolações.

Por ser baseado em interpolação global, o método baseado na parametrização do suporte do *kernel* gera modelos que apresentam contatos suaves que são geologicamente realistas. Os outros dois métodos baseados em simulação, apesar de apresentarem semelhanças, têm características distintas. A avaliação de incerteza usando campos de probabilidade e funções distância assinaladas é consideravelmente mais simples do que a simulação hierárquica de contatos. Entretanto, apresenta algumas limitações e desvantagens. Ao contrário da simulação hierárquica de contatos, a zona de incerteza deve ter a mesma espessura ao redor de todos os contatos. Além disso, também ao contrário da simulação hierárquica dos contatos, um único variograma para a simulação não condicional deve representar os contatos entre todas

as diferentes categorias do depósito mineral. Embora no estudo de caso apresentado nessa tese os contatos gerados pelos campos potenciais tenham sido suaves e realistas, algumas combinações de parâmetros podem gerar contatos desconexos e irreais. O mesmo não acontece com a simulação hierárquica de contatos, a comparação do valor simulado com o interpolado bloco a bloco faz com que os contatos sejam contínuos tornando o método robusto no sentido de criar contatos realistas.

O uso de uma zona de incerteza combinada como uma simulação não condicional, que apresente continuidade espacial, torna os métodos propostos mais fiéis no sentido de gerar estruturas geológicas realistas quando comparados à SISIM, que vai simular todos os nós do *grid*. A SISIM, quando pós processada usando MAPS, gera formas realistas em depósitos pouco erráticos e com alta densidade amostral, como foi demonstrado no estudo de caso para o depósito de ouro da [subseção 3.2.1](#), onde a SISIM - MAPS produziu formas que não divergem significativamente das formas interpretadas pelo geomodelador ou produzidas pela metodologia proposta. Já em depósitos com maior erraticidade e menor densidade amostral, como é o caso do depósito de cobre da [subseção 3.3.1](#), a definição de uma zona de incerteza ao redor dos contatos se mostra vantajosa no sentido de gerar estruturas realistas tendo em vista que os modelos gerados pela metodologia proposta apresentam estruturas com menor variabilidade e maior continuidade espacial em relação às geradas pela SISIM - MAPS. Entretanto, é importante frisar que diferentes metodologias devem ser aplicadas para diferentes objetivos. Se as proporções das categorias observadas nos dados e a continuidade espacial dos indicadores deve ser reproduzida, a SISIM deve ser aplicada. Se há restrição na posição relativa das categorias, algum método da família Gaussiana truncada deve ser aplicado. Se houver informação estrutural e a posição real dos contatos em subsuperfície, o método dos campos potenciais deve ser utilizado. E assim por diante...

Embora os métodos propostos criem modelos realistas de forma rápida e direta, eles não devem ser aplicados a todos os tipos de depósito de forma indiscriminada. Nenhum dos métodos leva em consideração elementos estruturais da geologia e também não dão tratamento diferenciados para estruturas específicas como dobrões, diques, falhas ou lentes. A reprodução desse tipo de estrutura em três dimensões por métodos de modelagem geológica automáticos geralmente não é satisfatória. Entretanto, os métodos propostos podem ser aplicados em fases iniciais da pesquisa mineral, para bancos de dados em 2D, para depósitos de baixa complexidade geológica, para depósitos de média complexidade geológica e alta densidade amostral ou podem ainda ter suas realizações refinadas manualmente por um geomodelador experiente poupano tempo e esforço do profissional.

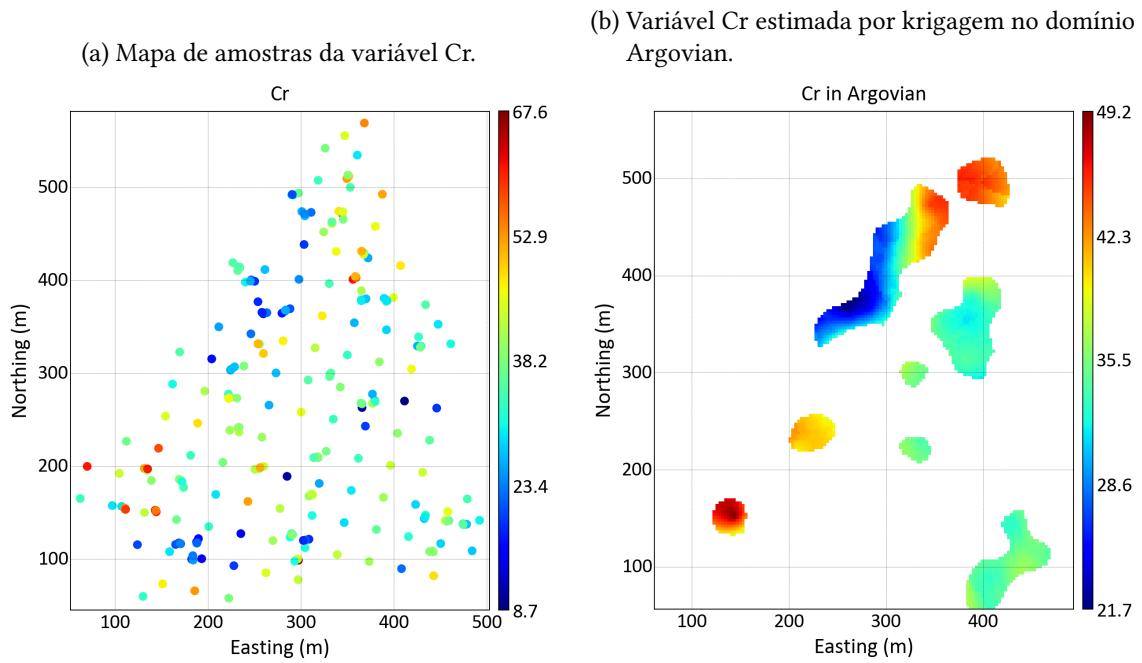
A validação de simulações para modelo geológicos multi-categóricos é especialmente difícil. Nessa tese apenas a reprodução das amostras e checagem visual de seções foram aplicadas. Caso haja confiança de que a amostragem seja representativa das proporções de cada uma das litologias do depósito mineral deve-se verificar a reprodução das proporções. A reprodução

dos variogramas dos indicadores para todas as categorias de um modelo multi-categórico de forma simultânea é um evento raro mesmo em bancos de dados de geometria simples em 2D. As estruturas geológicas, muitas vezes, são curvilineares e sua continuidade espacial não pode ser capturada por variogramas "tradicionais". A modelagem implícita tende a gerar estruturas em formatos esféricos/elípticos ("blobs") ao redor das amostras, e em forma de salsicha em torno de furos de sondagem. O surgimento desse tipo de estrutura indica que os modelos não são satisfatórios e alguma mudança nos parâmetros ou método utilizado para construção dos modelos deve ser feita.

Os métodos geoestatísticos de estimativa e simulação de teores requerem a delimitação prévia de domínios estacionários para modelagem do depósito mineral. Na indústria da mineração, modelos geológicos determinísticos são amplamente utilizados. No entanto, esses modelos não permitem a avaliação da incerteza em volumes e toneladas. A geometria dos domínios geológicos afeta diretamente a avaliação dos recursos minerais, pois determina os volumes de minério e estéril disponíveis e pode ser responsável pelas maiores incertezas em um projeto de mineração. Assim, o impacto é significativo na estimativa da tonelagem de minério, levando a estimativas enviesadas, erros no planejamento de lavra, custos inesperados na operação e desvios nas receitas esperadas ([SRIVASTAVA, 2005](#)).

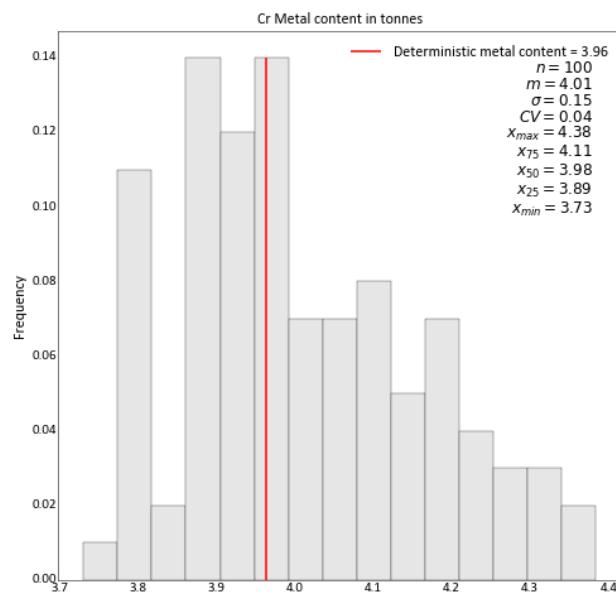
Além da variável categórica, o banco de dados *Swiss Jura* apresenta seis variáveis contínuas que representam o teor de diferentes metais (Cr, Ni, Zn, Cd, Cu e Pb) em ppm. A [Figura 68](#) mostra o mapa de amostras para a variável Cr em [68a](#) e mostra em [68b](#) os teores de Cr em ppm estimados por krigagem ordinária no domínio Argovian definido na [Figura 11](#). A linha vermelha da [Figura 69](#) mostra o conteúdo metálico de Cr considerando a espessura em Z dos blocos como um metro e a densidade 5g/cm³ para simplificar os cálculos.

Figura 68 – Variável Cr do Swiss Jura



Foram realizadas 100 realizações para o modelo geológico do *Swiss Jura* usando a metodologia apresentada na [seção 3.2](#). O conteúdo metálico de Cr foi calculado para cada uma das realizações para o domínio Argovian e o resultado é mostrado no histograma da [Figura 69](#).

Figura 69 – Análise de risco para o conteúdo metálico de Cr no domínio Argovian.



A partir da análise de risco em relação ao conteúdo metálico de Cr, levando em considera-

ção somente a incerteza em relação ao modelo geológico, no domínio Argovian há possibilidade de que haja disponível um mínimo de 3.7 toneladas de Cr e um máximo de 4.3 toneladas de Cr. Uma variação percentual de 17,5% em massa que deve ser levada em consideração na fase de planejamento mineiro. Ainda há a incerteza relacionada aos teores de Cr nos blocos, que não foi levada em conta nesse estudo, e deve ser somada à incerteza volumétrica do modelo geológico para se chegar à incerteza final do conteúdo metálico.

4 OBSERVAÇÕES FINAIS

Existem diversos métodos disponíveis para modelagem geológica automática e avaliação de incerteza em modelos geológicos, cada um deles com suas características e particularidades. Não obstante, a modelagem geológica implícita com funções distância assinaladas tem se tornado extremamente popular ao longo da última década, ganhando espaço nos principais softwares de mineração. Por esse motivo, essa tese se propôs a desenvolver e investigar métodos para avaliação de incerteza em modelos geológicos multi-categóricos usando funções distância assinaladas. Tendo em vista que o modelo geológico pode ser o responsável por incertezas cruciais em um projeto mineiro, a incerteza associada ao modelo geológico não deve ser negligenciada.

Nesse sentido, foram propostos três métodos: um baseado em estimativa global e dois baseados em simulações não condicionais no interior de uma zona de incerteza. A tese apresentou os métodos a partir de um conceito de prova no conhecido banco de dados *Swiss Jura*, e posteriormente, conduziu um estudo de caso em um banco de dados, sintético (simulando uma situação real) ou oriundo de um depósito mineral real, a fim de verificar a competência dos métodos propostos. Os resultados foram discutidos e comparados com métodos concorrentes. As vantagens, desvantagens e aplicabilidade de cada método foram apontadas.

4.1 Sumário das contribuições

- **Parametrização do suporte do *kernel*:** o método avalia a incerteza associada ao modelo geológico "simulando" diferentes interpretações para cada uma das litologias, contraindo ou expandindo seu volume ou área. O método é baseado em interpolação global e por isso gera formas geológicas suaves e realistas. A implementação do método ainda necessita de aprimoramentos e otimização de *software*.
- **Avaliação de incerteza usando funções distância assinaladas e campos de probabilidades:** o método é bastante simples e cômodo já que dissocia a simulação dos contatos do cálculo das probabilidades locais. As probabilidades locais precisam ser calculadas para a confecção dos modelos determinísticos, sendo assim, simular os contatos requer apenas a definição da zona de incerteza e simulações não condicionais Gaussianas. Em contrapartida, há limitações, já que a extensão da zona de incerteza e as simulações não condicionais devem ser as mesmas para todas as diferentes litologias. A combinação dos parâmetros ω e alcance do variograma das simulações pode gerar estruturas descontínuas.
- **Simulação hierárquica dos contatos:** a simulação hierárquica dos contatos, assim como a avaliação de incerteza usando funções distância assinaladas e campos de probabilidades,

também é baseada em simulações não condicionais no interior de uma zona de incerteza. Porém, é um método mais laborioso, já que exige que definição de grupos, a definição de zonas de incertezas para cada um dos grupos e posterior simulações Gaussianas não condicionais para cada um dos grupos. Todavia, isso também faz com que o método seja mais flexível já que é possível aplicar diferentes zonas de incertezas e diferentes simulações para cada um dos grupos. O método é robusto no sentido de gerar estruturas realistas já que a comparação entre distâncias interpoladas e valor simulado bloco a bloco gera limites contínuos. A definição dos grupos traz subjetividade ao método. Mesmo a implementação automática, apesar de poupar tempo de trabalho do geomodelador, depende de um proto-modelo: diferentes proto-modelos geram diferentes regras hierárquicas.

- **Software:** foi desenvolvida uma suite de *plugins* para o software geostatístico *AR2GeMS* que implementa todos os processos necessários para a modelagem geológica com distâncias assinaladas ou indicadores incluindo: cálculo das distâncias ou indicadores; criação de modelos determinísticos a partir das distâncias assinaladas, dos indicadores ou usando suporte de vetores de máquina (SVM); avaliação de incerteza dos modelos a partir da parametrização do *kernel* e usando campos de probabilidades; ferramentas auxiliares para criação automática de *grids* e exportação para visualização em softwares especializados. O método simulação hierárquica de contatos, por ser bastante dependente de ferramentas já implementadas na biblioteca *GSLib*, foi implementado a partir de *jupyter notebooks* e módulos auxiliares em Python disponibilizados em um repositório *git*. O [Apêndice A](#) apresenta o manual de instalação e de uso dos *softwares* desenvolvidos.

4.2 Sugestões para trabalhos futuros

Dada a limitação de tempo para conclusão da tese, algumas ideias e aprimoramentos foram deixados para trabalhos futuros.

- Definir uma metodologia objetiva para determinar os variogramas das simulações não condicionais nos métodos dos campos potenciais e simulação hierárquica dos contatos;
- Definir uma metodologia objetiva para calibrar a zona de incerteza nos métodos dos campos potenciais e simulação hierárquica dos contatos. A zona de incerteza não necessariamente deve ser simétrica ao redor dos contatos. Não deve haver amostras no interior da zona de incerteza;
- Aprimorar a parametrização do suporte do *kernel* para que a sensibilidade ao parâmetro *fmin* seja reduzida. A implementação computacional deve ser otimizada;
- Conduzir uma análise de risco mais robusta com simulação de teores e do modelo geológico em um banco de dados real comparando os resultados com os teores estimados por krigagem em um modelo geológico determinístico;

- Desenvolver uma metodologia para validação dos modelos geológicos. Usar redes neurais treinadas em seções de modelos determinísticos para identificar estruturas contínuas e realistas é uma avenida de investigação;
- Desenvolver metodologias, ou adaptações de metodologias existentes, para modelar estruturas geológicas específicas (dobras, diques, falhas, lentes) e avaliar sua incerteza.

REFERÊNCIAS

- AILLERES, L. et al. *Uncertainty in 3D models and implicit modeling of poly-deformed terranes*. 2015. Presentation. Disponível em: <http://www.tectonique.net/tectask/3DIGI/presentations/T1.2_ailleres_Goodbye2DEarthPresentation.pdf>. Acesso em: 23-07-2018. Citado na página 20.
- ALABERT, F. *Stochastic imaging of spatial distributions using hard and soft information*. Tese (Doutorado) – Stanford University Press, 1987. Citado 2 vezes nas páginas 19 e 43.
- AMARANTE, F.; ROLO, R.; COSTA, J. Incerteza associada ao modelamento implícito e hierárquico. In: CBMINA 2019. [S.l.]: CBMINA, 2018. p. xx-xx. Citado na página 78.
- AMARANTE, F. A. N.; ROLO, R. M.; COSTA, J. F. C. L. Assessing geologic model uncertainty-a case study comparing methods. *REM-International Engineering Journal*, SciELO Brasil, v. 72, n. 4, p. 643–653, 2019. Citado na página 78.
- AMARANTE, F. A. N.; ROLO, R. M.; COSTA, J. F. C. L. Boundary simulation—a hierarchical approach for multiple categories. *Applied Earth Science*, Taylor & Francis, p. 1–17, 2021. Citado 2 vezes nas páginas 56 e 78.
- ARMSTRONG, M. et al. *Plurigaussian simulations in geosciences*. [S.l.]: Springer Science & Business Media, 2011. Citado na página 44.
- AUÑÓN, J.; GÓMEZ-HERNÁNDEZ, J. J. Dual kriging with local neighborhoods: Application to the representation of surfaces. *Mathematical Geology*, Springer, v. 32, n. 1, p. 69–85, 2000. Citado na página 30.
- BEATSON, R. K.; CHERRIE, J. B.; MOUAT, C. T. Fast fitting of radial basis functions: Methods based on preconditioned gmres iteration. *Advances in Computational Mathematics*, Springer, v. 11, n. 2-3, p. 253–270, 1999. Citado na página 35.
- BLOOMENTHAL, J.; BAJAJ, C. *Introduction to implicit surfaces*. [S.l.]: Morgan Kaufmann, 1997. Citado na página 19.
- BOISVERT, J. Automatic geological modeling with iterative refinement. *CCG Annual Report*, 2013. Citado na página 27.
- BOUCHER, A. Considering complex training images with search tree partitioning. *Computers & Geosciences*, Elsevier, v. 35, n. 6, p. 1151–1158, 2009. Citado na página 44.
- CÁCERES, A. et al. Stochastic geological modeling using implicit boundary simulation. In: *Proceedings of the 2nd International Seminar on Geology for the Mining Industry GEOMIN*. [S.l.: s.n.], 2011. p. 21. Citado 2 vezes nas páginas 16 e 49.
- CALCAGNO, P. et al. Geological modelling from field data and geological knowledge: Part i. modelling method coupling 3d potential-field interpolation and geological rules. *Physics of the Earth and Planetary Interiors*, Elsevier, v. 171, n. 1-4, p. 147–157, 2008. Citado na página 20.
- CAUMON, G.; TERTOIS, A.-L.; ZHANG, L. Elements for stochastic structural perturbation of stratigraphic models. In: EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS. *EAGE Conference on Petroleum Geostatistics*. [S.l.], 2007. p. cp-32. Citado na página 44.

- CHILÈS, J.-P. et al. Modelling the geometry of geological units and its uncertainty in 3d from structural data: the potential-field method. In: *Proceedings of International Symposium on Orebody Modelling and Strategic Mine Planning, Perth, Australia*. [S.l.: s.n.], 2004. v. 22, p. 24. Citado na página [20](#).
- COWAN, E. et al. Rapid geological modelling. *Applied Structural Geology for Mineral Exploration and Mining, International Symposium*, 2002. Citado 2 vezes nas páginas [19](#) e [27](#).
- COWAN, E. et al. Practical implicit geological modelling. In: *Fifth International Mining Geology Conference*. [S.l.: s.n.], 2003. p. 17–19. Citado 3 vezes nas páginas [14](#), [15](#) e [19](#).
- DEUTSCH, C. A statistical resampling program for correlated data. *CCG Annual Report*, 2004. Citado na página [47](#).
- DEUTSCH, C. V. Cleaning categorical variable (lithofacies) realizations with maximum a-posteriori selection. *Computers & Geosciences*, v. 24, n. 6, p. 551–562, 1998. ISSN 0098-3004. Citado na página [72](#).
- DEUTSCH, C. V. A sequential indicator simulation program for categorical variables with point and block data: Blocksis. *Computers & Geosciences*, Elsevier, v. 32, n. 10, p. 1669–1681, 2006. Citado na página [43](#).
- DEUTSCH, C. V.; JOURNEL, A. G. Geostatistical software library and user's guide. *New York*, v. 119, n. 147, 1992. Citado na página [17](#).
- FASSHAUER, G. E. *Meshfree approximation methods with MATLAB*. [S.l.]: World Scientific, 2007. v. 6. Citado 2 vezes nas páginas [30](#) e [31](#).
- FROIDEVAUX, R. Probability field simulation. In: *Geostatistics tróia'92*. [S.l.]: Springer, 1993. p. 73–83. Citado na página [65](#).
- GONÇALVES, Í. G.; KUMAIRA, S.; GUADAGNIN, F. A machine learning approach to the potential-field method for implicit modeling of geological structures. *Computers & Geosciences*, Elsevier, v. 103, p. 173–182, 2017. Citado na página [44](#).
- GOOVAERTS, P. *Geostatistics for natural resources evaluation*. [S.l.]: Oxford University Press, 1997. Citado 2 vezes nas páginas [18](#) e [21](#).
- GUARDIANO, F. B.; SRIVASTAVA, R. M. Multivariate geostatistics: beyond bivariate moments. In: *Geostatistics Troia'92*. [S.l.]: Springer, 1993. p. 133–144. Citado na página [44](#).
- HOSSEINI, A.; DEUTSCH, C. V. A distance function based algorithm to quantify uncertainty in areal limits. *CCG Annual Report*, 2007. Citado na página [27](#).
- ISAAKS, E. H.; SRIVASTAVA R, M. *Applied geostatistics*. [S.l.]: Oxford University Press, 1989. Citado na página [30](#).
- JOURNEL, A. G.; HUIJBREGTS, C. J. *Mining geostatistics*. [S.l.]: Academic press, 1978. Citado na página [14](#).
- KARPEKOV, T.; DEUTSCH, C. A review of boundary modeling and one way of selecting the c parameter to model boundaries with uncertainty. *Center for Computational Geostatistics Annual Report*, v. 18, 2016. Citado na página [52](#).

- KLOECKNER, J. et al. Modelagem geológica implícita automática via tabela de covariância com funções distância assinaladas. In: *Anais do IX Congresso Brasileiro de Mineração, Belo Horizonte*. [S.l.: s.n.], 2018. p. xx-xx. Citado na página [24](#).
- LAJAUNIE, C.; COURRIOUX, G.; MANUEL, L. Foliation fields and 3d cartography in geology: principles of a method based on potential interpolation. *Mathematical Geology*, Springer, v. 29, n. 4, p. 571–584, 1997. Citado na página [20](#).
- LILLAH, M.; BOISVERT, J. B. Stochastic distance based geological boundary modeling with curvilinear features. *Mathematical Geosciences*, Springer, v. 45, n. 6, p. 651–665, 2013. Citado na página [51](#).
- LILLAH, M.; BOISVERT, J. B. Inference of locally varying anisotropy fields from diverse data sources. *Computers & Geosciences*, Elsevier, v. 82, p. 170–182, 2015. Citado na página [39](#).
- LINDSAY, M. D. et al. Locating and quantifying geological uncertainty in three-dimensional models: Analysis of the gippsland basin, southeastern australia. *Tectonophysics*, Elsevier, v. 546, p. 10–27, 2012. Citado na página [44](#).
- LORENSEN, W. E.; CLINE, H. E. Marching cubes: A high resolution 3d surface construction algorithm. In: ACM. *ACM siggraph computer graphics*. [S.l.], 1987. v. 21, n. 4, p. 163–169. Citado na página [36](#).
- MALLET, J.-L. Space–time mathematical framework for sedimentary geology. *Mathematical geology*, Springer, v. 36, n. 1, p. 1–32, 2004. Citado na página [20](#).
- MANCELL, S.; DEUTSCH, C. The problem with signed distance function. *Center for Computational Geostatistics Annual Report*, v. 21, 2019. Citado 3 vezes nas páginas [41](#), [42](#) e [43](#).
- MANCHUCK, J.; DEUTSCH, C. Geometric modeling of irregular tabular deposits. *Center for Computational Geostatistics Annual Report*, v. 17, 2015. Citado na página [52](#).
- MANCHUCK, J.; DEUTSCH, C. Mls for boundary interpretations with uncertainty. *CCG Annual Report*, 2018. Citado na página [27](#).
- MANCHUCK, J. A.; DEUTSCH, C. V. Geometric modeling of irregular tabular deposits. *CCG Annual Report*, 2015. Citado 2 vezes nas páginas [24](#) e [54](#).
- MARTIN, R.; BOISVERT, J. Review of radial basis functions and domain decomposition for implicit geological modelling. *CCG Annual Report*, 2015. Citado na página [35](#).
- MARTIN, R.; BOISVERT, J. Implicit modeling. *CCG Guidebook Series*, v. 22, 2017. Citado 12 vezes nas páginas [20](#), [24](#), [26](#), [27](#), [28](#), [31](#), [35](#), [38](#), [39](#), [40](#), [54](#) e [82](#).
- MARTIN, R.; BOISVERT, J. Implicit modeling. *University of Alberta – CCG, Guidebook series*, v. 22, 2017. Citado na página [93](#).
- MARTIN, R.; BOISVERT, J. B. Iterative refinement of implicit boundary models for improved geological feature reproduction. *Computers & Geosciences*, Elsevier, v. 109, p. 1–15, 2017. Citado 4 vezes nas páginas [35](#), [36](#), [40](#) e [41](#).
- MATHERON, G. *Traité de géostatistique appliquée*, tome i: Mémoires du bureau de recherches géologiques et minières. *Editions Technip, Paris*, v. 14, 1962. Citado na página [14](#).

- MATHERON, G. et al. Conditional simulation of the geometry of fluvio-deltaic reservoirs. In: SOCIETY OF PETROLEUM ENGINEERS. *Spe annual technical conference and exhibition*. [S.l.], 1987. Citado na página 44.
- MCCULLAGH, P.; NELDER, J. A. *Generalized linear models*. [S.l.]: Chapman and Hall, 1989. Citado na página 45.
- MCLENNAN, J. A. *The decision of stationarity*. Tese (Doutorado) — University of Alberta, 2007. Citado 6 vezes nas páginas 14, 47, 48, 49, 58 e 64.
- MCLENNAN, J. A.; DEUTSCH, C. V. Implicit boundary modeling (boundsim). *CCG Annual Report*, 2006. Citado 3 vezes nas páginas 14, 15 e 47.
- MUNROE, M.; DEUTSCH, C. Full calibration of c and β in the framework of vein-type deposit tonnage uncertainty. *Center for Computational Geostatistics Annual Report*, v. 10, 2008. Citado na página 52.
- NEUFELD, C.; WILDE, B. A global kriging program for artifact-free maps. *CCG Annual Report*, v. 7, p. 403–1, 2005. Citado na página 29.
- OSHER, S.; FEDKIW, R. *Level set methods and dynamic implicit surfaces*. [S.l.]: Springer-Verlag New York, Inc., 2003. Citado na página 20.
- OXFORD, U. *Oxford online dictionary*. 2018. Definição. Disponível em: <<http://www.oed.com>>. Acesso em: 15-06-2018. Citado na página 14.
- PAKYUZ-CHARRIER, E. J. et al. Monte carlo simulations for uncertainty estimation in 3d geological modeling, a guide for disturbance distribution selection and parameterization. *Solid Earth*, Copernicus Gesellschaft mbH, v. 9, p. 385–402, 2018. Citado na página 44.
- PYRCZ, M. J.; DEUTSCH, C. V. *Geostatistical reservoir modeling*. [S.l.]: Oxford university press, 2014. Citado na página 43.
- RADTKE, R. *Modelagem geológica implícita através de simulações de funções distância assinaladas*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, 2017. Citado 2 vezes nas páginas 49 e 50.
- REMY, N.; BOUCHER, A.; WU, J. *Applied geostatistics with SGeMS: A user's guide*. [S.l.]: Cambridge University Press, 2009. Citado na página 17.
- ROLO, R. *Modelagem geológica implícita com funções distância assinaladas*. Dissertação (Mestrado) — Universidade Federal do Rio Grande do Sul, 2017. Citado 2 vezes nas páginas 27 e 69.
- ROLO, R. M.; RADTKE, R.; COSTA, J. F. C. L. Signed distance function implicit geologic modeling. *REM-International Engineering Journal*, SciELO Brasil, v. 70, n. 2, p. 221–229, 2017. Citado na página 69.
- ROSSI, M.; DEUTSCH, C. *Mineral resource estimation*. [S.l.]: Springer Science & Business Media, 2013. Citado 3 vezes nas páginas 13, 14 e 88.
- ROYER, J.; VIEIRA, P. Dual formalism of kriging. In: *Geostatistics for natural resources characterization*. NATO advanced Study Institute. [S.l.: s.n.], 1984. p. 691–702. Citado na página 30.

- SAVCHENKO, V. V. et al. Function representation of solids reconstructed from scattered surface points and contours. In: WILEY ONLINE LIBRARY. *Computer Graphics Forum*. [S.l.], 1995. v. 14, n. 4, p. 181–188. Citado na página 19.
- SHANNON, C. E. A mathematical theory of communication. *Bell system technical journal*, Wiley Online Library, v. 27, n. 3, p. 379–423, 1948. Citado na página 45.
- SILVA, D.; DEUTSCH, C. Modeling multiple rock types with distance functions: methodology and software. In: *CCG Annual Report*. [S.l.: s.n.], 2012. Citado 2 vezes nas páginas 33 e 45.
- SILVA, D.; DEUTSCH, C. Correcting distance function models to correct proportions. In: *CCG Annual Report*. [S.l.: s.n.], 2013. Citado na página 45.
- SILVA, D.; DEUTSCH, C. Speeding up the signed distance implicit geologic modeling method for multiple rock types. *CCG Annual Report*, 2015. Citado na página 27.
- SILVA, D.; DEUTSCH, C. Multivariate categorical modeling with hierarchical truncated pluri-gaussian simulation. *Mathematical Geosciences*, Springer, v. 51, n. 5, p. 527–552, 2019. Citado na página 44.
- SILVA, D. M.; DEUTSCH, C. V. Speeding up the signed distance implicit geologic modeling method for mulple rock types. *CCG Annual Report*, 2015. Citado na página 36.
- SILVA, D. M. A. *Enhanced geologic modeling with data-driven training images for improved resources and recoverable reserves*. Tese (Doutorado) – University of Alberta, 2015. Citado 5 vezes nas páginas 15, 27, 34, 50 e 51.
- SMIRNOFF, A.; BOISVERT, E.; PARADIS, S. J. Support vector machine for 3d modelling from sparse geological information of various origins. *Computers & Geosciences*, Elsevier, v. 34, n. 2, p. 127–143, 2008. Citado 2 vezes nas páginas 19 e 114.
- SRIVASTAVA, M. Probabilistic modeling of ore lens geometry: An alternative to deterministic wireframes. *Mathematical Geology*, Springer, v. 37, n. 5, p. 513–544, 2005. Citado na página 96.
- STREBELLE, S. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, 2002. Citado na página 44.
- VARGA, M. de la; SCHAAF, A.; WELLMANN, F. Gempy 1.0: open-source stochastic geological modeling and inversion. *Geoscientific Model Development*, 2019. Citado na página 44.
- WENDLAND, H. *Scattered data approximation*. [S.l.]: Cambridge university press, 2004. v. 17. Citado na página 35.
- WILDE, B.; DEUTSCH, C. A new way to calibrate distance function uncertainty. In: *CCG Annual Report*. [S.l.: s.n.], 2011. Citado na página 31.
- WILDE, B.; DEUTSCH, C. Kriging and simulation in presence of stationary domains: developments in boundary modeling. In: *Geostatistics Oslo 2012*. [S.l.]: Springer, 2012. p. 289–300. Citado 9 vezes nas páginas 18, 51, 52, 53, 55, 57, 77, 82 e 88.
- YAMAMOTO, J. K. et al. Post-processing of sequential indicator simulation realizations for modeling geologic bodies. *Computational Geosciences*, Springer, v. 19, n. 1, p. 257–266, 2015. Citado na página 72.

YANG, L. et al. Assessing and visualizing uncertainty of 3d geological surfaces using level sets with stochastic motion. *Computers & Geosciences*, Elsevier, v. 122, p. 54–67, 2019. Citado 2 vezes nas páginas [51](#) e [92](#).

Apêndices

APÊNDICE A – SOFTWARE DESENVOLVIDO

Esse apêndice é um manual de instalação e uso dos *plugins* e *jupyter notebooks* desenvolvidos nessa tese. Os *plugins* foram desenvolvidos para o software geostatístico *SGeMS/AR2GeMS*. Recomenda-se a instalação e uso da suite de *plugins* em uma versão do *SGeMS/AR2GeMS* que inclua um interpretador python e a biblioteca *ar2gas* embarcados (*ar2gems-full-version-python3*).

A suite de *plugins* pode ser acessada no repositório: <https://github.com/robertorolo/LPM-Geomod_Suite>. Para instalá-la primeiro deve-se instalar as seguintes dependências: *ar2gas*, *numpy*, *math*, *itertools*, *scipy*, *pyevtk*, *sklearn*, *pyevtk*. O repositório conta com o arquivo *install_dependencies.py*. Ao copiar, colar e executar o conteúdo do arquivo no interpretador python embarcado encontrado no diretório do *SGeMS/AR2GeMS* as dependências (com exceção do *ar2gas* que não pode ser encontrado no repositório PyPI) são automaticamente instaladas.

Depois disso, os arquivos do diretório *ui* do repositório devem ser copiados e colados na pasta do *SGeMS/AR2GeMS* *plugins/Geostat*, os arquivos do diretório *py* do repositório devem ser copiados e colados na pasta do *SGeMS/AR2GeMS* *plugins/Geostat/python*, finalmente, o arquivo *helpers.py* do repositório deve ser copiado e colado na pasta *Lib* do *SGeMS/AR2GeMS*. Uma versão do *SGeMS/AR2GeMS* com a suite previamente instalada pode ser solicitada ao autor dessa tese.

Caso a instalação tenha sido bem sucedida os *plugins* serão mostrados no painel de algoritmos do *software*, como mostrado na [Figura 70](#).

Figura 70 – LPM-Geomod_Suite no painel de algoritmos do *SGeMS/AR2GeMS*.

Algorithms	
Algorithm	Description
> Utilities	
> LPM-UFRGS	
> Simulation	
> LPM-LVA_Suite	
> Estimation	
> Spectral	
> LPM-UFRGS-Sampling	
> LPM-Utilities	
>	
▼ LPM-Geomod_Suite	
SVM	support vector machines for geomodeling
auto_grid	creates a grid automatically
block_transform	softmax transformation and entropy calculation
boundary_simulation_pfields	creates different scenarios
deterministic	creates a deterministic geologic model
kernel_factor_uncertainty	creates different scenarios
pfields	probability fields simulation
point_transform	calculates indicators or signed distances properties
sd_perturbation_uncertainty	creates different scenarios
vtk_export	exports in VTK format

As próximas seções descrevem os diferentes *plugins* e seus parâmetros.

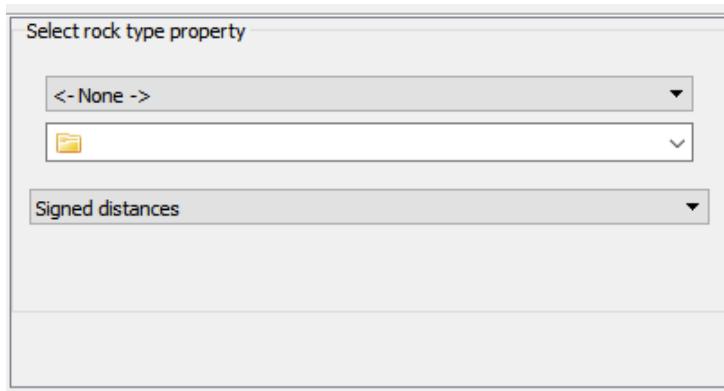
A.1 Transformações

A suite traz ferramentas para transformar variáveis em suporte de pontos ou blocos.

A.1.1 Em pontos

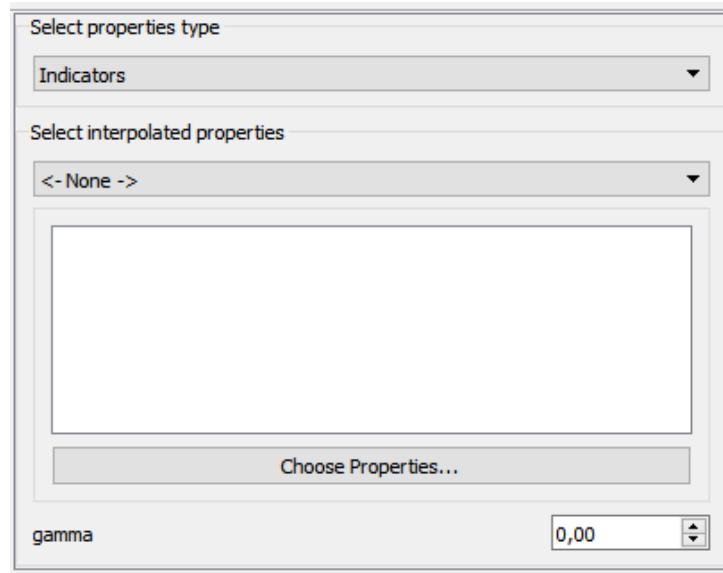
O *plugin point_transform* (Figura 71) recebe como *input* a propriedade categórica, que representa as diferentes litologias, e faz a transformação em distâncias assinaladas ou indicadores. O *output* são K propriedades, onde K é o número de litologias do banco de dados.

Figura 71 – Interface do *plugin point_transform*



A.1.2 Em blocos

O *plugin* de transformação em blocos, *block_transform* (Figura 72), recebe as propriedades distâncias assinaladas ou indicadores interpoladas, e o parâmetro *gamma* (ou *omega*) da Equação 2.20 como *input*. Como *output*, as propriedades probabilidade de ocorrência referente à cada distância ou indicador são criadas além da variável de incerteza U, calculada a partir da Equação 2.23.

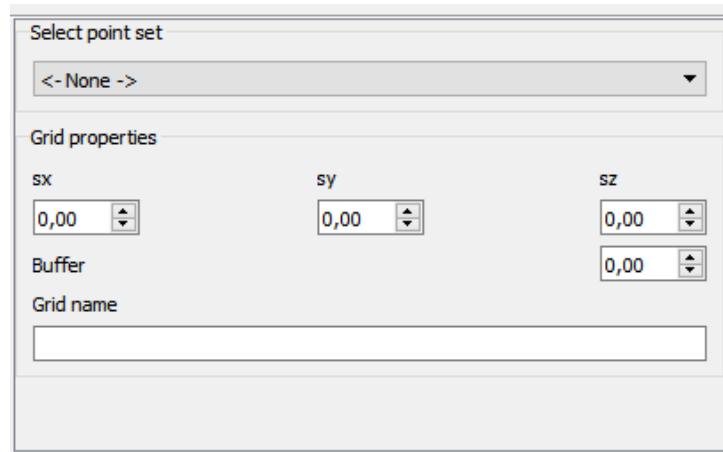
Figura 72 – Interface do *plugin block_transform*

A.2 Auxiliares

Há ferramentas, que embora não sejam para modelagem geológica, auxiliam no processo de modelagem e visualização dos resultados.

A.2.1 Criação automática do *grid*

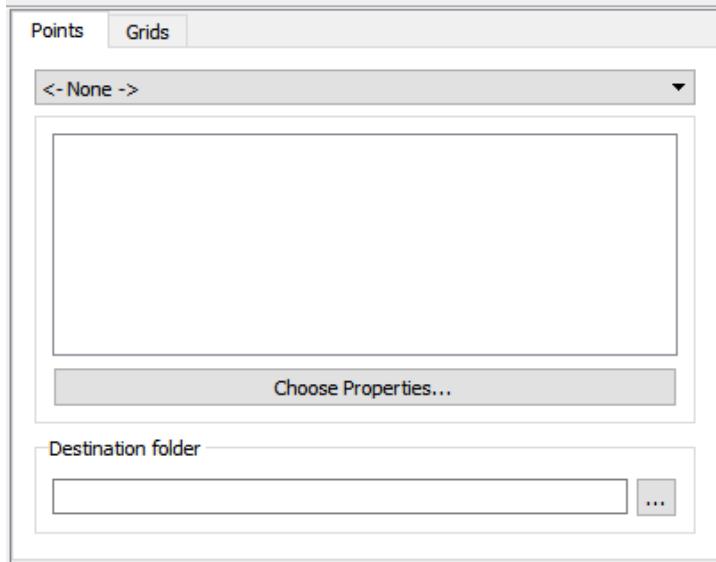
O *plugin auto_grid* (Figura 73) cria um *grid* que cobre toda área ou volume ocupado pelas amostras com dimensões dos blocos (*sx*, *sy*, *sz*) informadas pelo usuário. O parâmetro *buffer* controla a extensão do *grid* para além dos limites mínimos e máximos das amostras.

Figura 73 – Interface do *plugin auto_grid*

A.2.2 Exportação em formato VTK

O *plugin vtk_export* (Figura 74) exporta propriedades em suporte de ponto ou bloco em formato VTK para visualização em outros softwares, como o Paraview, por exemplo.

Figura 74 – Interface do *plugin vtk_export*



A.3 Modelagem determinística

O *plugin deterministic* (Figura 75) cria modelos geológicos uni ou multi categóricos a partir de distâncias assinaladas ou indicadores. Na aba *general* (Figura 75a), o usuário seleciona se as propriedades são distâncias assinaladas ou indicadores. Também devem ser selecionados o *grid* e o nome da propriedade que será criada. A opção *keep all variables* cria, além do modelo final, todas as propriedades distâncias ou indicadores interpolados.

Se apenas uma categoria de distâncias assinaladas for escolhida, a iso-superfície zero é extraída, se uma propriedade de indicadores for selecionada, a abordagem apresentada na subseção 2.1.1.8 é aplicada.

As propriedades interpoladas podem ser selecionadas em qualquer ordem, desde que tenham sido geradas pelo *plugin point_transform* (subseção A.1.1).

Em *refinement options*, o usuário insere os parâmetros para o refinamento de contatos (subseção 2.1.1.6.2): quantas iterações, zero gera um modelo sem refinamento, e os parâmetros de *downscaling* do *grid*, ou seja, em quantas partes as células serão divididas em x, y, e z.

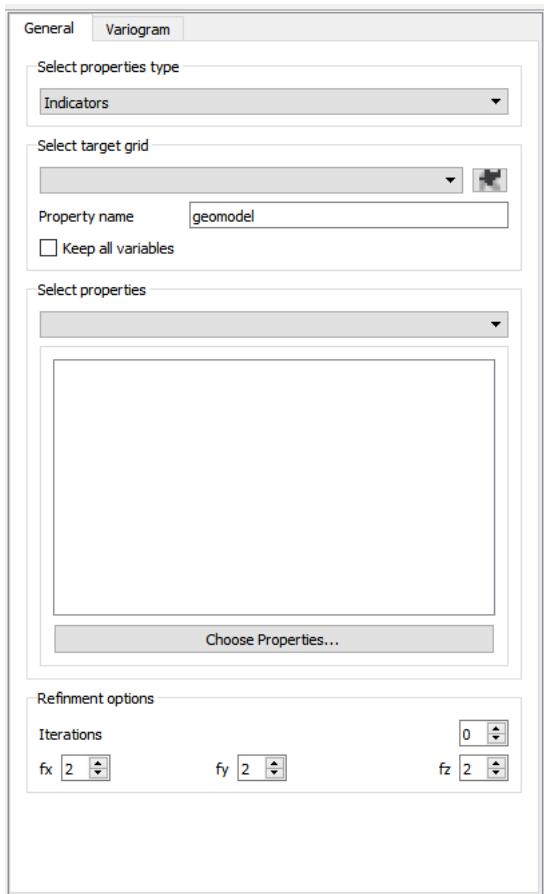
Na aba *variogram*, o usuário insere os variogramas na mesma ordem que selecionou as variáveis interpoladas na aba anterior. Caso o usuário selecione a opção *use variograms model file instead*, ele deve carregar um arquivo de texto, como mostrado na Figura 76. O arquivo deve conter o número que representa cada categoria acompanhado do modelo variográfico em

formato .xml, o mesmo formato usado pela ferramenta de variografia do SGeMS/AR2GeMS.

Caso um único variograma seja informado, todas as categorias serão interpoladas com esse mesmo modelo. Caso nenhum variograma seja informado, um modelo Gaussiano com o alcance igual a maior distância entre uma amostra codificada com aquele indicador e um nó do *grid* é usado para cada uma das propriedades selecionadas.

Figura 75 – Interface do *plugin deterministic*

(a) Aba general.



(b) Aba variogram.



Figura 76 – Arquivo de texto com os modelos variográficos.

```

1 1
2 <variogram_model nugget="0.001" structures_count="1" >
3   <structure_1 contribution="0.999" type="Gaussian" >
4     <ranges max="76" medium="76" min="76" />
5     <angles x="0" y="0" z="0" />
6   </structure_1>
7 </variogram_model>
8
9 2
10 <variogram_model nugget="0.001" structures_count="1" >
11   <structure_1 contribution="0.999" type="Gaussian" >
12     <ranges max="72" medium="72" min="72" />
13     <angles x="0" y="0" z="0" />
14   </structure_1>
15 </variogram_model>
16
17 3
18 <variogram_model nugget="0.001" structures_count="1" >
19   <structure_1 contribution="0.999" type="Gaussian" >
20     <ranges max="53" medium="53" min="53" />
21     <angles x="0" y="0" z="0" />
22   </structure_1>
23 </variogram_model>
24
25 4
26 <variogram_model nugget="0.001" structures_count="1" >
27   <structure_1 contribution="0.999" type="Gaussian" >
28     <ranges max="41" medium="41" min="41" />
29     <angles x="0" y="0" z="0" />
30   </structure_1>
31 </variogram_model>
32
33 5
34 <variogram_model nugget="0.001" structures_count="1" >
35   <structure_1 contribution="0.999" type="Gaussian" >
36     <ranges max="45" medium="45" min="45" />
37     <angles x="0" y="0" z="0" />
38   </structure_1>
39 </variogram_model>

```

A.3.1 Máquina de vetores de suporte (SVM)

A metodologia proposta por [Smirnoff, Boisvert e Paradis \(2008\)](#), que usa máquina de vetores de suporte (*support vector machine*) para a criação de modelos geológicos determinísticos multi-categóricos, foi implementada na suite a partir do *plugin SVM* (Figura 77).

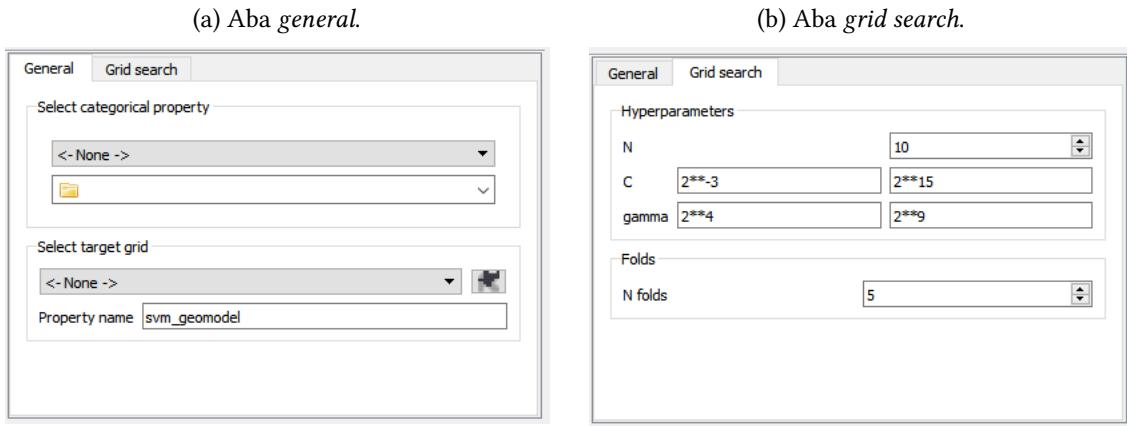
Na aba *general*, (Figura 77a) o usuário seleciona a propriedade categórica, um *grid* de estimativa e o nome da propriedade a ser criada.

Na aba *grid search*, (Figura 77b) o usuário seleciona os parâmetros que serão utilizados na função *GridSearchCV* da biblioteca *sklearn*. É necessário escolher valores mínimos e máximos para os hiper-parâmetros *C* e *gamma* e um número *N* de valores aleatórios que serão tomados nesse intervalo, combinados e testados e uma validação *k-folds*. Os valores padrão são os melhores intervalos encontrados em um teste de sensibilidade de hiper-parâmetros realizado por [Smirnoff, Boisvert e Paradis \(2008\)](#).

Em *N folds* o usuário seleciona quantas *folds* serão aplicadas na validação por *kfolds*. Para o caso da Figura 77b, o algoritmo vai remover 20% das amostras de forma aleatória do banco de dados cinco vezes e criar cinco bancos de dados de treino com 80% das amostras e cinco banco de dados de teste correspondentes com 20% das amostras. Então o algoritmo toma dez valores para *C* e dez valores para *gamma* aleatoriamente nos intervalos selecionados, esses

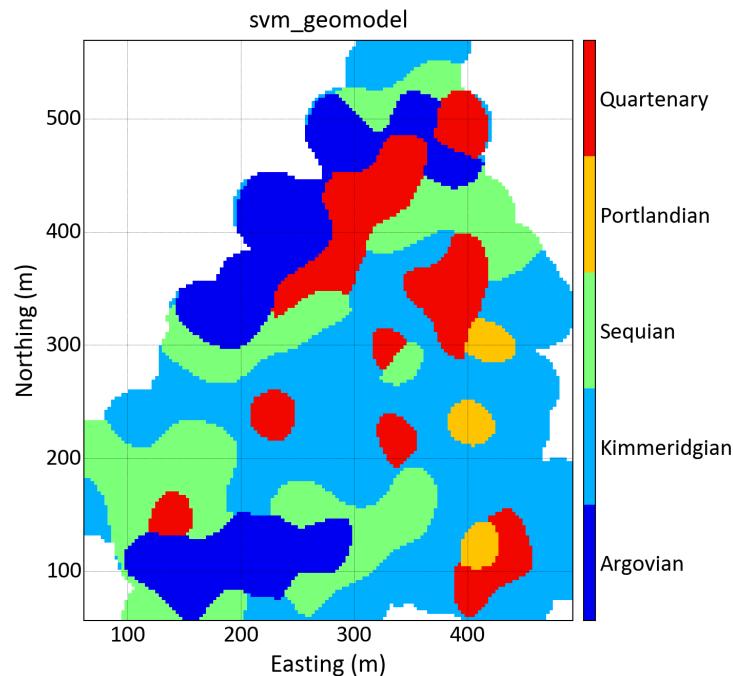
valores são combinados para gerar 100 diferentes modelos SVM. Cada um desses modelos é treinado em cada um dos cinco bancos de dados de treino e testado em seu respectivo banco de dados de teste. Com os parâmetros que apresentarem o melhor *score* de validação, um modelo é treinado usando a totalidade dos dados e a predição das categorias é feita no *grid*.

Figura 77 – Interface do *plugin svm*



A Figura 78 mostra o modelo geológico criado para o banco de dados *Swiss Jura* usando o *plugin* desenvolvido com os valores padrão.

Figura 78 – Modelo geológico para o dataset *Swiss Jura* criado com máquinas de vetores de suporte.



A.4 Modelagem estocástica

Dos algoritmos desenvolvidos nessa tese, dois deles (parametrização do suporte do *kernel*) e avaliação de incerteza usando funções distâncias assinaladas e campos potenciais) foram implementados como *plugins* para o *SGeMS/AR2GeMS*. A simulação hierárquica de contatos, por ser bastante dependente de *softwares* da biblioteca GSLib, foi implementando em *jupyter notebooks* usando a biblioteca *pygeostat* e um módulo desenvolvido.

A.4.1 Parametrização do suporte *kernel*

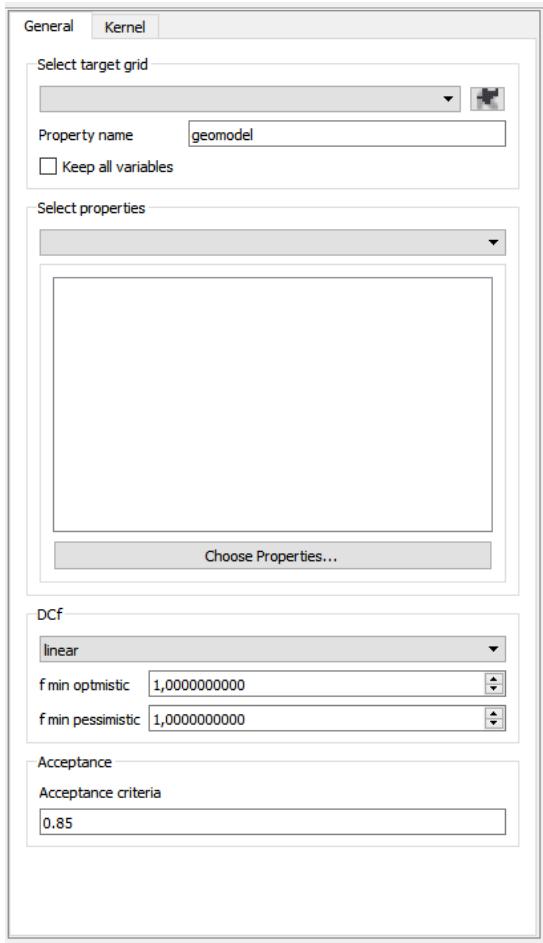
O *plugin kernel_factor_uncertainty* (Figura 79) implementa a metodologia proposta na seção 3.1. Na aba *general*, (Figura 79a) o usuário seleciona o *grid* de interpolação, as propriedades função distância assinaladas e os parâmetros do método. Parametrização linear ou quadrática deve ser selecionada, um valor de *fmin* otimista e um valor de *fmin* pessimista e o critério de aceitação de reprodução dos dados amostrais.

Na aba *kernel*, (Figura 79b) é necessário selecionar os parâmetros dos *kernels* para cada categoria. As funções Gaussiana e esférica estão disponíveis. Para os demais parâmetros, um único valor considera o mesmo para as múltiplas categorias, separando valores por vírgula, na mesma ordem que as distâncias assinaladas foram selecionadas na aba *general*, o algoritmo usará parâmetros diferentes para cada categoria. Selecionar o valor zero para o suporte fará com que o algoritmo calcule-o automaticamente como a maior distância entre uma amostra codificada com aquele indicador e um nó do *grid*.

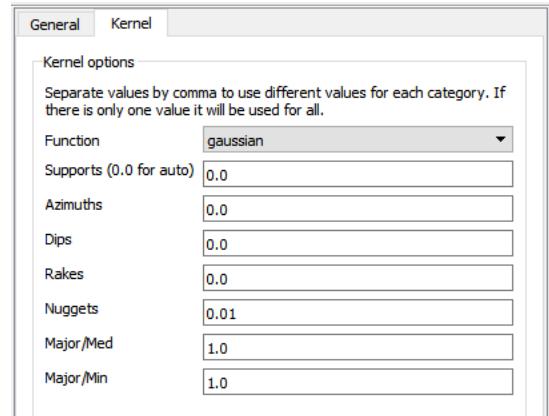
A implementação da anisotropia ainda é um trabalho em progresso e necessita de testes e ajustes.

Figura 79 – Interface do *plugin kernel_factor_uncertainty*.

(a) Aba general.



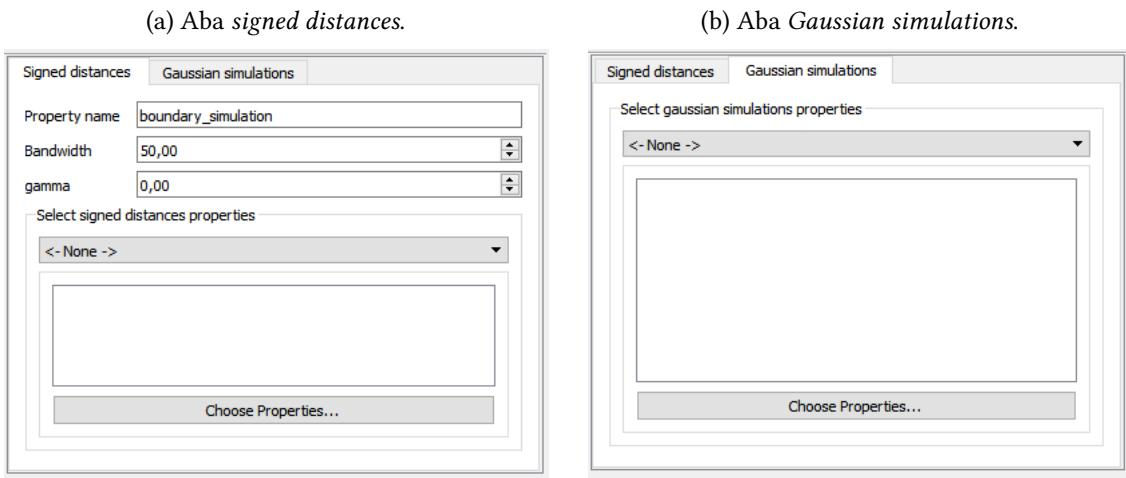
(b) Aba kernel.



A.4.2 Avaliação de incerteza usando funções distâncias assinaladas e campos potenciais

O *plugin boundary_simulation_pfields* (Figura 80) implementa o método apresentado na seção 3.2. Na aba *general*, (Figura 80a) o usuário deve informar o nome da propriedade que será criada, a largura da banda de incerteza e o parâmetro *gamma* (ou *omega*) que será usado na transformação das distâncias em probabilidades. Caso zero seja escolhido para o parâmetro *gamma*, o algoritmo usará o maior módulo entre entre todas as categorias estimadas bloco a bloco.

Na aba *gaussian simulations*, o usuário deve inserir as simulações Gaussianas que serão transformadas em campos de probabilidade e usadas para amostrar as distribuições locais.

Figura 80 – Interface do *plugin boundary_simulation_pfields*.

Se houver apenas uma categoria no banco de dados, o algoritmo estandardiza o valor simulado entre $[-bw, bw]$, onde bw é o valor escolhido para a banda de incerteza e compara, dentro da zona de incerteza, o valor simulado e o valor estimado. Os blocos são classificados de acordo com a [Equação 2.27](#).

A.4.3 Simulação de contatos hierárquica

Os *jupyter notebooks* utilizados para a prova de conceito no banco de dados *Swiss Jura* e no estudo de caso no banco de dados de cobre pôrfiro, apresentados na [seção 3.3](#), podem ser acessados no repositório: <https://github.com/robertorolo/hierarchical_boundary_simulation>. O repositório também conta com um módulo em Python necessário para aplicação do fluxo de trabalho.

- Prova de conceito no *Swiss Jura*: <https://github.com/robertorolo/hierarchical_boundary_simulation/blob/main/jura_case_study.ipynb>
- Estudo de caso no cobre pôrfiro: <https://github.com/robertorolo/hierarchical_boundary_simulation/blob/main/porphyr_copper_case_study.ipynb>