

Isaac Leonardo Santos Sacramento

Tese submetida ao Programa de Pós-graduação
em Ciência da Computação para a obtenção
do Grau de Doutorado.

Orientador

Universidade Federal de Santa Catarina:
Prof. Dr. Mauro Roisenberg

Florianópolis

2016

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

SobrenomeAutor, NomeAutor

Título do trabalho : Subtítulo do trabalho / NomeAutor

SobrenomeAutor ; orientador, NomeOrientador

SobrenomeOrientador ; coorientador, NomeCoorientador

SobrenomeCoorientador. - Florianópolis, SC, 2014.

100 p.

Tese (doutorado) - Universidade Federal de Santa
Catarina, Centro Tecnológico. Programa de Pós-Graduação em
Ciência da Computação.

Inclui referências

1. Ciência da Computação. 2. Exemplo de ficha
catalográfica. I. SobrenomeOrientador, NomeOrientador. II.
SobrenomeCoorientador, NomeCoorientador. III. Universidade
Federal de Santa Catarina. Programa de Pós-Graduação em
Ciência da Computação. IV. Título.

RESUMO

Muitos métodos de simulação multiponto (SMP) têm sido desenvolvidos na última década e são classificados em duas categorias. A primeira, aborda os métodos que mantêm a continuidade das estatísticas clássicas baseadas em variograma. A segunda categoria de algoritmos multiponto é baseada no cálculo de distâncias entre padrões em uma imagem de treinamento. Imagens de treinamento são bancos de dados de padrões a partir dos quais, estatísticas multiponto, incluindo o variograma, podem ser extraídas. Como ferramenta de modelagem, os algoritmos de aprendizado de máquina são universais, adaptativos, não-lineares, robustos e eficientes. Eles podem alcançar soluções aceitáveis para problemas de classificação, regressão e modelagem de densidade de probabilidade em espaço de alta dimensão e com características espacialmente referenciados. Neste sentido, se objetiva desenvolver um método, baseado em aprendizagem de máquina, capaz de gerar simulações a partir do treinamento prévio com imagens de treinamento.

Palavras-chave: Estatísticas Multiponto. Redes Neurais. Amostragem Direta. Simulação Estocástica. Aprendizagem de Máquina.

LISTA DE ABREVIATURAS E SIGLAS

SMP	Simulação Multiponto
IT	Imagem de Treinamento
FDP	Função Densidade de Probabilidade
FDCC	Função de Distribuição Condicional Cumulativa
DS	Direct Sample
IQ	Image Quilting
RNC	Redes Neurais Convolucionais
ReLU	Rectified Linear Unit
GA	Genetic Algorithm

SUMÁRIO

1	INTRODUÇÃO	9
1.1	ESTIMATIVA E SIMULAÇÃO GEOESTATÍSTICA	9
1.2	SIMULAÇÃO BASEADA EM VARIOGRAMA	12
2	SIMULAÇÃO MULTIPONTO	15
2.1	SIMULAÇÃO MULTIPONTO BASEADA EM PONTOS	17
2.2	SIMULAÇÃO MULTIPONTO BASEADA EM BLOCOS	21
3	REDES NEURAIS CONVOLUCIONAIS	23
3.1	DEEP LEARNING NA RESTAURAÇÃO DE IMAGENS	23
4	ABORDAGEM DE SOLUÇÃO	25
	REFERÊNCIAS	27

1 INTRODUÇÃO

De acordo com o glossário geoestatístico, geoestatística é a aplicação de métodos estatísticos nas ciências da terra. Journel (1990) definem geoestatística como um conjunto de ciências estatísticas que estudam fenômenos espaciais/temporais, e usam relacionamentos espaciais para modelar valores de variáveis em locais ou tempos não observados, não amostrados. Ou, de maneira mais curta, é a quantificação de correlação espacial para ser utilizada posteriormente em simulações (DEUTSCH, 2002). A estatística clássica é dedicada à análise e interpretação de incertezas causadas pela amostragem limitada da propriedade em estudo. A geoestatística, por outro lado, não é restrita a um modelo de distribuição de população; por exemplo, todas as amostras de uma população serem normalmente distribuídas e independentes. A maioria dos dados em ciências da terra, como propriedades de rocha, não satisfazem estas hipóteses, pois possuem correlação espacial. Ou seja, valores de dados de regiões que estão próximas tendem a ser mais similares que valores de regiões que estão mais afastadas.

1.1 ESTIMATIVA E SIMULAÇÃO GEOESTATÍSTICA

A geoestatística provê um conjunto de ferramentas para análise e interpretação de dados no processo de previsão de reservatórios de petróleo, aquíferos subterrâneos, meios porosos, materiais compósitos, mineração e hidrogeologia. As previsões obtidas podem ser de duas formas: estimativa e simulação. Na estimativa (figura 1), apenas uma realização, considerada a ideal, da ocorrência espacial é produzida. O resultado obtido é baseado na amostra de dados e em um modelo, admitido como a representação mais precisa da correlação espacial da amostra de dados, o variograma. Esta estimativa, também chamada de *maximum a posteriori (map)*, costuma ser obtida pela técnica de *kriging*. Por outro lado, na simulação (figura 2), muitas realizações semelhantes são produzidas usando o mesmo modelo de correlações espaciais utilizado na estimativa por *kriging*.

Uma propriedade é estimada para uma localização não amostrada com base nas características da correlação espacial desta propriedade e nos dados condicionantes (valores existentes em uma localização amostrada), isto é o que difere os métodos geoestatísticos de uma interpolação simples. Para ilustrar, considere a figura 3 como uma representação da permeabilidade em n locais amostrados. A permeabilidade no ponto desconhecido z_0 pode ser calculada por meio das equações 1.1 e 1.2 e usando a distância inversa. No

Figura 1 – Fluxo da estimativa geoestatística.

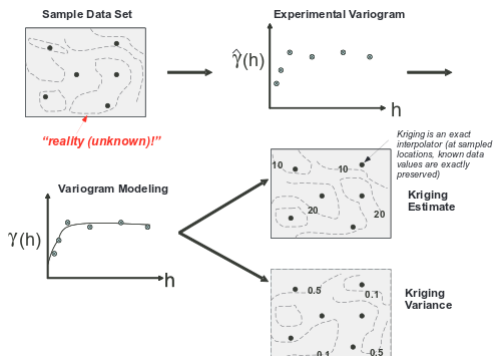
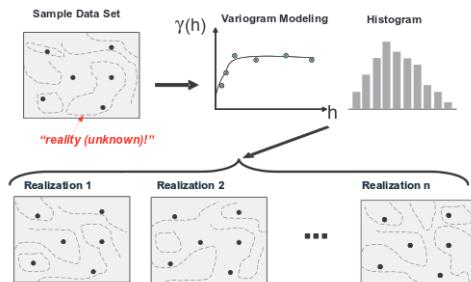


Figura 2 – Fluxo da simulação geoestatística.



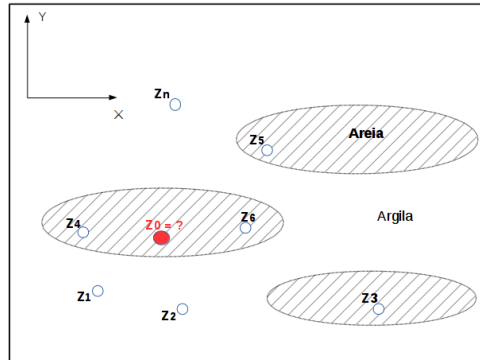
estimador linear z_0 , cada peso w_i é determinado pela distância do ponto conhecido até o ponto desconhecido.

$$z_0 = \sum_{i=1}^n w_i z_i \quad (1.1)$$

$$w_i = \frac{\frac{1}{d_i}}{\sum_{i=1}^n \left(\frac{1}{d_i}\right)} \quad (1.2)$$

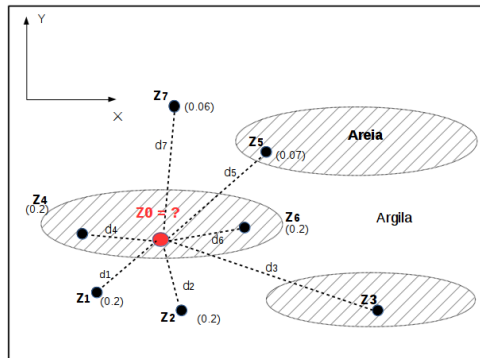
Tomando como exemplo $n = 7$, é possível obter a configuração apresentada na figura 4 (os valores entre parênteses representam os pesos dos pontos). Entretanto, do ponto de vista do conhecimento geológico, a permeabilidade dentro da estrutura de areia deve ser mais similar na direção lateral. Consequentemente, os pontos 4 e 6 devem ter pesos maiores que os pontos 1 e

Figura 3 – Cálculo da permeabilidade desconhecida z_0 baseado em um conjunto de valores em n locais conhecidos.



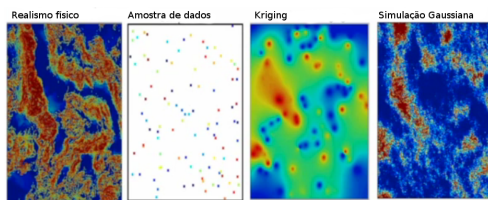
2, o que não acontece com o uso da distância invertida. Assim, a informação da correlação espacial não é incorporada na análise. De um modo geral a estimativa geoestatística consiste de três etapas: (1) examinar a similaridade entre as amostras conhecidas por meio da análise de um variograma experimental; (2) ajustar uma função matemática para o variograma experimental; (3) realizar uma interpolação por *kriging* baseada nesta função. Considerando o exemplo mostrado, nos passos 1 e 2 a correlação espacial é incorporada por meio da maior similaridade dos pontos z_4 e z_6 com o ponto a ser estimado. Por meio de *kriging*, os pesos dos pontos z_4 e z_6 se tornam maiores do que os pesos de z_1 e z_2 .

Figura 4 – Estimativa da permeabilidade desconhecida z_0 , dados 7 valores conhecidos.



Os problemas geoestatísticos podem ser sumarizados como ilustrado na figura 5. Para uma realidade física, da qual se dispõe de uma amostra limitada de dados, é possível obter uma estimativa média por *kriging* ou realizar simulações com estruturas multigaussianas, de forma a criar múltiplas realizações da realidade física. Ao comparar a imagem da simulação gaussiana com a imagem da realidade física é possível observar que a simulação não se assemelha com a realidade. Uma questão importante que emerge desta observação é: como adicionar realismo físico à modelagem estocástica?

Figura 5 – Processo básico de resolução de um problema geoestatístico.

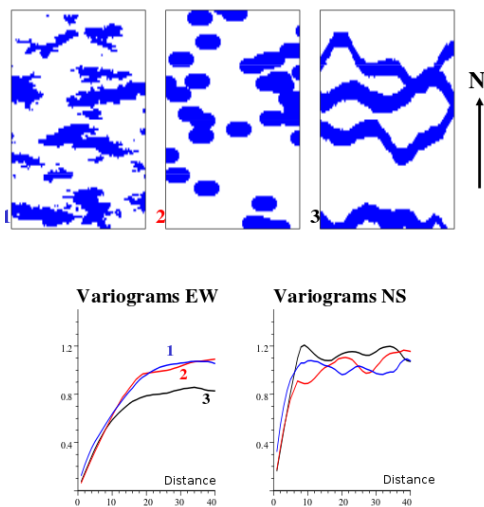


1.2 SIMULAÇÃO BASEADA EM VARIOGRAMA

O variograma é um dispositivo que armazena padrões, modelado a partir de dados experimentais e reproduzido em cada realização. O sucesso no processo de simulação geoestatística recai sobre o variograma, que, apesar de ser um conceito matemático, tem pouca conexão com a realidade. Vários tipos diferentes de heterogeneidades geológicas podem produzir variogramas experimentais semelhantes (CAERS; ZHANG, 2004), como visto na figura 6. As limitações em capturar a continuidade espacial dos modelos baseados em variograma são amplamente conhecidas na literatura. Por ser uma estatística que descreve a correlação entre dois pontos quaisquer no espaço, este dispositivo é incapaz de considerar estruturas curvilíneas e/ou estruturas contínuas em escalas maiores, como canais sinuosos (ARPAT; CAERS, 2007; TAHMASEBI; HEZARKHANI; SAHIMI, 2012).

É comum, em geoestatística, trabalhar com modelo de funções aleatórias, as quais são parte da teoria de probabilidades. Nesta abordagem é necessário lidar com conceitos considerados complicados como a noção de expectativa, estacionaridade, domínio infinito e ergodicidade, para então ser possível realizar uma formulação de que a realidade é uma realização de um processo estocástico. Por sua vez, esta realização é decomposta em dois outros problemas: um problema de média e outro de variação em torno desta

Figura 6 – Três heterogeneidades geológicas diferentes produzem três variogramas semelhantes.

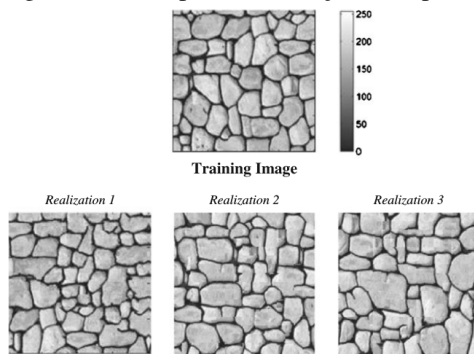


Fonte: (CAERS; ZHANG, 2004)

média; de modo que, diante de uma simulação geológica, separar o que é a média do que é dado residual pode ser um problema. A ferramenta utilizada para modelar sistemas aleatórios e fenômenos espaciais é denominada simulação estocástica condicional, seu objetivo é criar representações múltiplas e realísticas de um fenômeno espacial estudado, condicionadas a qualquer dado disponível. Cada realização deve refletir a continuidade espacial que se acredita existir, como ilustrado na figura 7. Os métodos de simulação multiponto (SMP) surgem da necessidade de capturar informações estruturais sobre continuidade espacial e tornam possível a comunicação entre física e estatística através do elemento chamado Imagem de Treinamento.

Na seção seguinte estão listados os principais trabalhos relacionados à simulação multiponto. Os algoritmos de simulação são agrupados baseados em *pixels* e baseados em blocos, e são apresentados em ordem cronológica. A partir da revisão literária dos algoritmos SMP, é observável a ausência de métodos de aprendizagem de máquina, capazes de reproduzir os padrões de estruturas espaciais nas realizações e, eventualmente, incorporar ao processo de simulação o conhecimento apresentado por um especialista.

Figura 7 – Exemplo de simulação multiponto.

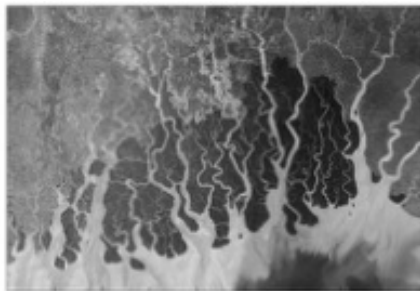


Fonte: (TAHMASEBI; HEZARKHANI; SAHIMI, 2012)

2 SIMULAÇÃO MULTIPONTO

Estatística multiponto engloba um conjunto de algoritmos de simulação sequencial que utilizam imagem de treinamento como dado de entrada para definir a estrutura espacial de um processo (GUARDIANO; SRIVASTAVA, 1993; STREBELLE, 2002b). As imagens de treinamento são descrições explícitas, 2D ou 3D, da continuidade espacial estudada ou, como definido por Jef Caers (CAERS; ZHANG, 2004), são bancos de dados de padrões a partir dos quais, estatísticas multiponto, incluindo o variograma, podem ser extraídas. Elas contêm informações de múltiplos pontos, são mais intuitivas para não-especialistas, não necessitam ser restritas localmente a quaisquer dados, não necessitam ter as mesmas dimensões da área estudada, podem ser rejeitadas baseado na análise visual do fenômeno em estudo e devem refletir um estilo da continuidade espacial similar ao fenômeno real. As figuras 8 e 9 mostram dois exemplos de imagens de treinamento obtidas da biblioteca de imagens disponibilizada por (MARIETHOZ; CAERS, 2014b). A figura 8 é uma imagem, em duas dimensões, do manguezal arbóreo da região de Sundarbans, Bangladeche. A figura 9 é a representação em 3 dimensões de um aquífero aluvial no vale Maules Creek, Austrália.

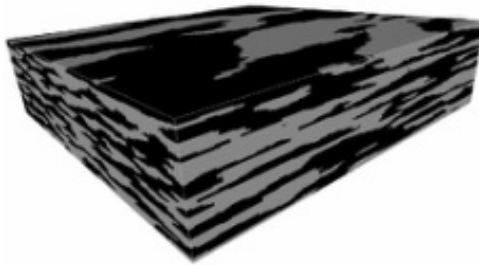
Figura 8 – Imagem de satélite 2D da região Sundarbans em Bangladesh.



Fonte: <http://trainingimages.org/training-images-library.html>

Os métodos tradicionais de simulação se apoiam na modelagem de estatísticas em dois pontos, geralmente as covariâncias e variogramas. Muitos fenômenos são complexos e inviabilizam a captura de seus padrões espaciais por meio de estatísticas de dois pontos. A simulação geoestatística multiponto é um método genérico e se baseia em três mudanças conceituais formalizadas por (GUARDIANO; SRIVASTAVA, 1993). A primeira, afirma que conjuntos de dados podem não ser suficiente para inferir todas as carac-

Figura 9 – Hidro-fácies em um aquífero aluvial.



Fonte: <http://trainingimages.org/training-images-library.html>

terísticas estatísticas que controlam o que se deseja modelar. A segunda é adotar uma estrutura estatística não-paramétrica para representar a heterogeneidade. A terceira mudança conceitual é avaliar a estatística de eventos de dados de múltiplos pontos. As estatísticas multipontos são expressas como funções densidades cumulativas para uma variável aleatória $Z(x)$ condicionadas a eventos de dados locais $d_n = Z(x_1), Z(x_2), \dots, Z(x_n)$, isto é, os valores de Z nos nós vizinhos x_i de x , equação 2.1.

$$f(z, x, d_n) = \text{Prob}(Z(x) \leq z | x) \quad (2.1)$$

Os diferentes algoritmos que compõem este método compartilham elementos comuns, como o uso de um caminho de simulação (geralmente aleatório), a amostragem em distribuições de probabilidades locais e o uso de grades múltiplas (MARIETHOZ; CAERS, 2014a). Deste modo, alguns algoritmos diferem apenas por um pequeno componente, mas mantêm um arquétipo relativamente consistente e que pode ser descrito como no pseudo-código 14.

Algoritmo 1: ALGORITMO MPS GENÉRICO

Entrada: Pelo menos uma variável Z a ser simulada

Grade de simulação SG , com nós denotados por x

Imagem de treinamento TI

[opcional] Conjunto de pontos de dados condicionais

[opcional] mapa local de probabilidade L

Conjunto de parâmetros específicos do algoritmo

Saída: SG , com Z conhecido em todos os nós

1 início

2 Migre os pontos condicionantes para nós correspondentes na grade

3 Analise TI para derivar o banco de dados de padrões DB

4 Defina um caminho de simulação

5 repeat

6 Defina o nó x a ser simulado

7 Encontre N_x , a vizinhança de x

8 Determine a distribuição condicional de $Z(x)$ condicionada a N_x e $L(x)$

9 Amostre um valor nesta distribuição

10 Coloque este valor em SG

11 até *até critério de parada ser encontrado*;

12 Pós-processamento

13 fim

14 retorna SG

2.1 SIMULAÇÃO MULTIPONTO BASEADA EM PONTOS

Caers e Journal (1998) apresentam um modelo de SMP baseado em redes neurais que utilizam imagens de treinamento como fonte de aprendizado. Deste modo, as redes neurais podem ser ensinadas a coletar estatísticas de múltiplos pontos e utilizá-las para gerar modelos probabilísticos condicionados aos dados reais. Para realizar esta tarefa, as distribuições condicionais aprendidas pela rede, relacionam o valor em qualquer região da imagem aos dados da sua vizinhança, dentro de um *template* centrado na região simulada, semelhante à ilustração da figura 10. A rede neural é treinada para determinar os valores de y em qualquer região, dados os valores da vizinhança x . Matematicamente, a rede neural modela a função densidade de probabilidade (FDP) local da equação 2.2, ou sua integral, a função de distribuição

Figura 10 – Definição de uma vizinhança.



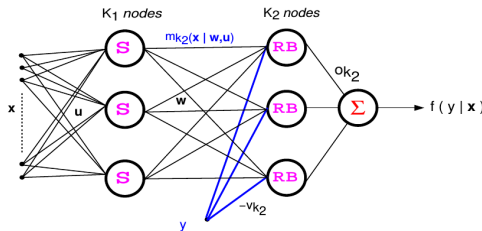
Fonte: (CAERS; JOURNAL, 1998)

condicional cumulativa (FDCC) .

$$f(y|x) = Pr(y < Y < y + dy|x) \quad (2.2)$$

Na fase de treinamento, são inseridos os dados y condicionados a x , presentes na imagem de treinamento, e a rede determina o modelo $f(y|x)$. Na fase de simulação, é percorrido cada nó i dos N nós da grade de simulação e um valor simulado y_i é desenhado, condicionado aos dados x_i encontrados dentro do *template* definido. A figura 11 apresenta a proposta de rede neural para simulação estocástica multiponto. Os parâmetros w e u definem a conexão não-linear entre as médias das funções de base radial RB e os dados condicionantes x , k_1 é o número de neurônios da primeira camada com função de ativação não-linear S , os k_2 é o número de neurônios da segunda camada escondida da rede, a qual recebe adicionalmente os pontos y ponderados pelos pesos $-(V_{k_2})$. A arquitetura proposta permite interpretar a rede como uma distribuição de mistura de gaussianas e a tarefa de otimização da rede pode ser realizada com algoritmo de maximização de expectativa.

Figura 11 – Rede neural que retorna a PDF como saída.



Fonte:(CAERS; JOURNAL, 1998)

A abordagem multiponto baseada em redes neurais (CAERS; JOURNAL, 1998), fortaleceu o interesse no estudo dos métodos multipontos. Em 2002, o algoritmo *SNESIM* (STREBELLE, 2002a) é apresentado como uma alternativa para realizar simulações sequenciais livres de variogramas e ca-

pazes de lidar com a presença de padrões não-estacionários nas imagens de treinamento. Os autores elucidam que, no lugar de utilizar algum tipo de *kriging* para construir um modelo de probabilidade para a propriedade a ser simulada, o *SNESIM* determina diretamente o modelo de probabilidades a partir da imagem de treinamento sem o uso de variogramas.

O método desenvolvido por Strebelle (2002a) realizam simulações com base na computação da distribuição de probabilidade condicional, derivada a partir da imagem de treinamento e alocada em uma árvore de busca. Em 2011, Straubhaar et al. (2011) apresentaram o método *IMPALA*, que consiste na implementação do método multiponto original, com a substituição da árvore de busca por uma estrutura de lista. Este, mais simples, permite reduzir a quantidade de memória RAM utilizada pelo algoritmo e, consequentemente, utilizar diferentes listas para guardar dados adicionais igualmente utilizados durante a simulação. Adicionalmente, o uso de diferentes listas facilita a tarefa de paralelização do algoritmo em *clusters* e processadores multinúcleos, e permite lidar com a não-estacionaridade na imagem de treinamento e na grade de simulação. *GROWTHSIM* é um algoritmo similar ao *SNESIM*, exceto pelo uso de um caminho aleatório de vizinhança (ESKANDARIDALVAND; SRINIVASAN, 2010).

O número limitado de dados condicionantes ao redor do ponto simulado leva a negligenciar dados que, embora afastados, possuem informações importantes em escala maior. Com isto, anomalias como desconexões de canais podem ser geradas nas realizações simuladas. O método proposto em 2005 (STREBELLE; REMY, 2005) utiliza mais dados para condicionamento, especialmente dados localizados distante do nó simulado. Uma medida de consistência entre as realizações simuladas e a IT é definida baseada no número de vezes que cada dado simulado teve de ser ignorado para ser capaz de inferir a distribuição de probabilidade condicional no nó. Uma nova simulação é realizada para os valores de nós mais inconsistentes, o que permite melhorar a reprodução de padrões de treinamento sem aumento computacional significativo.

Em 2010 o algoritmo *simulated annealing* foi implementado com intuito de gerar realizações estocásticas de variáveis categóricas por reprodução de estatísticas multipontos (PEREDO; ORTIZ, 2011). A imagem de treinamento é utilizada para determinar as frequências de ocorrência das configurações de cada nó. Estas frequências são usadas como estatística alvo que deve combinar com as imagens estocásticas geradas com o algoritmo. Inicialmente é criada uma imagem aleatória da distribuição espacial das categorias, em seguida, os nós são perturbados aleatoriamente e a diferença entre a estatística alvo e a estatística atual da imagem é calculada.. A perturbação é aceita se as estatísticas são mais próximas do alvo, ou condicionalmente rejeitada. A

principal relevância deste trabalho está no ganho de performance alcançado com a implementação paralelizada do algoritmo e sua execução sob um esquema especulativo.

Os algoritmos, *SNESIM* e *IMPALA*, se baseiam no armazenamento dos eventos de dados encontrados na imagem de treinamento, são restritos à simulação de variáveis categóricas e necessitam de um modelo para a vizinhança do ponto simulado. Uma nova abordagem para SMP procede com amostragens diretamente sobre a imagem de treinamento para um determinado evento, de modo que o uso de um banco de dados de evento se torna dispensável (MARIETHOZ; RENARD; STRAUBHAAR, 2010). O método de amostragem direta permite estender a aplicação de geoestatística multiponto para variáveis contínuas e categóricas e se destaca por respeitar a distribuição de probabilidade condicional sem realizar a sua computação. Na utilização de múltiplas variáveis, uma função d é escolhida apropriadamente para variáveis categóricas ou contínuas. É possível utilizar a função de distância ($d(\cdot)$) para controlar as proporções globais e locais, ou impor uma média local. Para tanto, é adicionado um fator de erro (E_p) à fdp, que quantifica a diferença entre a probabilidade alvo $L(x)$ e a proporção atual $P(x)$ na grade SG parcialmente simulada, na forma da equação 2.3.

$$E_p |P(x) - L(x)| \quad (2.3)$$

Em 2012, é apresentado o guia prático para realizações de simulação multiponto com o algoritmo de amostragem direta (DS) (MEERSCHMAN et al., 2013). Neste trabalho, os autores apresentam uma análise quantitativa sobre os parâmetros de entrada do algoritmo, listam aqueles que são mais importantes, avaliam seu impacto nas simulações realizadas e demonstram a capacidade do algoritmo em realizar simulações 3D. O trabalho vai além e apresenta o resultado qualitativo obtido com pós-processamento e o potencial do algoritmo para simular problemas com duas variáveis e reproduzir dados condicionantes. Os autores apresentam estratégias para definir os valores adequados para os parâmetros t (*threshold*) e n (número de vizinhos) nos casos contínuo e categórico. Para este, se recomenda realizar uma etapa de pós-processamento para remoção de ruído. Em 2013, este método sofreu modificações para simular em bloco, no lugar de simular ponto a ponto. Esta modificação acelerou a execução do DS em duas ordens de magnitude e permitiu realizar aplicações com simulações ainda maiores (REZAEE et al., 2013).

2.2 SIMULAÇÃO MULTIPONTO BASEADA EM BLOCOS

Uma observação comum na comparação entre a IT e as realizações geradas é a existência de padrões exatamente reproduzidos da IT para a grade de simulação. Este fenômeno é uma consequência de coerência de textura e o tamanho limitado da imagem de treinamento. Algoritmos de simulação baseados em fragmentos ou blocos (*patch*) tendem a agrupar valores de simulação que estejam próximos uns dos outros na imagem de treinamento. O algoritmo de simulação baseada em filtro, *FILTERSIM*, foi concebido para superar a limitação do *SNESIM* em funcionar apenas para variáveis categóricas. O *FILTERSIM* agrupa todos os padrões obtidos da IT dentro de um conjunto de classes de padrões e, em cada local de simulação, identifica a classe de padrão que mais se assemelha ao evento de dado condicionante. Em seguida, um padrão é amostrado dentro do protótipo de padrões e copiado para a grade de simulação (ZHANG; SWITZER; JOURNAL, 2006). Na versão original do algoritmo, a escolha da classe de padrão é baseada na distância de pixel entre cada classe de padrão e o evento de dado condicionante, o que torna o algoritmo custoso computacionalmente para simulações 3D. Entretanto, uma nova abordagem é adotada, na qual o cálculo da distância é substituído por comparação de pontuação de filtro, ou seja, a diferença da pontuação de filtro entre o padrão condicionante e cada protótipo de padrão (WU; ZHANG; JOURNAL, 2008). O cálculo baseado em pontuação permite reduzir o custo computacional por conta da redução dimensional dos dados.

O algoritmo *SIMPA* (ARPAT; CAERS, 2007) abandona a abordagem por reprodução de estatísticas e reprodução probabilística, e se baseia em reconhecimento de padrões. Neste sentido, a imagem de treinamento é utilizada como um banco de dados de padrões, nos quais é possível identificar entidades com algum significado. Percorrendo um caminho aleatório, o algoritmo atribui a cada localização visitada, um padrão compatível com os dados locais disponíveis e quaisquer padrões previamente simulados. Assim, todo o padrão é copiado para a grade de simulação.

Cronologicamente próximo ao *SIMPA*, é apresentado à comunidade científica o *Patchwork Simulation* (OUASSINI et al., 2008). Algoritmo de simulação sequencial estocástica, o *SIMPAT* é limitado a variáveis binárias e reproduz propriedades de conectividade e a geometria de agrupamentos de escala pequena, ou blocos. São estimadas as probabilidades de transição entre blocos adjacentes na imagem de treinamento. As probabilidades são usadas para simular cada bloco da imagem e cada bloco é conectado ao bloco anteriormente simulado. O procedimento de conexão dos blocos reproduz estruturas de pequena escala e propriedades de conectividade local. O método apresentou performance considerada boa em imagens com padrão de tabu-

leiro de damas ou imagens compostas de pequenos discos. Para estas imagens, a simulação com *patchwork simulations* alcançou resultados melhores que o algoritmo de Strebelle (STREBELLE, 2002b, 2002a).

Diferente dos algoritmos que constroem banco de dados de padrões e pequenos eventos de dados, o *CCSIM* utiliza função de similaridade entre o padrão gerado e a imagem de treinamento, de modo a obter realizações que combinem com um evento de dado existente na imagem, isto é, correlação cruzada (TAHMASEBI; HEZARKHANI; SAHIMI, 2012). As medidas de correlação cruzada são combinadas com estratégias para reproduzir a continuidade e a reprodutibilidade dos padrões. O método *image quilting* (EFROS; FREEMAN, 2001) é idêntico ao *CCSIM*, exceto pelo uso de programação dinâmica para realizar, na área de sobreposição, a correção de padrões que não combinam.

3 REDES NEURAIIS CONVOLUCIONAIS

As redes neurais convolucionais , também conhecidas como *Deep Learning*, estão no centro das pesquisas relacionadas a aprendizagem de máquina. As RNC são muito populares nas áreas de classificação de imagens (KRIZHEVSKY; SUTSKEVER; HINTON, ; HE et al., 2014), visão computacional (ZHANG et al., 2014), reconhecimento de faces (SUN et al., 2014), à segmentação de imagem e reconhecimento de ações em vídeos; Os fatores que conferem às RNC a importância atual são: (i) a eficiência no treinamento em GPUs modernas, (ii) a proposta das Unidades Lineares Retificadas (em tradução livre do termo em inglês Rectified Linear Unit - ReLU) as quais tornam a convergência mais rápida, e (iii) o acesso amplo aos dados para treinamento de modelos extensos (ImageNet)(KRIZHEVSKY; SUTSKEVER; HINTON,).

3.1 DEEP LEARNING NA RESTAURAÇÃO DE IMAGENS

De acordo com (DONG et al., 2016), processo de extrair uma imagem de alta resolução a partir de outra imagem de menor resolução consiste, conceitualmente, em um mapeamento que segue os três passos a seguir:

- Extração e representação de blocos: consiste na obtenção, a partir da imagem Y , dos blocos que possuem sobreposição e representar cada bloco como um vetor de alta dimensão. Estes vetores representam um conjunto de mapas de características.
- Mapeamento não-linear: mapeamento não-linear de cada vetor de alta dimensão em outro vetor também de alta dimensão. Cada vetor mapeado é a representação de um bloco de alta resolução.
- Reconstrução: agrega as representações de alta resolução obtida no passo 2 para gerar a imagem de alta resolução final. Esta imagem deve ser similar à realidade geológica X .

É possível sumarizar o problema da super-resolução em imagens da seguinte forma: seja a imagem Y , é possível obter uma imagem $F(Y)$ que seja tão similar quanto possível à imagem da realidade geológica de alta resolução X ?

4 ABORDAGEM DE SOLUÇÃO

A simulação multiponto gera realizações que reproduzam padrões estatísticos inferidos a partir de alguma fonte, usualmente uma imagem de treinamento. Como ferramenta de modelagem, os algoritmos de aprendizagem de máquina são universais, adaptativos, não-lineares, robustos e eficientes. Eles podem alcançar soluções aceitáveis para problemas de classificação, regressão e modelagem de densidade de probabilidade em espaço de alta dimensão e com características espacialmente referenciados. Com base no levantamento literário, se observa a possibilidade de explorar a SMP por meio da implementação baseada em aprendizagem de máquina. Dentre os raros trabalhos encontrados estão o uso de algoritmos genéticos (GA) na simulação de variáveis categóricas para reproduzir estatísticas multipontos (PEREDO; ORTIZ, 2012). Este algoritmo requer alto desempenho computacional e, portanto, depende da disponibilidade de computadores com múltiplos núcleos, bem como unidades de processamentos gráficos, características inerentes aos métodos de GA. A simulação multiponto com redes neurais apresentada por Jef Caers (CAERS; JOURNAL, 1998) explora uma solução neural para a simulação pixel a pixel e, embora haja diversas citações a este trabalho, nenhuma delas está relacionada com a melhoria, expansão ou aplicação do método proposto. Este fato se dá, possivelmente, por conta do interesse no desenvolvimento de novos algoritmos, os quais estão discriminados na seção anterior.

REFERÊNCIAS

- ARPAT, G. B.; CAERS, J. Conditional simulation with patterns. **Mathematical Geology**, v. 39, n. 2, p. 177–203, 2007. ISSN 1573-8868.
- CAERS, J.; JOURNEL, A. Stochastic reservoir simulation using neural networks trained on outcrop data. **SPE Annual Technical Conference and Exhibition**, 1998.
- CAERS, J.; ZHANG, T. Multiple-point geostatistics: A quantitative vehicle for integrating geologic analogs into multiple reservoir models. **AAPG Memoir 80**, p. 383–394, 2004.
- DEUTSCH, C. **Geostatistical Reservoir Modeling**. Oxford University Press, 2002. (Applied geostatistics series). ISBN 9780195138061. Disponível em: <<https://books.google.com.br/books?id=1WvxAAAAMAAJ>>.
- DONG, C. et al. Image super-resolution using deep convolutional networks. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 38, n. 2, p. 295–307, Feb 2016.
- EFROS, A. A.; FREEMAN, W. T. Image quilting for texture synthesis and transfer. In: **Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques**. New York, NY, USA: ACM, 2001. (SIGGRAPH '01), p. 341–346. ISBN 1-58113-374-X. Disponível em: <<http://doi.acm.org/10.1145/383259.383296>>.
- ESKANDARIDALVAND, K.; SRINIVASAN, S. Reservoir modelling of complex geological systems - a multiple-point perspective. **Journal of Canadian Petroleum Technology**, v. 49, n. 8, p. 59–68, 2010.
- GUARDIANO, F. B.; SRIVASTAVA, R. M. **Geostatistics Tróia 92: Volume 1**. Dordrecht: Springer Netherlands, 1993. 133–144 p.
- HE, K. et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. In: _____. **Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part III**. Cham: Springer International Publishing, 2014. p. 346–361.
- JOURNEL, A. G. Fundamentals of geostatistics in five lessons. In: _____. **Fundamentals of Geostatistics in Five Lessons**. [S.l.]: American Geophysical Union, 1990. p. 40.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.]. p. 2012.

MARIETHOZ, G.; CAERS, J. **Multiple-Point Geostatistics Algorithms**. John Wiley & Sons, Ltd, 2014. 155–171 p. ISBN 9781118662953. Disponível em: <<http://dx.doi.org/10.1002/9781118662953.ch9>>.

MARIETHOZ, G.; CAERS, J. Training image construction. In: _____. **Multiple-Point Geostatistics**. [S.l.]: John Wiley & Sons, Ltd, 2014. p. 221–238. ISBN 9781118662953.

MARIETHOZ, G.; RENARD, P.; STRAUBHAAR, J. The direct sampling method to perform multiple-point geostatistical simulations. **Water Resources Research**, v. 46, n. 11, p. n/a–n/a, 2010. ISSN 1944-7973. W11536. Disponível em: <<http://dx.doi.org/10.1029/2008WR007621>>.

MEERSCHMAN, E. et al. A practical guide to performing multiple-point statistical simulations with the direct sampling algorithm. **Computers and Geosciences**, v. 52, p. 307–324, 2013. ISSN 0098-3004. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0098300412003299>>.

OUASSINI, A. E. et al. A patchwork approach to stochastic simulation: A route towards the analysis of morphology in multiphase systems. **Chaos, Solitons & Fractals**, v. 36, n. 2, p. 418 – 436, 2008. ISSN 0960-0779. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0960077906006680>>.

PEREDO, O.; ORTIZ, J. M. Parallel implementation of simulated annealing to reproduce multiple-point statistics. **Computers & Geosciences**, v. 37, n. 8, p. 1110 – 1121, 2011. ISSN 0098-3004. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0098300410003407>>.

PEREDO, O.; ORTIZ, J. M. Geostatistics oslo 2012. In: _____. Dordrecht: Springer Netherlands, 2012. cap. Multiple-Point Geostatistical Simulation Based on Genetic Algorithms Implemented in a Shared-Memory Supercomputer, p. 103–114.

REZAEI, H. et al. Multiple-point geostatistical simulation using the bunch-pasting direct sampling method. **Computers & Geosciences**, v. 54, p. 293 – 308, 2013. ISSN 0098-3004.

STRAUBHAAR, J. et al. An improved parallel multiple-point algorithm using a list approach. **Mathematical Geosciences**, v. 43, n. 3, p. 305–328, 2011. ISSN 1874-8953.

STREBELLE, S. Conditional simulation of complex geological structures using multiple-point statistics. **Mathematical Geology**, v. 34, n. 1, p. 1–21, 2002. ISSN 1573-8868.

STREBELLE, S. **Sequential simulation drawing structures from training images**. Tese (Doutorado) — Stanford University, 2002. Disponível em: <<http://www-sul.stanford.edu/depts/branner/collections/theses.html>>.

STREBELLE, S.; REMY, N. Geostatistics banff 2004. In: _____. Dordrecht: Springer Netherlands, 2005. cap. Post-processing of Multiple-point Geostatistical Models to Improve Reproduction of Training Patterns, p. 979–988. ISBN 978-1-4020-3610-1.

SUN, Y. et al. Deep learning face representation by joint identification-verification. In: GHAHRAMANI, Z. et al. (Ed.). **Advances in Neural Information Processing Systems 27**. Curran Associates, Inc., 2014. p. 1988–1996. Disponível em: <<http://papers.nips.cc/paper/5416-deep-learning-face-representation-by-joint-identification-verification.pdf>>.

TAHMASEBI, P.; HEZARKHANI, A.; SAHIMI, M. Multiple-point geostatistical modeling based on the cross-correlation functions. **Computational Geosciences**, v. 16, n. 3, p. 779–797, 2012. ISSN 1573-1499.

WU, J.; ZHANG, T.; JOURNAL, A. Fast filtersim simulation with score-based distance. **Mathematical Geosciences**, v. 40, n. 7, p. 773–788, 2008. ISSN 1874-8953.

ZHANG, N. et al. Part-based r-cnns for fine-grained category detection. In: _____. **Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I**. Cham: Springer International Publishing, 2014. p. 834–849.

ZHANG, T.; SWITZER, P.; JOURNAL, A. Filter-based classification of training image patterns for spatial simulation. **Mathematical Geology**, v. 38, n. 1, p. 63–80, 2006.