



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

Dissertação de Mestrado

**ANÁLISE DE INCERTEZAS EM RESERVATÓRIOS DE  
PETRÓLEO:  
UMA ABORDAGEM ESTOCÁSTICA USANDO  
MÉTODOS NÃO INTRUSIVOS**

Apresentada por:

**André Menezes Oliveira Guimarães**

Orientador: Ramiro Brito Wilmersdorf, Ph.D.

Co-orientador: Ézio da Rocha Araújo, DSc.

Recife, Pernambuco  
Fevereiro de 2016

**Análise de Incertezas em Reservatórios de  
Petróleo**  
*Uma Abordagem Estocástica Usando Métodos Não  
Intrusivos*

André Menezes Oliveira Guimarães

Dissertação de Mestrado submetida ao Corpo Docente do Curso de Pós-Graduação em Engenharia Civil da Universidade Federal de Pernambuco, como parte dos requisitos necessários à obtenção do grau de Mestre em Ciências em Engenharia Civil.

Área de Concentração: Engenharia de Petróleo

Ramiro Brito Willmersdorf, Ph.D.

Orientador

Ézio da Rocha Araújo, DSc.

Coorientador

Recife, Pernambuco, Brasil

©André Menezes Oliveira Guimarães, Fevereiro de 2016

Catálogo na fonte  
Bibliotecária Valdicéa Alves, CRB-4 / 1260

G963a Guimaraes, André Menezes Oliveira.  
Análise de incertezas em reservatórios de petróleo: uma abordagem estocástica usando métodos não intrusivos / André Menezes Oliveira. - 2016.  
85folhas, Il., e Tab.  
  
Orientador: Prof. Ph.D. Ramiro de Brito Willmersdorf.  
Coorientadora: Prof<sup>o</sup> DSc Ézio da Rocha Araújo.  
  
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia Civil, 2016.  
Inclui Referências.  
  
1. Engenharia Civil. 2. Engenharia de petróleo. 3. Simulação de reservatórios análise de incertezas. 4. Caracterização de reservatórios. I. Willmersdorf, Ramiro de Brito (Orientador). II. Araújo, Ézio da Rocha (Coorientador). III. Título.

UFPE

624 CDD (22. ed.)

BCTG/2016-257



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

A comissão examinadora da Defesa de Dissertação de Mestrado

**ANÁLISE DE INCERTEZAS EM RESERVATÓRIOS DE PETRÓLEO:  
UMA ABORDAGEM ESTOCÁSTICA USANDO MÉTODOS NÃO INTRUSIVOS**

defendida por

André Menezes Oliveira Guimarães

Considera o candidato APROVADO

Recife, 29 de fevereiro de 2016

Orientadores:

---

Prof. Dr. Ramiro Brito Willmersdorf - UFPE  
(orientador)

---

Prof. Dr. Ézio da Rocha Araújo - UFPE  
(coorientador)

Banca Examinadora:

---

Prof. Dr. Ramiro Brito Willmersdorf - UFPE  
(orientador)

---

Prof.<sup>a</sup> Dr.<sup>a</sup> Liliane de Allan Fonseca – UFPE  
(examinadora externa)

---

Prof. Dr. Darlan Karlo Elisiário de Carvalho - UFPE  
(examinador interno)

## **Dedicatória**

Dedico este trabalho aos meus filhos Guilherme e Gabriela e a Roberta, minha eterna namorada.

## Agradecimentos

Ao amigo e orientador Prof. Ézio, pela orientação valiosa, segura e incansável, pelas inúmeras demonstrações de confiança e palavras de estímulo durante o desenvolvimento deste trabalho.

Aos meus filhos Guilherme e Gabriela, minhas fortalezas nos momentos de insegurança. À minha querida esposa, Roberta, pelo apoio incondicional e compreensão nos vários momentos de ausência. Aos meus pais João Soares e Clotildes pelos valores e princípios transmitidos.

À amiga Liliane Fonseca pela valiosa companhia, palavras de incentivo e por sempre acreditar na conclusão desse trabalho.

Aos colegas do programa de pós-graduação em Engenharia Civil (PPGEC) com quem tive o prazer de conviver, Manuela, Felipe, Marlon, Gustavo, Adriano, Darlan, Diego, Luiz Gustavo, Fabiana.

Aos professores do PPGEC pela oportunidade de aprender valiosas lições para além da vida acadêmica, Paulinho, Bernardo, Silvana, Ramiro, Leonardo.

Aos funcionários do PPGEC, em especial a Andrea pela presteza e competência demonstradas em várias situações.

À PETROBRAS, na pessoa da minha "chefe" Gislene Aparecida da Silva, pela confiança, oportunidade e suporte necessário durante a execução desse trabalho.

Aos colegas do ATP-SM/RES, Marcos Aurélio e Felipe que, nas conversas de corredor e nas pausas do cafezinho, sempre me motivaram a concluir a pesquisa.



## Resumo

Os projetos de desenvolvimento no segmento de Exploração e Produção (E&P) da indústria do petróleo exigem altos dispêndios de capital, e possuem intrinsecamente um elevado grau de risco associado aos diversos tipos de incertezas presentes nas variáveis do problema. A tomada de decisão nesses projetos, seja ela baseada em maximização de lucro, de reservas ou de produção acumulada, é fundamentada na previsão sobre o comportamento futuro do reservatório. Entretanto, estas previsões apresentam um alto grau de variabilidade, sobretudo, devido às incertezas originadas em diversos fatores do processo de modelagem, dentre os quais podemos citar: (1) incertezas na modelagem geológica (geometria, estrutura, distribuição espacial das propriedades, etc.), (2) incertezas na caracterização dos fluidos do reservatório, (3) incertezas sobre o comportamento do sistema rocha-fluido e (4) incertezas econômicas.

A necessidade de quantificar as incertezas em projetos de E&P surge naturalmente a partir do processo de modelagem, pois é simplesmente impraticável, ou mesmo impossível, representar com precisão todas as propriedades de um sistema dinâmico real. Para obter uma maior confiabilidade nas curvas de previsão de produção de fluidos, resultantes do modelo de simulação numérica, é de extrema importância uma modelagem que consiga caracterizar adequadamente propriedades do meio poroso, tais como a permeabilidade e porosidade. Contudo, estas propriedades são inerentemente heterogêneas e exibem um alto grau de variabilidade espacial o que, somada à natureza esparsa dos dados disponíveis, contribuem para o surgimento de incertezas na caracterização dessas propriedades e, conseqüentemente, incertezas na previsão do comportamento futuro do reservatório.

A proposta deste trabalho é apresentar um modelo substituto para alguma variável dinâmica de interesse que leve em conta as incertezas existentes nas propriedades estáticas do reservatório. A abordagem seguida neste trabalho é baseada na expansão espectral dos campos randômicos de entrada por meio da expansão de *Karhunen-Loeve* (em sua versão linear ou não linear) e das variáveis aleatórias de saída por meio da expansão do caos polinomial (*PCE*). Esta abordagem é geralmente uma ou mais ordens de grandeza mais rápida do que técnicas baseadas em Monte Carlo e tem a vantagem de que o modelo de simulação pode ser tratado como caixa-preta, implicando numa implementação relativamente direta.

**Palavras Chave:** Engenharia de Petróleo. Simulação de Reservatórios Análise de Incertezas. Caracterização de Reservatórios.

## Abstract

The production development projects in Exploration and Production (E&P) segment of the oil industry require high capital expenditures, and inherently have a high degree of risk associated with various types of uncertainties present in the problem variables. Decision making in these projects, be it based on maximizing profits, reserves or cumulative oil production is based on the forecast of the future reservoir behavior. However, these predictions have a high degree of variability, especially because of the uncertainties arising in various factors in the modeling process, among which include: (1) uncertainties in the geological modeling (geometry, structure, spatial distribution of the properties, ...), (2) uncertainty in the characterization of reservoir fluids, (3) uncertainties about the behavior of the rock-fluid system and (4) economical assumptions.

The need to quantify uncertainties in E&P projects arises naturally from the modeling process, because it is simply impractical, if not impossible, to accurately represent all the properties of a real dynamic system. For greater reliability in the fluid production forecast resulting from numerical simulation model it is of utmost importance as modeling that can adequately characterize properties of the porous medium such as permeability and porosity. However, these properties are inherently heterogeneous and exhibit a high degree of spatial variability which, in addition to the sparse nature of the data available, contribute to the emergence of uncertainties in the characterization of these properties and, consequently, uncertainties in predicting reservoir behavior.

The purpose of this thesis is to present a proxy model for some dynamic variable of interest that takes into account uncertainties in the static properties of the reservoir. The approach followed in this work is based on the spectral expansion of the input random fields through Karhunen- Loeve expansion (in its linear or non-linear version) and the random variables output by the polynomial chaos expansion (PCE). This approach is typically one or more orders of magnitude faster than techniques based on Monte Carlo, and has advantage that the simulation model can be treated as black-box, implying in a relatively straightforward implementation.

**Keywords:** Petroleum Engineer. Reservoir Simulation. Uncertainty Analysis. Reservoir Characterization

# Lista de Tabelas

4.1 Resultados da Análise de Risco para o Caso 1 .....	59
4.2 Resultados da Análise de Risco para o Caso 2 .....	74

# Lista de Figuras

2.1 Idéia básica do método KPCA (adaptado de SARMA, 2006) . . . . .	28
2.2 Exemplo de uma RL com 3 Fácies dispostas sem ordenamento sequencial	33
2.3 Distribuição de probabilidades e determinação do ponto de corte para um exemplo com 2 fácies. Adaptado de CHEN ET. AL., 2015. . . . .	35
3.1 Característica do método PCM: Englobar as regiões de alta probabilidade.	42
4.1 Modelo de Simulação do Caso 1 sobre o mapa de permeabilidade. . . . .	47
4.2 Realizações originais do logaritmo da Permeabilidade. . . . .	49
4.3 Magnitude dos autovalores da matriz de covariância e o % de variância explicada por cada um deles. . . . .	50
4.4 Realizações do campo de permeabilidades: (A) Realizações originais; (B) Reconstrução utilizando expansão K-L com 120 autovalores; (C) Reconstrução utilizando expansão K-L com 30 autovalores. . . . .	52
4.5 Curva de Risco para o VPL obtido via PCE+PCM: (A) Histograma do VPL (milhões \$); (B) Distribuição Acumulada do VPL (milhões \$). . . . .	53
4.6 Convergência da média e da variância do VPL a partir das realizações geoestatísticas: (A) Média do VPL em (\$ milhões); (B) Variância do VPL em (\$ milhões). . . . .	54
4.7 Comparação da curva de risco para o VPL (em \$ milhões) obtidos pelos métodos PCE+PCM e Monte Carlo. . . . .	55
4.8 Exemplo de amostra usando sequência de Sobol. . . . .	56
4.9 Coeficientes da expansão PCE pelo método NISP. . . . .	57
4.10 Resultados NSIP após validação cruzada: (A) Coeficientes da expan- são; (B) comparação entre valores preditos pela expansão versus valores simulados. . . . .	57
4.11 Resultados do NISP: (A) Histograma do VPL (\$ milhões); (B) Dis- tribuição Acumulada do VPL (\$ milhões). . . . .	58

## LISTA DE FIGURAS

4.12	Comparação dos resultados para o VPL obtidos via NISP e pelo Método de Monte Carlo. . . . .	58
4.13	Exemplo de campo de permeabilidades usado no Caso 2. . . . .	60
4.14	Imagem de Treinamento utilizada para geração das realizações do campo de permeabilidade do caso 2 (Caers e Zhang, 2004). . . . .	60
4.15	Exemplos de realizações do logaritmo da permeabilidade - Caso 2. . . . .	61
4.16	Autovalores da Matriz Kernel usada no KPCA: (A) Magnitude do autovalores; (B) Percentual da variabilidade dos componentes principais. . . . .	62
4.17	Realizações da permeabilidade do caso 2 usando KPCA: (A) Considerando 400 autovalores (80% da variabilidade); (B) Considerando 30 autovalores (40% da variabilidade). . . . .	63
4.18	Exemplos de realizações reconstruídas com Pluri-PCA - Caso 2. . . . .	65
4.19	Caso 2 - exemplo de regra de litologia e ponto de corte determinados por uma das realizações originais. . . . .	66
4.20	Caso 2 - autovalores da matriz de covariância a partir do vetor de indicadores de fácies expandido: (A) Magnitude dos autovalores; (B) Percentual da variabilidade explicada pelos componentes principais. . . . .	67
4.21	Caso 2 - Diagrama esquemático exemplificando a geração de realizações aleatórias usando o pluri-PCA. . . . .	68
4.22	Curva de risco para o VPL obtida pelo método PCE+PCM: (A) Histograma do VPL; (B) Distribuição Acumulada do VPL. . . . .	69
4.23	Curva de risco do VPL a partir das 1000 realizações geoestatísticas: (A) Histograma do VPL (\$ milhões); (B) Distribuição Acumulada do VPL (\$ milhões). . . . .	70
4.24	Comparação do VPL (\$ milhões) entre PCE+PCM e Monte Carlo. . . . .	70
4.25	Exemplo de amostra usando sequência de Sobol - Caso 2. . . . .	71
4.26	Coefficientes da expansão PCE obtidos por NISP. . . . .	72
4.27	Caso 2 - resultados do método NISP: (A) coeficientes da expansão após validação cruzada; (B) comparação entre valores preditos pela expansão versus valores simulados. . . . .	73
4.28	Caso 2 - curva de risco do VPL (\$ milhões) obtida via NISP: (A) histograma; (B) curva de distribuição acumulada. . . . .	73
4.29	Caso 2 - comparação da curva de risco para o VPL (\$ milhões) obtidas por NISP e pelas realizações geoestatísticas. . . . .	74

# Sumário

<b>1</b>	<b>Introdução</b>	<b>14</b>
1.1	Motivação . . . . .	14
1.2	Incertezas na modelagem de reservatórios de petróleo . . . . .	15
1.3	Escopo da Dissertação . . . . .	19
1.4	Organização dos capítulos . . . . .	19
<b>2</b>	<b>Parametrização dos Campos Randômicos de Entrada</b>	<b>21</b>
2.1	Expansão de Karhunen-Loeve . . . . .	22
2.1.1	Interpretação Geométrica da Expansão de Karhunen-Loeve . . . . .	24
2.2	<i>KPCA (Kernel Principal Component Analysis)</i> . . . . .	25
2.2.1	Expansão de Karhunen-Loeve usando formulação de kernel . . . . .	26
2.2.2	<i>PCA</i> não linear usando kernel . . . . .	27
2.2.3	O problema da pré-imagem . . . . .	30
2.3	Método Pluri-PCA . . . . .	32
2.3.1	Simulação Plurigaussiana . . . . .	32
2.3.2	Formulação mista Pluri-PCA . . . . .	33
2.3.3	Regra de Litologia . . . . .	34
<b>3</b>	<b>Parametrização da Função Objetivo</b>	<b>38</b>
3.1	Expansão em Caos Polinomial e Polinômios de Hermite . . . . .	38
3.2	Método de Colocação Probabilística . . . . .	40
3.3	Método Espectral não Intrusivo . . . . .	43
<b>4</b>	<b>Aplicações</b>	<b>45</b>
4.1	Descrição dos Casos Analisados . . . . .	45
4.2	Caso 1 . . . . .	46
4.2.1	Parametrização do <i>VPL</i> via <i>PCM</i> . . . . .	51
4.2.2	Parametrização do <i>VPL</i> com <i>NISP</i> . . . . .	55

## SUMÁRIO

4.3 Caso 2 .....	59
4.3.1 Parametrização do campo de permeabilidade .....	62
4.3.2 Parametrização do <i>VPL</i> via <i>PCM</i> .....	64
4.3.3 Parametrização do <i>VPL</i> com <i>NISP</i> .....	71
<b>5 Conclusões e Trabalhos Futuros</b> .....	<b>76</b>
5.1 Conclusões .....	76
5.2 Trabalhos Futuros .....	78

# Capítulo 1

## Introdução

### 1.1 Motivação

Os projetos de desenvolvimento da produção no segmento de Exploração e Produção da indústria do petróleo exigem altos dispêndios de capital, e possuem intrinsecamente um elevado grau de risco associado aos diversos tipos de incertezas presentes nas variáveis do problema. A tomada de decisão nesses projetos, seja ela baseada em maximização de lucro, de reservas ou de produção acumulada, é fundamentada na previsão sobre o comportamento futuro do reservatório. Entretanto, estas previsões apresentam um alto grau de variabilidade, sobretudo, devido às incertezas originadas em diversos fatores do processo de modelagem numérica para simulação de fluxo, dentre os quais podemos citar: (1) incertezas na modelagem geológica (geometria, estrutura, distribuição espacial das propriedades, etc.), (2) incertezas na caracterização dos fluidos do reservatório, (3) incertezas sobre o comportamento do sistema rocha-fluido e (4) incertezas econômicas.

Assim, a necessidade de quantificar as incertezas em projetos de E&P surge naturalmente a partir do processo de modelagem, pois é simplesmente impraticável, ou mesmo impossível, representar com absoluta exatidão todas as propriedades de um sistema dinâmico real. Para conseguir uma maior confiabilidade nas curvas de previsão de produção de fluidos, resultantes do modelo de simulação numérica, é de extrema importância uma modelagem que descreva, detalhadamente, propriedades do meio poroso tais como a permeabilidade e porosidade. Contudo, estas propriedades são inerentemente heterogêneas e exibem um alto grau de variabilidade espacial. Heterogeneidade espacial significativa e a natureza esparsa dos dados disponíveis contribuem para o surgimento de incertezas na caracterização dessas propriedades e, conseqüentemente,

incertezas na previsão do comportamento do reservatório.

Dos tipos de incertezas listadas acima, este trabalho foca nas incertezas decorrentes do processo de modelagem das fácies e das propriedades petrofísicas (porosidade e permeabilidade) do reservatório. No contexto da análise de incertezas utilizando um modelo substituto, os parâmetros oriundos do processo de modelagem geológica, descritos por técnicas de simulação geoestatística, são considerados como campos randômicos de entrada. São apresentadas e discutidas técnicas de parametrização dessas variáveis visando tornar o processo mais eficiente. Os resultados das simulações numéricas, tratados como funções objetivo, são caracterizados através de sua função densidade de probabilidade (*fdp*) a partir da qual é possível avaliar o grau de risco do projeto.

## 1.2 Incertezas na modelagem de reservatórios de petróleo

A abordagem probabilística dos cenários de previsão de produção com uso de simuladores numéricos de escoamento somente teve início no final dos anos 80 com o trabalho de ØVREBERG ET AL., 1990 [35]. Desde então, diversas abordagens têm sido propostas, dentre as quais destacam-se as técnicas de planejamento de experimentos, simulações de Monte Carlo e o método da superfície de resposta, [4], [5]. Diversos trabalhos abordaram o tema da propagação de incertezas e quantificação do risco na curva de produção utilizando o método da árvore de derivação [25], [45], [39], [3], que executa todas as combinações possíveis através da discretização na curva de probabilidade de ocorrência de cada atributo incerto, e a simulação de Monte Carlo [5], [49], que sorteia os cenários combinando os valores dos atributos segundo as respectivas probabilidades de ocorrência. Uma forma de tornar o processo mais eficiente, de modo que não seja preciso simular todas as combinações possíveis, considerada em [4], [5], [29] e [36] utiliza a metodologia de planejamento de experimentos aliada ao método da superfície de respostas para encontrar um polinômio que represente as respostas obtidas pelas simulações de fluxo, e que substitua o simulador na obtenção de novas respostas. Até então, as metodologias propostas consideravam as propriedades estáticas do reservatório como variáveis determinísticas. No trabalho de MANCEAU ET AL., 2001 [29], é proposto o uso do *Joint Modeling Method* para a análise do impacto das incertezas estocásticas envolvendo várias realizações geoestatísticas. Contudo, a consideração de um número reduzido de realizações por essas metodologias leva a uma avaliação incompleta das

incertezas nas respostas de interesse [47].

Na engenharia de reservatórios, o comportamento dinâmico do escoamento de fluidos em meios porosos é modelado através de equações diferenciais parciais que, devido à sua complexidade, são resolvidas numericamente e constituem o cerne do simulador numérico. Dentre os parâmetros de entrada do modelo numérico, estão as propriedades usadas para representar a rocha-reservatório, tais como tipo de fácies, porosidade, permeabilidade, razão *net-to-gross*, entre outras, que são oriundas do processo de modelagem geológica e determinadas através de simulação geoestatística, geralmente condicionada a alguma informação dita *hard* (análise de dados de poços, por exemplo). A escassez de dados disponíveis e a variabilidade espacial que estes parâmetros exibem, conduzem a incertezas no processo de modelagem. Assim, os parâmetros estáticos usados como entrada na simulação numérica devem ser considerados como campos randômicos (que exibem algum grau de correlação espacial modelada pela geoestatística) e as equações que regem o fluxo em meios porosos se tornam estocásticas. Trabalhos envolvendo abordagens estocásticas apareceram inicialmente na área de hidrogeologia [54], e depois surgiram algumas aplicações na engenharia de petróleo [55].

Grande parte dos trabalhos assumiram que as propriedades estáticas seguem uma distribuição normal ou log-normal e, assim, as abordagens foram desenvolvidas com base nas características de um campo randômico gaussiano, que necessita apenas da média e da covariância para ser completamente determinado. No contexto da geoestatística, diversos algoritmos baseados na estatística de dois pontos estão disponíveis e são bastante utilizados na geração de realizações equiprováveis de campos randômicos gaussianos. Abordagens tradicionais como simulação de Monte Carlo e suas variantes (Hipercubo Latino, Quase Monte Carlo, etc.), apesar de serem de fácil implementação, são em geral, computacionalmente ineficientes e podem ser inacurados para modelos mais realistas em função da grande quantidade de simulações necessárias por estes métodos [53]. Neste contexto, a decomposição de campos randômicos por meio da expansão de *Karhunen-Loeve* (*expansão K-L*), ou análise de componentes principais (*PCA*), tem sido bastante aplicada em diversas áreas do conhecimento para representação do espaço probabilístico por meio de séries polinomiais truncadas reduzindo o problema estocástico a uma sequência de problemas determinísticos [13], [27], [41], [16], [23]. Desta forma, é possível aproximar o modelo geológico de alta resolução através de poucas variáveis aleatórias não correlacionadas, reduzindo a dimensão do problema ao mesmo tempo em que preserva as principais características dos modelos geológicos a priori. A técnica foi utilizada em [14] e [10] aplicada a um modelo de reservatório

real, no contexto de otimização (das vazões de produção e injeção e locação de poços, respectivamente) sob incertezas nas propriedades estáticas do reservatório e consistem nas primeiras aplicações do método no Brasil.

Apesar de conseguir capturar e preservar as principais estruturas de correlação presentes no modelo geológico, as realizações geradas pelo método não são, devido ao efeito de suavização decorrente da técnica, visualmente satisfatórias para os geocientistas principalmente quando o campo randômico de entrada não for gaussiano. O método chamado *Kernel Principal Component Analysis (KPCA)*, que consiste na forma não linear da expansão de *Karhunen-Loeve*, foi utilizado em [42] para preservar os momentos de ordem superior de campos randômicos não gaussianos e reconstruir estruturas geológicas canalizadas. Para capturar as características de feições canalizadas, uma generalização do método *k-means clustering* chamada de *K-SVD* foi proposta em [19] e [20]. Os resultados obtidos demonstram que é possível reproduzir estruturas geológicas complexas a partir de uma quantidade limitada de dados dinâmicos apesar do custo computacional necessário pela técnica ser um fator limitante. Outra abordagem derivada do *PCA* foi proposta em [50] visando a obtenção de resultados mais próximos a valores binários através da adição de um termo de regularização.

De fato, os modelos resultantes do método *PCA* ou *KPCA* provocam uma distorção na *fdp* dos modelos geológicos a priori pois a técnica tende a filtrar a informação de baixa frequência. Diversos trabalhos [57], [58], [60] e [59] utilizaram a técnica de transformação normal em problemas de ajuste de histórico em modelos cuja distribuição a priori não era gaussiana. Inspirado pelo método da gaussiana truncada, CHEN ET AL., 2014 [2] propuseram adicionar uma técnica de mapeamento da função de distribuição acumulada (*cdf*) após o *PCA* convencional para honrar uma *fdp* não gaussiana com múltiplos picos. O método foi aplicado em dois casos, sendo um modelo de reservatório com duas fácies e outro com três. A qualidade do modelo reconstruído com *CDF-PCA* para o modelo com duas fácies foi considerado bastante satisfatório. Já o modelo com três fácies somente mostrou bons resultados quando as fácies seguem um ordenamento sequencial fixo. Um método que combina *PCA* com o modelo plurigaussiano, denominado *pluri-PCA*, foi proposto em [1] para reconstruir modelos de reservatórios com múltiplas fácies. Ao contrário do método *CDF-PCA*, nesta abordagem a ordem das fácies não impacta o resultado final dos modelos reconstruídos.

Para obter uma boa aproximação das variáveis de saída do modelo de forma eficiente, ou seja com o menor número de avaliações (rodadas do simulador), é importante escolher cuidadosamente os valores dos parâmetros incertos. O Método de Galerkin é

geralmente usado para encontrar os coeficientes de uma expansão polinomial [13], [53], [30] e leva a um conjunto de equações desacopladas que determinam estes coeficientes. Para problemas de fluxo multifásico, onde as equações que regem o problema são não-lineares, a abordagem torna-se computacionalmente proibitiva [23]. Uma abordagem alternativa é o método da colocação probabilística (*PCM*), o qual foi introduzido por TATANG ET AL., 1997 [48]. A *expansão K-L* e o *PCM* foram acoplados com sucesso para resolver problemas de fluxo em sub-superfície onde parâmetros incertos são correlacionados [41], [22]. No método *PCM*, as variáveis aleatórias dependentes são representadas empregando-se polinômios bi-ortogonais (expansão em caos polinomial) como as bases do espaço de incertezas. Os polinômios empregados no caos polinomial são ortogonais entre si e também em relação às *fdp*'s das variáveis de entrada. O método é capaz de capturar possíveis relações não-lineares existentes entre as variáveis de entrada e saída e podem, portanto, ser usado como modelo substituto ao modelo de simulação para uma quantificação de incertezas eficiente.

A expansão em caos polinomial (*PCE*) é uma técnica bastante utilizada para quantificação de incertezas em processos dinâmicos. O uso do método foi primeiro proposto por GHANEM e SPANOS, 1991 [11] e tem sido desde então aplicado a vários problemas de engenharia. O método *PCE* foi usado em [42] para quantificação de incertezas num contexto de otimização da produção em uma abordagem *closed-loop*. O método *PCE* possui uma vantagem significativa sobre outros modelos substitutos ao garantir convergência à distribuição real (das respostas de interesse, como, por exemplo, a produção acumulada de óleo) quando a ordem do polinômio e o número de simulações usados para obter a expansão é aumentado [8]. No método *PCM*, os coeficientes da expansão são calculados via colocação ou regressão. Assim, como a maior parte dos métodos de superfície de resposta, o simulador também é tratado como caixa preta. Dois métodos analíticos recentes e computacionalmente eficientes que usam expansão em caos polinomial são o método dos elementos finitos estocásticos [53], [28], [18] e sua combinação com método da perturbação [56], [27]. Estes métodos, contudo, exigem acesso às equações do modelo matemático, ou seja, não é possível tratar o modelo como caixa-preta. Alguns desses métodos são discutidos em detalhes em [17].

Há, contudo, algumas desvantagens no método *PCM* que limitam a aplicabilidade prática em problemas de larga escala. A principal delas é que o número de simulações necessárias para criar o polinômio é diretamente dependente do número de termos da expansão, e este aumenta exponencialmente com o grau e o número de variáveis aleatórias. Uma abordagem similar, denominada de projeção espectral não intrusiva

(*NISP*) [40], também usa *PCE* e se propõe a eliminar alguns dos principais problemas do *PCM*. A principal diferença é que ao invés de utilizar colocação para estimar os coeficientes da expansão, ele utiliza uma abordagem baseada em amostragem. Dependendo da técnica de amostragem empregada, o número de simulações exigido para estimar os coeficientes da expansão *PCE* pode ser independente do número de variáveis aleatórias e da ordem da expansão, resultando em maior eficiência na aplicação a problemas de larga escala.

### 1.3 Escopo da Dissertação

Embora a maioria dos algoritmos aqui apresentados tenham sido desenvolvidos sob a ótica da otimização e/ou ajuste de histórico assistido, os casos estudados se enquadram no contexto da análise de incertezas em reservatórios de petróleo.

A proposta deste trabalho é apresentar um modelo substituto para alguma variável dinâmica de interesse (aqui foi considerado o valor presente líquido - VPL) que leve em conta as incertezas existentes nas propriedades estáticas do reservatório (porosidade, permeabilidade e tipo de fácies). Não foram consideradas variações no cenário do projeto, ou seja o número e tipo dos poços bem como sua locação e o modelo econômico utilizado para calcular o VPL foram considerados determinísticos.

A abordagem seguida neste trabalho é baseada na expansão espectral dos campos randômicos de entrada por meio da expansão de *Karhunen-Loeve* (em sua versão linear ou não linear) e das variáveis aleatórias de saída por meio da expansão do caos polinomial (*PCE*). Esta abordagem é geralmente uma ou mais ordens de grandeza mais rápida do que técnicas baseadas em Monte Carlo, e possui complexidade de implementação computacional similar ao método dos elementos finitos estocásticos, mas tem a vantagem de que o modelo de simulação pode ser tratado como caixa-preta, implicando numa implementação relativamente direta.

### 1.4 Organização dos capítulos

O Capítulo 2 é dedicado à apresentação das metodologias empregadas para a parametrização dos parâmetros incertos (fácies e/ou permeabilidade) que constituem a entrada para o modelo substituto. A depender da classificação do campo randômico em gaussiano ou não gaussiano, um determinado método, conforme será discutido, se mostrará

mais adequado que os outros.

O Capítulo 3 contém as técnicas empregadas para a representação das variáveis de saída do modelo substituto que, neste caso, são as respostas do simulador numérico de fluxo.

O Capítulo 4 descreve os casos utilizados para validar e avaliar as metodologias descritas nos capítulos anteriores bem como discute e compara os resultados alcançados por cada método tendo como referência os resultados obtidos através do Método de Monte Carlo. A performance de cada método para alcançar a solução também é comparada.

O Capítulo 5 apresenta as conclusões deste trabalho de pesquisa e fornece sugestões de trabalhos futuros para aprofundamento dos resultados e continuidade dessa linha de pesquisa

## Capítulo 2

# Parametrização dos Campos Randômicos de Entrada

Neste capítulo, serão apresentadas técnicas de parametrização dos campos randômicos, tipicamente associados às propriedades petrofísicas (porosidade e permeabilidade), que servem de entrada para a simulação numérica de reservatórios de petróleo. A consequência direta dessa parametrização é a possibilidade de redução da dimensionalidade do problema, o que implica em diminuir o esforço computacional envolvido no processo de análise de incertezas cuja principal característica é a necessidade de um grande número de avaliações das funções objetivo, obtidas via simulação numérica. Em muitos casos, uma única simulação pode demorar horas e, assim, a utilização do modelo em sua dimensão completa para fins de análise de incertezas torna-se proibitiva. Neste trabalho, os métodos investigados farão uso de uma função de covariância conhecida, associada a uma função densidade de probabilidade  $f_{dp}$ , ou determinada a partir de um conjunto de amostras (ensemble de realizações). Na primeira parte do capítulo, será apresentada a expansão de *Karhunen-Loeve* como forma eficiente de representar campos randômicos gaussianos que, por possuírem estacionariedade de 2ª ordem (frequentemente estacionários), podem ser caracterizados por sua média e variância (estatística de dois pontos). Em seguida, serão descritas técnicas mais adequadas à parametrização de campos não-gaussianos (estatística multiponto) cuja utilização na modelagem numérica de reservatórios tem crescido bastante nos últimos tempos devido à sua capacidade de representar com maior fidelidade os corpos e feições geológicas.

## 2.1 Expansão de Karhunen-Loeve

A expansão de *Karhunen-Loeve* (expansão K-L) [24] permite representar campos ou processos estacionários e quase-estacionários com covariância conhecida. Qualquer campo ou processo randômico pode ser representado como uma série envolvendo um conjunto de funções determinísticas e seus correspondentes coeficientes aleatórios [15]. O uso da expansão K-L com bases ortogonais determinísticas e coeficientes aleatórios não correlacionados tem gerado grande interesse por causa de sua propriedade de bi-ortogonalidade, ou seja, tanto as bases determinísticas quanto os coeficientes aleatórios correspondentes são ortogonais [15]. Dessa forma, é possível incorporar a informação contida no processo randômico dentro de um conjunto discreto de variáveis aleatórias não correlacionadas.

A expansão K-L de um campo estocástico  $y(x, \theta)$  é baseada na expansão espectral de sua função de covariância  $C(x_1, x_2)$ , onde  $x_1$  e  $x_2$  são as coordenadas espaciais e  $\theta \in \Theta$  (espaço probabilístico) representa a natureza aleatória da determinada propriedade  $y$  [13]. Por definição, a função de covariância é limitada, simétrica e positiva definida. Sendo assim, todas as suas autofunções são mutuamente ortogonais, e formam um conjunto completo cobrindo o espaço ao qual  $y(x, \theta)$  pertence. Se este conjunto determinístico é usado para representar o processo  $y(x, \theta)$  então os coeficientes aleatórios usados na expansão são também ortogonais [13]. Para um conjunto discreto de realizações centradas (média zero) condicionadas ou não de um campo randômico  $y_k, k = 1, \dots, N_R, (y_k \in R^{N_c})$ , a matriz de covariâncias pode ser calculada com a seguinte expressão [42]:

$$C = \frac{1}{N_R} \sum_{j=1}^{N_R} y_j y_j^T \quad (2.1)$$

onde  $N_R$  representa o número de realizações, que deve ser grande o suficiente para garantir a convergência da matriz de covariância  $C$ . O desenvolvimento da expansão K-L resulta no seguinte problema de autovalores [42]:

$$\lambda v = C v \quad (2.2)$$

onde  $\lambda$  são os autovalores e  $v$  os autovetores de  $C$  de tamanho  $(N_c \times N_c)$ . A solução da Equação 2.2 pode ser obtida utilizando algoritmos como o *SVD* (*Singular Value Decomposition*). Contudo, esse é um processo que envolve alto custo computacional, com ordem de complexidade  $O(N_c^3)$  [42]. Em problemas com dimensões reais, fica impraticável resolver esta equação numericamente. Para superar esta dificuldade, outra

formulação da expansão  $K-L$  que será apresentada na próxima seção, denominada *Kernel Karhunen-Loeve* ou análise dos componentes principais não linear, têm sido sugerida na literatura.

A forma discreta da expansão  $K-L$  que gera realizações do campo randômico honrando a matriz de covariância  $C$  é dada por [15]:

$$y = E\Lambda^{1/2}\xi \quad (2.3)$$

onde  $E$  é a matriz dos autovetores da matriz de covariância  $C$ ,  $\Lambda$  é uma matriz diagonal dos autovalores de  $C$  e  $\xi$  é um vetor de variáveis aleatórias não correlacionadas e também independentes se o campo randômico for gaussiano. Se os elementos de  $\xi$  forem amostrados segundo a distribuição normal padronizada, o  $y$  obtido usando a equação 2.3 será um campo randômico gaussiano correlacionado com covariância  $C$ . A expansão  $K-L$  proporciona uma parametrização da forma  $y = f(\xi)$ , sendo a relação funcional linear. O tamanho máximo das matrizes  $E$  e  $\Lambda$  é  $Nc \times Nc$  e do vetor  $\xi$  é  $Nc \times 1$ , pois, por questões de ordem prática, é comum utilizar a forma truncada a  $N_M$  termos da expansão, o que significa reter os  $N_M$  maiores autovalores da matriz  $C$ . Neste caso o tamanho de  $E$  fica  $Nc \times N_M$ ,  $\Lambda$ ,  $N_M \times N_M$  e  $\xi$  fica  $N_M \times 1$ .

A expansão  $K-L$  possui convergência quadrática na média desde que o processo estocástico seja gaussiano [11]. Os autovalores  $\lambda_i$  são monotonicamente decrescentes, fato que é garantido pela simetria da função de covariância. A taxa de decaimento de  $\lambda_i$  determina o número  $N_M$  de termos que se faz necessário reter na expansão, o que, por sua vez, está relacionado ao comprimento de correlação do processo randômico que está sendo expandido. Quanto maior for o comprimento de correlação, menor o número de termos necessários na expansão. Descartar os  $(Nc - N_M)$  menores autovalores implica em descartar os menores comprimentos de correlação do campo randômico. Contudo, a expansão  $K-L$ , dentre todas as possíveis parametrizações para campos multigaussianos, é a que minimiza o erro quadrático médio da aproximação [42]. Mesmo quando o campo randômico não for multigaussiano, a expansão  $K-L$  minimiza o erro quadrático médio da aproximação.

O aspecto primordial dessa decomposição espectral é que o campo randômico pode ser representado por meio de funções determinísticas das variáveis espaciais multiplicadas por coeficientes aleatórios independentes. Ou seja, por meio da expansão  $K-L$ , o espaço probabilístico (envoltória de incertezas) é decomposto em séries polinomiais truncadas, reduzindo o problema estocástico a uma seqüência de problemas determinísticos representados por um número muito menor de parâmetros. A grande vantagem é

que os parâmetros podem ser continuamente variados pois, uma vez que os coeficientes da expansão são não correlacionados, qualquer conjunto deles proporciona um campo randômico que honra a geoestatística de dois pontos implícita.

### 2.1.1 Interpretação Geométrica da Expansão de Karhunen-Loeve

A expansão  $K-L$  pode ser entendida como uma operação de mudança de coordenadas no sentido em que permite mudar a representação de um conjunto de vetores dentro de um subespaço vetorial, utilizando outro conjunto de vetores chamados de base (*kernel*) que são as componentes principais. A transformação utilizando a base ortonormal dos autovetores da sua matriz de covariância proporciona a decomposição do campo randômico em componentes não correlacionados projetados sobre essa base, que é ordenada de modo que um subconjunto reduzido de componentes preserve o máximo da variância presente nos dados originais.

Um subespaço  $N_M$ -dimensional proporcionará a melhor aproximação para um conjunto de vetores se, e somente se, for um subespaço de *Karhunen-Loeve* constituído pelos primeiros  $N_M$  autovalores de *Karhunen-Loeve* [33]. Desta forma, a utilização dos autovetores calculados pela expansão  $K-L$  como base de representação do campo randômico de entrada, constitui-se numa condição necessária e suficiente para a formação de uma base que fornecerá a melhor aproximação deste campo, ou seja, aquela que fornecerá menor erro quadrático médio entre o campo original e o campo reconstruído utilizando a nova base. Esta importante característica se deve ao fato de que as bases são obtidas a partir das estatísticas do campo randômico original.

A reconstrução do campo randômico original a partir dos vetores  $\xi$  é outro importante aspecto da análise. Utilizando a forma discreta da expansão e admitindo que  $[E]^{-1} = [E]^T$  pois as colunas desta matriz são formadas por vetores ortonormais, podemos escrever:

$$\xi = \Lambda^{-1/2} E^T (y - \bar{y}) \quad (2.4)$$

Suponha-se, contudo, que ao invés de utilizar todos os autovetores da matriz de covariância  $C$ , seja usada a matriz formada por  $N_M$  autovetores correspondentes aos  $N_M$  maiores autovalores de  $C$  (ordenando os autovalores de forma que  $\lambda_1 > \lambda_2 > \dots \lambda_{N_M}$ ). Os vetores  $\xi$  teriam então a dimensão  $N_M$ , e a reconstrução dada pela Equação 2.4 não reconstruiria mais a matriz  $y$  de forma exata.

Pode ser mostrado [12] que o erro quadrático médio,  $ems$ , entre o campo original e o reconstituído pode ser dado pela expressão:

$$ems = \sum_{j=1}^{Nc} \lambda_j - \sum_{j=1}^{N_M} \lambda_j = \sum_{j=N_M+1}^{Nc} \lambda_j \quad (2.5)$$

A primeira parte da Equação 2.5 indica que o erro é zero se  $N_M = Nc$ . isto é, se todos os autovalores forem preservados. Como os  $\lambda_i$  decrescem monotonicamente, a Equação 2.5 também mostra que o erro pode ser minimizado pela escolha dos  $N_M$  autovalores associados com os maiores valores de  $\lambda$ . Portanto, a expansão  $K-L$  é ótima no sentido que minimiza o erro quadrático médio entre os campos randômicos original e o reconstituído.

## 2.2 KPCA (*Kernel Principal Component Analysis*)

Ao contrário dos campos randômicos gaussianos, cujas funções densidade de probabilidade ( $fdp$ ) conjunta podem ser totalmente caracterizadas pelos momentos de primeira e segunda ordem (média e variância), a caracterização de campos randômicos não-gaussianos exige a determinação de momentos de ordem superior que são geralmente difíceis de obter devido à escassez de dados disponíveis. Assim, na prática, são utilizadas a  $fdp$  marginal e a função de covariância para caracterizar os campos randômicos não-gaussianos. Do ponto de vista geológico, a representação das propriedades petrofísicas por meio de campos não-gaussianos resulta em modelos visualmente mais realistas, pois reproduzem as geometrias e feições dos corpos (canais, *levee*, leques aluviais, leques deltaicos, etc.) mais condizentes com a concepção idealizada inicialmente. As técnicas discutidas a seguir, buscam justamente preservar estas características geológicas mas sem abrir mão da redução da dimensionalidade.

Conforme foi visto na seção anterior, a expansão  $K-L$  (também conhecida com análise dos componentes principais -  $PCA$ ) possui dois aspectos que limitam bastante a sua utilização prática. O primeiro aspecto é que ela preserva apenas a covariância do campo randômico (estatística de dois pontos) e por isso é adequada apenas a campos multigaussianos e não pode ser aplicada para modelar estruturas geológicas complexas com geometrias curvilíneas. O segundo, está relacionado com o elevado custo computacional uma vez que o algoritmo pressupõe a decomposição em valores singulares da matriz de covariância de tamanho  $Nc \times Nc$  e, para modelos grandes, esta operação pode ser bastante demorada mesmo com os recursos computacionais atuais. Ambos os

problemas podem ser resolvidos com uma técnica desenvolvida inicialmente na área de reconhecimento de padrões denominada *kernel principal component analysis (KPCA)* ou análise dos componentes principais não linear [42]. O segundo fator limitante mencionado acima, pode ser facilmente superado ao utilizar uma formulação baseada em funções *kernel* para a solução do problema de autovalores que surge da aplicação da expansão *K-L* ao invés de realizar a decomposição em valores singulares de uma matriz  $N_c \times N_c$ , conforme veremos a seguir. Deste modo, será realizada a decomposição de uma matriz diferente, denominada matriz *kernel* de dimensões  $N_R \times N_R$  ( $N_R$  é o número de realizações necessárias para a convergência da matriz de covariância) que é muito mais eficiente haja visto  $N_R \ll N_c$ . Já a solução para o primeiro aspecto, deriva da utilização da formulação baseada em *kernels*, mas agora utilizando funções de ordem superior.

### 2.2.1 Expansão de Karhunen-Loeve usando formulação de kernel

Usando a definição da matriz de covariância dada pela Equação 2.1 e do problema de autovalores dado pela Equação 2.2, a seguinte equação pode ser obtida [42]:

$$Cv = \frac{1}{N_R} \sum_{j=1}^{N_R} (y_j \cdot v) y_j \quad (2.6)$$

A Equação 2.6 declara que os vetores  $v$ , com  $\lambda \neq 0$ , solução do problema de autovalores dado pela Equação 2.2, devem estar contidos no subespaço gerado pelas  $N_R$  realizações  $y_k, k = 1, \dots, N_R$ . Como também não pode haver mais que  $N_R$  direções ortogonais no subespaço das  $N_R$  realizações, existem apenas  $N_R$  autovalores diferentes de zero associados com estes  $N_R$  autovetores [42]. Desta constatação surgem duas importantes consequências [44]. A primeira é que a Equação 2.2 pode ser escrita como:

$$\lambda(y_k \cdot v) = (y_k \cdot Cv) \quad \forall k = 1, \dots, N_R \quad (2.7)$$

A segunda consequência é que existem coeficientes  $\alpha_j$  tais que

$$v = \sum_{j=1}^{N_R} \alpha_j y_j \quad (2.8)$$

Combinando as Equações 2.7 e 2.8 obtém-se [44]:

$$\lambda \sum_{i=1}^{N_R} \alpha_j (y_k \cdot y_i) = \frac{1}{N_R} \sum_{i=1}^{N_R} \alpha_j \left( y_k \cdot \sum_{j=1}^{N_R} y_j \right) (y_j \cdot y_i) \quad \forall k = 1, \dots, N_R \quad (2.9)$$

Definindo uma matriz  $K$  de tamanho  $N_R \times N_R$  onde seus elementos sejam  $K_{ij} = (y_i \cdot y_j)$ , ou seja, o produto interno das realizações  $i$  e  $j$ , e seguindo os desenvolvimentos contidos em [44], a Equação 2.9 pode ser escrita como:

$$N_R \lambda \alpha = K \alpha \quad (2.10)$$

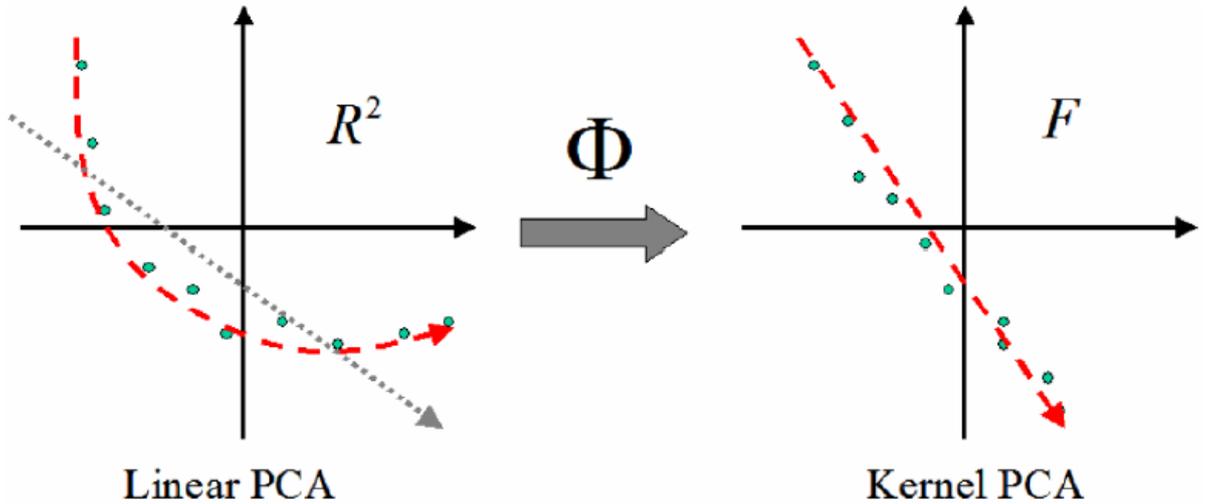
A matriz  $K$  é chamada de matriz *kernel*, de tamanho  $N_R \times N_R$ , ao passo que a matriz de covariância  $C$  é de tamanho  $N_C \times N_C$ . A Equação 2.10 é conhecida como problema de autovalor do *kernel*. Os autovalores são dados por  $N_R \lambda$  e os autovetores são  $\alpha$ . Resolver esta equação é o mesmo que resolver a Equação 2.2 pois os autovalores diferentes de zero da Equação 2.2 são os mesmos da Equação 2.10 escalados por  $N_R$  e os autovetores  $v$ , associados com os autovalores não nulos  $\lambda$  da Equação 2.2, podem ser obtidos a partir de  $\alpha$  usando a Equação 2.8 [44]. A grande atratividade da formulação usando *kernel* é que, para problemas práticos,  $N_R \ll N_C$  e portanto a decomposição em valores singulares da matriz  $K$  pode ser conseguida de forma muito mais eficiente, se comparada à matriz de covariância  $C$ .

### 2.2.2 PCA não linear usando kernel

Como o *PCA* linear, no caso de campos não-gaussianos, não é capaz de capturar a relação não-linear existente entre as realizações no espaço original, o *KPCA* pressupõe o mapeamento (geralmente através de uma função não linear  $\Phi$ ) das realizações num espaço de dimensão superior (espaço característico)  $F$  onde as realizações serão linearmente relacionadas e então realizar a expansão K-L nos dados mapeados em  $F$ . A Figura 2.1 ilustra a idéia do *KPCA*.

Considere um campo randômico arbitrário no  $R^2 (N_C = 2)$ . Cada realização é um vetor de dimensão dois,  $y = (y_1, y_2)^T$ . No contexto da modelagem geológica,  $y_1$  e  $y_2$  pode ser interpretado como valores do *log* da permeabilidade em duas células específicas. Cada realização (o par de *log* da permeabilidade) pode ser plotada num gráfico cartesiano como mostrado na Figura 2.1. Se  $y$  for não-gaussiana, então  $y_1$  e  $y_2$  não são relacionados linearmente em  $R^2$ . Caso a expansão K-L seja aplicada, a relação obtida entre  $y_1$  e  $y_2$  será modelada como linear (na direção da componente principal, ilustrada como a seta cinza na imagem da esquerda). Neste caso, como esperado, o *PCA* linear claramente não é capaz de capturar a relação não linear entre as realizações.

Figura 2.1: Idéia básica do método KPCA (adaptado de SARMA, 2006)



A expansão K-L é apenas capaz de capturar relações lineares entre as realizações no espaço onde é executado.

Considerando um mapeamento não linear  $\Phi$  que relaciona o espaço original  $R^{N_C}$  a outro espaço  $F$  :

$$\Phi : R^{N_C} \longrightarrow F; y \longrightarrow Y = \Phi(y); y \in R^{N_C}, Y \in F \quad (2.11)$$

O espaço  $F$  é chamado de espaço característico e pode assumir uma dimensionalidade arbitrariamente grande. Após a transformação, as realizações que eram relacionadas não-linearmente em  $R^2$  tornam-se relacionadas linearmente em  $F$ . A *PCA* linear pode então ser realizada em  $F$  para determinar os componentes principais neste espaço. As conclusões aqui obtidas podem ser generalizadas para o caso real onde dispõe-se de um grande número de realizações (cada uma de comprimento  $N_C$ ) de um campo de permeabilidades obtidas usando algum algoritmo de geoestatística. O espaço original é  $R^{N_C}$ , e as realizações podem ser entendidas como pontos neste espaço. Estes pontos (realizações) são relacionados não-linearmente em  $R^{N_C}$ . O *Kernel PCA* captura as relações não-lineares entre estas realizações, permitindo assim, obter novas realizações que satisfazem estas relações criando imagens muito parecidas com aquelas geradas usando os *softwares* de geoestatística usuais.

Assim, o método *kernel PCA* consiste essencialmente em realizar *PCA* linear no espaço característico  $F$ , ao invés de  $R^{N_C}$ , e todos os resultados discutidos na seção anterior podem ser facilmente generalizados. A transformação das realizações  $y_k, k =$

$1, \dots, N_R$  no espaço  $F$  será  $\Phi(y_k)$  e, assumindo que  $\Phi(y_k)$  é centrada (caso contrário é possível centralizá-la conforme [44] e [21]), a matriz de covariância no espaço  $F$  será dada por [43]:

$$\bar{C} = \frac{1}{N_R} \sum_{j=1}^{N_R} \Phi(y_j) \Phi(y_j)^T \quad (2.12)$$

A dimensão desta matriz de covariância não é mais  $N_C X N_C$  e sim  $N_F X N_F$ , onde  $N_F$  é comprimento de  $\Phi(y)$ , que pode ser extremamente grande. Assim como na expansão K-L, um problema de autovalores novamente deve ser resolvido, agora usando  $\bar{C}$  ao invés de  $C$ .

$$\lambda V = \bar{C} V \quad (2.13)$$

onde  $\lambda$  são os autovalores e  $V$  os autovetores de  $\bar{C}$ . Ao invés de resolver este problema diretamente (o que é impossível com a atual tecnologia computacional, dado o tamanho de  $N_F$ ), um novo problema de autovalores, com funções *kernels*, é formulado [44] como a seguir:

$$N_R \lambda \alpha = K \alpha \quad (2.14)$$

A matriz *kernel*  $K$  é definida como [44]:

$$K : K_{ij} = (\Phi(y_j) \cdot \Phi(y_j)) \quad (2.15)$$

Cada elemento da matriz *kernel* é um produto interno de dois vetores no espaço  $F$  e não um produto de vetores no espaço original  $R^{N_C}$ . A dimensão da matriz *kernel* é ainda  $N_R X N_R$  semelhante ao caso da expansão K-L. A formulação do problema de autovalores baseada em *kernels* é exatamente equivalente ao problema de autovalores da Equação 2.13 sem suposições. Assim como no método *PCA* linear, é possível obter todos os autovalores  $\lambda$  e autovetores  $V$  de  $\bar{C}$  a partir dos autovalores  $N_R \lambda$  e autovetores  $\alpha$  de  $K$ .

Uma vez que o espaço característico pode ser de dimensão muito grande, o método *KPCA* emprega funções *kernels* ao invés de determinar o mapeamento  $\Phi$  explicitamente e, assim, apenas é necessário calcular o produto interno de dois vetores em  $F$  para obter a matriz *kernel* [31]. Isto é extremamente importante porque pode não ser possível calcular e armazenar  $\Phi(y)$  na memória dos computadores atuais devido ao tamanho de  $F$ . As funções *kernel* são a forma eficiente de calcular produtos internos no espaço característico [43], [44].

$$\Phi(x) \cdot \Phi(y) = k(x, y) \quad (2.16)$$

A função *kernel* calcula o produto interno no espaço  $F$  diretamente a partir dos elementos no espaço original das realizações  $R^{Nc}$ . Há vários tipos de funções *kernel* disponíveis, mas a função de interesse nesta aplicação é a função *kernel* polinomial definida como [44]:

$$(\Phi(x) \cdot \Phi(y)) = k(x, y) = (x \cdot y)^d \quad (2.17)$$

onde  $d$  é o grau do *kernel* polinomial. Cada função *kernel* está associado unicamente a um mapeamento  $\Phi$  [43], [44]. Executar a expansão K-L no espaço  $F$  utilizando o *kernel* polinomial  $(x \cdot y)^d$  corresponde a preservar os momentos de ordem  $2d$  em  $R^{Nc}$ . Porém, deseja-se preservar não apenas os momentos de ordem  $2d$ , mas sim todos os momentos até o de ordem  $2d$ . Assim, usa-se a seguinte formulação para a função *kernel* polinomial.

$$(\Phi(x) \cdot \Phi(y)) = k(x, y) = \sum_{i=1}^d (x \cdot y)^i \quad (2.18)$$

### 2.2.3 O problema da pré-imagem

Uma vez que a expansão de *Karhunen-Loeve* é realizada no espaço  $F$ , as novas realizações geradas com a expansão também estarão contidas neste espaço, ou seja,  $Y \in F$ , onde  $Y$  é uma realização simulada. Porém, o que se espera é que as realizações estejam no espaço das realizações originais  $R^{Nc}$  pois o objetivo é obter uma parametrização do campo randômico de entrada. Para obter uma realização  $y$  no espaço original das realizações  $R^{Nc}$  que corresponda à realização simulada  $Y \in F$ , é necessário um mapeamento inverso  $y = \Phi^{-1}(Y)$ , o que é conhecido como problema da pré-imagem [44] e [21]. Devido à dimensionalidade do espaço  $F$ , pode não ser possível calcular esta pré-imagem ou mesmo ela pode nem existir e, se existir, pode não ser única [44] e [21]. Estas questões são resolvidas através de um problema de minimização no qual o vetor  $y$  é determinado tal que o erro quadrático médio entre  $\Phi(y)$  e  $Y$  seja mínimo [44] e [21].

$$\min_y \rho(y) = \|\Phi(y) - Y\|^2 = \Phi(y) \cdot \Phi(y) - 2Y \cdot \Phi(y) + Y \cdot Y \quad (2.19)$$

Usando as Equações 2.3 e 2.8 no espaço característico  $F$ , é possível mostrar que qualquer realização  $Y$  é uma combinação linear das realizações originais mapeadas  $\Phi(y_k)$ ,  $y_k, k = 1, \dots, N_R$ , conforme Equação 2.20.

$$Y = \sum_{i=1}^{N_R} \beta_i \Phi(y_i) \quad (2.20)$$

Os coeficientes  $\beta_i$  são funções das variáveis independentes  $\xi$  e são dados por:

$$\beta_i = \frac{1}{\sqrt{N_R}} \sum_{j=1}^{N_R} \alpha_{ij} \delta_j^{1/2} \xi_j \quad (2.21)$$

onde  $\alpha_i = [\alpha_{1i}, \dots, \alpha_{N_R i}]^T$  e  $\delta_i$  são os  $i$ -ésimos autovetor e autovalor da matriz *kernel*  $K$ , respectivamente. Usando a Equação 2.20 e, substituindo os produtos internos em 2.19 por funções de *kernel*, a Equação 2.19 pode ser reescrita conforme a Equação 2.22 [42].

$$\rho(y) = k(y, y) - 2 \sum_{i=1}^{N_R} \beta_i k(y_i, y) + \sum_{i=1}^{N_R} \sum_{j=1}^{N_R} \beta_i \beta_j k(y_i, y_j) \quad (2.22)$$

O mínimo da função  $\rho(y)$  é obtido fazendo  $\frac{d\rho(y)}{d(y)} = 0$ , o que resulta na seguinte equação:

$$\frac{dk(y, y)}{dy} - 2 \sum_{i=1}^{N_R} \beta_i \frac{dk(y_i, y)}{dy} \quad (2.23)$$

simplificando esta equação, considerando a função kernel dada por 2.18, surge a Equação 2.24.

$$\sum_{j=1}^d j(y \cdot y)^{j-1} y - \sum_{i=1}^{N_R} \beta_i \sum_{j=1}^d j(y_i \cdot y)^{j-1} y_i = 0 \quad (2.24)$$

Esta é uma parametrização relacionando uma realização  $y$  no espaço original das realizações  $R_{N_C}$  a um conjunto de variáveis aleatórias  $\xi$ , já que  $\beta_i$  é função de  $\xi$ . Esta é uma parametrização não-linear implícita da forma  $f(y, \xi) = 0$ . Uma maneira eficiente para encontrar  $y$  para um dado  $\xi$  é conseguida aplicando o método iterativo do ponto fixo [44]. Um esquema mais estável e que resulta em uma combinação não-linear das realizações originais com os fatores de ponderação somando um é dado pela Equação 2.25 [42].

$$y^{k+1} = \frac{\sum_{i=1}^{N_R} \beta_i \sum_{j=1}^d j(y_i \cdot y^k)^{j-1} y_i}{\sum_{i=1}^{N_R} \beta_i \sum_{j=1}^d j(y_i \cdot y^k)^{j-1}} \quad (2.25)$$

Obtendo a realização  $y$  conforme a Equação 2.25, preservam-se as estruturas, padrões e até mesmo os dados condicionantes *hard* que porventura estejam incorporados às realizações originais.

## 2.3 Método Pluri-PCA

Os modelos gerados através da utilização das técnicas descritas anteriormente, expansão de *Karhuen-Loeve* e *KPCA*, por vezes são considerados inaceitáveis pelos geocientistas pois elas distorcem as *fdp* dos modelos a priori das fácies, permeabilidade e/ou porosidade. O problema da modelagem de fácies é mais grave porque, em função da perda dos menores comprimentos de correlação provocado pelo truncamento da expansão, o *PCA* converte valores discretos em valores reais suavizando a fronteira entre as diferentes fácies. Para os parâmetros petrofísicos, há uma distorção das *fdp*'s, sempre tendendo a suavizá-las, especialmente quando os modelos a priori são multigaussianos com vários picos.

Para superar esta dificuldade, CHEN ET. AL., 2015 [1] propuseram o método *Pluri-PCA* que combina o *PCA* com o método geoestatístico denominado de simulação plurigaussiana que será apresentado a seguir.

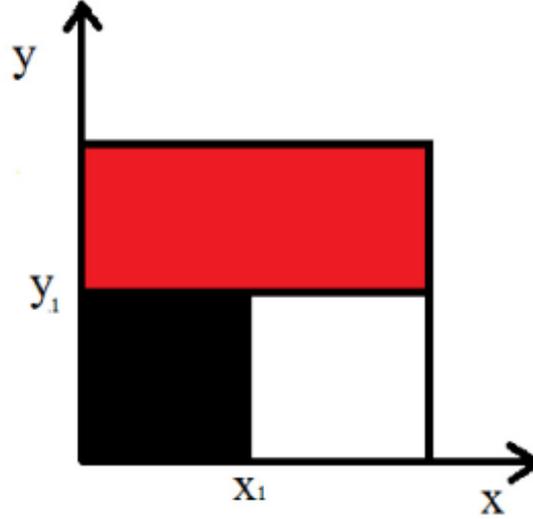
### 2.3.1 Simulação Plurigaussiana

A simulação plurigaussiana é aplicada a modelos com múltiplas fácies onde não exista um ordenamento sequencial entre elas e pode ser entendido como uma extensão natural do método Gaussiano truncado. Ao invés de se executarem os truncamentos sobre apenas uma gaussiana, executa-se sobre uma combinação  $N_G$  ( $N_G = N_F - 1$  onde  $N_F$  é o número de fácies) de funções gaussianas que podem estar correlacionadas ou não. Ao amostrar uma campo randômico multigaussiano com  $N_G$  variáveis aleatórias, o método plurigaussiano primeiramente gera um vetor real de dimensão  $N_G$  para cada célula e então converte o vetor real em indicadores discretos de fácies de acordo com alguma regra de litologia (*RL*) especificada por cada usuário. A *RL* determina o mapeamento do vetor de valores reais para os indicadores de fácies e reflete as frações de fácies e conexões de vizinhança entre as diferentes fácies [1].

A Figura 2.2 mostra um exemplo de *RL* onde as diferentes cores representam diferentes fácies sem que haja uma ordem sequencial entre elas. Um ponto  $(x, y)$  na Figura 2.2 representa os valores de duas variáveis aleatórias e  $x_1$  e  $y_1$  são valores de corte para  $x$  e  $y$  que diferenciam as três fácies. Neste exemplo, se  $x > x_1$  e  $y < y_1$ , o indicador de

fácies será 2; se  $x < x_1$  e  $y < y_1$ , o indicador de fácies será 1; caso contrário será 0.

Figura 2.2: Exemplo de uma RL com 3 Fácies dispostas sem ordenamento sequencial



### 2.3.2 Formulação mista Pluri-PCA

Para aplicação da metodologia, a formulação do método *PCA* visto nas seções anteriores deve ser levemente modificada. Seja  $m$  um vetor de dimensão  $N_m$ , composto por propriedades das células do modelo, tais como fácies, permeabilidade, porosidade, etc. Desse modo,  $m$  pode ser escrito como  $m = [F, P]^T$ , onde  $F$  e  $P$  são vetores compostos por indicadores de fácies e propriedades petrofísicas das células do modelo respectivamente. O comprimento deste vetor é  $N_m = (N_P + 1) \times N_C$ , onde  $N_C$  é o número de células do modelo e  $N_P$  é o número de propriedades petrofísicas utilizadas no modelo. Usando a formulação tradicional do *PCA*, é possível gerar uma realização das propriedades estáticas  $m$  como uma combinação linear dos vetores de base ortogonais,

$$\begin{bmatrix} \widehat{F} \\ \widehat{P} \end{bmatrix} = \begin{bmatrix} \overline{F} \\ \overline{P} \end{bmatrix} + \begin{bmatrix} \Phi_F \\ \Phi_P \end{bmatrix} \xi \quad (2.26)$$

onde  $\overline{F}$  e  $\overline{P}$  são os valores médios das propriedades, obtidos a partir de um conjunto de realizações de treinamento,  $\Phi_F$  e  $\Phi_P$  são matrizes  $N_C \times N_\xi$  e  $(N_P N_C) \times N_\xi$  compostas dos  $N_\xi$  maiores vetores de base ortogonais e  $\xi$  é o vetor dos coeficiente *PCA* de dimensão  $N_\xi$ . Os modelos de fácies e de propriedades petrofísicas reconstituídos a partir do *PCA* são compostos de variáveis contínuas e, evidentemente, não são exatamente as mesmas dos modelos originais do conjunto de treinamento. Para aumentar a acurácia do modelo de

fácies reconstruído, o vetor de indicadores de fácies de dimensão  $N_C$  é expandido para um novo vetor de números inteiros  $U = [u_1, \dots, u_{N_F-1}]^T$  composto de  $N_F - 1$  sub-vetores independentes  $u_i$  ( $i = 1, 2, \dots, N_F - 1$ ), onde cada  $u_i$  tem dimensão  $N_C$  e é composto por valores binários 0 ou 1. Por exemplo, se  $F^{(j)} = i$  (ou seja, o indicador de fácies da  $j$ -ésima célula do modelo é  $i$ ), então  $u_i^{(j)} = 1$  e  $u_k^{(j)} = 0 \forall k \neq i$ . Cada vetor indicador  $u_i$  é gerado através da recategorização dos indicadores originais de fácies em cada célula [1].

O propósito da transformação do vetor de indicadores de fácies original  $F$  para o vetor expandido  $U$  é superar a desvantagem da formulação tradicional do *PCA* onde o vetor de fácies reconstruído é dependente do ordenamento das fácies no modelo, o que pode ser facilmente comprovado conforme explicado em CHEN ET. AL., 2015. Esta abordagem aumenta o realismo geológico do modelo reconstruído, reduzindo o potencial de geração de fácies incorretas devido o efeito de suavização provocado pelo método *PCA*.

Usando o vetor de fácies expandido  $U$  e o vetor de propriedades petrofísicas  $P$ , a Equação 2.26 pode ser generalizada para mapear  $U$  (ao invés de  $F$ ) e  $P$  em um vetor real  $\xi$  [1]:

$$\begin{bmatrix} \widehat{Z} \\ \widehat{P} \end{bmatrix} = \begin{bmatrix} \overline{Z} \\ \overline{P} \end{bmatrix} + \begin{bmatrix} \Phi_U \\ \Phi_P \end{bmatrix} \xi \quad (2.27)$$

onde  $\overline{Z} = \overline{U}$  e  $\overline{P}$  são os vetores médios obtidos a partir das  $N_R$  realizações;  $\Phi_U$  e  $\Phi_P$  são matrizes  $[(N_F - 1) \times N_C] \times N_\xi$  e  $(N_P \times N_C) \times N_\xi$  compostas dos  $N_\xi$  maiores vetores de base ortogonais, respectivamente. Observe que  $\widehat{Z}$  contém variáveis contínuas e  $U$  apenas variáveis discretas. O vetor obtido após a aplicação do *PCA* é  $\widehat{Z} = [Z^{(1)}, \dots, Z^{(j)}, \dots, Z^{(N_C)}]^T$ , onde  $Z^{(j)}$  é o resultado da aplicação do *PCA* sobre  $U$  correspondente à  $j$ -ésima célula. Na  $j$ -ésima célula, o vetor de fácies reconstruído fica  $Z = [Z_1, \dots, Z_{N_F-1}]^T$  e é composto de valores reais sendo, assim, diferente do vetor original descrito por valores inteiros  $U = [u_1, \dots, u_{N_F-1}]^T$  [1].

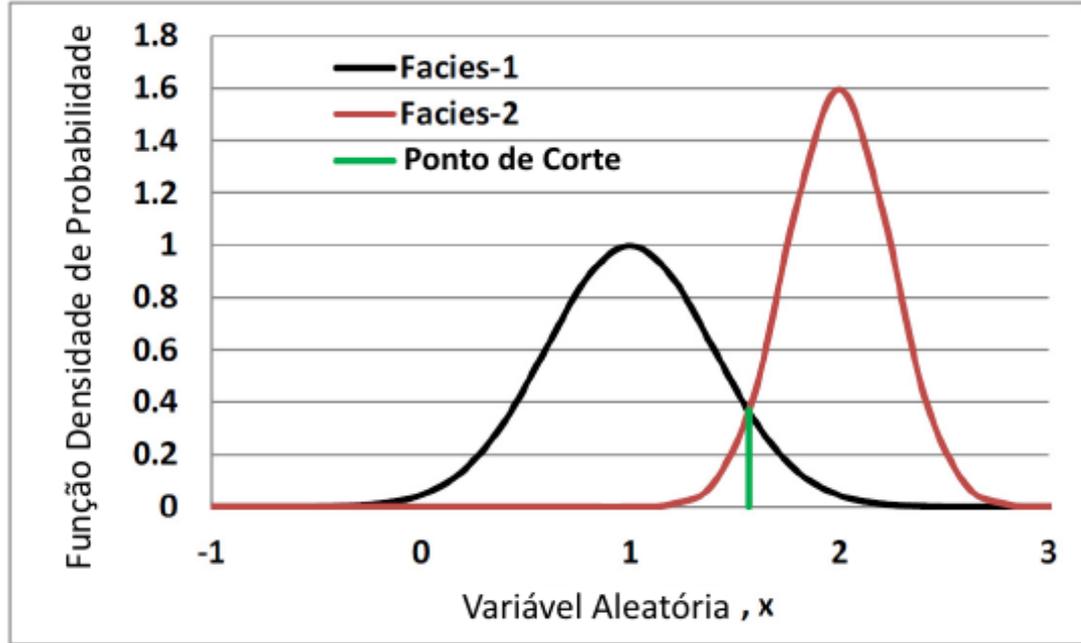
### 2.3.3 Regra de Litologia

A correta obtenção das fácies pela metodologia pluri-*PCA* passa pela definição da RL apropriada. Mesmo nos caso onde a metodologia pluri-*PCA* esteja sendo usada para reconstrução de realizações idênticas às originais, a RL não é totalmente conhecida.

Para uma maior acurácia na definição dos indicadores de fácies, a RL deve ser selecionada de forma que minimize a probabilidade de atribuir fácies incorretas às

realizações originiais do conjunto de treinamento. A Figura 2.3 ilustra, usando um exemplo com duas fácies, como determinar o ponto de corte usando as *fdp's* de cada uma.

Figura 2.3: Distribuição de probabilidades e determinação do ponto de corte para um exemplo com 2 fácies. Adaptado de CHEN ET. AL., 2015.



Sejam  $f_1(x)$  e  $f_2(x)$  as respectivas *fdp's* das fácies 1 e 2 (linhas preta e vermelha na Figura 2.3). Seja  $\Omega_i$  o conjunto das células que são classificadas como fácies  $i$ , ou seja, se  $j$  pertence a  $\Omega_i$ , então o indicador de fácies da  $j$ -ésima célula é  $i$ , para uma dada realização. Seja  $n_i$  o número de células do modelo que pertencem à  $i$ -ésima fácies, ou seja o comprimento do vetor  $\Omega_i$ . Para qualquer modelo obtido utilizando o *PCA*, cada célula em  $\Omega_i$  terá um valor real associado  $x_i^{(j)}$  (a aproximação do indicador original discreto  $i$ ). Os  $n_i$  valores de  $x_i^{(j)}$  ( $\forall j \in \Omega_i$ ) estão geralmente espalhados ao redor de  $i$  e sua *fdp*  $f_i(x)$  pode então ser aproximada por uma gaussiana, conforme mostra a Figura 2.3. Por simplicidade, assume-se que o gráfico da *fdp* de cada fácies possui apenas um pico se esta for não-gaussiana [1].

Para o exemplo da Figura 2.3, se  $x^*$  for o valor de corte que separa as duas fácies, então o número de células com atribuições incorretas da fácies 1 e da fácies 2 pode ser determinado por [1]:

$$\tilde{n}_1 = n_1 \int_{x^*}^{\infty} f_1(x) dx \quad e \quad \tilde{n}_2 = n_2 \int_{-\infty}^{x^*} f_2(x) dx \quad (2.28)$$

O erro relativo de atribuição incorreta para a fácies  $i$  pode ser expresso como sendo o número de células com atribuições incorretas da fácies  $i$  dividido pelo número de células com a fácies  $i$ . Assim, é possível escrever a seguinte expressão

$$E_{fr}(x^*) = \int_{x^*}^{\infty} f_1(x) dx + \int_{-\infty}^{x^*} f_2(x) dx \quad (2.29)$$

A escolha natural para o valor de corte  $x^*$  é o que minimize o erro relativo de atribuição de fácies. O valor de  $x^*$  que minimiza  $E_{fr}(x^*)$  é obtido fazendo-se  $\frac{dE_{fr}(x^*)}{dx^*} = 0$  o que leva a  $f_1(x^*) = f_2(x^*)$ , conforme indicado na Figura 2.3 pela linha verde. Dessa forma, o valor real  $x$  será mapeado para a fácies 2 quando  $x > x^*$  (ou  $f_2(x) > f_1(x)$ ) ou para a fácies 1 quando  $x < x^*$  (ou  $f_1(x) > f_2(x)$ ). É possível generalizar a formulação dada pela Equação 2.29 e considerar o erro relativo ponderado de todas as fácies através da expressão [1]:

$$E_{fw}(x^*) = w_1 \int_{x^*}^{\infty} f_1(x) dx + w_2 \int_{-\infty}^{x^*} f_2(x) dx \quad (2.30)$$

onde  $w_i$  é o fator de ponderação para a  $i$ -ésima fácies. Novamente o valor de corte pode ser encontrado ao minimizar a expressão dada pela Equação 2.30. Observe que a Equação 2.30 se torna idêntica a Equação 2.29 se  $w_1 = w_2 = 1$ . Generalizando o resultado para o caso de múltiplas fácies, seja  $f_i(Z)$  a *fdp* da  $i$ -ésima fácies, o vetor de valores reais  $Z = [Z_1, \dots, Z_{N_F-1}]^T$  será mapeado como fácies 1 se [1],

$$w_i f_i(Z) > w_j f_j(Z) \quad \forall j \neq i \quad (2.31)$$

A desigualdade dada pela Equação 2.31 pode então ser usada como RL. Dados a *fdp*  $f_i(Z)$  para cada fácies, os fatores de ponderação  $w_i$  e os vetores de valores reais  $Z = [Z_1, \dots, Z_{N_F-1}]^T$ , obtidos para cada célula através do *PCA*, aplica-se a Equação 2.31 para encontrar o indicador de fácies inteiro.

Os fatores de ponderação são escolhidos de modo que as fácies do modelo reconstruído sejam iguais às fácies do modelo correspondente no conjunto de treinamento [1]. Geralmente, as frações de cada fácies  $\alpha_i^* = n_i/N_C$  dependem não-linearmente dos fatores de ponderação. É necessário um procedimento iterativo para determinar estes fatores resolvendo as equações não-lineares  $\alpha_i^*(w) - \alpha_i = 0$  numericamente.

As *fdp's* de cada fácies do modelo reconstruído pode ser aproximado utilizando o conjunto de treinamento [1]. Como já mencionado, usa-se a distribuição gaussiana para aproximar as *fdp's* das fácies. A formulação gaussiana necessita apenas da estimativa do valor médio de  $z_k (k = 1, \dots, N_F - 1)$  no conjunto de valores em torno da  $i$ -ésima fácies,  $\bar{z}_{k|i}$ , e da covariância entre  $z_k$  e  $z_l$  no conjunto de valores em torno da  $i$ -ésima fácies,  $\sigma_{k,l|i} = cov(z_{k|i}, z_{l|i})$  para  $(k, l = 1, 2, \dots, N_F - 1)$ . A equação a seguir calcula  $\bar{z}_{k|i}$  a partir da distribuição de fácies da realização de treinamento,

$$\bar{z}_{k|i} = \frac{1}{n_i} \sum_{n \in \Omega_i} z_k^{(n)} \quad (2.32)$$

onde  $n_i$  é o número de células do modelo que pertencem à  $i$ -ésima fácies e  $\Omega_i$  representa o conjunto de índices das células com indicador de fácies igual a  $i$ , ou seja, e  $n$  pertence a  $\Omega_i$ , então o indicador de fácies da  $n$ -ésima célula é  $i$ , para uma dada realização. De modo similar, pode-se estimar  $\sigma_{k,l|i}$

$$\sigma_{k,l|i} = \frac{1}{n_i} \sum_{n \in \Omega_i} (z_k^{(n)} - \bar{z}_{k|i})(z_l^{(n)} - \bar{z}_{l|i}) \quad (2.33)$$

Como vimos acima, para uma dada realização, os indicadores de fácies em cada bloco são conhecidos e as *fdp's* para cada fácies  $f_i(Z)$  do modelo reconstruído para esta realização em particular pode ser aproximada a partir da frequência relativa do número da fácies, conforme Equações 2.32 e 2.33. Contudo, caso o objetivo seja gerar uma nova realização aleatória, não há fácies de referência para aproximar a distribuição e determinar o ponto de corte na RL. Neste caso, as médias  $\bar{z}_{k|i} (k = 1, \dots, N_F - 1; i = 1, \dots, N_F)$  as covariâncias  $\sigma_{k,l|i} = cov(z_{k|i}, z_{l|i})$  para  $(k, l = 1, 2, \dots, N_F - 1)$  e também as frações das fácies  $\alpha_i$  são igualmente obtidos aleatoriamente a partir de uma distribuição gaussiana onde os parâmetros são determinados a partir da média e covariância destas variáveis calculadas a partir dos resultados do *PCA* no conjunto de treinamento [1]. Assim, define-se o seguinte vetor:

$$R_{RL|i} = [\bar{z}_{1|i}, \dots, \bar{z}_{N_F-1|i}, \sigma_{1,1|i}, \dots, \sigma_{1,N_F-1|i}, \dots \quad (2.34a)$$

$$\dots, \sigma_{2,2|i}, \sigma_{2,N_F-1|i}, \sigma_{N_F-1,N_F-1|i}, \alpha_i]^T \quad (i = 1, 2, \dots, N_F) \quad (2.34b)$$

O comprimento do vetor  $R_{RL|i}$  é  $N_F X(N_F + 1)/2$ . O vetor médio  $\bar{R}_{RL|i}$  e a matriz de covariância  $C_{RL|i}$  de  $R_{RL|i}$  podem ser estimados a partir dos resultados do *PCA* sobre o conjunto de  $N_R$  realizações.

## Capítulo 3

# Parametrização da Função Objetivo

Uma vez obtida a representação do campo randômico de entrada, as estatísticas de interesse da função objetivo serão aproximadas usando o método da expansão em Caos Polinomial associado a alguma técnica eficiente para obter os coeficientes do polinômio. Neste trabalho foram estudados o método de colocação probabilística (*PCM*) e o método espectral não-intrusivo (*NISP*). Ambos os métodos se destinam a encontrar uma parametrização das respostas do modelo na forma,  $y = f(\xi)$ , onde  $\xi$  é o vetor de variáveis aleatórias (coeficientes da expansão K-L), por alguma expressão polinomial função dos parâmetros incertos.

### 3.1 Expansão em Caos Polinomial e Polinômios de Hermite

A projeção em Caos Polinomial (*PCE*) proporciona uma forma de expandir um processo estocástico de segunda ordem em termos de polinômios Hermitianos em variáveis aleatórias gaussianas [7]. Processos estocásticos de segunda ordem são processos com variância finita, onde podem ser enquadrados a maioria dos processos físicos [42]. Assim, uma variável aleatória de segunda ordem genérica  $Y(\xi(\theta))$ , onde  $\theta$  é um evento aleatório independente, pode ser representada na seguinte forma [53]:

$$\begin{aligned} Y(\xi(\theta)) = & a_0\Gamma_0 + \sum_{i_1=1}^{\infty} a_{i_1}\Gamma_1(\xi_{i_1}(\theta)) + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} a_{i_1 i_2}\Gamma_2(\xi_{i_1}(\theta), \xi_{i_2}(\theta)) \quad (3.1) \\ & + \sum_{i_1=1}^{\infty} \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} a_{i_1 i_2 i_3}\Gamma_3(\xi_{i_1}(\theta), \xi_{i_2}(\theta), \xi_{i_3}(\theta)) + \dots \end{aligned}$$

onde  $\Gamma_n(\xi_{i_1}, \dots, \xi_{i_n})$  representa o polinômio de Hermite de grau  $n$  em termos do vetor  $n$ -dimensional das variáveis aleatórias independentes gaussianas padronizadas  $(\xi_{i_1}, \dots, \xi_{i_n})$  e os coeficientes  $a_0, a_{i_1}, \dots, a_{i_1 i_2 \dots i_n}$  são funções determinísticas. A utilização dos polinômios de Hermite se deve à característica de bi-ortogonalidade (são ortogonais entre si e também com relação à distribuição de probabilidade de  $\xi$ , aqui assumida gaussiana) necessária para a expansão. A Equação 3.1 constitui a versão discreta da expansão *PCE* original proposta por WIENER, 1938 [52], onde as integrais estão substituídas por somatórios. A expressão geral dos polinômios de Hermite de ordem  $n$  é dada por [53]:

$$\Gamma_n(\xi_{i_1}, \dots, \xi_{i_n}) = (-1)^n \exp\left(\frac{1}{2}\xi^T \xi\right) \frac{\partial}{\partial \xi_{i_1}, \dots, \partial \xi_{i_n}} \exp\left(-\frac{1}{2}\xi^T \xi\right) \quad (3.2)$$

sendo  $\xi = (\xi_{i_1}, \dots, \xi_{i_n})^T$  o vetor  $n$ -dimensional das variáveis aleatórias Gaussianas padronizadas  $\xi$ . Para o caso unidimensional, os polinômios de Hermite são dados por:

$$\begin{aligned} \Gamma_1(\xi) &= \xi \\ \Gamma_2(\xi) &= \xi^2 - 1 \\ \Gamma_3(\xi) &= \xi^3 - 3\xi \\ \Gamma_4(\xi) &= \xi^4 - 6\xi^2 + 3 \\ \Gamma_5(\xi) &= \xi^5 - 10\xi^3 + 15\xi \end{aligned} \quad (3.3)$$

Por conveniência notacional, a Equação 3.1 pode ser reescrita da seguinte forma:

$$Y(\xi(\theta)) = \sum_{j=1}^{\infty} \beta_j \Psi_j(\xi) \quad (3.4)$$

Na Equação 3.4, os polinômios Hermitianos multidimensionais são essencialmente produtos tensoriais dos polinômios unidimensionais [40]. Existe uma correspondência entre as funções  $\Gamma_n(\xi_{i_1}, \dots, \xi_{i_n})$  e  $\Psi_j(\xi)$ , assim como entre os coeficientes. Na Equação 3.1, o somatório é realizado seguindo a ordem crescente do grau do polinômio Hermitiano, ao passo que na Equação 3.4, ocorre uma reordenação para contabilizar primeiro os polinômios de ordem mais baixa.

A expansão em caos polinomial possui a melhor taxa de convergência no critério de mínimos quadrados para campos randômicos multigaussianos, principalmente devido ao critério de ortogonalidade mencionado [42]. Para trabalhar com campos randômicos genéricos, uma abordagem possível é usar o caos polinomial generalizado conforme

descrito por XIU ET. AL., 2003 [53]. Outra abordagem é usar a transformação para variáveis aleatórias com distribuição normal e depois aplicar a expansão do caos polinomial [17]. Para fins práticos, a série pode ser truncada para um número finito de termos e a Equação 3.4 pode ser reescrita como:

$$Y(\xi(\theta)) = \sum_{j=0}^{N-1} \beta_j \Psi_j(\xi) \quad (3.5)$$

O número de termos  $N$  da expansão é função do número  $n$  de variáveis independentes usadas para representar o campo randômico de entrada e do grau  $p$  do polinômio empregado e cresce muito rápido a medida em que se aumenta qualquer um destes parâmetros. Assim, temos a seguinte expressão:

$$N = \frac{(n+p)!}{n!p!} \quad (3.6)$$

Uma vantagem significativa que a expansão *PCE* possui sobre outros modelos substitutos é que eles garantem convergência para a distribuição real (das respostas de interesse, como por exemplo a produção acumulada de óleo) quando a ordem do polinômio e o número de simulações usados para calcular o *PCE* é aumentado [8].

Maiores detalhes sobre a determinação da expansão em caos polinomial podem ser obtidos em [13].

## 3.2 Método de Colocação Probabilística

O método da colocação probabilística (*PCM*) foi introduzido por TATANG ET. AL., 1997 [48] e tem sido aplicado com sucesso em diversos campos do conhecimento. A expansão em Caos Polinomial pode ser empregada em conjunção com o método da colocação probabilística (*PCM*) para realizar análise de incertezas em modelos computacionalmente complexos de forma bastante eficiente a custo computacional reduzido [48].

O *PCM* destina-se a encontrar uma aproximação das respostas do modelo,  $y = f(\xi)$ , onde  $\xi$  é o vetor de variáveis aleatórias independentes, por alguma expressão polinomial função dos parâmetros incertos, tal como representado pela Equação 3.5. Em termos práticos, a expansão é truncada a  $N$  termos que determinam o grau do polinômio utilizado na aproximação. Para obter tal aproximação, é necessário apenas encontrar os coeficientes determinísticos  $\beta_j, j = 0, \dots, N - 1$  da expansão.

Utilizando as respostas obtidas por  $N$  avaliações do modelo numérico, obtém-se, a partir da Equação 3.5, um conjunto de equações lineares que são resolvidas para encontrar  $\beta_j$ . A grande atração do método é o fato de que, dada as distribuições das variáveis aleatórias de entrada, pode-se encontrar expansões em caos polinomial com a melhor taxa de convergência. Por esta razão, para um determinado nível de acurácia, a abordagem *PCM* exige um pequeno número de avaliações do modelo se comparado a outras possíveis expansões [22].

A formulação do *PCM* deriva do método dos resíduos ponderados onde os pesos são escolhidos como sendo funções delta de Dirac [22]:

$$w_j = \delta(\xi - \xi_k) \quad (3.7)$$

onde  $\xi_k$  é um conjunto particular do vetor aleatório  $\xi$  escolhido segundo uma determinada regra heurística que será abordada mais adiante. Os elementos em  $\xi_k$  são chamados de pontos de colocação. A equação do resíduo pode, então, ser expressa por [22]:

$$\mathcal{L}\hat{y} - u = R(\{\beta_j\}, \xi_k) = 0 \quad (3.8)$$

onde  $\mathcal{L}$  é um operador que envolve derivadas no espaço e no tempo,  $\hat{y}$  é a aproximação PCE da função objetivo e  $u$  a função objetivo avaliada. Observe que o número de pontos de colocação é igual ao número de termos da expansão *PCE*. Assim, são necessários  $N$  conjuntos de pontos de colocação para determinar os coeficientes da expansão  $\beta_j, j = 0, \dots, N - 1$ . A aplicação da Equação 3.8 resulta em um conjunto de equações independentes que devem ser avaliadas nos pontos de colocação dados,  $\xi_k, k = 1, \dots, N$ . Desse modo,

$$u(x, t, \xi_k) = \sum_{j=0}^{N-1} \beta_j \Psi_j(\xi_k), \quad k = 1, 2, \dots, N \quad (3.9)$$

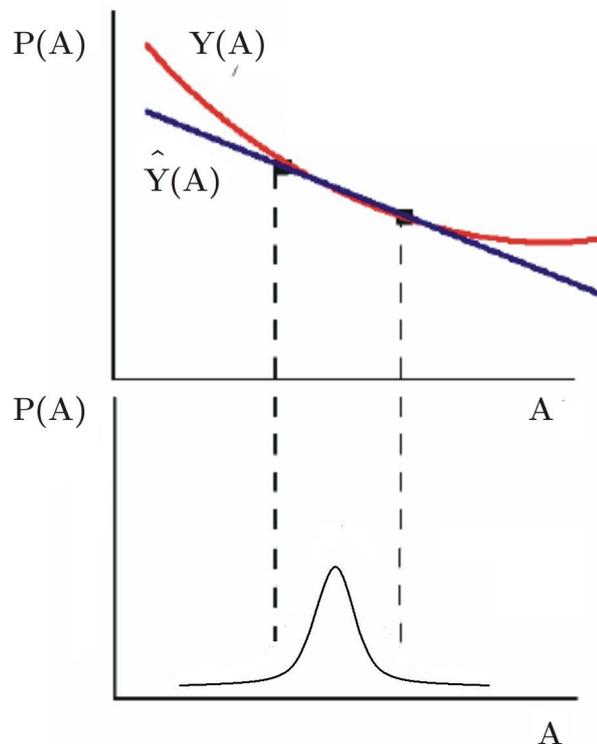
Os coeficientes da expansão podem então ser obtidos resolvendo o seguinte sistema de equações lineares:

$$H\beta = u \quad (3.10)$$

sendo  $H$  uma matriz de dimensões  $N \times N$  e formada pelos polinômios de Hermite avaliados nos pontos de colocação, ou seja,  $H_{i,j} = \Psi_j(\xi_i)$ ;  $\beta$  é o vetor dos coeficientes da expansão, ou seja,  $\beta = [\beta_0, \beta_1, \dots, \beta_{N-1}]^T$  e  $u$  os pontos de colocação avaliados no modelo de simulação,  $u = [u_1, \dots, u_N]^T$ .

A etapa crucial em estimar eficientemente as respostas do modelo é a escolha dos valores das variáveis aleatórias para os quais o modelo será avaliado. Estes pontos de colocação devem ser escolhidos de tal maneira que a Equação 3.5 capture o máximo possível do comportamento de  $y = f(\xi)$ . Uma vez que as variáveis  $\xi$  possuem *fdp's* associadas, os pontos de colocação devem ser selecionados de tal forma que a região de alta probabilidade dessas *fdp's* seja coberta, conforme mostrado na Figura 3.1. Para isso, pode-se aplicar uma técnica baseada na idéia da quadratura Gaussiana [48] para estimar integrais. Usando esta abordagem, a integral de um polinômio pode ser aproximada a partir de um somatório usando as raízes do polinômio de ordem imediatamente superior ao do polinômio considerado na expansão. Similarmente, no *PCM*, são usadas as raízes do polinômio Hermitiano de ordem imediatamente superior como os pontos nos quais a aproximação será resolvida. Para maiores detalhes, consultar [48] e [17]. Espera-se que os maiores desvios da aproximação em relação às respostas do modelo ocorram na região de baixa probabilidade e, portanto, produzem erros pequenos.

Figura 3.1: Característica do método PCM: Englobar as regiões de alta probabilidade.



### 3.3 Método Espectral não Intrusivo

A diferença entre o *PCM* e método espectral não intrusivo (*NISP*) está essencialmente na forma de calcular os coeficientes da expansão *PCE*.

A idéia básica por trás da metodologia é que, usando as propriedades de ortogonalidade da expansão *PCE*, é possível mostrar que os coeficientes são obtidos por [37]:

$$\beta_k = \frac{E[g(\xi)\Psi_k(\xi)]}{E[\Psi_k(\xi)\Psi_k(\xi)]} \quad (3.11)$$

O denominador da equação acima pode ser calculado analiticamente pois significa o valor esperado de polinômios *PCE* conhecidos e não envolve avaliações do modelo,  $g(\xi)$ . O valor esperado no numerador envolve  $g(\xi)$  e portanto, a estimativa dos coeficientes  $\beta_k$  requer o cálculo acurado desta esperança. O cálculo deste termo pode ser colocado como um problema de amostragem ou um problema de integração numérica  $N$ -dimensional, em ambos os casos pode ser escrito como [40]:

$$E[g(\xi)\Psi_k(\xi)] \approx \frac{1}{N_S} \sum_{i=1}^{N_S} w_i g(\xi_i) \Psi_k(\xi_i) \quad (3.12)$$

onde  $N_S$  é o número de amostras (número de simulações), e  $w_i$  são os pesos para cada ponto amostrado que depende do método usado. Deve ser observado que  $N_S$  não é dependente diretamente do número de termos da expansão. O número necessário deve ser suficiente apenas para reduzir o erro da estimativa dos coeficiente a um nível aceitável. Outro aspecto favorável é que cada coeficiente é determinado de maneira independente sem precisar resolver um sistema de equações lineares ou mesmo usar um método de regressão. No *PCM*, os coeficientes são dependentes e a estimativa dos coeficientes associados aos termos de ordem superior (geralmente menos precisos) afeta a estimativa dos termos de ordem inferior, afetando também a sua precisão [40].

O esquema mais eficiente para estimar o valor esperado indicado pelo numerador da Equação 3.11 é através de técnicas de quasi-Monte Carlo, usando sequências de baixa discrepância [32] que, para problemas de larga escala, tem apresentado convergência na ordem de  $O(1/N_S)$  [40]. Dentre as sequências de baixa discrepância, uma técnica, denominada sequência de Sobol, tem apresentado bons resultados para diversas classes de problemas [40].

Já foi mencionado que os coeficientes associados aos termos de ordem superior apresentam uma estimativa menos precisa. A grosso modo, quanto maior for a ordem de  $\Psi_k(\xi)$ , maior será o erro na estimativa de  $\beta_k$  para um dado número de simulações

$N_S$  [40]. Manter estes termos na expansão acaba por afetar a precisão da expansão como um todo. Uma das vantagens do método é que, como os coeficientes são obtidos por amostragem, as mesmas simulações podem ser usadas num processo de validação cruzada para determinar se um coeficiente  $\beta_k$  estimado é suficientemente preciso para permanecer na expansão, ou se deve ser descartado.

De forma geral, as etapas do método *NISP* são:

1. Calcular os coeficientes da expansão para um grau razoavelmente elevado (6<sup>a</sup> ordem, por exemplo), usando as Equações 3.11 e 3.12 e alguma técnica de amostragem conforme discutido acima;
2. Usando apenas os termos da expansão *PCE* de primeira ordem (com os coeficientes de terminados no passo 1), calcular, usando todas as  $N_S$  amostras, o erro da aproximação  $E^j = \sum \|y - \hat{y}\|$ , onde  $j$  é o índice da iteração;
3. Adicionar os termos de ordem imediatamente superiores e calcular novamente o erro da aproximação,  $E^{j+1}$ ;
4. Se  $E^{j+1} < E^j$  mantenha o termo de ordem superior, caso contrário deve-se descartá-los definindo  $\beta_k = 0$ ;
5. Repetir o passo 3 até que todos os coeficientes tenham sido validados.

# Capítulo 4

## Aplicações

### 4.1 Descrição dos Casos Analisados

As técnicas apresentadas nos capítulos anteriores foram aplicadas a dois modelos sintéticos adequadamente elaborados para demonstrar, não apenas as potencialidades mas também os aspectos negativos de cada método estudado. Os casos aqui analisados representam modelos sintéticos de reservatórios de petróleo cujas propriedades estáticas (porosidade, permeabilidade, etc.) foram simuladas a partir de algoritmos de geoestatística disponíveis gratuitamente através do *software SGeMS*<sup>®</sup> [38]. Os modelos gerados foram submetidos às técnicas de quantificação e propagação de incertezas estocásticas e análise do risco associado ao *VPL* apresentadas nos capítulos anteriores.

O modelo econômico adotado para a estimativa do *VPL* em ambos os casos analisados seguiu os mesmos parâmetros que foram utilizados em OLIVEIRA, 2006 [34], onde podem ser encontrados maiores detalhes sobre as premissas utilizadas.

Em vista do que foi apresentado nos capítulos anteriores, para determinação da curva de risco do *VPL*, foi seguida a seguinte sequência de etapas:

- Representação do campo randômico de entrada por meio de um conjunto de variáveis aleatórias padronizadas;
- Aproximação das respostas de interesse (função objetivo) através da expansão em caos polinomial das variáveis aleatórias decorrentes da etapa anterior;
- Estimativa dos coeficientes da expansão usando as técnicas discutidas no Capítulo 2;

- Avaliação das estatísticas de interesse - geração da curva de distribuição acumulada (curva de risco) para o *VPL*

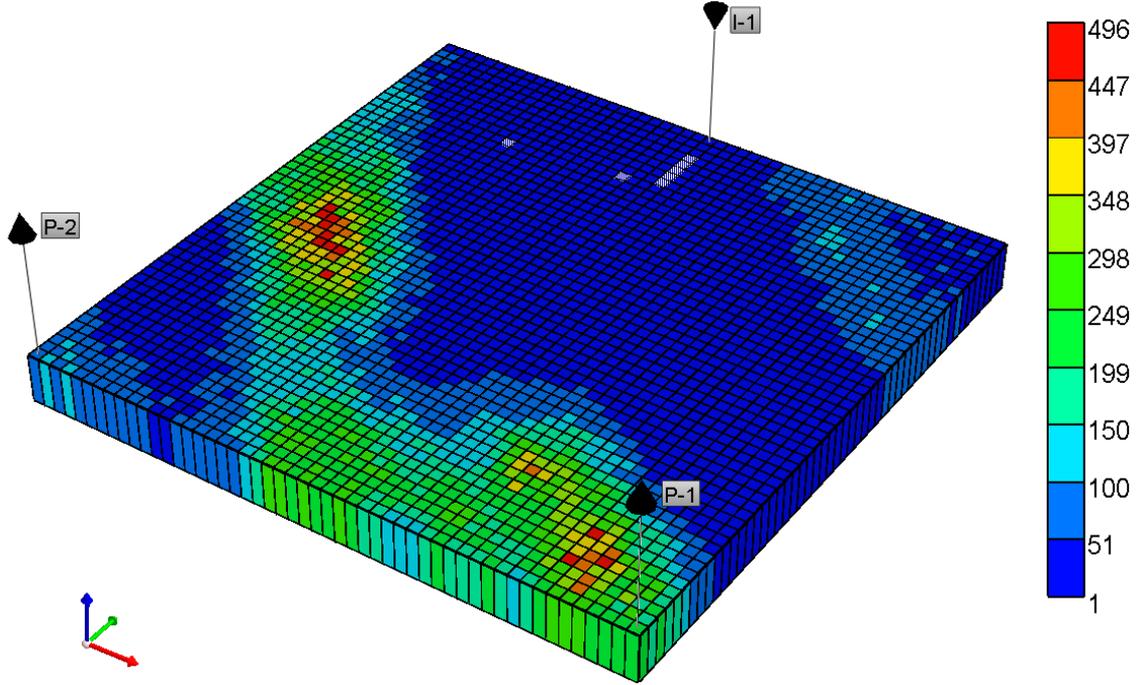
## 4.2 Caso 1

O primeiro caso estudado consiste em um modelo sintético de geometria e características simples, similar ao que foi utilizado em [34] e [9]. A escolha desse modelo justifica-se pela sua simplicidade e rápido tempo de simulação uma vez que o processo de propagação de incertezas é uma tarefa bastante onerosa computacionalmente pois exige muitas simulações completas para avaliação da função objetivo. Esta característica, portanto, permite explorar com profundidade os diversos aspectos que envolvem as técnicas apresentadas, o que seria inviável diante de um modelo com características mais realistas.

A estratégia de desenvolvimento adotada para o reservatório em questão envolve a injeção de água através de um poço injetor, I-1, e produção de petróleo em outros dois poços, P-1 e P-2 conforme Figura 4.1. Para tornar o problema mais realista foram admitidas restrições operacionais tanto no poço injetor quanto no poços produtores, representando limitações na capacidade de produção de líquidos e na capacidade de tratamento de água da plataforma. Conforme apresentado em [9], os controles de poço foram inseridos com intuito de gerar um modelo com grande variabilidade na esperança matemática do valor presente líquido (*VPL*) que representa a função objetivo de interesse. Assim, os poços produtores podem produzir individualmente até  $50m^3/d$  de líquidos e, adicionalmente, estão submetidos a um limite na capacidade de processamento de líquidos na plataforma de  $80m^3/d$ . O poço injetor está limitado a uma vazão máxima de  $70m^3/d$  de água. Neste problema foi utilizada uma única etapa de controle da vazão no tempo e a distribuição das vazões de produção por poço foi fixada em 25% para o poço P-1 e 75% para o poço P-2, inversamente proporcionais ao *IP* dos poços para se conseguir um melhor varrido e, conseqüentemente, uma drenagem mais eficiente do reservatório. A malha de simulação (tamanho e geometria), o tipo e a propriedades dos fluidos do modelo são os mesmos do Caso 1 de OLIVEIRA, 2006 [34].

O *caso 1*, portanto, consiste em um modelo de reservatório bidimensional com dois poços produtores e um injetor de água. OLIVEIRA, 2006 [34] analisou o problema da alocação dinâmica das vazões de produção visando maximizar o *VPL* em um

Figura 4.1: Modelo de Simulação do Caso 1 sobre o mapa de permeabilidade.



reservatório cujas propriedades estáticas (permeabilidade e porosidade) foram caracterizadas deterministicamente. FONSECA, 2010 [9] utilizou uma versão estocástica do mesmo problema onde foram geradas realizações do campo de permeabilidade a partir do método geoestatístico de Simulação Sequencial Gaussiana (*SGSIM*). Assim como em FONSECA, 2010 [9], os parâmetros de entrada, tais como média e variância da permeabilidade, foram escolhidos de modo a gerar uma grande variabilidade no valor esperado do *VPL*.

O campo de permeabilidades foi modelado segundo uma distribuição lognormal com média 100 e desvio-padrão 48. Por simplicidade, optou-se por trabalhar com o logaritmo da propriedade, uma vez que este possui distribuição normal cuja média e desvio padrão podem ser obtidos pelas equações a seguir:

$$\mu = \log(m^2 / \sqrt{v + m^2}) \quad (4.1)$$

$$\sigma = \sqrt{\log\left(\frac{v}{m^2 + 1}\right)} \quad (4.2)$$

sendo  $m$  e  $v$  respectivamente a média e a variância da distribuição lognormal correspondente. Com estes parâmetros e assumindo um modelo de covariância exponencial com efeito pepita de 0, 1 e comprimento de correlação de 3/5 da dimensão espacial do

modelo, foram geradas 1000 realizações equiprováveis do campo de permeabilidade horizontal. Foi admitida a existência de uma correlação entre porosidade e permeabilidade dada pela Equação 4.3:

$$\phi = 0.13390 + 0.01305 \cdot \log(K) \quad (4.3)$$

onde  $\phi$  é porosidade e  $K$  a permeabilidade. Para fins de simplificação, tendo em vista que o objetivo do trabalho é avaliar o impacto de incertezas nas variáveis estocásticas, todas as outras propriedades do modelo foram consideradas determinísticas. A Figura 4.2 apresenta algumas destas realizações.

O modelo foi construído em uma malha bidimensional com 51 células de dimensões 10 metros em cada direção e quatro metros na vertical representando a espessura do reservatório. Isto significa que o modelo contém 2601 variáveis aleatórias correlacionadas uma vez que o único atributo incerto deste problema é o campo de permeabilidade horizontal e todas as outras propriedades do reservatório em estudo são conhecidas deterministicamente. Por se tratar de um campo randômico gaussiano, foi utilizada a expansão de *Karhunen-Loeve* para parametrização da variável de entrada. Serão apresentadas e comparadas duas soluções para parametrização da função objetivo, utilizando expansão em caos polinomial: a) via método da colocação probabilística (*PCM*) e b) pelo método de projeção espectral não-intrusiva (*NISP*).

A expansão de *Karhunen-Loeve* foi aplicada à matriz de covariância gerada a partir das 1000 realizações do campo de permeabilidade. Os autovalores e o percentual da variância explicada por cada par autovalor/autovetor está apresentado na Figura 4.3. Pode-se observar que a maior parte da variância é explicada por poucos pares. Neste modelo, os 30 primeiros pares são responsáveis por quase 60% da variabilidade total. Isto significa que com apenas 30 variáveis aleatórias independentes, ao invés de 2601 variáveis aleatórias correlacionadas, é possível gerar realizações do campo de permeabilidade com variabilidade próxima de 60% da variabilidade original.

O percentual de variância explicada por cada par autovalor/autovetor é dependente do alcance do variograma (comprimento de correlação) utilizado na modelagem geoestatística e não do número de células da malha. De modo geral, quanto maior o alcance, maior será a redução no número de pares necessários para representar a variável de entrada, o que significa uma economia significativa no esforço computacional requerido para a propagação e quantificação de incertezas. Conforme explicado no Capítulo 2, a decomposição do campo randômico original através da expansão de *Karhunen-Loeve* torna possível a representação do mesmo em termos de um conjunto

Figura 4.2: Realizações originais do logaritmo da Permeabilidade.

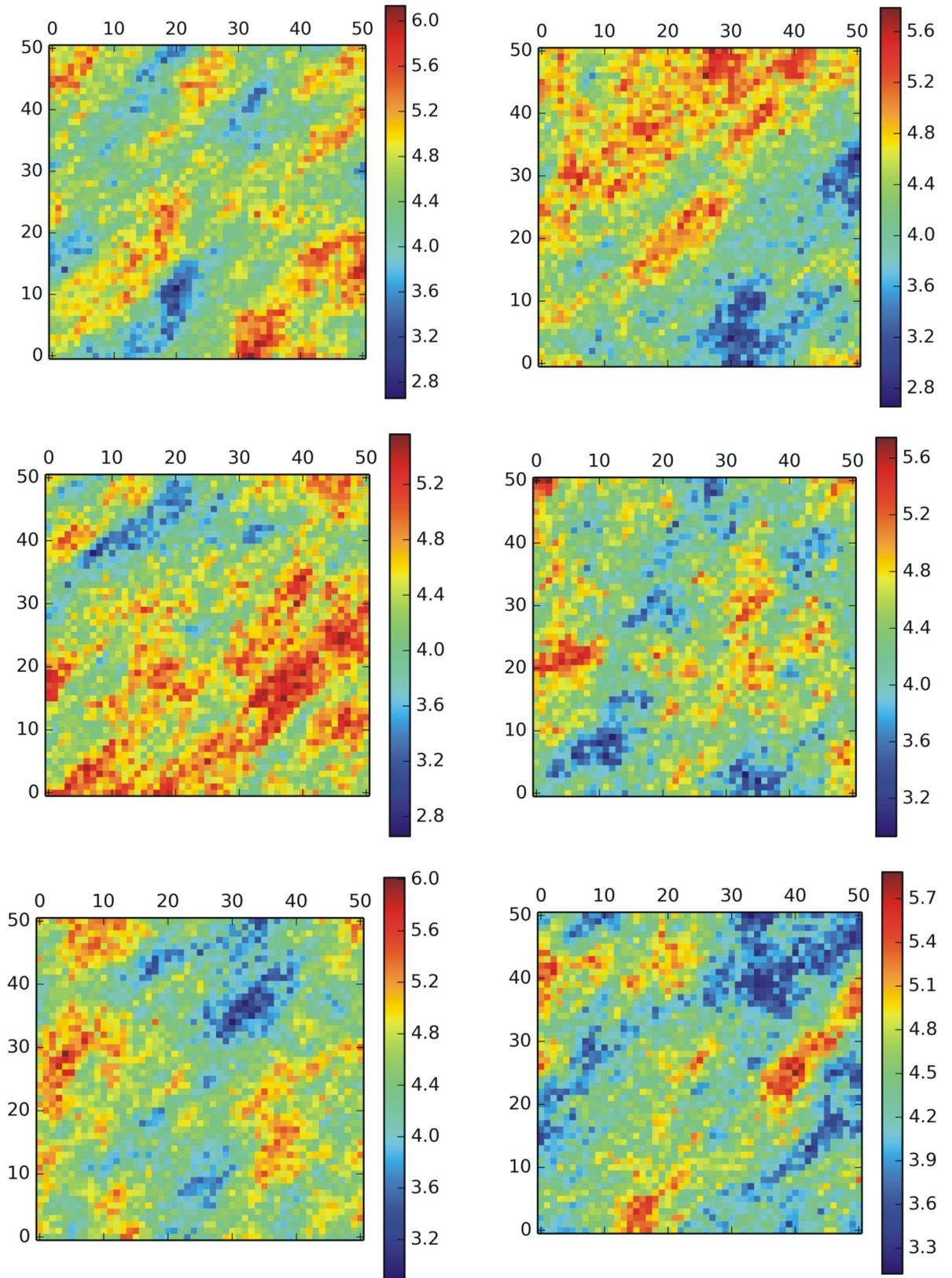
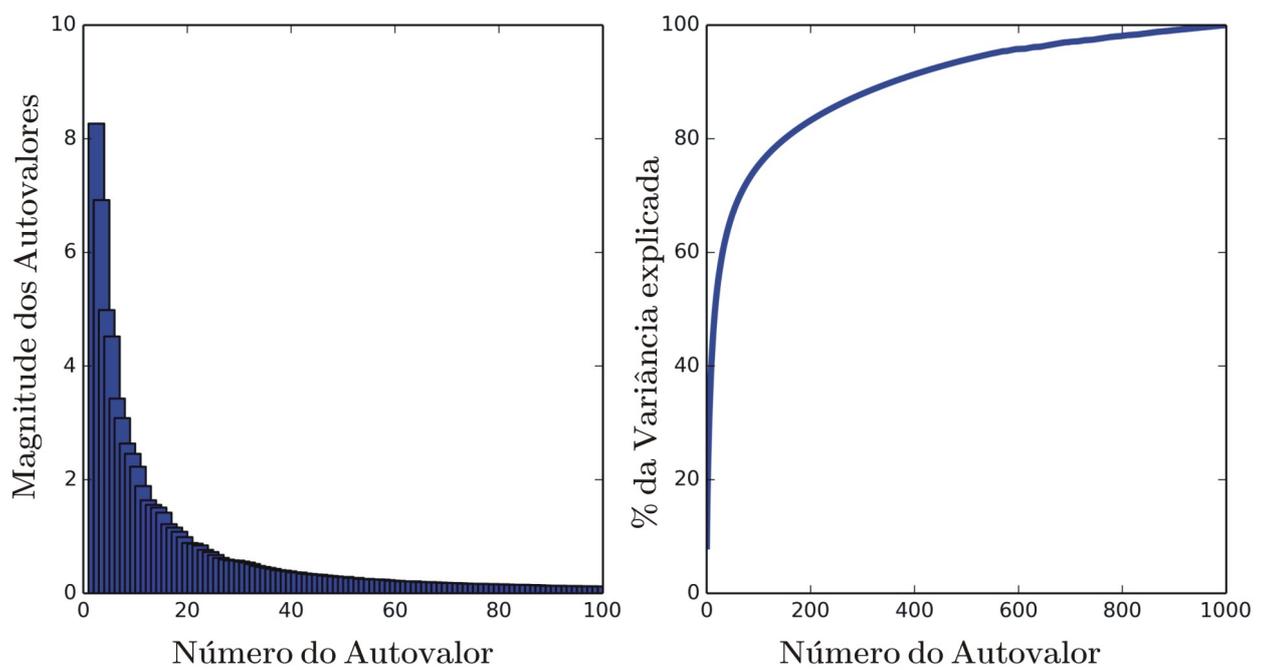


Figura 4.3: Magnitude dos autovalores da matriz de covariância e o % de variância explicada por cada um deles.



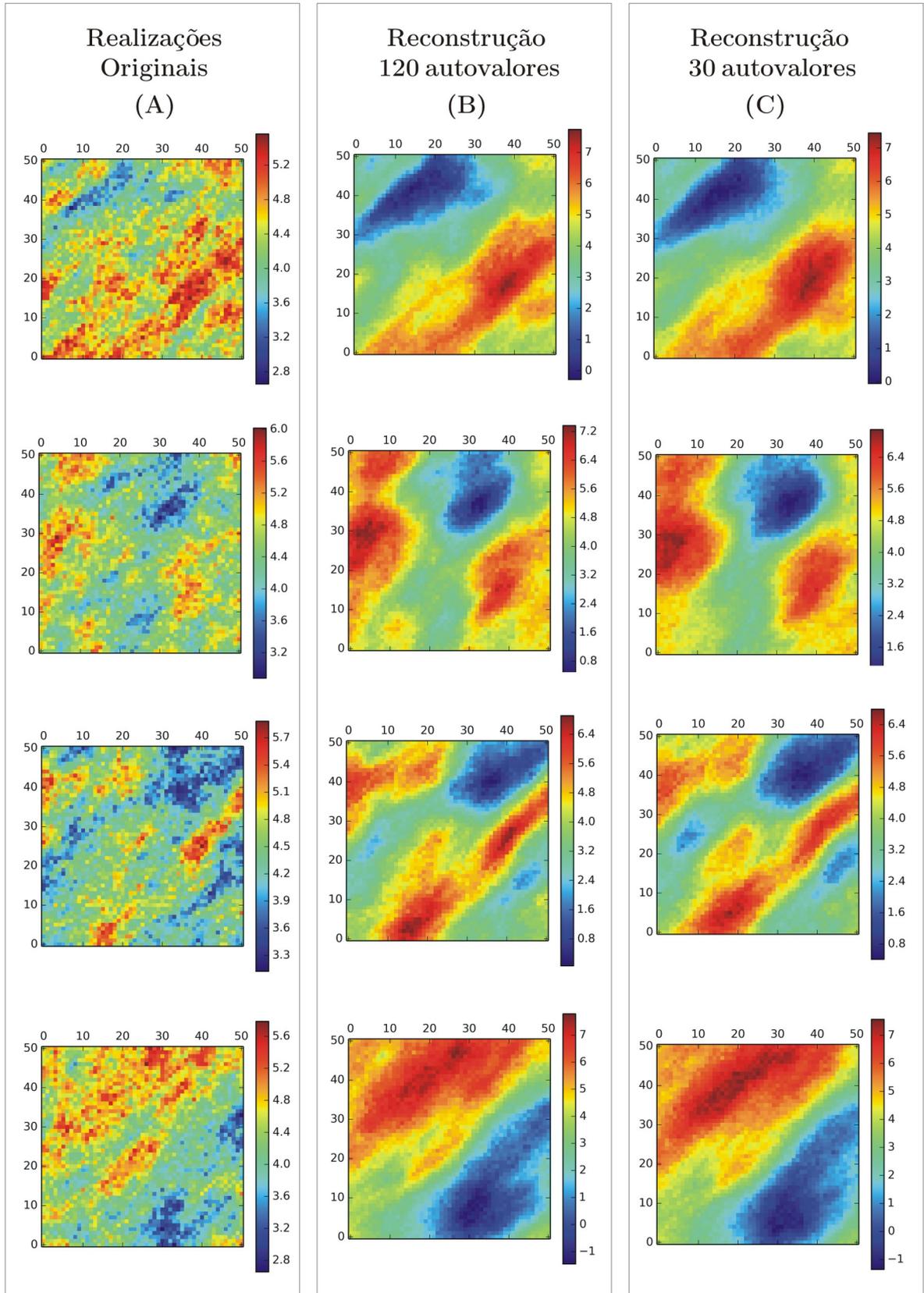
de variáveis aleatórias independentes ao mesmo tempo que mantém a estrutura da matriz de covariância. Para exemplificar, a Figura 4.4 apresenta quatro realizações do campo de permeabilidade original e as realizações construídas a partir da expansão de *Karhunen-Loeve* utilizando 120 (75% da variância total) e 30 (60% da variância total) componentes principais.

As realizações do campo de permeabilidade construídas a partir da expansão de *Karhunen-Loeve* na sua forma truncada, mesmo retendo-se cerca de 75% da variabilidade total conforme a Figura 4.4, apresentam padrões suavizados caracterizando a preservação da correlação de longo alcance em detrimento das estruturas de curta correlação (baixa frequência). Essas representações não são campos de permeabilidade reais, mas são capazes de capturar a estrutura principal dos campos randômicos. Adicionalmente, uma vez que o comportamento do fluxo e, portanto, o *VPL* são principalmente dependentes das estruturas geológicas de correlação longa [42], tal aproximação das variáveis de entrada parece ser razoável, ao menos para campos gaussianos.

### 4.2.1 Parametrização do *VPL* via *PCM*

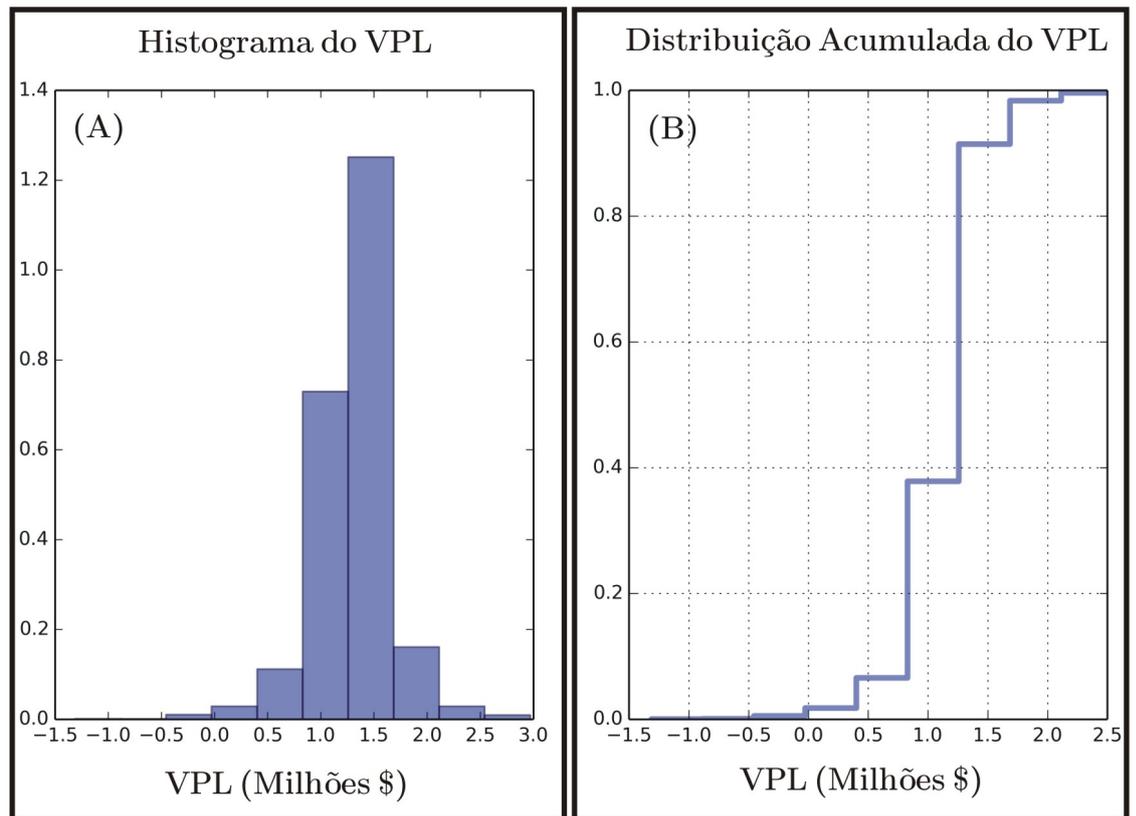
Já foi visto que a expansão de *Karhunen-Loeve* representa uma forma bastante eficiente de parametrização do campo randômico de entrada. Porém, como a tarefa de propagação e quantificação de incertezas é geralmente uma tarefa que demanda um grande número de simulações, é preciso encontrar uma forma mais eficiente de representar, mesmo que de forma aproximada, os resultados que se deseja analisar. Neste estudo foi utilizada uma expansão em Caos Polinomial (*PCE*) em termos de polinômios de Hermite. Conforme discutido no Capítulo 2, o número de termos da expansão, intrínsecamente relacionado ao número de simulações requeridas pelo método, aumenta exponencialmente à medida em que aumenta o grau da expansão e o número de variáveis de entrada. A expansão do Caos Polinomial foi aplicada aos modelos construídos através da expansão de *Karhunen-Loeve* truncada a 30 autopares. Com 30 variáveis representando a incerteza na entrada, foi possível gerar uma aproximação polinomial do 2º grau para o *VPL* com apenas 496 simulações, utilizando o método de Colocação Probabilística (*PCM*) para determinar os valores atribuídos ao vetor aleatório para cada simulação. Em seguida foi aplicado o Método de Monte Carlo com 2000 amostras sobre o polinômio gerado pela expansão *PCE* para determinar as estatísticas da variável de interesse, nesse caso o *VPL*. A Figura 4.5 apresenta o histograma e o gráfico de frequências acumuladas para o *VPL*, expresso em milhões de unidades monetárias

Figura 4.4: Realizações do campo de permeabilidades: (A) Realizações originais; (B) Reconstrução utilizando expansão K-L com 120 autovalores; (C) Reconstrução utilizando expansão K-L com 30 autovalores.



(milhões \$).

Figura 4.5: Curva de Risco para o VPL obtido via PCE+PCM: (A) Histograma do VPL (milhões \$); (B) Distribuição Acumulada do VPL (milhões \$).



Para avaliar a qualidade dos resultados utilizando-se o modelo substituto composto por expansões  $PCE + PCM$  com parametrização do campo de permeabilidade usando 30 autovalores, foi executada uma simulação de Monte Carlo a partir das realizações geoestatísticas originais. A Figura 4.6 mostra a convergência da média e da variância do  $VPL$  a partir das realizações geoestatísticas. O resultado obtido indicou que 1000 realizações seriam suficientes para atingir a convergência na média e na variância.

Assim, as 1000 realizações geradas pelo método de Monte Carlo foram simuladas e os resultados, comparado aos resultados obtidos através do modelo substituto, estão apresentados na Figura 4.7.

Apesar da semelhança no valor médio do  $VPL$  obtidos pelos dois métodos, no desvio padrão o resultado foi bastante diferente. Isto pode ser notado ao observar na Figura 4.7 que a curva de distribuição gerada por  $PCE+PCM$  apresenta uma longa cauda em contraposição à curva gerada por Monte Carlo. Este fenômeno ocorre pelo fato de

Figura 4.6: Convergência da média e da variância do VPL a partir das realizações geostatísticas: (A) Média do VPL em (\$ milhões); (B) Variância do VPL em (\$ milhões).

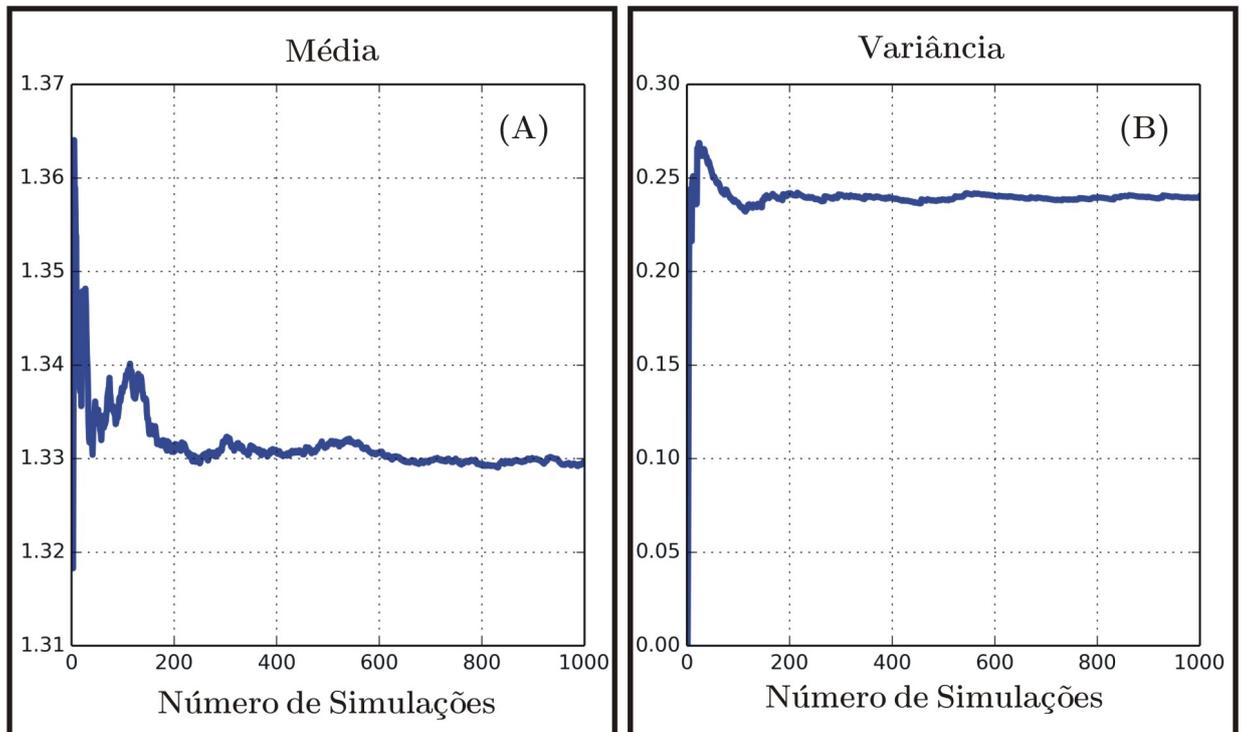
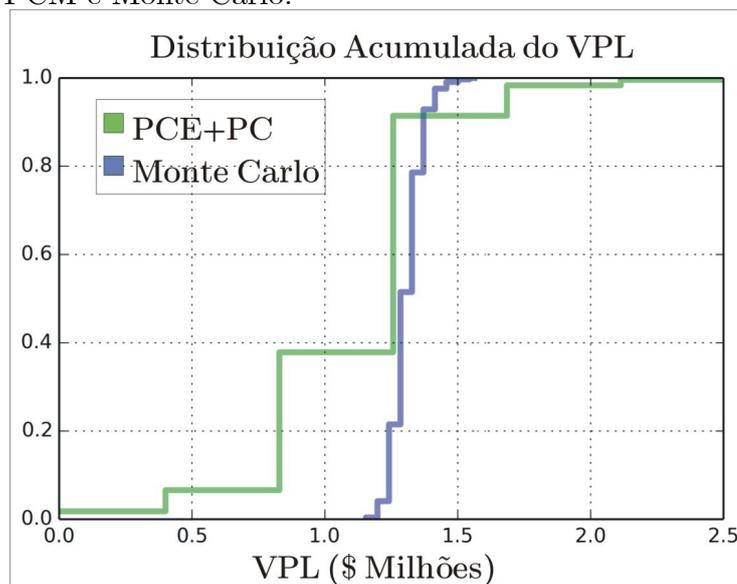


Figura 4.7: Comparação da curva de risco para o VPL (em \$ milhões) obtidos pelos métodos PCE+PCM e Monte Carlo.



que a curva de distribuição, no caso  $PCE+PCM$ , ser gerada através de amostragens com base em funções polinomiais e não diretamente baseadas nas equações que regem o modelo. Conseqüentemente, certas restrições físicas podem ser violadas durante processo de amostragem. Além disso, é sabido que a expansão  $PCE$  construída com  $PCM$  captura preferencialmente as informações contidas nas regiões de alta probabilidade de ocorrência dos parâmetros de entrada. As regiões de baixa probabilidade (caudas da distribuição normal), responsáveis pela variabilidade da variável, deixam de ser bem representadas e daí vem o desvio na variância do  $VPL$ . Outro fator que também contribui para esse resultado é o fato de estar sendo utilizada uma parametrização do campo de entrada com dimensionalidade reduzida, que provoca um efeito de suavização no mapa da propriedade, provocada pela perda das estruturas de curta correlação.

#### 4.2.2 Parametrização do $VPL$ com $NISP$

O método  $NISP$  utiliza amostragens por meio de técnicas de quasi-Monte Carlo com sequências de baixa discrepância para estimar os coeficientes da expansão  $PCE$ . Neste trabalho, após algumas tentativas foi verificado que 150 simulações seriam suficientes para produzir estimativas com boa precisão. A sequência de Sobol [40] foi então utilizada para gerar 150 amostras das variáveis aleatórias que representam o campo randômico. A Figura 4.8 apresenta a disposição dos pontos para duas das variáveis aleatórias usadas para parametrizar a permeabilidade.

Com base nas 150 amostras geradas acima e, utilizando-se a formulação descrita no Capítulo 3, foram obtidos os coeficientes da expansão. Foi gerada, a princípio, uma expansão  $PCE$  de 5ª ordem. A título de ilustração, a Figura 4.9 apresenta os coeficientes calculados.

Após o processo de validação cruzada necessário para a decisão de manter ou descartar os termos de alta ordem da expansão, restaram apenas os termos lineares, conforme pode ser visto na Figura 4.10. Na mesma figura apresenta-se o gráfico de qualidade da aproximação onde são confrontados os valores preditos pelo polinômio e os valores simulados.

Após o processo de validação, a expansão  $PCE$  fica totalmente definida. A fim de avaliar as estatísticas de interesse para o  $VPL$ , foram geradas 2000 amostras aleatórias e foi calculado o  $VPL$  através do polinômio  $PCE$ . Os resultados obtidos estão apresentados na Figura 4.11.

Por fim, a Figura 4.12 apresenta a comparação dos resultados obtidos pelo  $NISP$  e por 1000 amostras geradas por geoestatística que, conforme vimos anteriormente,

Figura 4.8: Exemplo de amostra usando sequência de Sobol.

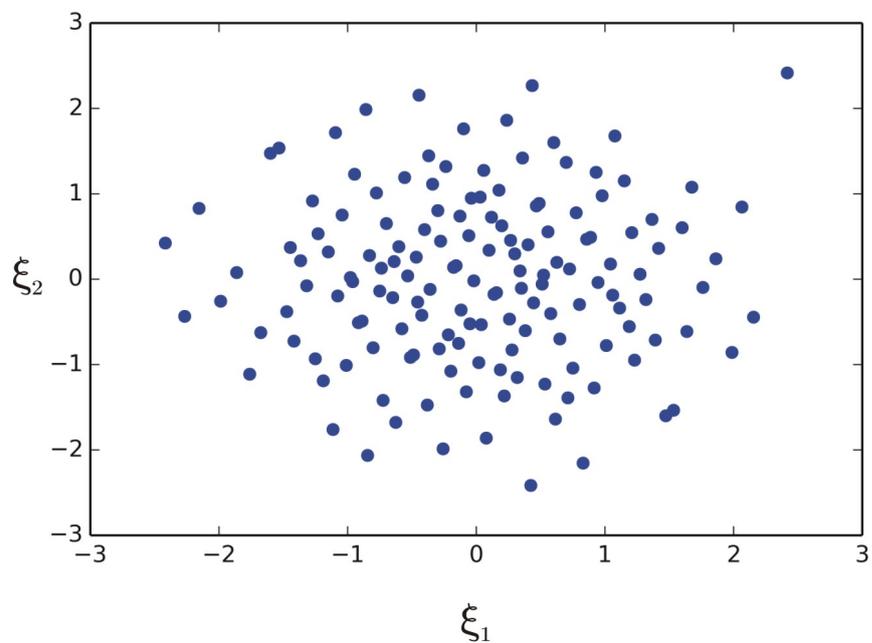


Figura 4.9: Coeficientes da expansão PCE pelo método NISP.

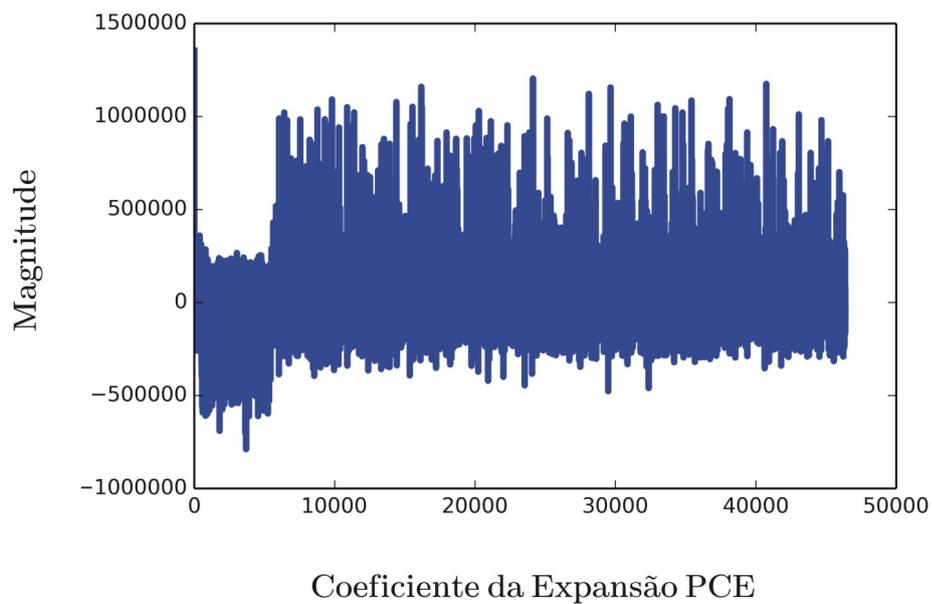


Figura 4.10: Resultados NSIP após validação cruzada: (A) Coeficientes da expansão; (B) comparação entre valores preditos pela expansão versus valores simulados.

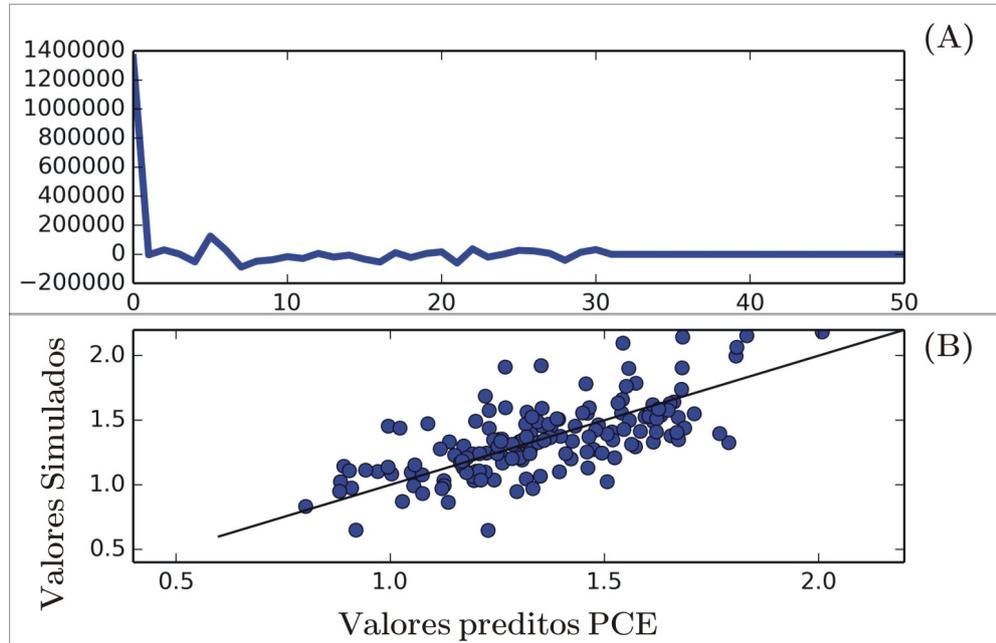
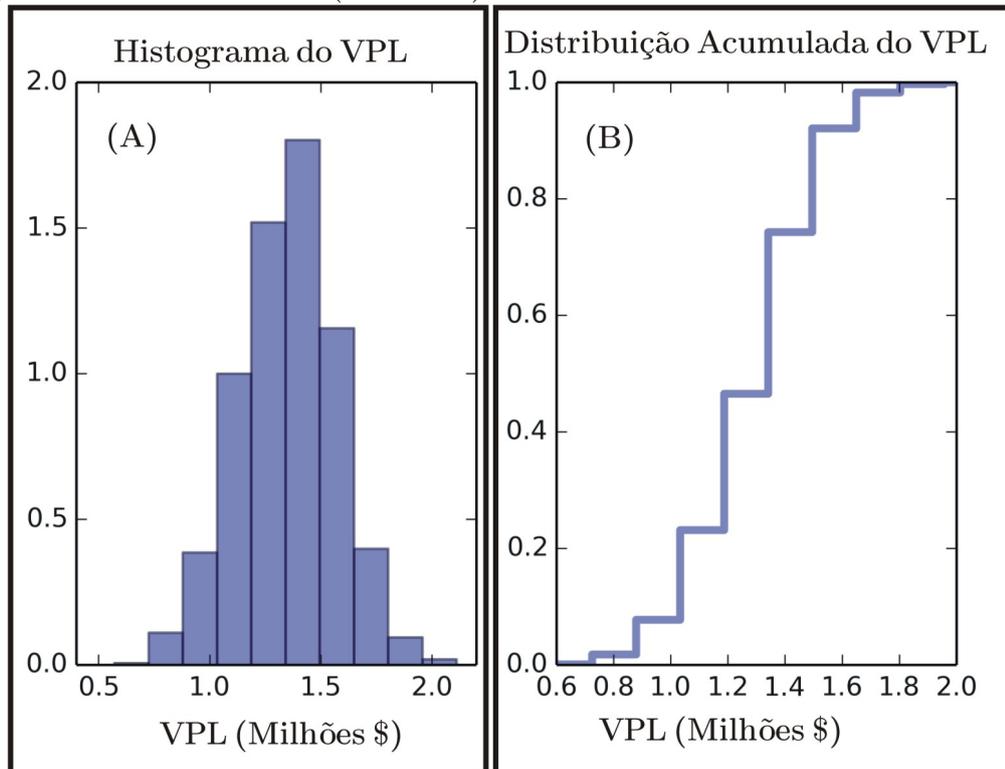
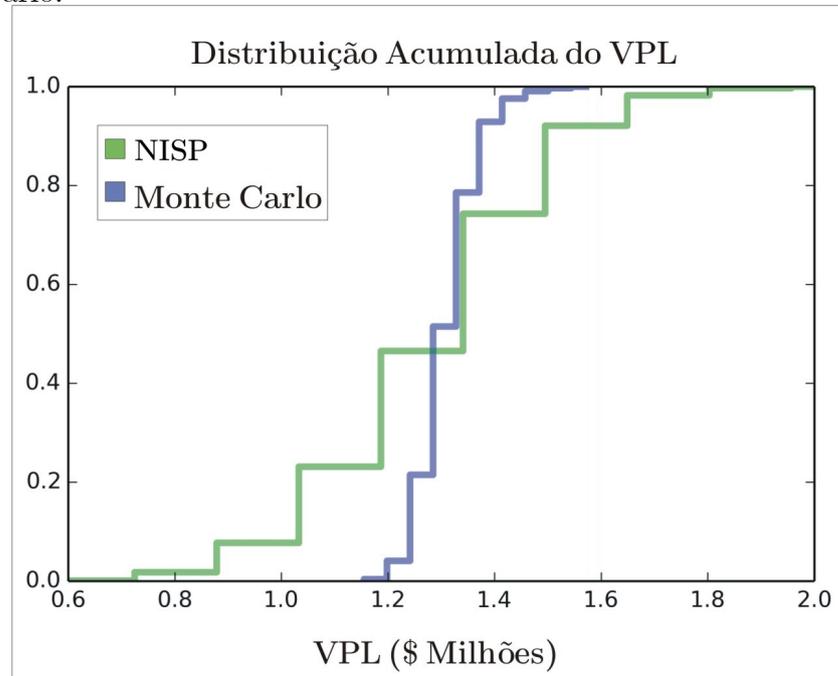


Figura 4.11: Resultados do NISP: (A) Histograma do VPL (\$ milhões); (B) Distribuição Acumulada do VPL (\$ milhões).



podem ser usadas para aproximar a distribuição verdadeira.

Figura 4.12: Comparação dos resultados para o VPL obtidos via NISP e pelo Método de Monte Carlo.



Ao comparar as Figuras 4.7 e 4.12, observa-se que a distribuição do *VPL* obtida pelo método *NISP* ficou ligeiramente mais próxima da distribuição dita verdadeira. Adicionalmente, o método *NISP* leva vantagem também no número de simulações necessárias (150 contra 496 do *PCM*). Porém, o fato de descartar os termos de alta ordem e os termos cruzados da expansão pode ser uma limitação importante, principalmente nos casos em que o modelo substituto assim gerado for utilizado para processos de otimização e ajuste de histórico que são problemas fortemente não-lineares. Este certamente é um fator que merece ser investigado no futuro. A Tabela 4.1 (valores em \$ milhões) apresenta a comparação entre os resultados das principais métricas obtidos pelos dois casos, tendo como referência os valores obtidos a partir das realizações geoestatísticas simuladas pelo método de Monte Carlo.

### 4.3 Caso 2

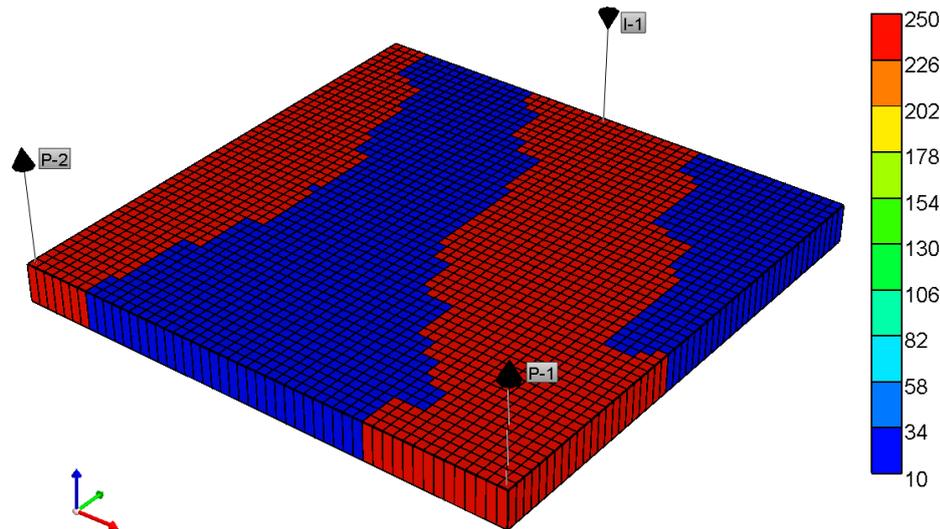
O caso 2 é idêntico ao primeiro caso estudado em termos de geometria da malha de simulação, propriedades dos fluidos, quantidade de poços e restrições operacionais. A

Tabela 4.1: Resultados da Análise deRisco para o Caso 1

Estatística	Monte Carlo	$PCE+PCM$	$PCE+NISP$
Média	1.29	1.17	1.19
Variância	0.36	0.82	0.63
P10	1.22	0.47	0.87
P90	1.35	1.27	1.38

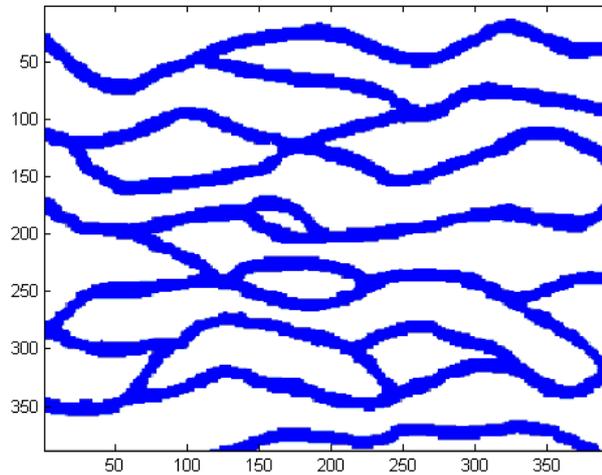
diferença fundamental reside na caracterização do campo de permeabilidade, conforme Figura 4.13. Esse caso foi elaborado com a intenção de testar a performance dos métodos de parametrização adequados a campos randômicos não gaussianos, *Kernel Principal Component Analysis (KPCA)* e *Pluri-PCA*

Figura 4.13: Exemplo de campo de permeabilidades usado no Caso 2.



As realizações geoestatísticas foram obtidas através do algoritmo *SNESIM* [46], também integrante do software *SGeMS*<sup>®</sup>. Para execução da simulação geoestatística, o algoritmo necessita de uma imagem de treinamento a partir da qual são extraídas as estruturas e padrões geológicos concebidos pelos geocientistas. A partir da imagem de treinamento mostrada na Figura 4.14, foram geradas 1000 realizações não condicionadas em uma malha bidimensional com 51x51 células do campo de permeabilidade representando um ambiente canalizado. Alguns exemplos de realização da permeabilidade encontram-se apresentadas na Figura 4.15. Vale salientar que, neste exemplo, a permeabilidade segue uma distribuição discreta com valores 250 mD dentro do canal e 10 mD fora.

Figura 4.14: Imagem de Treinamento utilizada para geração das realizações do campo de permeabilidade do caso 2 (Caers e Zhang, 2004).



### 4.3.1 Parametrização do campo de permeabilidade

#### Método *KPCA*

As 1000 realizações geoestatísticas geradas conforme mencionado anteriormente, foram usadas para calcular a matriz de covariância do campo randômico. A Figura 4.16 apresenta os autovalores da matriz de covariância ordenados segundo sua magnitude. Neste modelo, os 30 primeiros pares são responsáveis por quase 40% da variabilidade total. Isto significa que, com apenas 30 variáveis aleatórias independentes, ao invés de 2601 variáveis aleatórias correlacionadas, é possível gerar realizações do campo de permeabilidade com variabilidade próxima de 40% da variabilidade original. Em contrapartida, os 400 maiores autopares são responsáveis por 80% da variabilidade.

Seguindo a metodologia explicada no Capítulo 2, é possível gerar realizações do campo de permeabilidade usando *KPCA* em sua forma truncada ou não. Para exemplificar, a Figura 4.17 apresenta quatro realizações construídas a partir da expansão de *Karhunen-Loeve* utilizando 400 (80% da variância total) e 30 componentes principais (40% da variância total).

Deve-se notar que as realizações de permeabilidade geradas através do *KPCA* apresentam valores reais contínuos para uma variável que originalmente era discreta. Esse justamente é um dos pontos negativos do método para representar variáveis discretas (como as fácies, por exemplo) e que é explorado pelo método *Pluri-PCA*.

Figura 4.15: Exemplos de realizações do logaritmo da permeabilidade - Caso 2.

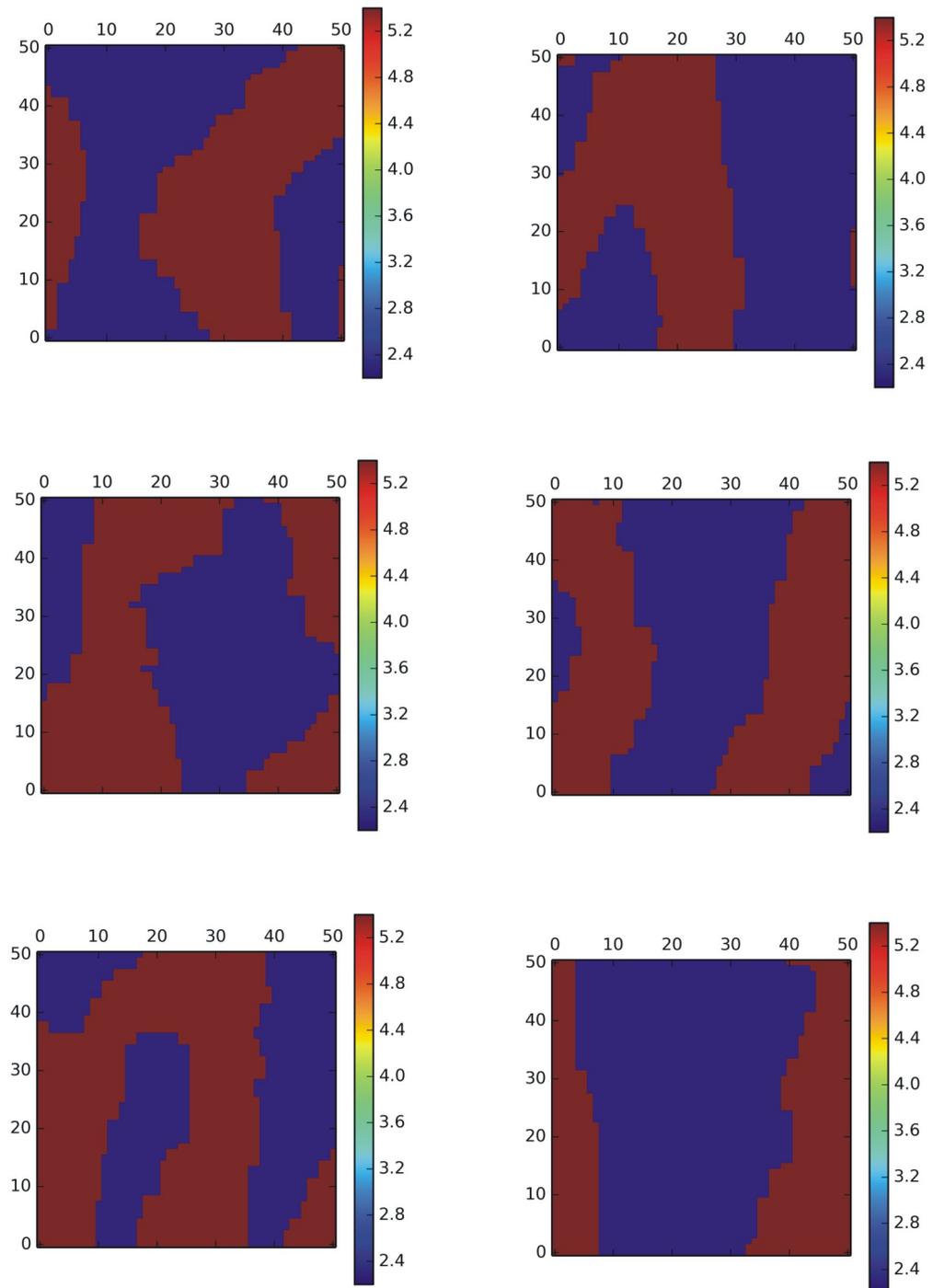
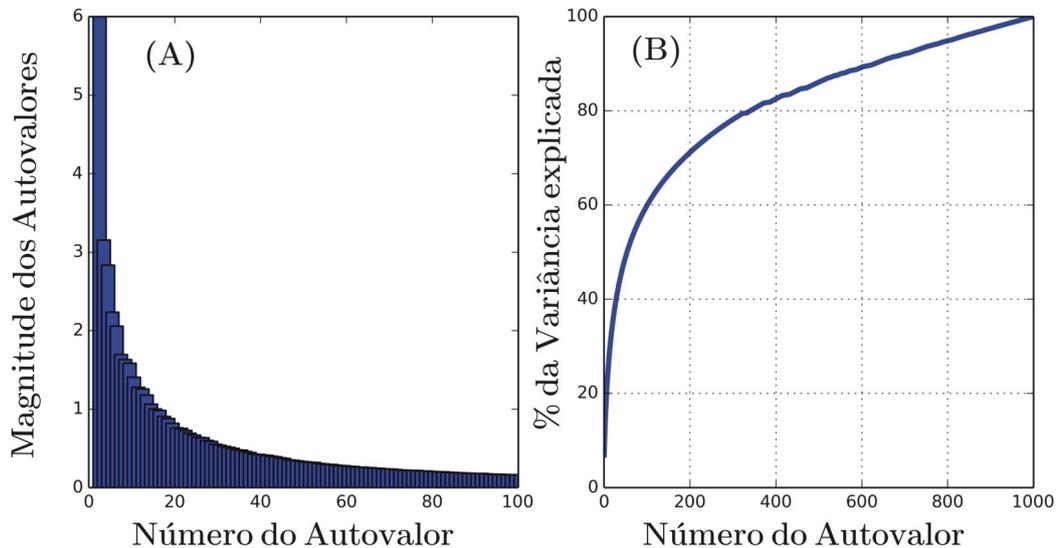


Figura 4.16: Autovalores da Matriz Kernel usada no KPCA: (A) Magnitude do autovalores; (B) Percentual da variabilidade dos componentes principais.



### Método Pluri-PCA

O método *Pluri-PCA* foi utilizado aqui para reproduzir as realizações originais e simular novas realizações não do campo de permeabilidade mas sim das fácies que, em termos práticos, condicionam a realização da permeabilidade. Exemplos de realizações de fácies reconstruídas com essa técnica estão apresentadas na Figura 4.18

Conforme descrito no Capítulo 2, inicialmente é aplicada a expansão  $K-L$  sobre o vetor de fácies expandido, o que gera como resultado um vetor de fácies real e não discreto como o original. As realizações originais são então utilizadas para determinar as regras de litologia que serão empregadas para determinar o ponto de corte através do qual será atribuído o indicador de fácies correspondente. A Figura 4.19 ilustra as regras de litologia usadas na reconstrução das realizações mostradas na Figura 4.18

No contexto da análise de incertezas, o método *pluri-PCA* foi utilizado para gerar novas realizações das fácies, parametrizadas por 30 variáveis aleatórias após aplicação prévia da expansão  $K-L$ . A Figura 4.20 apresenta o resultado da decomposição em valores singulares da matriz de indicadores de fácies expandido.

Neste exemplo, foi atribuído o valor de 250 mD para as células com indicador de fácies 1 e 10 mD para a fácies 0 e seguiu-se com a parametrização da função objetivo e determinação da curva de risco. Em virtude da notória vantagem em relação ao *KPCA*

Figura 4.17: Realizações da permeabilidade do caso 2 usando KPCA: (A) Considerando 400 autovalores (80% da variabilidade); (B) Considerando 30 autovalores (40% da variabilidade).

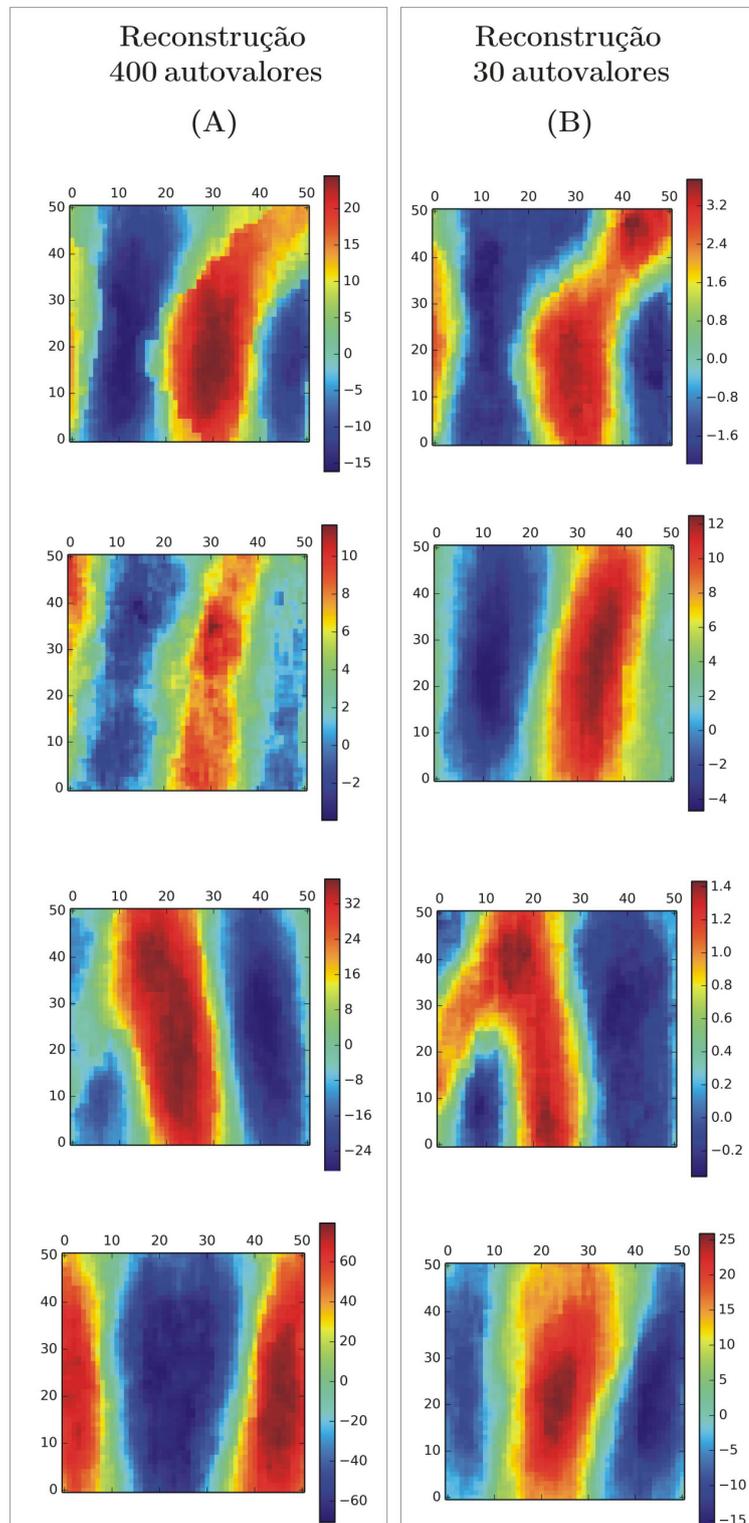


Figura 4.18: Exemplos de realizações reconstruídas com Pluri-PCA - Caso 2.

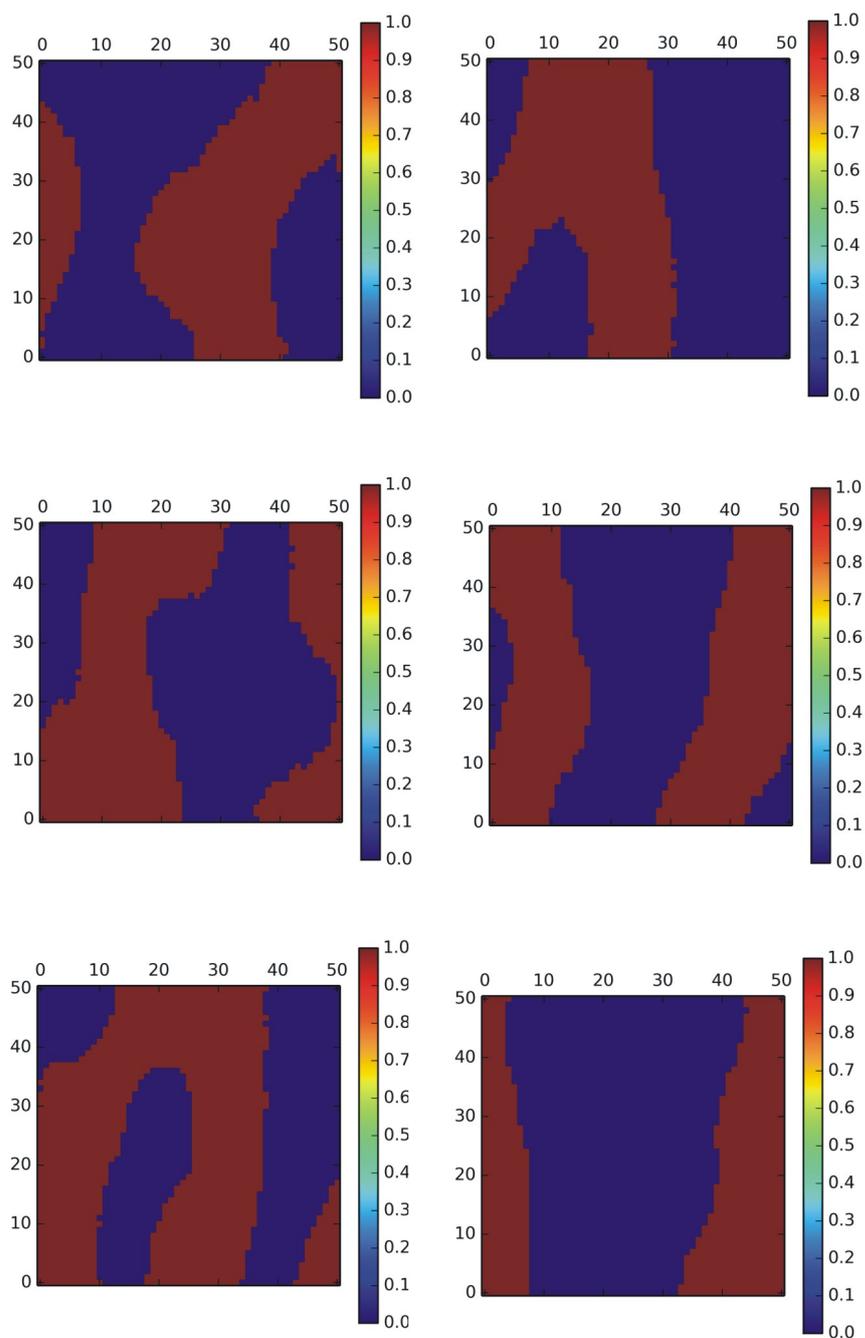


Figura 4.19: Caso 2 - exemplo de regra de litologia e ponto de corte determinados por uma das realizações originais.

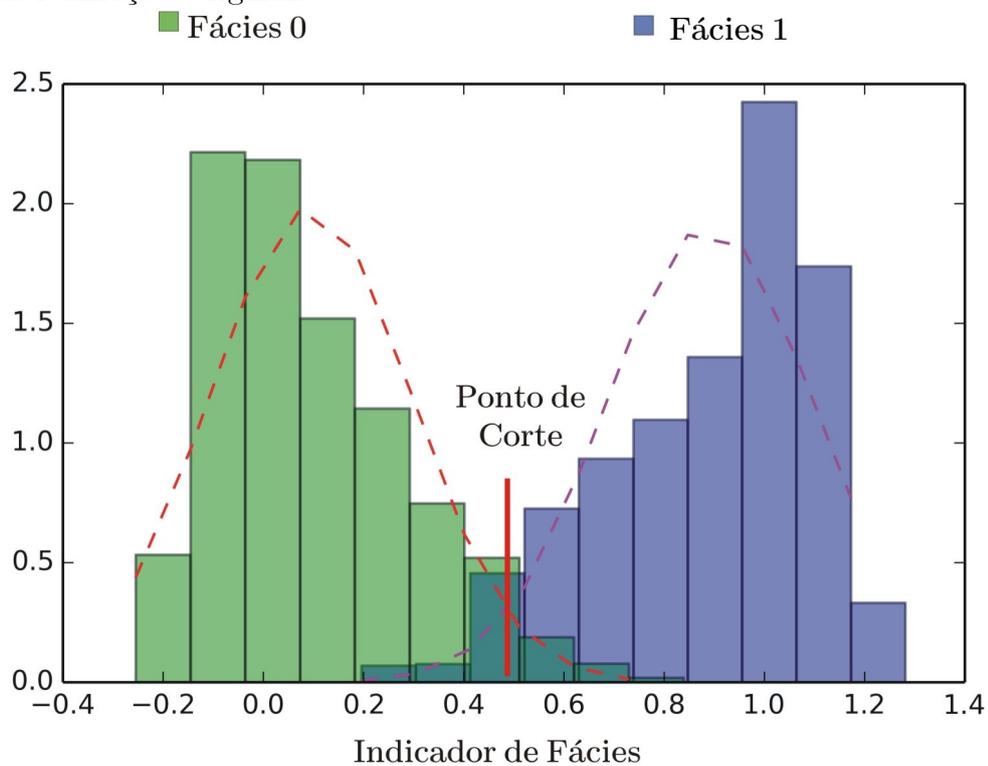
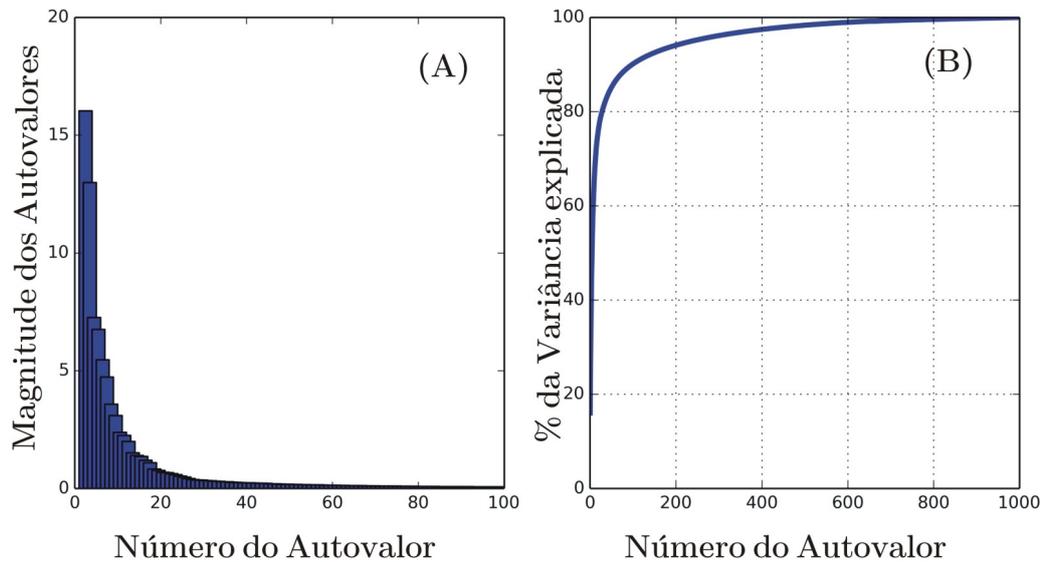


Figura 4.20: Caso 2 - autovalores da matriz de covariância a partir do vetor de indicadores de fácies expandido: (A) Magnitude dos autovalores; (B) Percentual da variabilidade explicada pelos componentes principais.



no que diz respeito a reproduzir variáveis discretas, as técnicas de parametrização da função objetivo foram aplicadas apenas ao caso de parametrização do campo randômico de entrada por meio do método *pluri-PCA*. A Figura 4.21 apresenta um exemplo de uma realização aleatória de fácies gerada pelo método *pluri-PCA*.

### 4.3.2 Parametrização do *VPL* via *PCM*

Assim como no Caso 1 a expansão do Caos Polinomial foi aplicada aos modelos construídos através do método *pluri-PCA* associado à expansão de *Karhunen-Loeve* truncada a 30 autopares do vetor de indicador de fácies. Com 30 variáveis representando a incerteza na entrada, foi possível gerar uma aproximação polinomial do 2º grau para o *VPL* com apenas 496 simulações, utilizando o método Colocação Probabilística (*PCM*) para determinar os valores atribuídos ao vetor aleatório para cada simulação. Em seguida foi aplicado o Método de Monte Carlo com 2000 amostragens sobre o polinômio gerado pela expansão *PCE* para determinar as estatísticas da variável de interesse, nesse caso o *VPL*. A Figura 4.22 apresenta o histograma e o gráfico de frequências acumuladas para o *VPL*.

Figura 4.21: Caso 2 - Diagrama esquemático exemplificando a geração de realizações aleatórias usando o pluri-PCA.

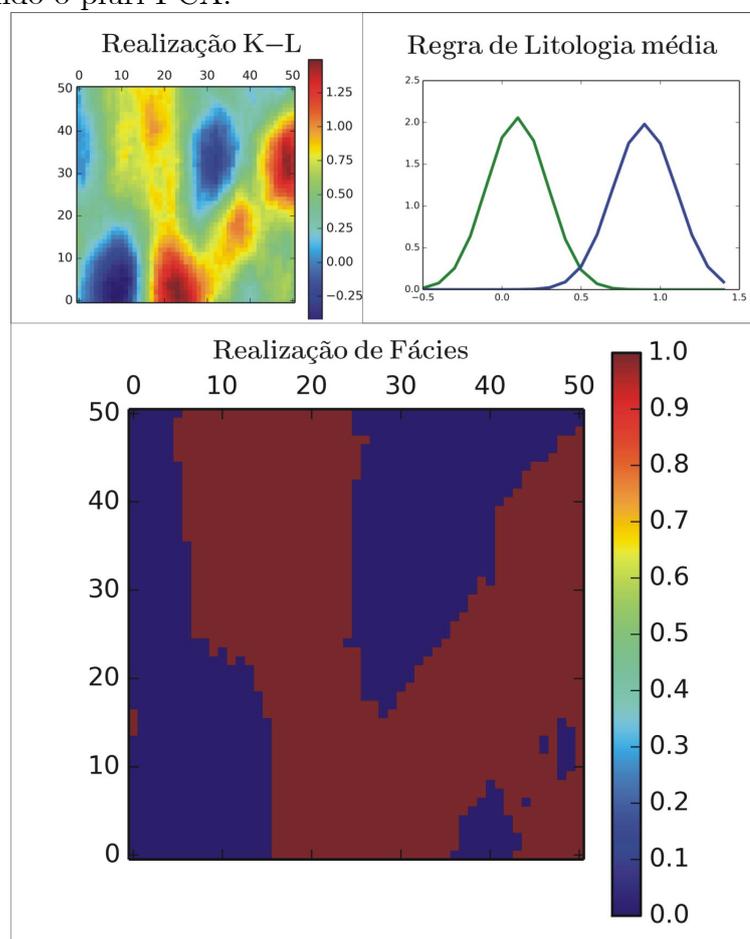


Figura 4.22: Curva de risco para o VPL obtida pelo método PCE+PCM: (A) Histograma do VPL; (B) Distribuição Acumulada do VPL.

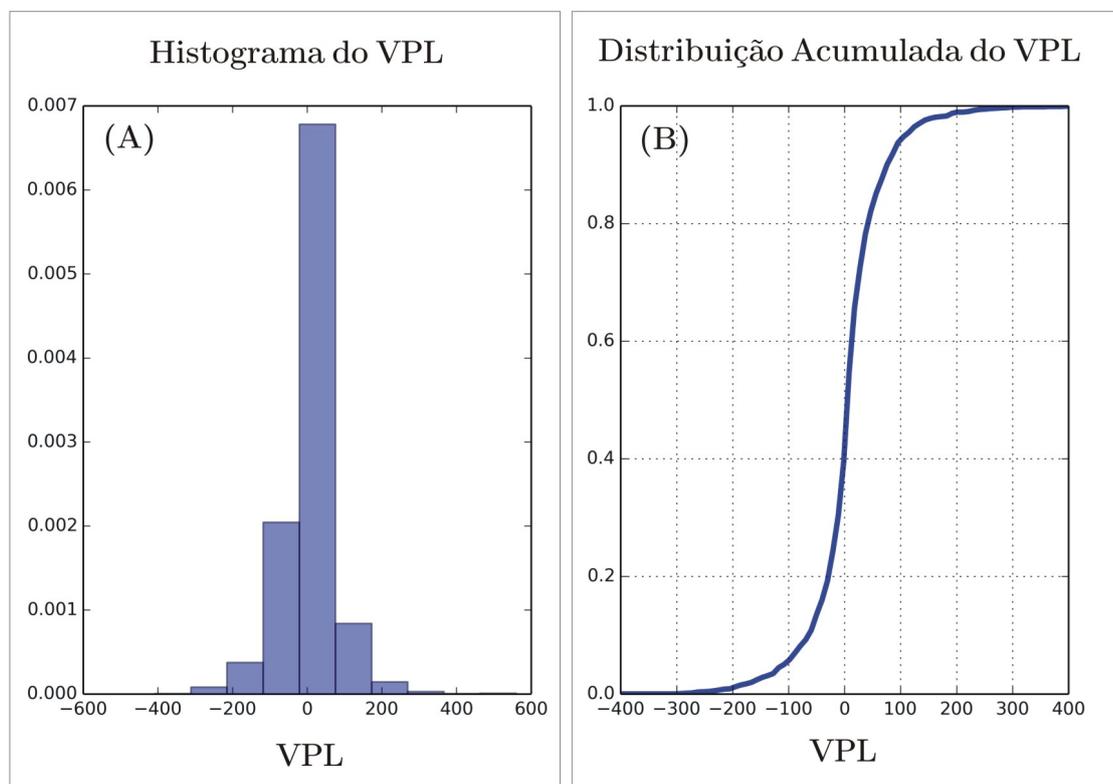
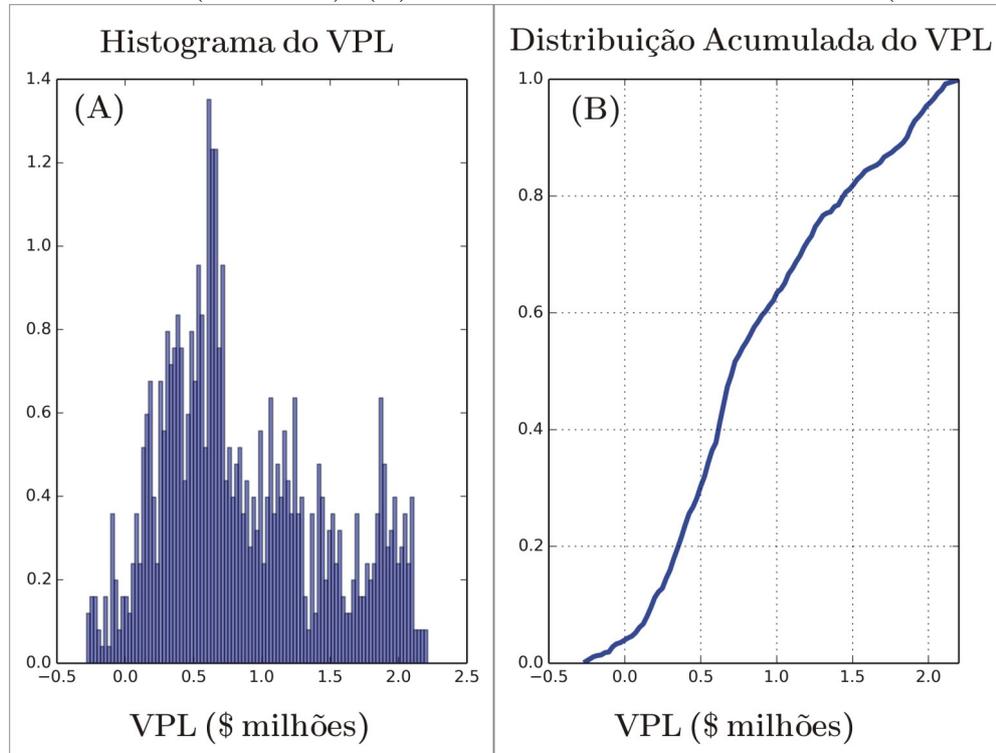


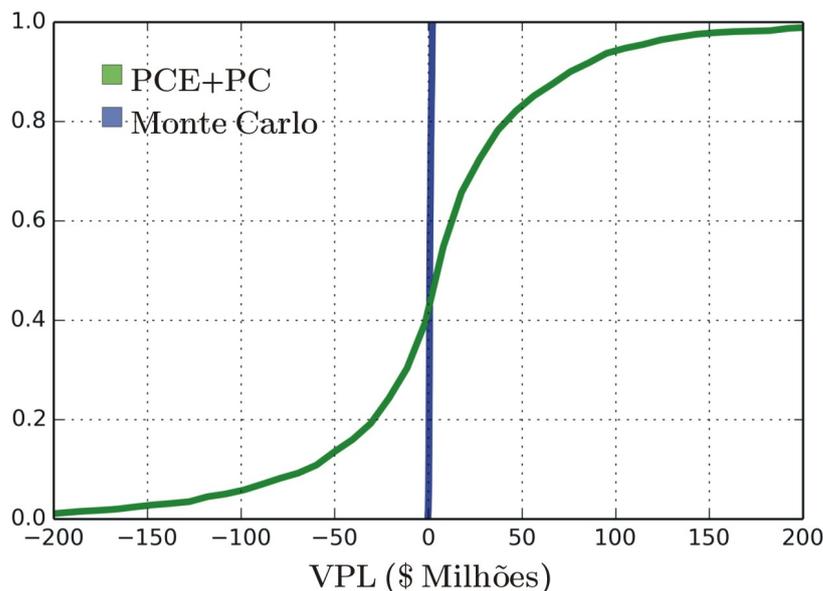
Figura 4.23: Curva de risco do VPL a partir das 1000 realizações geoestatísticas: (A) Histograma do VPL (\$ milhões); (B) Distribuição Acumulada do VPL (\$ milhões).



As 1000 realizações geoestatísticas foram simuladas e os resultados estão apresentados na Figura 4.23. Na Figura 4.24 é apresentada a comparação entre os resultados oriundos das realizações originais e o resultado a partir do método *PCE+PCM*.

Assim como no Caso 1, a média do *VPL* obtido pelos dois métodos é bastante semelhante mas o mesmo não ocorre com a variância. Pode ser observado na Figura 4.24 que a curva de distribuição gerada por *PCE+PCM* apresenta uma longa cauda em contraposição à curva gerada por Monte Carlo, que possui uma variância pequena e aparece como uma linha reta em função da escala adotada na figura. Este fenômeno ocorre pelo fato de que a curva de distribuição, no caso *PCE+PCM*, ser gerada através de amostragens com base em funções polinomiais e não diretamente baseadas nas equações que regem o modelo. Assim certas restrições físicas podem ser violadas durante processo de amostragem. Além disso, é sabido que a expansão *PCE* construída com *PCM* captura preferencialmente as informações contidas nas regiões de alta probabilidade de ocorrência dos parâmetros de entrada. As regiões de baixa probabilidade (caudas da distribuição normal), responsáveis pela variabilidade da variável, deixam de ser bem representadas e daí vem o desvio na variância do *VPL*. Outro fator que

Figura 4.24: Comparação do VPL (\$ milhões) entre PCE+PCM e Monte Carlo.  
Distribuição Acumulada do VPL



também contribui para esse resultado é o fato de se utilizar uma parametrização do campo de entrada com dimensionalidade reduzida o que, conforme já foi mencionado, provoca um efeito de suavização no mapa da propriedade, provocada pela perda das estruturas de curta correlação.

### 4.3.3 Parametrização do VPL com NISP

O método *NISP* foi também utilizado para estimar os coeficientes da expansão *PCE*. Neste trabalho, a sequência de Sobol foi então utilizada para gerar 150 amostras das variáveis aleatórias que representam o campo randômico. A Figura 4.25 abaixo apresenta a disposição dos pontos para duas das variáveis aleatórias usadas para parametrizar a permeabilidade.

Com base nas 150 amostras geradas acima e, utilizando-se a formulação descrita no Capítulo 3, foram obtidos os coeficientes da expansão. Foi gerada, a princípio, uma expansão *PCE* de 5ª ordem. A título de ilustração, a Figura 4.26 apresenta os coeficientes calculados.

Após o processo de validação cruzada utilizado para decidir manter ou descartar os termos de alta ordem da expansão, neste caso, restaram apenas os termos lineares, conforme pode ser visto na Figura 4.27 a seguir. Na mesma figura apresenta-se o gráfico

Figura 4.25: Exemplo de amostra usando sequência de Sobol - Caso 2.

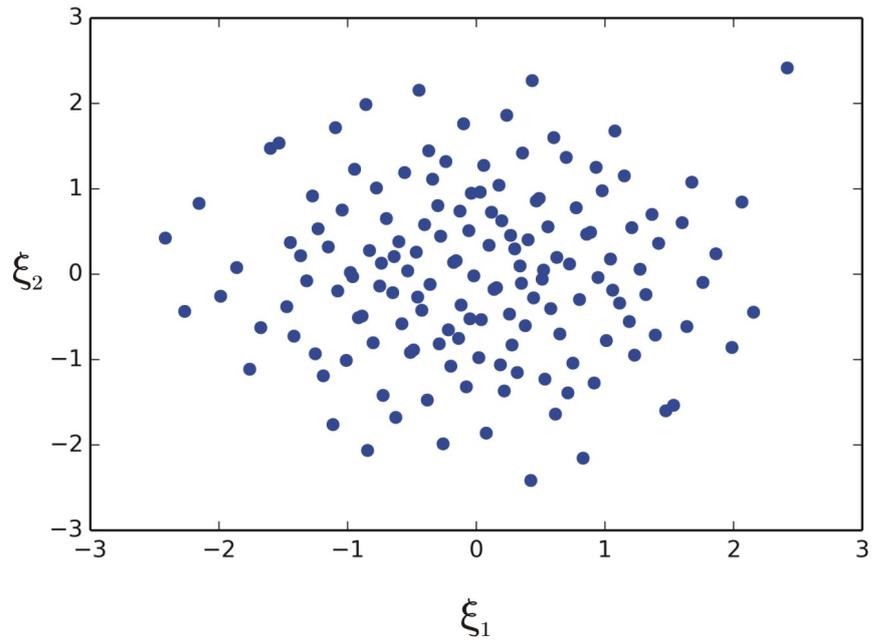
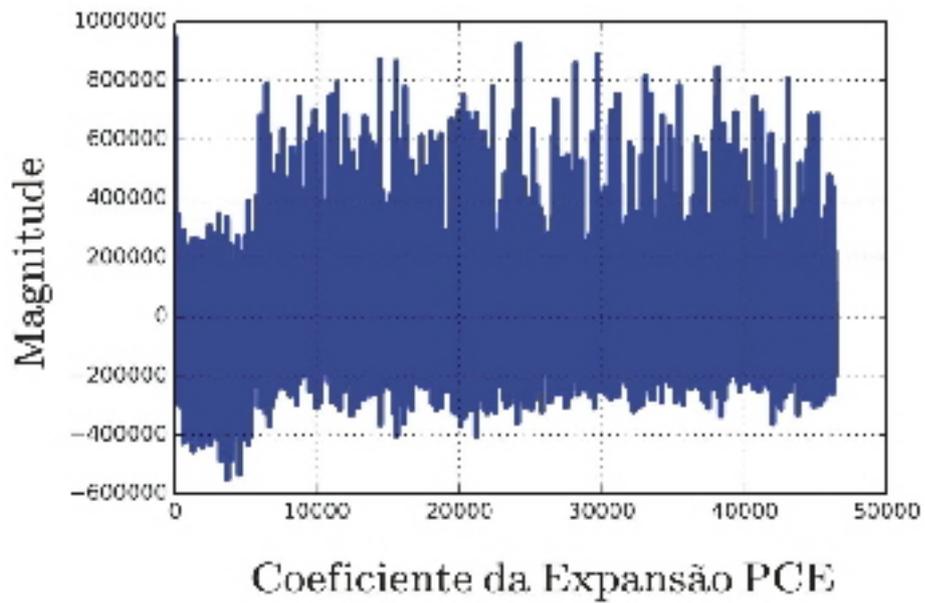
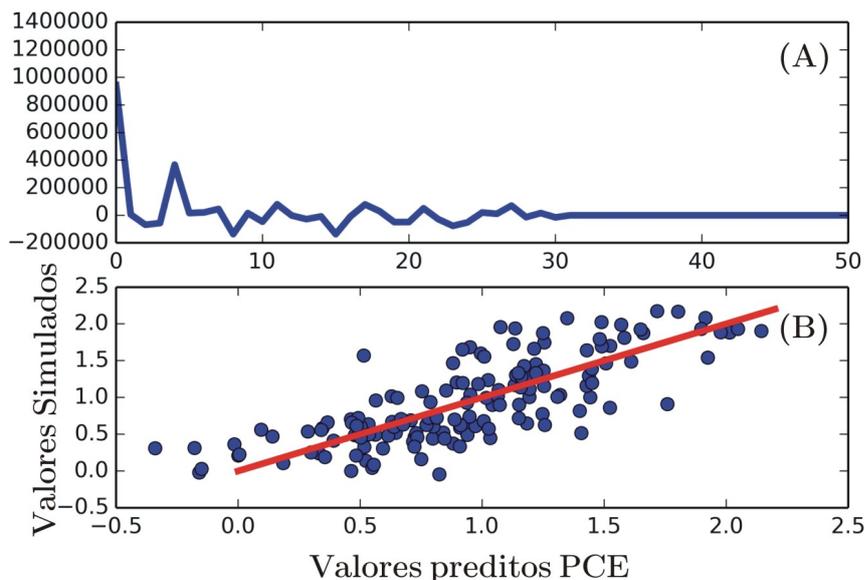


Figura 4.26: Coeficientes da expansão PCE obtidos por NISP.



de qualidade da aproximação onde são confrontados os valores preditos pelo polinômio e os valores simulados.

Figura 4.27: Caso 2 - resultados do método NISP: (A) coeficientes da expansão após validação cruzada; (B) comparação entre valores preditos pela expansão versus valores simulados.



Após o processo de validação a expansão *PCE* fica totalmente definida. A fim de avaliar as estatísticas de interesse para o *VPL*, foram geradas 2000 amostras aleatórias e calculado o *VPL* através do polinômio *PCE*. Os resultados obtidos estão apresentados na Figura 4.28.

A comparação dos resultados obtidos pelo *NISP* e por 1000 amostras geradas por geoestatística que, conforme vimos anteriormente, podem ser usadas para aproximar a distribuição verdadeira está apresentada na Figura 4.29. Pode ser observado, comparando-se a Figura 4.24 com a Figura 4.29 que a distribuição do *VPL* obtida pelo método *NISP* ficou muito mais próxima da distribuição resultante das 1000 realizações geoestatísticas. A Tabela 4.2 (valores em \$ milhões) apresenta a comparação das principais métricas obtidas pelos dois casos, tendo como referência os valores obtidos a partir das realizações geoestatísticas simuladas pelo método de Monte Carlo.

Os resultados apresentados na Tabela 4.2, demonstram que, para este caso, o método *PCE+PCM* superestimou os riscos associados ao projeto em estudo. Os prováveis motivos que levaram a esse resultado já foram expostos nos comentários dos resultados do Caso 1. Na prática da indústria, um resultado como o obtido pelo

Figura 4.28: Caso 2 - curva de risco do VPL (\$ milhões) obtida via NISP: (A) histograma;(B) curva de distribuição acumulada.

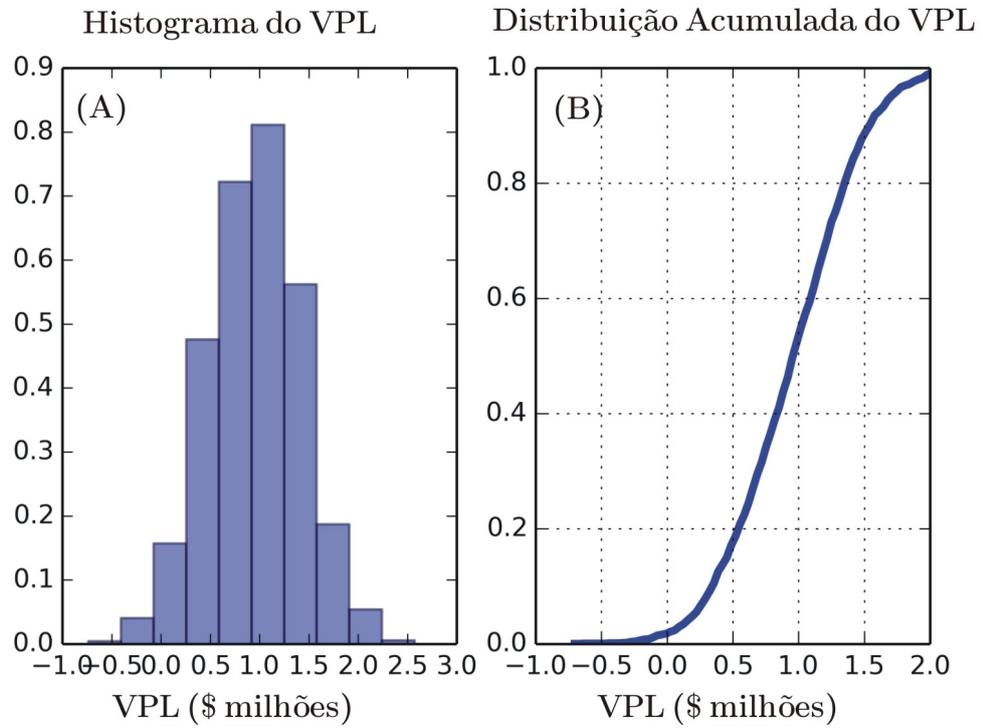


Figura 4.29: Caso 2 - comparação da curva de risco para o VPL (\$ milhões) obtidas por NISP e pelas realizações geoestatísticas.

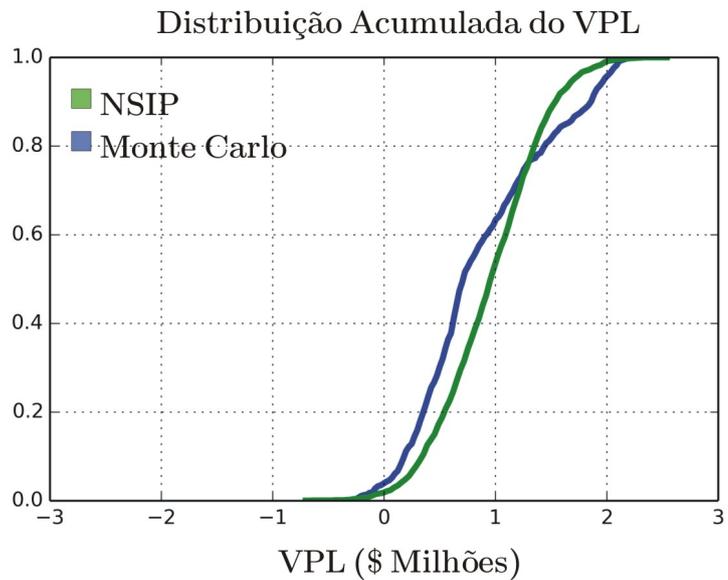


Tabela 4.2: Resultados da Análise de Risco para o Caso 2

Estatística	Monte Carlo	<i>PCE+PCM</i>	<i>PCE+NISP</i>
Média	0.87	1.02	0.99
Variância	$7.7 \times 10^{-4}$	11.5	$7.91 \times 10^{-4}$
P10	0.12	-68.76	0.36
P90	1.92	84.22	1.48

método *PCE+PCM* inviabilizaria a aprovação do projeto por decisores mais avessos a riscos.

# Capítulo 5

## Conclusões e Trabalhos Futuros

### 5.1 Conclusões

No contexto de problemas de quantificação de incertezas utilizando um modelo substituto (*proxy*), foram apresentadas e testadas metodologias para a parametrização, tanto das variáveis de entradas, aqui tratadas como campos randômicos, quanto da função objetivo que no caso dos problemas elaborados foi o *VPL*. Apesar de não ter sido o foco deste trabalho, os métodos apresentados possuem um bom potencial de utilização para processos de otimização e/ou ajuste de histórico que, reconhecidamente, demandam múltiplas avaliações de uma ou mesmo várias funções objetivo.

No que diz respeito a representação de campos randômicos gaussianos, a expansão de *Karhunen-Loeve* se mostrou eficiente ao conseguir reproduzir as realizações geoestatísticas originais com poucas variáveis aleatórias não correlacionadas. Porém, o número de variáveis aleatórias utilizadas na expansão é dependente do comprimento de correlação adotado, modelado pelo alcance do variograma nos algoritmos de geoestatística. Pelo fato de utilizar uma parametrização do campo de entrada com dimensionalidade reduzida ocorre um efeito de suavização nas novas realizações da propriedade, provocada pela perda das estruturas de curta correlação.

Para campos não gaussianos, a formulação *KPCA* se apresenta como uma evolução em relação a expansão de *Karhunen-Loeve* (*PCA* linear) mas se o parâmetro de entrada for originalmente modelado por meio de uma distribuição discreta, como é caso das fácies no modelo geológico, a utilização do método produzirá realizações com distribuição contínua que são geologicamente inaceitáveis por não respeitar os contornos de cada feição idealizada. Outra desvantagem observada diz respeito ao problema da pré-imagem que se mostrou bastante ineficiente. O método pluri-*PCA* se mostrou su-

perior ao *KPCA* em relação à representação de propriedades discretas uma vez que utiliza técnicas de truncamento para transformar resultados reais em valores inteiros. Porém, na versão implementada neste trabalho, o método admite que o parâmetro incerto possui distribuição com apenas um pico e pode ser modelada por uma distribuição normal. No caso estudado, o erro médio de atribuição de fácies, para novas realizações geradas pelo método ficou em torno de 5%, o que pode ser considerado insatisfatório pois, no processo de modelagem geológica, a modelagem de fácies é importante porque ela condiciona a modelagem das propriedades petrofísicas.

A parametrização da função objetivo foi realizada por meio da expansão em caos polinomial (*PCE*) usando dois métodos distintos para determinação de seus coeficientes, o método de colocação probabilística (*PCM*) e o método espectral não intrusivo (*NISP*). A curva de distribuição do *VPL*, gerada por *PCE+PCM*, apresentou, em todos os casos, longas caudas, ao contrário do apresentado pelo método de Monte Carlo. Este fenômeno ocorre pelo fato de que a curva de distribuição, no caso *PCE+PCM*, é gerada através de amostragens com base em funções polinomiais e não diretamente baseadas nas equações que regem o modelo. Assim, certas restrições físicas podem ser violadas durante processo de amostragem. Além disso, é sabido que a expansão *PCE* construída com *PCM* captura preferencialmente as informações contidas nas regiões de alta probabilidade de ocorrência dos parâmetros de entrada. As regiões de baixa probabilidade (caudas da distribuição normal), responsáveis pela variabilidade da variável, deixam de ser bem representadas e daí vem o desvio na variância do *VPL*. Outro fator que também contribui para esse resultado é o fato de se estar utilizando uma parametrização do campo de entrada com dimensionalidade reduzida o que, conforme já foi mencionado, provoca um efeito de suavização no mapa da propriedade, provocada pela perda das estruturas de curta correlação.

A distribuição do *VPL* obtida pelo método *NISP* ficou mais próxima da distribuição dita verdadeira, obtida através de simulações de Monte Carlo com algoritmos de geostatística. Adicionalmente, o método *NISP* levou vantagem também no número de simulações necessárias (150 contra 496 do *PCM*). Porém, o fato de descartar os termos de alta ordem e os termos cruzados da expansão pode ser uma limitação importante, principalmente nos casos em que o modelo substituto assim gerado for utilizado para processos de otimização e ajuste de histórico que são problemas sabidamente altamente não-lineares. Este certamente é um fator que merece ser investigado no futuro.

Os métodos que apresentaram melhores resultados e, portanto, constituem a sugestão deste trabalho para a solução eficiente de problemas desta classe, são o método

pluri-PCA para parametrização dos campos randômicos de entrada e o método espectral não intrusivo para obtenção dos coeficientes da expansão do caos polinomial para a representação das respostas de interesse.

## 5.2 Trabalhos Futuros

A linha de pesquisa seguida neste trabalho é bastante ativa e potencialmente muito promissora no sentido de produzir resultados concreto e de ordem prática para a indústria do petróleo. Algumas etapas precisam de um estudo mais aprofundado e os algoritmos implementados precisam ser testados em casos mais complexos e, principalmente mais próximos da realidade da indústria. A seguir, encontram-se listadas algumas sugestões para desenvolvimentos futuros visando dar prosseguimento à linha de pesquisa.

- As metodologias apresentadas precisam ser testadas em problemas de geometria mais complexa para que seja avaliada a robustez da formulação. Além disso, os próximos testes devem ser feitos considerando o processo de modelagem geológica mais próximo do aplicado na prática no qual a simulação dos parâmetros petrofísicos é condicionada à simulação das fácies, que por sua vez pode utilizar uma medida indireta (dado sísmico) como tendência para distribuição dos valores no modelo. Há evidências na literatura que o método pluri-PCA apresenta bons resultados em modelagens com mais de duas fácies onde não haja um ordenamento sequencial entre elas.
- Existem outras técnicas para parametrização dos campos randômicos de entrada que merecem ser investigados como alternativas aos métodos aqui empregados, dentre elas podemos citar o *Independent Component Analysis (ICA)*, o método de *escalonamento multidimensional (MDS)*, etc. Mesmo o *KPCA*, aqui implementado conforme sugerido por SARMA, 2006 [42], apresenta derivações em sua formulação (tipo do kernel, método de obtenção da pré-imagem, etc.) que merecem ser mais aprofundadas.
- A formulação do método NISP mostrou bom potencial para o problema da análise de incertezas mas sua performance deve ser também avaliada em problemas com um grau de não-linearidade maior como os problemas de otimização e ajuste de histórico sob incertezas. Neste contexto, uma possível abordagem seria a utilização das técnicas apresentadas neste trabalho para gerar um modelo substituto

e em seguida acoplar com algum método como o *Markov Chain Monte Carlo* (*MCMC*) para amostrar a *fdp* a posteriori dos parâmetros de entrada.

# Bibliografia

- [1] CHEN, C.; GAO, G.; RAMIREZ, B. A ; VINK, J. C ; GIRARDI, A. M. **Assisted History Matching of Channelized Models Using Pluri-Principal Component Analysis**. SPE Reservoir Simulation Symposium, Society of Petroleum Engineers, 2015.
- [2] CHEN, C.; GAO, G.; HONORIO, J.; GELDERBLOM, P.; JIMENEZ, E.; JAAKKOLA, T. **Integration of principal-component-analysis and streamline information for the history matching of channelized reservoirs**. SPE Annual Technical Conference and Exhibition, Society of Petroleum Engineers, 2014.
- [3] COSTA, A. P. A. **Quantificação do impacto de incertezas e análise de risco no desenvolvimento de campos de petróleo**. Tese de Doutorado, Campinas: Faculdade de Engenharia Mecânica e Instituto de Geociências, Universidade Estadual de Campinas, 2003.
- [4] DAMSLETH, E.; HAGE, A.; VOLDEN, R. **Maximum information at minimum cost**. 1992.
- [5] DEJEAN, J. P.; BLANC, G. **Managing uncertainties on production predictions using integrated statistical methods**. SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers, 1999.
- [6] DONGXIAO, Z.; TCHELEPI, H. **Stochastic analysis of immiscible two-phase flow in heterogeneous media**. SPE journal 4.04 (1999): 380-388.
- [7] ELDRED, M. S.; WEBSTER, C. G.; CONSTANTINE, P. **Evaluation of non-intrusive approaches for Wiener-Askey generalized polynomial chaos**. Proceedings of the 10th AIAA Non-Deterministic Approaches Conference, number AIAA-2008-1892, Schaumburg, IL. Vol. 117. 2008.

- [8] FIELD, R. V.; GRIGORIU, M. **Convergence properties of polynomial chaos approximations for L2 random variables**. Public report, Sandia National Laboratories, Albuquerque, 2007.
- [9] FONSECA, L. A. **Otimização para simulação estocástica de reservatórios de petróleo: Algoritmos de diferenças randômicas simultâneas**. Dissertação de Mestrado, Recife: UFPE, 2010.
- [10] FONSECA, L. A.; GUIMARÃES, A. M. O.; OLIVEIRA, D. F. B.; ARAÚJO, E. R.; WILMERSDORF, R. B. **Otimização sob Incertezas de Locação de Poços com Alocação Dinâmica das Vazões de Produção e Injeção**. Rio Oil & Gas Expo and Conference 2010, Rio de Janeiro/RJ, Brasil, IBP - Instituto Brasileiro de Petróleo, Gás e Biocombustíveis, 2010.
- [11] GHANEM, R. G.; SPANOS, P. D. **Spectral stochastic finite-element formulation for reliability analysis**. Journal of Engineering Mechanics 117.10, 2351-2372. 1991.
- [12] GHANEM, R. G.; SPANOS, P. D. **Stochastic finite elements: a spectral approach**. 2003.
- [13] GHANEM, R. G.; DHAM, S. **Stochastic finite element analysis for multi-phase flow in heterogeneous porous media**. Transport in Porous Media 32.3, 239-262. 1998.
- [14] GUIMARÃES, A. M. O.; ARAÚJO, E. R.; **Método Eficiente para a Solução do Problema de Otimização das Vazões de Injeção e Produção Considerando Incertezas nas Propriedades Estáticas do Reservatório**, Seminário de Reservas e Reservatórios 2009, Rio de Janeiro - RJ, Brasil, 22 a 25 de setembro, Petrobras, 2009.
- [15] HUANG, S. P.; QUEK, S. T.; PHOON, K. K. **Convergence study of the truncated Karhunen-Loeve expansion for simulation of stochastic processes**. International journal for numerical methods in engineering 52.9 (2001): 1029-1043.
- [16] HUANG, S. P.; MAHADEVAN, S.; REBBA, R. **Collocation-based stochastic finite element analysis for random field problems**. Probabilistic Engineering Mechanics 22.2 (2007): 194-205.

- [17] ISUKAPALLI, S. S. **Uncertainty analysis of transport-transformation models**. PhD Thesis, The State University of New Jersey, USA, 1999.
- [18] JARDAK, M.; SU, C-H.; KARNIADAKIS, G. E. **Spectral polynomial chaos solutions of the stochastic advection equation**. *Journal of Scientific Computing* 17.1-4 (2002): 319-338.
- [19] KHANINEZHAD, M. M.; JAFARPOUR, B.; LI, L. **Sparse geologic dictionaries for subsurface flow model calibration: Part I. Inversion formulation**. *Advances in Water Resources* 39 (2012): 106-121.
- [20] KHANINEZHAD, M. M.; JAFARPOUR, B.; LI, L. **Sparse geologic dictionaries for subsurface flow model calibration: Part II. Robustness to uncertainty**. *Advances in Water Resources* 39 (2012): 122-136.
- [21] KWOK, J. T-Y.; TSANG, I. W-H. **The pre-image problem in kernel methods**. *Neural Networks, IEEE Transactions on* 15.6 (2004): 1517-1525.
- [22] LI, H.; ZHANG, D. **Probabilistic collocation method for flow in porous media: Comparisons with other stochastic methods**. *Water Resources Research* 43.9, 2007.
- [23] LI, H.; CHANG, H.; ZHANG, D. **Stochastic collocation methods for efficient and accurate quantification of uncertainty in multiphase reservoir simulations**. *SPE reservoir simulation symposium*. Society of Petroleum Engineers, 2009.
- [24] LOÈVE, M. **Probability theory**. 1977.
- [25] LOSCHIAVO, R. **Estimativa de Incertezas na Previsão de Comportamento de Reservatórios**. *Dissertação de Mestrado*, Campinas: Faculdade de Engenharia Mecânica e Instituto de Geociências, Universidade Estadual de Campinas, 1999.
- [26] LU, Z.; ZHANG, D. **A comparative study on uncertainty quantification for flow in randomly heterogeneous media using Monte Carlo simulations and conventional and KL-based moment-equation approaches**. *SIAM Journal on Scientific Computing* 26.2 (2004): 558-577.

- [27] LU, Z.; ZHANG, D. **Accurate, efficient quantification of uncertainty for flow in heterogeneous reservoirs using the KLME approach**. SPE Reservoir Simulation Symposium. Society of Petroleum Engineers, 2005.
- [28] LUCOR, D.; XIU, D.; SU, C-H.; KARNIADAKIS, G. E. **Predictability and uncertainty in CFD**. International Journal for Numerical Methods in Fluids 43.5 (2003): 483-506.
- [29] MANCEAU, E.; MEZGHANI, M.; MEZGHANI, I. Z.; ROGGERO, F. **Combination of experimental design and joint modeling methods for quantifying the risk associated with deterministic and stochastic uncertainties-An integrated test study**. SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers, 2001.
- [30] MATHELIN, L.; HUSSAINI, M. Y.; ZANG, T. A. **Stochastic approaches to uncertainty quantification in CFD simulations**. Numerical Algorithms 38.1-3 (2005): 209-236.
- [31] MIKA, S.; SCHOLKOPF, B.; SMOLA, A. J.; MULLER, K. R.; SCHOLZ, M.; RATSCH, G. **Kernel PCA and De-Noising in Feature Spaces**. NIPS. Vol. 4. No. 5. 1998.
- [32] NIEDERREITER, H. **Low-discrepancy and low-dispersion sequences**. Journal of number theory 30.1 (1988): 51-70.
- [33] OGAWA, H. **Karhunen-Loeve subspace**. Pattern Recognition, 1992. Vol. II. Conference B: Pattern Recognition Methodology and Systems, Proceedings., 11th IAPR International Conference on. IEEE, 1992.
- [34] OLIVEIRA, D. F. B. **Técnicas de Otimização da Produção para Reservatórios de Petróleo: Abordagens Sem Uso de Derivadas para Alocação Dinâmica das Vazões de Produção e Injeção**. Dissertação de Mestrado, Universidade Federal de Pernambuco-UFPE, Recife, PE , 2006.
- [35] OVREBERG, O.; DAMSLETH, E.; HALDORSEN, H. H. **Putting Error Bars on Reservoir Engineering Forecast**. SPE 20512. "Annual Technical Conference and Exhibition. New Orleans, Louisiana. 1990.
- [36] PORTELLA, R. C. M.; SALOMÃO, M. C.; BLAUTH, M.; DUARTE, R. L. B. **Uncertainty quantification to evaluate the value of information in a**

- deepwater reservoir**. SPE Reservoir Simulation Symposium. Society of Petroleum Engineers, 2003.
- [37] REAGANA, M. T.; NAJM, H. N.; GHANEM, R. G.; KNIO, O. M. **Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection**. Combustion and Flame 132.3 (2003): 545-555.
- [38] REMY, N.; BOUCHER, A.; WU, J. **Applied geostatistics with SGeMS: A user's guide**. Cambridge University Press, 2009.
- [39] SANTOS, J. P. M.; SCHIOZER, D. J. **Determinação de metodologia de ajuste automatizado de histórico**. Rio Oil & Gas Expo and Conference. 2000.
- [40] SARMA, P.; XIE, J. **Efficient and robust uncertainty quantification in reservoir simulation with polynomial chaos expansions and non-intrusive spectral projection**. SPE Reservoir Simulation Symposium. Society of Petroleum Engineers, 2011.
- [41] SARMA, P.; DURLOFSKY, L. J.; AZIZ, K. **Efficient Closed-Loop Production Optimization under Uncertainty**. SPE94241 67th EAGE Conference & Exhibition. 2005.
- [42] SARMA, P. **Efficient closed-loop optimal control of petroleum reservoirs under uncertainty**. PhD Thesis, Stanford University, USA, 2006.
- [43] SCHOLKOPF, B.; SMOLA, A. J. **Learning with kernels**. 2002.
- [44] SCHOLKOPF, B.; MIKA, S.; SMOLA, A. J.; RATSCH, G.; MULLER, K. R. **Kernel PCA pattern reconstruction via approximate pre-images**. ICANN 98 (1998): 147-152.
- [45] STEAGALL, D. E. **Análise de risco nas previsões de produção com simulação numérica de fluxo—Exemplo de um campo na fase de delimitação**. Dissertação de Mestrado. Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas, 2001.
- [46] STREBELLE, S. B.; JOURNAL, A. G. **Reservoir modeling using multiple-point statistics**. SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers, 2001.

- [47] STRIPE, J. A.; ARISAKA, K.; DURANDEAU, M. **Integrated field development planning using risk and decision analysis to minimise the impact of reservoir and other uncertainties: a case study.** Middle East Oil Show (1993).
- [48] TATANG, M. A.; PAN, W.; PRINN, R. G.; MCRAE, G. J. **An efficient method for parametric uncertainty analysis of numerical geophysical models.** Journal of Geophysical Research: Atmospheres 102.D18 (1997): 21925-21932.
- [49] VENKATARAMAN, R. **Application of the method of experimental design to quantify uncertainty in production profiles.** SPE Asia Pacific Conference on Integrated Modelling for Asset Management. Society of Petroleum Engineers, 2000.
- [50] VO, H. X.; DURLOFSKY, L. J. **A new differentiable parameterization based on principal component analysis for the low-dimensional representation of complex geological models.** Mathematical Geosciences 46.7 (2014): 775-813.
- [51] WEBSTER, M. D.; TATANG, M. A.; MCRAE, G. J. **Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model.** (1996).
- [52] WIENER, N. **The homogeneous chaos.** American Journal of Mathematics 60.4 (1938): 897-936.
- [53] XIU, D.; KARNIADAKIS, G. E. **Modeling uncertainty in flow simulations via generalized polynomial chaos.** Journal of computational physics 187.1 (2003): 137-167.
- [54] ZHANG, D.; LU, Z. **An efficient, high-order perturbation approach for flow in random porous media via Karhunen–Loeve and polynomial expansions.** Journal of Computational Physics 194.2 (2004): 773-794.
- [55] ZHANG, D.; LI, L.; TCHELEPI, H. A. **Stochastic formulation for uncertainty analysis of two-phase flow in heterogeneous reservoirs.** SPE Journal 5.01 (2000): 60-70.
- [56] ZHANG, D. **Stochastic methods for flow in porous media: coping with uncertainties.** (2001).

- [57] ZHAO, Y.; REYNOLDS, A. C.; LI, G. **Generating facies maps by assimilating production data and seismic data with the ensemble Kalman filter.** SPE Symposium on Improved Oil Recovery. Society of Petroleum Engineers, 2008.
- [58] ZHOU, H.; GOMEZ-HERNANDEZ, J. J.; FRANSSEN, H. J. H.; LI, L. **An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering.** Advances in Water Resources 34.7 (2011): 844-864.
- [59] ZHOU, H.; LI, L.; FRANSSEN, H. J. H.; GOMEZ-HERNANDEZ, J. J. **Pattern recognition in a bimodal aquifer using the normal-score ensemble Kalman filter.** Mathematical Geosciences 44.2 (2012): 169-185.
- [60] ZHOU, H.; LI, L.; GOMEZ-HERNANDEZ, J. J. **Characterizing curvilinear features using the localized normal-score ensemble Kalman filter.** Abstract and Applied Analysis. Vol. 2012. Hindawi Publishing Corporation, 2012.