



Alejandra Camacho Solano

**Quantificação de Incertezas na Simulação de
Reservatórios de Petróleo via Programação
Genética e Caos Polinomial**

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós-Graduação
em Engenharia Elétrica da PUC-Rio como requisito parcial
para obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Prof. Marco Aurélio Cavalcanti Pacheco

Rio de Janeiro
Agosto de 2015



Alejandra Camacho Solano

Quantificação de Incertezas na Simulação de Reservatórios de Petróleo via Programação Genética e Caos Polinomial

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Marco Aurélio Cavalcanti Pacheco

Orientador

Departamento de Engenharia Elétrica – PUC-Rio

Dr. Alexandre Anozé Emerick

Petróleo Brasileiro – Rio de Janeiro - Matriz

Prof. Carlos Kubrusly

Departamento de Engenharia Elétrica – PUC-Rio

Dr. Douglas Mota Dias

Departamento de Engenharia Elétrica – PUC-Rio

Prof. José Eugenio Leal

Coordenador Setorial do Centro

Técnico Científico

Rio de Janeiro, 17 de agosto de 2015

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, da autora e do orientador.

Alejandra Camacho Solano

Graduou-se em Ciências Atuariais pela Universidade de Costa Rica (UCR) em 2010.

Ficha Catalográfica

Solano, Alejandra Camacho

Quantificação de incertezas na simulação de reservatórios de petróleo via programação genética e caos polinomial / Alejandra Camacho Solano; orientador: Marco Aurélio Cavalcanti Pacheco. – 2015.

132 f. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2015.

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. Qualificação de incertezas. 3. Simulação de reservatórios. 4. Caos polinomial. 5. Quadratura de Smolyak. 6. Regressão. 7. Programação genética. I. Pacheco, Marco Aurélio Cavalcanti. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

Agradecimentos

À Deus por tudo, porque sem Ele nada teria acontecido.

Ao meu orientador Professor Marco Aurélio, pela confiança e pelas importantes contribuições. Estarei sempre agradecida pela ajuda constante, tanto acadêmica como pessoal.

Ao meu esposo, pelo seu apoio incondicional, sua paciência e seu grande amor que foi a minha força para continuar.

À minha mãe, pela educação, pelo seu amor e porque a distância nunca foi um obstáculo para estar presente em cada momento.

Ao meu pai e minha família Maynard Guillén, pelo estímulo e constante apoio.

Ao Emerick, ao Alvaro e ao João, pela disposição em me ajudar sempre neste trabalho.

Aos Professores William, Alexander e Roldán da Universidade de Costa Rica, pela motivação e pelo apoio, sem os quais não estaria aqui.

À todos os meus amigos e familiares da Costa Rica e do Brasil que de uma ou de outra maneira me estimularam e me ajudaram. Especialmente Karen, Adriana e Rafael.

Aos professores que participaram da Comissão examinadora.

Ao CAPES, ao CNPq, à PUC-RIO e à Petrobrás, pelos auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado.

Resumo

Solano, Alejandra Camacho; Pacheco, Marco Aurélio Cavalcanti (Orientador). **Quantificação de Incertezas na Simulação de Reservatórios de Petróleo via Programação Genética e Caos Polinomial**. Rio de Janeiro, 2015. 132p. Dissertação de Mestrado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Os modelos de simulação de reservatórios estão sujeitos à incerteza presente em uma grande variedade de seus parâmetros de entrada. Esta incerteza é o resultado da heterogeneidade das formações geológicas, erros nas medições dos dados e da modelagem petrofísica, estrutural e do transporte dos fluidos no meio poroso. Uma quantificação precisa da incerteza requer, na maioria dos casos, uma quantidade elevada de simulações, o que é usualmente inviável se considerarmos o tempo consumido para simular modelos de grande escala. Por outro lado, uma avaliação adequada da incerteza aumenta a qualidade e robustez das decisões tomadas para o gerenciamento dos campos de petróleo. Com esta motivação, foi investigado o método das Expansões por Caos Polinomial (PCE, por suas siglas em inglês). PCE é uma técnica de convergência rápida utilizada para analisar como se propaga, na saída de um modelo, a incerteza presente nos parâmetros de entrada. Mediante PCE, pode-se representar a resposta aleatória de um modelo de simulação de reservatórios de petróleo como um polinômio, construído a partir de uma base de funções que dependem da distribuição de probabilidade das variáveis incertas de entrada. Por outro lado, quando a relação entre os parâmetros de entrada e a saída do modelo têm um componente não polinomial, o algoritmo de Programação Genética (PG) pode ser utilizado para representar esta dependência utilizando funções ou operadores mais complexos. PG é um algoritmo de regressão simbólica capaz de encontrar uma expressão aleatória explícita, que aproxime a saída de um modelo de simulação de reservatórios de petróleo, conhecendo-se a priori a distribuição de probabilidade dos parâmetros de entrada. Neste trabalho foram aplicadas as duas técnicas, antes mencionadas, num modelo de simulação de reservatórios baseado no campo PUNQ-S3, considerando até vinte e três parâmetros incertos durante um período de produção de 13 anos. Foi feita uma análise de incerteza, calculando-se a distribuição de probabilidade completa da saída do simulador. Os resultados foram comparados com o método de Monte Carlo, indicando um alto desempenho em termos de custo computacional e acurácia. Ambas as técnicas conseguem níveis de ajuste superiores a 80% com uma quantidade de simulações consideravelmente baixa.

Palavras-chave

Quantificação de Incertezas; Simulação de Reservatórios; Caos Polinomial; Quadratura de Smolyak; Regressão; Programação Genética.

Abstract

Solano, Alejandra Camacho; Pacheco, Marco Aurélio Cavalcanti (Orientador). **Uncertainty Quantification in Oil Reservoir Simulation via Genetic Programming and Chaos Polynomial**. Rio de Janeiro, 2015. 132p. MSc Dissertation - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Reservoir simulation models are subject to uncertainty in a wide variety of its inputs. This uncertainty is a result of the heterogeneity of the geological formations, data measurement errors, and petrophysical, structural, and fluid transport in porous media modelling. An accurate uncertainty quantification requires, in most cases, a large number of simulations, which is unviable considering the time it takes to simulate large scale models. On the other hand, a proper uncertainty assessment, increases the robustness of the decision making process for the oil field management. To this end, the method of Polynomial Chaos Expansions (PCE) was studied. PCE is a fast paced convergence technique, used to analyze the uncertainty propagation of the input parameters all the way to the output of the model. Through PCE is possible to represent the response of an oil reservoir simulation model as a polynomial, built from a function basis, that depend on the probability distribution of the uncertain input variables. Furthermore, when the relationship between the input and output parameters of the model has a non-polynomial component, the algorithm of Genetic Programming (GP) can be used to represent this dependency by more elaborate functions or operators. GP is a symbolic regression algorithm, capable of finding an explicit expression that approximates the output of a reservoir simulation model, with prior knowledge of the probability distribution of the input parameters. In this work, the two previously mentioned techniques were applied in a reservoir simulation model, based on the oil field PUNQ-S3, considering up to twenty three uncertain parameters during a simulation period of 13 years. An uncertainty analysis of the output of the simulator was conducted, calculating the entire probability distribution. The results were compared to the Monte Carlo simulation method, presenting a satisfactory performance in terms of accuracy and computational cost. Both techniques show adjustment levels higher than 80%, with a considerable small amount simulations.

Keywords

Uncertainty Quantification; Reservoir Simulation; Chaos Polynomial; Smolyak Quadrature; Regression; Genetic Programming.

Sumário

1	Introdução	14
1.1.	Motivações	14
1.2.	Objetivos	17
1.3.	Organização do Trabalho	18
2	Fundamentos Matemáticos	20
2.1.	Introdução	20
2.2.	Teoria Básica de Probabilidade	20
2.3.	Polinômios Ortogonais	24
2.4.	Teoria de Produto Tensor	30
3	Quantificação de Incertezas em Simulação de Reservatórios	34
3.1.	Introdução	34
3.2.	Simuladores de Reservatórios de Petróleo	35
3.3.	Quantificação de Incertezas	36
3.3.1.	Parâmetros incertos	37
3.3.2.	Técnicas de Quantificação de Incertezas	41
3.3.3.	Maldição da Dimensionalidade	43
4	Métodos de Integração Numérica	45
4.1.	Introdução	45
4.2.	Quadraturas Gaussianas Unidimensionais	46
4.2.1.	Quadratura de Gauss-Hermite	48
4.2.2.	Quadratura de Gauss-Laguerre	49
4.2.3.	Quadratura de Gauss-Legendre	49
4.3.	Quadraturas Genz-Keister	51
4.4.	Quadraturas Multivariadas	53
4.4.1.	Quadratura Produto Tensor	53
4.4.2.	Quadratura de Smolyak	54
5	Expansões por Caos Polinomial	59
5.1.	Introdução	59
5.2.	Expansões por Caos Polinomial Generalizado	60
5.3.	Aproximação por Projeção Espectral Ortogonal	64
5.4.	Aproximação por Regressão	67
5.5.	Análise Estatística e Probabilística	70
5.6.	Tratamento de variáveis com múltiplas distribuições	71
5.7.	Quantificação do Erro de Aproximação	72
6	Programação Genética	74
6.1.	Introdução	74
6.2.	Programação Genética	75
6.2.1.	Representação	75
6.2.2.	Avaliação	77
6.2.3.	Operadores	77
6.3.	Programação Genética Multi-Gene	79

6.4. Formulação para o Problema de Quantificação de Incerteza	81
7 Estudo de Casos	84
7.1. Introdução	84
7.2. Experimento Numérico	84
7.3. Modelos de Simulação de Reservatórios de Petróleo	87
7.3.1. Descrição do campo	88
7.3.2. Estudo de Caso 1	91
7.3.3. Estudo de Caso 2	101
7.3.4. Estudo de Caso 3	112
8 Conclusões e Trabalhos Futuros	121
8.1. Conclusões e Contribuições	121
8.2. Trabalhos Futuros	123
Referências Bibliográficas	124
Apêndice	130

Lista de Figuras

Figura 2.1: Distribuição de probabilidade Normal com parâmetros μ e σ^2	22
Figura 2.2: Distribuição de probabilidade Uniforme no intervalo a, b	22
Figura 2.3: Esquema Askey para polinômios ortogonais hipergeométricos, tomado de (XIU, 2010).....	26
Figura 2.4: Primeiros cinco polinômios de Hermite probabilísticos.	27
Figura 2.5: Primeiros cinco polinômios de Laguerre.....	29
Figura 2.6: Primeiros cinco polinômios de Legendre.....	30
Figura 4.1: Gride completo (esquerda) e o gride de Smolyak (direita), baseados em regras de quadratura trapezoidal de 1, 2 e 7 pontos, tomando de (HOLTZ, 2010).....	57
Figura 5.1: Esquema de propagação de incerteza.....	59
Figura 6.1: Representação em forma de árvore de um indivíduo de PG.	76
Figura 6.2: Exemplo de Cruzamento. Os indivíduos da esquerda são cópias dos pais, portanto o material genético pode ser utilizado sem realizar alterações nos indivíduos originais.	78
Figura 6.3: Modelo PG Multi-Gene.....	80
Figura 6.4: PGMG no processo de quantificação de incertezas.	82
Figura 7.1: Porosidade do campo PUNQ-S3, na data 1967-01-01.....	90
Figura 7.2: Variações na produção acumulada de óleo, gás e água, quando varia o multiplicador da porosidade no intervalo $[0.85, 1.15]$	92
Figura 7.3: Variações na produção acumulada de óleo, gás e água, quando varia o multiplicador da permeabilidade horizontal no intervalo $[0.01, 10]$	92
Figura 7.4: Variações na produção acumulada de óleo, gás e água, quando varia o multiplicador da permeabilidade vertical no intervalo $[0.01, 10]$	92
Figura 7.5: Solução simbólica de PG para a produção acumulada de óleo.	97
Figura 7.6: Função distribuição de probabilidade empírica da produção acumulada de óleo.....	98
Figura 7.7: Função distribuição de probabilidade empírica da produção acumulada de gás.	98

Figura 7.8: Função distribuição de probabilidade empírica da produção acumulada de água.	99
Figura 7.9: Coeficientes de determinação e número de simulações para cada aproximação, modelo de três parâmetros.	99
Figura 7.10: Variações na produção acumulada de óleo, gás e água, quando varia a compressibilidade da rocha no intervalo [0.000003, 0.000006].....	102
Figura 7.11: Variações na produção acumulada de óleo, gás e água, quando varia o raio do aquífero no intervalo [1, 6000]......	103
Figura 7.12: Variações na produção acumulada de óleo, gás e água, quando varia o expoente de Corey da permeabilidade relativa ao gás, no intervalo [1.5, 4.5]......	103
Figura 7.13: Variações na produção acumulada de óleo, gás e água, quando varia o valor máximo da permeabilidade relativa ao gás, no intervalo [0.05, 0.35]......	103
Figura 7.14: Variações na produção acumulada de óleo, gás e água, quando varia o expoente de Corey da permeabilidade relativa à água no intervalo [2.5, 5.5]......	103
Figura 7.15: Variações na produção acumulada de óleo, gás e água, quando varia o valor máximo da permeabilidade relativa à água no intervalo [0.3, 0.9]......	104
Figura 7.16: Variações na produção acumulada de óleo, gás e água, quando varia a profundidade de contato óleo-água no intervalo [2392, 2398].....	104
Figura 7.17: Variações na produção acumulada de óleo, gás e água, quando varia a profundidade de contato óleo-gás no intervalo [2352, 2358]......	104
Figura 7.18: Efeito dos parâmetros incertos nas saídas do modelo.	105
Figura 7.19: Função distribuição de probabilidade empírica da produção acumulada de óleo.....	109
Figura 7.20: Função distribuição de probabilidade empírica da produção acumulada de gás.	109
Figura 7.21: Função distribuição de probabilidade empírica da produção acumulada de água.	110
Figura 7.22: Coeficientes de determinação e número de simulações para cada aproximação, modelo de onze parâmetros.	110

Figura 7.23: Função distribuição de probabilidade empírica da produção acumulada de óleo.....	117
Figura 7.24: Função distribuição de probabilidade empírica da produção acumulada de gás.	117
Figura 7.25: Função distribuição de probabilidade empírica da produção acumulada de água.	117
Figura 7.26: Coeficientes de determinação e número de simulações para cada aproximação, modelo de vinte e três parâmetros.....	118

Lista de Tabelas

Tabela 4.1: Nodos e pesos da Quadratura Genz-Keister.....	52
Tabela 5.1: Correspondência entre o tipo de distribuição e a base de polinômios.	61
Tabela 5.2: Bijeção entre os índices β e k da soma truncada na equação (5.6). ...	62
Tabela 5.3: Número de coeficientes na expansão por caos polinomial.	63
Tabela 5.4: Correspondência entre o número de avaliações e o nível para as regras GH, SGH e SGK com dimensão 7.....	66
Tabela 5.5: Correspondência entre o número de avaliações e o nível para as regras GH, SGH e SGK com dimensão 15.....	66
Tabela 7.1: Aproximação da integral da eq. (7.1).	86
Tabela 7.2: Aproximação da integral da eq. (7.3)	87
Tabela 7.3: Coeficientes de determinação para cada solução obtida mediante expansões por caos polinomial com três dimensões.	94
Tabela 7.4: Valores dos parâmetros e coeficiente de determinação para cada programa genético no caso de três dimensões.	96
Tabela 7.5: Média e variâncias para cada saída do modelo de 3 parâmetros incertos.	100
Tabela 7.6: Variação das estimações da média e a variância no modelo de 3 parâmetros incertos.	101
Tabela 7.7: Parâmetros da modelagem probabilística das variáveis incertas no modelo.	102
Tabela 7.8: Coeficientes de determinação para cada solução obtida mediante expansões por caos polinomial com onze dimensões.	106
Tabela 7.9: Valores dos parâmetros e coeficiente de determinação para cada programa genético no caso de onze dimensões.	108
Tabela 7.10: Média e variâncias das saídas do modelo de 11 parâmetros incertos.	111
Tabela 7.11: Variação das estimações da média e da variância.	112
Tabela 7.12: Dependência entre o número de simulações e os parâmetros da PCE com dimensão 23, utilizando o enfoque da projeção.	113

Tabela 7.13: Relação entre o número de coeficientes da PCE de dimensão 23 com o grau selecionado.	114
Tabela 7.14: Coeficientes de determinação para a solução obtida mediante expansões por caos polinomial com 23 dimensões.....	114
Tabela 7.15: Valores dos parâmetros e coeficiente de determinação para cada programa genético no caso de vinte e três dimensões.....	116
Tabela 7.16: Média e variâncias das saídas do modelo de 23 parâmetros incertos.	119
Tabela 7.17: Variação das estimações da média e da variância.	119

1 Introdução

1.1. Motivações

A quantificação de incertezas é um elemento fundamental no desenvolvimento e no gerenciamento dos campos de petróleo. A qualidade das decisões tomadas depende da análise correta da incerteza na saída do modelo. A variedade, a quantidade e a incerteza dos parâmetros de um modelo de simulação de reservatórios faz com que tais modelos sejam notoriamente complexos. Esta incerteza é o resultado da heterogeneidade das formações geológicas, deficiências nas medições dos dados e a modelagem petrofísico e estrutural. Como resultado, a resposta do modelo, tal como a produção acumulada, pode apresentar um nível de incerteza importante que deve ser analisado.

A necessidade de determinar uma resposta robusta frente à estocasticidade presente nestes modelos fizeram com que o estudo do tratamento de incertezas ganhasse bastante destaque na literatura atual de petróleo, (WEBSTER, et al., 1996), (SARMA, et al., 2006), (SARMA, et al., 2005) ou (HENG LI, et al., 2011).

Uma das técnicas para quantificar a incerteza é a Expansão por Caos Polinomial (PCE) introduzida por Wiener (WIENER, 1938) no início do século XX. Apesar disso, a aplicação desta ferramenta somente veio cerca de 60 anos depois com o trabalho pioneiro de Ghanem e Spanos (GHANEM, et al., 1991), onde os autores apresentam o uso da PCE na quantificação de incerteza aplicado à engenharia mecânica. Posteriormente, Xiu em (XIU, 2010) estende o trabalho do Wiener, para o uso de diferentes tipos de polinômios ortogonais, oferecendo uma maneira eficiente de representar processos não gaussianos. Esta nova metodologia é aplicada por Xiu a equações diferenciais estocásticas de baixa dimensão, com resultados satisfatórios em termos de custo computacional e precisão, se comparado ao método de Monte Carlo tradicional. O destaque da PCE se dá principalmente por sua rápida convergência e por expressar a solução final como um processo aleatório, e não meramente como um conjunto de estatísticas.

Trabalhos aplicados na área de petróleo, podem-se ver em (SARMA, et al., 2011), onde é aplicada a técnica PCE com projeção espectral e o método de colocação probabilística, para quantificar a incerteza presente na simulação de reservatórios. Os autores mostram em dois casos de estudo de baixa dimensão, que mediante PCE é possível estimar a função distribuição de probabilidade da saída do modelo, isto, de maneira eficiente em termos de exatidão e custo computacional. Barzagan em (BAZARGAN, et al., 2012) utiliza em duas estruturas geológicas complexas, a expansão por caos polinomial como um *proxy* ou modelo substituto do simulador de reservatórios, para resolver o problema de ajuste de histórico eficientemente. Também, em (BAZARGAN, et al., 2013), Christie, Bazargan e Tchelepi mostram em um modelo de canal de areia sintético, que a técnica de PCE com um algoritmo de Monte Carlo e cadeias de Markov consegue aproximar, com precisão e em poucas rodadas do simulador de reservatórios, a distribuição a posterior das zonas de alta permeabilidade.

Por outro lado, quando a modelagem faz necessário aumentar o número de variáveis propagadas, também aumenta significativamente o custo computacional, este problema é conhecido como *Curse of Dimensionality* (KUO, et al., 2005). Devido a que a maior parte dos sistemas físicos ou mecânicos reais têm envolvidos uma grande quantidade de parâmetros, este problema de dimensionalidade em caos polinomial é muito investigado na atualidade (HAGAN, 2013), (SUDRET, et al., 2013) e (PRANGA, et al., 2012), com resultados promissores. Na área de petróleo é de grande interesse a solução deste problema.

Para otimizar um projeto de desenvolvimento de um campo de petróleo é necessário determinar o efeito dos parâmetros incertos na resposta do modelo de simulação. Devido a que são muitas as variáveis necessárias para descrever o modelo, busca-se encontrar técnicas de aproximação que permitam obter os coeficientes da expansão em um tempo computacional viável. Mediante a técnica de projeção espectral, metodologia que será abordada neste trabalho, os coeficientes da expansão são definidos mediante integrais multidimensionais, portanto, é preciso adotar técnicas de integração numérica para estimar estes valores. Kaarnioja em (KAARNIOJA, 2013) desenvolve fundamentos teóricos e práticos no uso da quadratura de Smolyak ou *sparse grid* para estimar integrais de alta dimensão. Esta quadratura é construída a partir de produtos tensores de diferenças

consecutivas, permitindo que o número de pontos de integração seja menor ao produto clássico completo. Em (HOLTZ, 2010) também se mostra a eficiência de este método em problemas de alta dimensão, observando que o efeito no custo computacional é mais pronunciado conforme aumenta o número de parâmetros.

Outra alternativa possível para calcular os coeficientes da expansão é adotar um enfoque de regressão. Neste caso, é preciso construir um planejamento de experimentos a partir de um conjunto de realizações. Este planejamento pode ser gerado de maneira aleatória, com técnicas amostrais como o Hipercubo Latino ou pesquisas mais recentes mostraram que é possível construir um planejamento eficiente, com valores que apresentem uma alta probabilidade de incidência e portanto, representem melhor o espaço de busca. Os autores Sudret, Blatman e Berveiller mostram em (BLATMAN, et al., 2008), (SUDRET, 2008) e (SUDRET, et al., 2013) que para problemas mecânicos e de análises de sensibilidade de alta dimensão, esta metodologia de regressão com um planejamento de experimentos otimizado é eficiente, em termos de precisão e número de rodadas, para o cálculo dos coeficientes da PCE.

Outra alternativa para quantificar incerteza, em contraparte com as expansões por caos polinomial, é o algoritmo de Programação Genética (PG), que pertence à família de técnicas de inteligência computacional. Este método permite obter a partir de um conjunto de dados, uma expressão explícita que represente a relação entre as entradas e a saída de um modelo. Este algoritmo de regressão simbólica, se vem destacando cada vez mais em aplicações, onde se tenha pouca informação e seja preciso uma expressão matemática explícita para descrever um problema (YU, et al., 2008) (BURKE, et al., 2009).

PG é uma ferramenta de simulação e otimização que tenta obter a melhor predição com o mínimo erro entre os dados observados e os valores previstos. Na área de engenharia de petróleo, (KAYDANI, et al., 2014) (FALLAH-MEHDIPOUR, et al., 2012) e (YU, et al., 2008) aplicaram PG na estimação da permeabilidade, em problemas de operação em tempo real, na construção de proxies para ajuste de histórico e análises de incertezas na previsão da produção, respectivamente. Em todos os casos se mostraram resultados satisfatórios ao ser comparados com outras ferramentas.

Adicionalmente, o algoritmo de programação genética permite representar a saída de um modelo mediante funções ou operadores complexos, como por

exemplo, uma combinação de logaritmos, exponenciais, somas, raízes, mínimos, máximos, operadores lógicos, etc. De esta maneira é possível capturar com maior precisão, um comportamento não polinomial nos dados.

No contexto de quantificação de incertezas em modelos de simulação de reservatórios, das maiores vantagens de aplicar os algoritmos de PCE e PG é obter uma função explícita, polinomial ou não, que pode ser manipulada e utilizada como um *proxy* sob incerteza da saída do modelo. Uma vez construída esta expressão, é possível fazer análises estatísticas e probabilísticas da resposta do simulador, sem necessidade de simular novamente o campo de petróleo, diminuindo consideravelmente o custo computacional de realizar estes estudos.

Finalmente, considerando as metodologias antes mencionadas, será tratado o problema da quantificação de incertezas em modelos de simulação de reservatórios de alta dimensão, utilizando expansões por caos polinomial e programação genética. No caso da PCE, é proposto nesta dissertação, estimar os coeficientes da expansão com a quadratura de Smolyak e o método de regressão, com um planejamento de experimentos eficiente e otimizado. Vai-se trabalhar com um modelo de simulação de reservatórios que tem por saída a produção acumulada de óleo, água e gás. Utilizando PCE e PG se quantificará a incerteza presente nestas saídas por meio de uma estimação, precisa e com baixo custo computacional, da função distribuição de probabilidade inteira.

1.2. Objetivos

O objetivo geral deste trabalho é desenvolver e validar o desempenho do algoritmo de programação genética e da técnica de caos polinomial, para o estudo do problema da quantificação de incertezas, dentro da abrangência de reservatórios de petróleo.

Os objetivos específicos são:

- Estudo e implementação da técnica de projeção espectral ortogonal e o enfoque de regressão, para aproximar os coeficientes da expansão com um erro de estimação aceitável e um custo computacional viável.

- Estudo e implementação das quadraturas Gaussianas, Genz-Keister e Smolyak para o processo de integração numérica, necessário no enfoque da projeção espectral ortogonal.
- Estudo e implementação da técnica de programação genética para a estimação de saídas em modelos estocásticos, com a finalidade de quantificar incerteza.
- Aproximação da saída do modelo de simulação de reservatórios de petróleo, por meio de uma função analítica ou um programa de computador. Isto é, gerar um *proxy* sob incerteza.
- Estimação estatística da resposta do modelo de simulação de reservatórios de petróleo. Cálculo da função distribuição de probabilidade completa.

1.3. Organização do Trabalho

Esta dissertação contém mais seis capítulos, cujos conteúdos estão descritos a seguir.

No Capítulo 2, são apresentados os fundamentos teóricos matemáticos, necessários para o desenvolvimento dos restantes capítulos deste trabalho. Tais fundamentos são: teoria de probabilidade, polinômios ortogonais e produtos tensores. Esta base teórica é essencial para: a modelagem estocástica do problema, a definição das regras de quadratura, utilizadas como métodos de integração numérica e finalmente para a construção da PCE.

No Capítulo 3 é desenvolvido o principal problema nesta dissertação, que é a quantificação de incertezas em modelos de simulação de reservatórios de petróleo. Faz-se uma descrição dos modelos de simulação numérica, as variáveis aleatórias que podem ser envolvidas nestes modelos e finalmente se enunciam quais técnicas serão abordadas para resolver o problema em questão.

No Capítulo 4 é apresentada uma descrição dos métodos de integração numérica considerados neste trabalho, para resolver o problema da aproximação

dos coeficientes das PCE. Para o caso de alta dimensionalidade é proposta a quadratura de Smolyak, que permite diminuir consideravelmente o custo computacional de obter a estimação.

No Capítulo 5 se expõe a técnica das expansões por caos polinomial. São consideradas duas aproximações para obter os coeficientes da expansão, mediante a projeção espectral e por meio do enfoque da regressão. Para qualquer dos dois casos, o resultado pode ser utilizado para estimar estatísticas da saída do modelo considerado.

No Capítulo 6 é definido o algoritmo de programação genética e sua variante, o algoritmo de programação genética multi-gene. São descritos os dois métodos e é apresentada a modelagem do problema de quantificação de incertezas, utilizando esta metodologia.

O Capítulo 7, desenvolve o estudo de casos realizado e a descrição dos modelos de simulação de reservatórios considerados. Inicialmente são expostas as técnicas das quadraturas, com o fim de valorar sua acurácia e custo computacional. Seguidamente, são apresentados os resultados com um modelo baseado no campo PUNQ-S3 (IMPERIAL COLLEGE LONDON).

Finalmente, no Capítulo 8 são apresentadas as conclusões, contribuições realizadas e os possíveis direcionamentos para trabalhos futuros.

2 Fundamentos Matemáticos

2.1. Introdução

Neste capítulo se enunciaram alguns conceitos básicos que serviram como base, para o desenvolvimento desta pesquisa. Na primeira seção vão ser descritas as propriedades básicas da teoria de probabilidade. A justificativa desta seção reside no fato de que, se vai trabalhar com funções distribuição de probabilidade de variáveis aleatórias, que conformam um sistema físico. A segunda seção fornece uma introdução aos fundamentos e as propriedades da teoria de polinômios ortogonais, que é o foco principal das expansões por caos polinomial e a construção das regras de quadraturas. Por último, na seção final são descritas as características principais da teoria dos produtos tensores, a qual é necessária para a construção do *grid sparse*, que vai ser utilizado na aproximação dos coeficientes da PCE. Esta técnica de integração numérica permite aproximar integrais multidimensionais, com uma quantidade de avaliações consideravelmente menor, se comparado com as quadraturas tradicionais.

2.2. Teoria Básica de Probabilidade

O resultado de um experimento, um jogo ou um evento é comumente aleatório. Para o tratamento matemático de estes tipos de experimentos aleatórios é necessário assignar um número para cada resultado, este valor é chamado de *variável aleatória* ou *parâmetro incerto* e o espaço de resultados é denotado por Ω . Por outro lado, o conceito de *probabilidade* é utilizado para medir a veracidade da ocorrência destes experimentos ou eventos. Por exemplo, no caso de lançamentos de moedas, é assignado um valor baseado em observações ou na evidencia empírica: se lançamos uma moeda justa uma grande quantidade de vezes, esperamos que próximo ao 50% dos resultados sejam cara e próximo ao 50% sejam coroa. Em teoria de probabilidade, a *lei dos grandes números* nos

oferece a justificação teórica para esta observação empírica. Vamos definir alguns conceitos básicos da teoria de probabilidade que são de importância para o trabalho feito nos seguintes capítulos. As definições foram recopiladas de (XIU, 2010) e (ROSS, 2012).

Definição 2.1: (função distribuição) A coleção de probabilidades

$$F_X(x) = P(X \leq x) = P(\{\omega: X(\omega) \leq x\}), \quad x \in \mathbb{R} \quad (2.1)$$

É a *função distribuição de probabilidade* F_X de X . Denotada por *cdf* pelas siglas em inglês (*cumulative distribution function*).

Para variáveis aleatórias contínuas, a função distribuição pode ser definida mediante a *função densidade de probabilidade* f_X :

$$F_X(x) = \int_{-\infty}^x f_X(y) dy, \quad x \in \mathbb{R} \quad (2.2)$$

Onde,

$$f_X(x) \geq 0, \quad \forall x \in \mathbb{R} \quad e \quad \int_{-\infty}^{\infty} f_X(y) dy = 1 \quad (2.3)$$

Denotada por *pdf* pelas siglas em inglês (*probability distribution function*).

Um conceito associado à *cdf* de uma variável aleatória é percentil que definimos a continuação.

Definição 2.2: (Percentil) O α – *percentil* de \mathbf{X} é o nível \mathbf{x}_α tal que a probabilidade de que \mathbf{X} tome um valor superior que \mathbf{x}_α é α . Assim,

$$x_\alpha = \inf\{x; F_X(x) \geq \alpha\} \quad (2.4)$$

Onde $F_X(x)$ é a distribuição de probabilidade de X

Alguns exemplos de funções de distribuição são:

- *Distribuição Normal ou Gaussiana com parâmetros* $\mu \in \mathbb{R}$ e $\sigma^2 > 0$: $\mathcal{N}(\mu, \sigma^2)$

Sua função de densidade é,

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R} \quad (2.5)$$

Na seguinte Figura 2.1 pode-se observar a *cdf*, comumente denotada por Φ , para distintos valores de parâmetros.

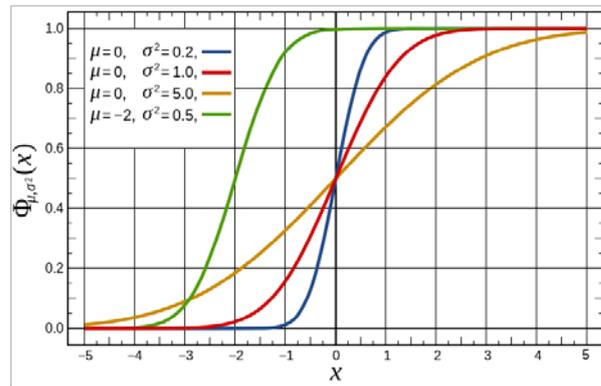


Figura 2.1: Distribuição de probabilidade Normal com parâmetros μ e σ^2 .

- *Distribuição Uniforme com constantes no intervalo (a, b) :*

$$\mathcal{U}(a, b)$$

Sua função de densidade é,

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b) \\ 0, & \text{outro caso} \end{cases} \quad (2.6)$$

Na seguinte Figura 2.2 pode-se observar sua cdf para o intervalo (a, b) .

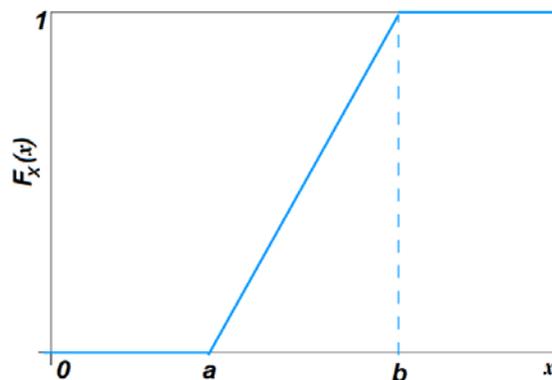


Figura 2.2: Distribuição de probabilidade Uniforme no intervalo (a, b) .

Propriedades importantes das variáveis aleatórias são o *valor esperado ou esperança*, a *variância* e os *momentos*. Estes valores permitem-nos ter um conhecimento do comportamento do parâmetro.

Para uma variável aleatória X com função de densidade f_X , podemos definir estes conceitos como,

- Esperança de X :

$$\mu_X = \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx \quad (2.7)$$

- Variância de X :

$$\sigma_X^2 = \text{var}(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx \quad (2.8)$$

- m -ésimo momento ($m \in \mathbb{N}$) de X :

$$\mathbb{E}[X^m] = \int_{-\infty}^{\infty} x^m f_X(x) dx \quad (2.9)$$

Outro conceito relevante na área de probabilidade é a independência de variáveis aleatórias. Intuitivamente, dois variáveis são independentes se o resultado de uma no influi no resultado da outra, mais precisamente,

Definição 2.3: (Independência) As variáveis aleatórias X_1, X_2, \dots, X_n são independentes, se e somente se, suas funções de distribuição conjuntas podem-se escrever como,

$$F_{X_1, X_2, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n), \quad (2.10)$$

$$(x_1, \dots, x_n) \in \mathbb{R}^n$$

Se segue imediatamente que, se X_1, X_2, \dots, X_n são independentes, então para quaisquer funções de valor real g_1, \dots, g_n ,

$$\mathbb{E}[g_1(X_1), \dots, g_n(X_n)] = \mathbb{E}[g_1(X_1)], \dots, \mathbb{E}[g_n(X_n)] \quad (2.11)$$

Finalizamos esta seção, com o celerado teorema do limite central (CLT) que tem um papel fundamental em diversos aspectos da teoria de probabilidade.

Teorema 2.1: (CLT) Sejam X_1, X_2, \dots, X_n variáveis aleatórias independentes identicamente distribuídas (i.i.d.), com $\mathbb{E}[X_i] = \mu$ e $\text{var}(X_i) = \sigma^2 < \infty$. Seja além disso,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (2.12)$$

E também,

$$U_n = \sqrt{n} \left(\frac{\bar{X} - \mu}{\sigma} \right) \quad (2.13)$$

Então a função distribuição de U_n converge para uma função distribuição $\mathcal{N}(0,1)$ quando $n \rightarrow \infty$.

2.3. Polinômios Ortogonais

Nesta seção é fornecida uma introdução rápida da teoria de polinômios ortogonais, a qual tem um papel importante na teoria de integração e aproximação que vamos desenvolver nos capítulos seguintes, no caso de ter mais interesse pode-se consultar (SZEGO, 1939). Primeiramente se vai definir as relações de ortogonalidade, seguido das series hipergeométricas, o esquema Askey e finalizamos com alguns exemplos de polinômios ortogonais que têm relação com o trabalho realizado. As seguintes definições são baseadas em (XIU, 2010).

Em geral podemos expressar um polinômio de grau n como:

$$Q_n(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0, \quad a_n \neq 0 \quad (2.14)$$

Onde a_n é chamado *coeficiente principal* do polinômio e denotamos por

$$P_n(x) = \frac{Q_n(x)}{a_n} = x^n + \frac{a_{n-1}}{a_n} x^{n-1} + \dots + \frac{a_1}{a_n} x + \frac{a_0}{a_n} \quad (2.15)$$

a versão mônica do polinômio, é dizer, aquele com o coeficiente principal igual a 1.

Um sistema de polinômios $\{Q_n, n \in \mathbb{N}\}$ é chamado um sistema ortogonal de polinômios com respeito a alguma medida real positiva θ , se é mantida a seguinte relação de ortogonalidade:

$$\int_S Q_n(x) Q_m(x) d\theta(x) = \gamma_n \delta_{mn}, \quad m, n \in \mathbb{N} \quad (2.16)$$

Onde $\delta_{mn} = 0$ se $m \neq n$ e $\delta_{mn} = 1$ se $m = n$, é a função delta de Kronecker, S é o suporte da medida θ e γ_n são valores reais denominados *constantes de normalização*. Se $\gamma_n = 1$ dizemos que o sistema é ortonormal.

Por outro lado se definimos o produto interno,

$$(u, v)_{d\theta} = \int_S u(x)v(x)d\theta(x) \quad (2.17)$$

e a medida θ tem uma densidade $\omega(x)$, então, a relação de ortogonalidade pode ser expressada como,

$$(Q_m, Q_n)_\omega = \gamma_n \delta_{mn} \quad m, n \in \mathbb{N} \quad (2.18)$$

A densidade $\omega(x)$, é comumente referida como *função de peso* na teoria de polinômios ortogonais. Vamos ver mais adiante que as funções de peso de alguns polinômios ortogonais são idênticas a funções de probabilidade. Por exemplo, a

função de peso para os polinômios de Hermite probabilísticos é a mesma que a função de probabilidade de uma variável aleatória Gaussiana. Este evento, tem um papel importante na representação de processos estocásticos com polinômios ortogonais.

Trabalhos mais recentes mostram que uma classe importante de polinômios ortogonais pertence a um esquema específico dos polinômios hipergeométricos (ASKEY, et al., 1985). Por isto, vamos fazer uma breve revisão da teoria de polinômios ortogonais hipergeométricos.

Primeiro introduzimos o símbolo Pochhammer $(a)_n$ como

$$(a)_n = \begin{cases} 1 & \text{se } n = 0 \\ a(a+1) \dots (a+n-1) & \text{se } n = 1, 2, 3, \dots \end{cases} \quad (2.19)$$

Então assim, definimos a *serie hipergeométrica generalizada* ${}_rF_s$ como

$${}_rF_s(a_1, \dots, a_r; b_1, \dots, b_s; z) = \sum_{k=0}^{\infty} \frac{(a_1)_k \dots (a_r)_k z^k}{(b_1)_k \dots (b_s)_k k!} \quad (2.20)$$

Onde $b_i \neq 0, -1, -2, \dots$ para $i = \{1, \dots, s\}$ nos garanta que os fatores do denominador das series nunca sejam zero. Se algum parâmetro do numerador a_i é negativo, digamos $a_1 = -n$, a serie hipergeométrica da eq. (2.20) termina no n -ésimo termo e torna-se um polinômio em z .

Por outro lado, o raio de convergência ρ da série hipergeométrica (2.20) é,

$$\rho = \begin{cases} \infty & \text{se } r < s + 1 \\ 1 & \text{se } r = s + 1 \\ 0 & \text{se } r > s + 1 \end{cases} \quad (2.21)$$

Alguns casos básicos das series hipergeométricas são: serie exponencial ${}_0F_0$ e a serie binomial ${}_1F_0$.

Os polinômios ortogonais hipergeométricos podem ser hierarquizados mediante a metodologia proposta por Richard Askey em (ASKEY, et al., 1985). Este esquema de organização que leva por nome Askey, classifica polinômios ortogonais hipergeométricos contínuos e discretos que obedecem certas diferenças e equações diferentes. Esta estrutura tem forma de árvore e é mostrada na Figura 2.3 obtida na referência (XIU, 2010). Pode-se observar que inicia com os polinômios de Wilson e os polinômios de Racah na parte superior, ambos pertencem ao tipo de serie hipergeométrica ${}_4F_3$. Os primeiros são contínuos por enquanto os segundos são discretos. As linhas que conectam os diferentes polinômios representam as relações de transição limite entre eles, é dizer, os

polinômios no extremo inferior podem ser obtidos considerando o limite de um dos parâmetros dos polinômios no nível superior.

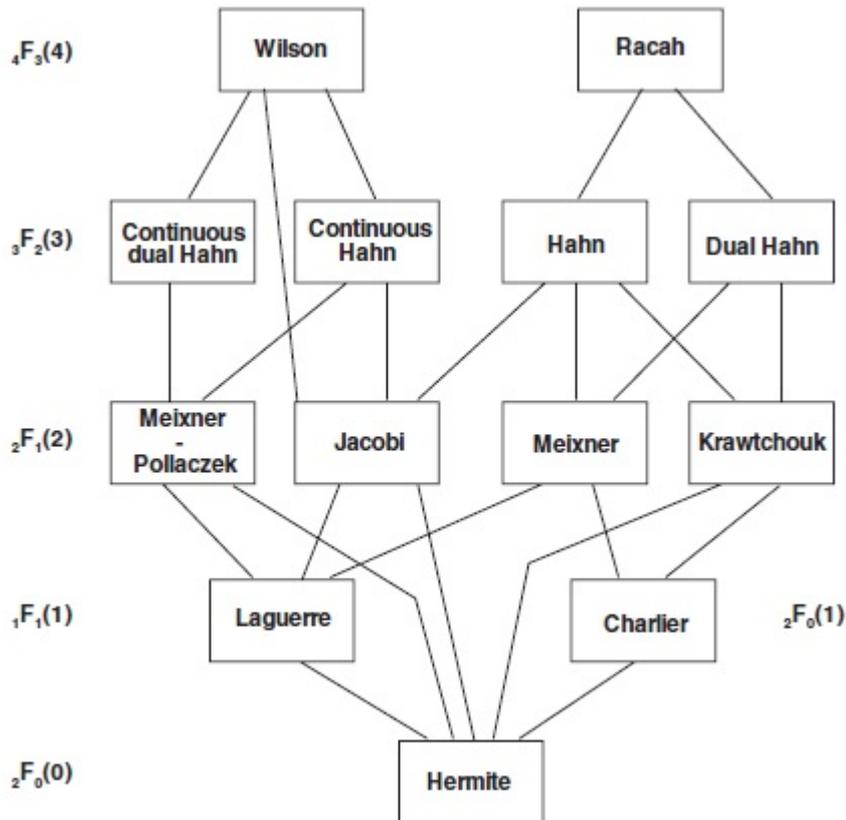


Figura 2.3: Esquema Askey para polinômios ortogonais hipergeométricos, tomado de (XIU, 2010).

Encerramos esta seção com alguns exemplos de polinômios ortogonais do esquema Askey, associados com caos polinomial generalizado, também chamado caos-Askey. Devido a que nosso interesse está nos polinômios contínuos, vamos mencionar algumas características importantes dos polinômios de Hermite definidos sobre \mathbb{R} , Laguerre definidos sobre o intervalo $[0, +\infty)$ e Jacobi sobre $[-1, 1]$, (SÜLI, et al., 2003) e (PEIXOTO, 2008).

Os polinômios de Hermite probabilísticos denotados por $H_n(x)$ estão definidos mediante series hipergeométricas como,

$$H_n(x) = (\sqrt{2}x)^n {}_2F_0\left(-\frac{n}{2}, -\frac{n-1}{2}; ; -\frac{2}{x^2}\right) \quad (2.22)$$

A relação de ortogonalidade está dada por,

$$\int_{-\infty}^{\infty} H_m(x)H_n(x)\omega(x)dx = n! \delta_{mn} \quad (2.23)$$

Onde a função de peso $\omega(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ satisfaz a relação de ortogonalidade anterior, note-se que é a mesma que a função densidade de probabilidade das variáveis aleatórias Gaussianas com média zero e desvio padrão um. Os primeiros cinco polinômios de Hermite probabilísticos são,

- $H_0(x) = 1$
- $H_1(x) = x$
- $H_2(x) = x^2 - 1$
- $H_3(x) = x^3 - 3x$
- $H_4(x) = x^4 - 6x^2 + 3$

Na seguinte Figura 2.4 mostram-se os gráficos dos cinco primeiros polinômios de Hermite probabilísticos.

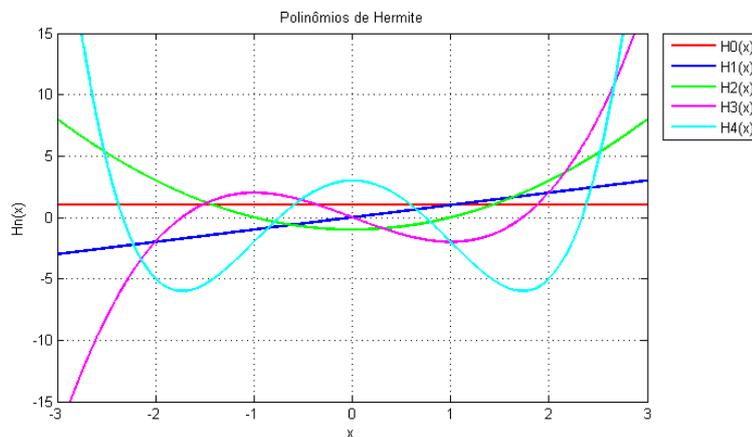


Figura 2.4: Primeiros cinco polinômios de Hermite probabilísticos.

Por outro lado, existem também polinômios de Hermite físicos, que vamos denotar por $\widehat{H}_n(x)$. Estes são ortogonais com respeito à função de peso $\omega(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$. Assim,

$$\int_{-\infty}^{\infty} \widehat{H}_m(x) \widehat{H}_n(x) \omega(x) dx = 2^n n! \delta_{mn} \quad (2.24)$$

Os primeiros cinco polinômios de Hermite físicos são,

- $\widehat{H}_0(x) = 1$
- $\widehat{H}_1(x) = 2x$
- $\widehat{H}_2(x) = 4x^2 - 2$
- $\widehat{H}_3(x) = 8x^3 - 12x$
- $\widehat{H}_4(x) = 16x^4 - 48x^2 + 12$

Existe uma relação entre os polinômios de Hermite probabilísticos e físicos. A continuação mostramos as expressões,

$$\hat{H}_n(x) = 2^{\frac{n}{2}} H_n(\sqrt{2}x) \quad \& \quad H_n(x) = 2^{-\frac{n}{2}} \hat{H}_n\left(\frac{x}{\sqrt{2}}\right) \quad (2.25)$$

Nosso interesse vai estar nos polinômios de Hermite probabilísticos, devido a que vamos trabalhar com variáveis que apresentam incerteza, por tanto vai ser considerado em alguns casos distribuições de probabilidade gaussiana, ou seja, funções peso da forma $\omega(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Agora vamos definir os polinômios contínuos de Laguerre generalizados no intervalo $[0, +\infty)$, os quais são ortogonais com respeito à função densidade de probabilidade das variáveis aleatórias Gamma. Lembre-se que a distribuição Gamma tem função de densidade de probabilidade,

$$f(x) = \frac{x^\alpha e^{-x/\beta}}{\beta^{\alpha+1} \Gamma(\alpha+1)}, \quad \alpha > -1, \beta > 0 \quad (2.26)$$

Onde $\Gamma(t) = \int_0^\infty e^{-x} x^{t-1} dx$ é a função Gamma. Se consideramos na formula (2.26) o parâmetro $\beta = 1$, temos a função de peso $\omega(x) = \frac{e^{-x} x^\alpha}{\Gamma(\alpha+1)}$ com a qual os polinômios de Laguerre formam um conjunto ortogonal.

Esta classe de polinômios estão definimos como,

$$L_n^{(\alpha)}(x) = \frac{(\alpha+1)_n}{n!} {}_1F_1(-n; \alpha+1; x) \quad (2.27)$$

São ortogonais satisfazendo a seguinte relação,

$$\int_0^\infty L_m^{(\alpha)} L_n^{(\alpha)} \omega(x) dx = \frac{(\alpha+1)_n}{n!} \delta_{mn}, \quad \alpha > -1 \quad (2.28)$$

Para $\alpha = 0$ temos os polinômios de Laguerre ortogonais com respeito à função de peso positiva $\omega(x) = e^{-x}$. Os primeiros cinco polinômios de Laguerre são,

- $L_0(x) = 1$
- $L_1(x) = -x + 1$
- $L_2(x) = \frac{1}{2}x^2 - 2x + 1$
- $L_3(x) = \frac{-1}{6}x^3 + \frac{3}{2}x^2 - 3x + 1$
- $L_4(x) = \frac{1}{24}x^4 - \frac{2}{3}x^3 + 3x^2 - 4x + 1$

A Figura 2.5 apresenta os gráficos dos polinômios acima mencionados.

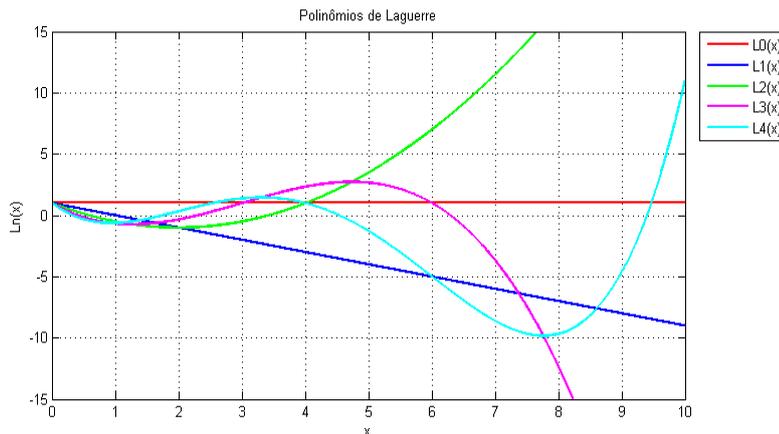


Figura 2.5: Primeiros cinco polinômios de Laguerre.

Finalmente, vamos definir os polinômios de Jacobi associados à distribuição de probabilidade Beta.

$$P_n^{(\alpha, \beta)}(x) = \frac{(\alpha + 1)_n}{n!} {}_2F_1\left(-n; n + \alpha + \beta + 1; \alpha + 1; \frac{1-x}{2}\right) \quad (2.29)$$

Sua relação de ortogonalidade está dada por,

$$\int_{-1}^1 P_m^{(\alpha, \beta)}(x) P_n^{(\alpha, \beta)}(x) \omega(x) dx = h_n^2 \delta_{mn}, \quad \alpha, \beta > -1 \quad (2.30)$$

Onde,

$$h_n^2 = \frac{(\alpha + 1)_n (\beta + 1)_n}{n! (2n + \alpha + \alpha + 1) (\alpha + \beta + 2)_{n-1}} \quad (2.31)$$

$$\omega(x) = \frac{\Gamma(\alpha + \beta + 2)}{2^{\alpha + \beta + 1} \Gamma(\alpha + 1) \Gamma(\beta + 1)} (1-x)^\alpha (1+x)^\beta \quad (2.32)$$

Quando $\alpha = \beta = 0$ os polinômios de Jacobi se tornam os polinômios de Legendre e a função de peso é uma constante $\omega(x) = 1$, a qual corresponde à importante função de distribuição uniforme. Os primeiros cinco polinômios de Legendre são,

- $P_0(x) = 1$
- $P_1(x) = x$
- $P_2(x) = \frac{3}{2}x^2 - \frac{1}{2}$
- $P_3(x) = \frac{5}{2}x^3 - \frac{3}{2}x$
- $P_4(x) = \frac{35}{8}x^4 - \frac{15}{4}x^2 + \frac{3}{8}$

A Figura 2.6 apresenta os gráficos dos primeiros cinco polinômios de Legendre no intervalo $[-1,1]$. Pode-se observar que os zeros são simétricos pois a função peso é par e o intervalo é também simétrico com respeito à origem. Além disso, é possível notar que os zeros destes polinômios se entrelaçam.

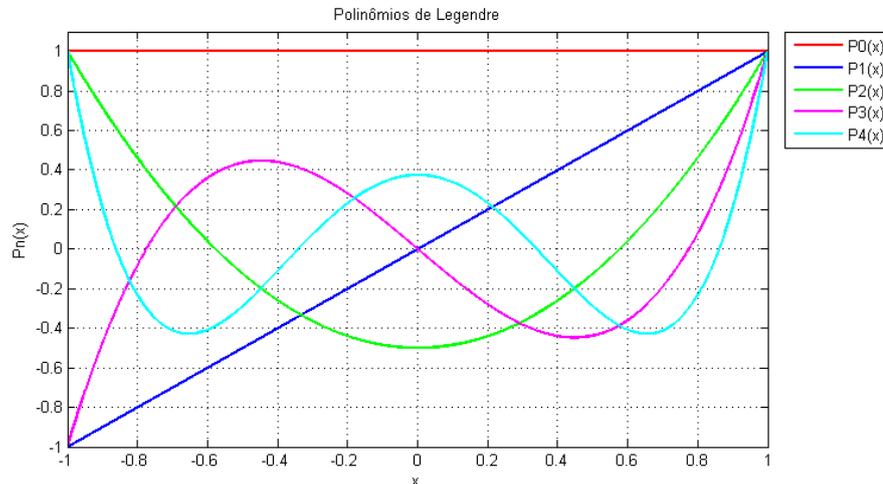


Figura 2.6: Primeiros cinco polinômios de Legendre.

Finalizamos esta seção expressando a importância que têm os polinômios ortogonais neste trabalho. Primeiramente, uma expansão por caos polinomial não é mais que uma combinação linear de polinômios, estes são ortogonais com respeito a uma função de peso associada à distribuição de probabilidade dos parâmetros envolvidos no modelo considerado. Além disso, os zeros de alguns polinômios ortogonais são utilizados na construção das regras de quadratura, que serão os métodos de integração numérica considerados na estimação dos coeficientes da expansão por caos polinomial.

Na próxima seção é enunciada a teoria básica dos produtos tensores, que é a base fundamental para os métodos de quadraturas multidimensionais.

2.4. Teoria de Produto Tensor

Nesta seção são enunciadas algumas definições e resultados da teoria dos produtos tensores. O interesse é gerado pelo uso do produto tensor nas regras de quadratura compostas e na construção do *sparse grid* ou quadratura de Smolyak, que se vai desenvolver no próximo capítulo.

O produto tensor é uma ferramenta matemática conveniente para fatorizar entidades com uma estrutura de produto. A fatoração tensorial pode ser aplicada a

funções, operadores lineares ou espaços vetoriais, mas nossa linha de desenvolvimento está orientada a definir os produtos tensores sobre funcionais lineares, para poder utilizar eles diretamente nas regras de quadratura. No caso de ter mais interesse neste tema, pode-se consultar (RYAN, 2002) e (KUBRUSLY, 2006).

Vamos começar introduzindo o seguinte espaço de funções para uma região fixa $\emptyset \neq \Omega \subseteq \mathbb{R}^d$,

$$H^r(\Omega) = \left\{ f: \Omega \rightarrow \mathbb{R}; \frac{\partial^\alpha f(x)}{\partial x^\alpha} \exists e \text{ é limitado em } \Omega \forall |\alpha|_\infty \leq r \right\} \quad (2.33)$$

São as funções com derivadas mistas limitadas de ordem r , esta classe de funções são chamadas de *regularidade r* e contam com a seguinte norma,

$$\|f\|_{H^r(\Omega)} = \max_{\alpha \in \mathbb{N}^d, |\alpha|_\infty \leq r} \left\{ \left| \frac{\partial^\alpha f(x)}{\partial x^\alpha} \right|; x \in \Omega \right\} \quad (2.34)$$

Vamos considerar funcionais lineares nesse espaço para definir o produto tensor. O dual deste espaço nos permitirá mais adiante derivar o erro de integração na quadratura de Smolyak, metodologia utilizada para aproximar os coeficientes da expansão em caos polinomial.

Definição 2.4: Suponha que $\emptyset \neq \Omega \subseteq \mathbb{R}^{d_1}$ e $\emptyset \neq \Upsilon \subseteq \mathbb{R}^{d_2}$ e sejam $\mathbf{S}: H^r(\Omega) \rightarrow \mathbb{R}$ e $\mathbf{T}: H^r(\Upsilon) \rightarrow \mathbb{R}$ funcionais. Suponha que admitem as seguintes representações,

$$Sf = \sum_{i=1}^m a_i f(x_i) \quad e \quad Tg = \sum_{i=1}^n b_i g(y_i) \quad (2.35)$$

Para a seleção de pesos positivos $(\mathbf{a}_i)_{i=1}^m$, $(\mathbf{b}_i)_{i=1}^n$ e vetores $(\mathbf{x}_i)_{i=1}^m$, $(\mathbf{y}_i)_{i=1}^n$ nos domínios Ω e Υ . Então, $\Omega \times \Upsilon \subseteq \mathbb{R}^{d_1+d_2}$ e o *produto tensor* de \mathbf{S} e \mathbf{T} é o funcional linear $\mathbf{S} \otimes \mathbf{T}: H^r(\Omega \times \Upsilon) \rightarrow \mathbb{R}$ definido como,

$$\mathbf{S} \otimes \mathbf{T} f = \sum_{i=1}^m \sum_{j=1}^n a_i b_j f(x_i, y_j) \quad (2.36)$$

Em geral, sejam $\Omega_k \neq \emptyset$ regiões em espaços Euclidianos \mathbb{R}^{d_k} e sejam $T_k: H^r(\Omega_k) \rightarrow \mathbb{R}$ funcionais para $k = 1, 2, 3, \dots$ tais que,

$$T_k f = \sum_{i=1}^{m_k} \omega_i^{(k)} f(x_i^{(k)}) \quad (2.37)$$

Onde $(\omega_i^{(k)})_{i=1}^{m_k}$ são pesos positivos e $(x_i^{(k)})_{i=1}^{m_k}$ é uma sequência de vetores em Ω_k . Então, podemos enumerar as seguintes propriedades:

$$\text{i.} \quad \bigotimes_{i=1}^1 T_i = T_1 \quad (2.38)$$

$$\text{ii.} \quad \bigotimes_{i=1}^n T_i = \bigotimes_{i=1}^{n-1} T_i \otimes T_n, \quad n = 2, 3, \dots \quad (2.39)$$

iii. $T_1 \otimes T_2 \otimes \dots \otimes T_n$, para $n = 1, 2, 3, \dots$ define um funcional linear $H^r(\Omega_1 \times \dots \times \Omega_n) \rightarrow \mathbb{R}$ tal que,

$$\bigotimes_{i=1}^n T_i f = \sum_{i_1=1}^{m_1} \dots \sum_{i_n=1}^{m_n} \omega_{i_1}^{(1)} \dots \omega_{i_n}^{(n)} f(x_{i_1}^{(1)}, \dots, x_{i_n}^{(n)}) \quad (2.40)$$

iv. O produto tensor não é comutativo, é dizer,

$$T_i \otimes T_j \neq T_j \otimes T_i, \quad \forall i \neq j \quad (2.41)$$

v. Associatividade,

$$(T_i \otimes T_j) \otimes T_k = T_i \otimes (T_j \otimes T_k), \quad \forall i, j, k \quad (2.42)$$

vi. Distributividade com respeito à soma ordinária ponto a ponto,

$$(T_i + T_j) \otimes T_k = T_i \otimes T_k + T_j \otimes T_k, \quad \forall i, j, k \quad (2.43)$$

Encerramos esta seção com um exemplo de avaliação orientado a nosso interesse.

Exemplo 2.1: Seja o conjunto de d operadores,

$$\left(U_k^{(j)} g \right)_{j=1}^d = \sum_{i=1}^k \omega_i^{(j)} g(x_i^{(j)})$$

Onde $\omega_i^{(j)}$ e $x_i^{(j)}$ são os pesos e pontos de avaliação respectivamente, para $j = 1, 2, \dots, d$. Vamos considerar o seguinte problema clássico de integração numérica,

$$\mathcal{J}_W^d f = \int_{I_1} \dots \int_{I_d} W_1(z_1) \dots W_d(z_d) f(z_1, \dots, z_d) dz_d \dots dz_1 \quad (2.44)$$

Suponha que, $U_k^{(j)} p = \mathcal{J}_{W_j}^1 p$, para todos os polinômios p de grau máximo m_k em I_j . Então, empregando cada operador coordenada a coordenada no problema anterior, temos

$$\begin{aligned}
& \mathcal{J}_W^d f \\
& \approx \int_{I_1} \dots \int_{I_{d-1}} W_1(z_1) \dots W_{d-1}(z_{d-1}) \sum_{i_d=1}^k \omega_{i_d}^{(d)} f(z_1, \dots, z_{d-1}, x_{i_d}^{(d)}) dz_{d-1} \dots dz_1 \\
& \approx \sum_{i_1=1}^k \dots \sum_{i_d=1}^k \omega_{i_1}^{(1)} \dots \omega_{i_d}^{(d)} f(x_{i_1}^{(1)}, \dots, x_{i_d}^{(d)})
\end{aligned}$$

Note-se que isto pode ser expressado mediante um produto tensor, como segue,

$$\mathcal{J}_W^d f \approx \bigotimes_{i=1}^d U_k^{(i)} f$$

Este tipo de expressão é conhecido na literatura como quadratura produto tensor. No capítulo 4 se vai desenvolver com mais detalhe este problema das quadraturas unidimensionais, cuja composição é exatamente esta expressão. No próximo capítulo serão definidos alguns métodos de integração numérica entre os quais estão as regras de quadratura, que são de nosso interesse.

3 Quantificação de Incertezas em Simulação de Reservatórios

3.1. Introdução

A modelagem numérica (determinística) e a simulação de sistemas complexos, são continuamente desenvolvidos em vários campos da engenharia e as ciências. A simulação numérica, atualmente é uma ferramenta essencial para os engenheiros durante todo o processo de desenho de um sistema. Esta permite em muitos casos, minimizar a necessidade de custosos experimentos físicos, que podem até ser impossíveis em etapas iniciais do processo. No entanto, os modelos de simulação devem ser cuidadosamente desenhados e verificados, para obter a partir deles informações úteis e confiáveis do sistema estudado. Com efeito, os simuladores inerentemente envolvem erros que devem ser entendidos e quantificados para avaliar as diferenças entre as previsões numéricas e o comportamento real do sistema simulado (LE MAÎTRE, et al., 2010).

Por outro lado, a modelagem estocástica é mais difícil do que a modelagem determinística, e a implementação dos modelos é muito mais custosa. No caso de modelagem determinística, os parâmetros do modelo são constantes, seu equivalente em modelagem estocástica é transformar essas constantes em variáveis aleatórias. Para cada realização do parâmetro aleatório o valor da saída é uma constante, mas essa constante pode ser diferente para outras realizações e as realizações seguem uma distribuição de probabilidade. Ou seja, os objetos mais simples de um modelo estocástico, as variáveis aleatórias, são funções, portanto objetos de dimensão infinita (SOUZA DE CURSI, et al., 2012).

Em um modelo determinístico, se ele é coerente, temos uma solução uma vez fixadas as condições suplementares. Em um modelo estocástico não é assim. Para cada realização dos parâmetros do modelo estocástico, temos um modelo determinístico que tem uma solução, e devem ser realizadas estatísticas dessas

soluções para conhecer como elas se distribuem; em outras palavras, o que interessa em modelagem estocástica é a distribuição de probabilidade.

Num modelo com parâmetros de entrada incertos, os valores obtidos para uma realização não representam o comportamento completo do sistema estocástico. Para que verdades desse tipo fossem absorvidas pela indústria foram necessários anos de controle de qualidade, que não funcionava, até que a modelagem estocástica fosse promovida à ferramenta essencial de engenharia (SOUZA DE CURSI, et al., 2012). Daí o grande interesse atual em métodos que considerem as incertezas presentes nos modelos estudados.

Na área de reservatórios de petróleo, as empresas vêm utilizando para o gerenciamento, a simulação de reservatórios como peça-chave na propagação de incertezas e análise de riscos. A simulação de reservatórios influencia muito o processo de tomada de decisão, relacionado à exploração e produção de petróleo, já que é através dela, que são testadas várias alternativas para esses processos.

A precisão das estimativas obtidas através dos simuladores de reservatórios, depende da qualidade dos dados que os alimentam e dos modelos utilizados. Em anos anteriores, muitas das propriedades do reservatório eram definidas por um único valor, embora apresentassem variações, provocando previsões imprecisas. Atualmente, estimativas mais precisas são necessárias pelos altos investimentos alocados nos campos de petróleo, para isto, as propriedades do reservatório cujos comportamentos não são conhecidos completamente, são parametrizadas usando variáveis aleatórias e tratadas por metodologias específicas que chamaremos de técnicas de propagação ou quantificação de incertezas.

As incertezas associadas aos dados que descrevem um sistema de simulação de reservatórios de petróleo constituem o tema central desta dissertação. Especificamente, tentamos descrever métodos e análises para caracterizar os efeitos da variabilidade nos dados, ou, o impacto do conhecimento impreciso nos dados de entrada, para especificar a resposta do sistema.

3.2. Simuladores de Reservatórios de Petróleo

Um modelo de simulação de reservatórios é um modelo computacional capaz de emular o comportamento de um reservatório, predizendo a sua produção

sob condições de operação corrente, ou sua reação mediante alteração dessas condições (ERTEKIN, et al., 2001).

Uma das funções da simulação é auxiliar no entendimento do comportamento da pressão e produção de um reservatório e, conseqüentemente, na predição das taxas de produção como uma função do tempo. A previsão futura, quando expressa em termos de rendimentos e comparada com custos e investimentos, determina a viabilidade da operação (ADAMSON, et al., 1996).

Os simuladores de reservatórios são tipicamente classificados em duas categorias: os simuladores *black-oil* e os simuladores térmicos e composicionais. Os primeiros consideram um sistema com apenas três componentes –óleo, água e gás– e três fases, também designadas por óleo, água e gás. Assume-se que o componente óleo só existe na fase óleo, a componente água só existe na fase água e o componente gás pode-se encontrar como gás livre ou dissolvido no óleo. A segunda categoria, são modelos para reservatórios que requerem descrição mais detalhada da composição do fluido. Uma descrição composicional poderia abranger as quantidades e propriedades do pentano, butano e outros hidrocarbonetos, e pode ser usada quando a composição de fluidos muda durante a simulação. Um simulador térmico, considera a mudança de temperatura para modificar a composição do fluido (ADAMSON, et al., 1996).

Das principais dificuldades na utilização de qualquer tipo de simulador de reservatório, é a aquisição dos dados de entrada quando não se tem um total conhecimento do comportamento destas informações. A imprecisão destes dados, geram incerteza, que se não é tratada adequadamente, pode provocar erros nas estimativas produzidas pelos simuladores. A maneira correta de tratar esse tipo de informações e com métodos de quantificação de incertezas.

3.3. Quantificação de Incertezas

No contexto de modelos de simulação de reservatórios de petróleo, a incerteza presente nos parâmetros de entrada, é resultado da heterogeneidade das formações geológicas, as propriedades dos fluidos e a imprecisão nos dados medidos (SARMA, et al., 2011). Como resultado, as saídas do modelo, tal como a produção acumulada de óleo, água ou gás apresentam também um nível de incerteza significativo, a quantificação da qual, é usualmente realizada usando

propriedades estatísticas tais como a média, o desvio padrão, alguns percentis (P10, P50, P90) ou, mais rigorosamente, a função distribuição de probabilidades completa. A quantificação desta incerteza, tem um papel importante no gerenciamento dos campos de reservatórios. Quando se tem conhecimento da incerteza e da sua propagação através do modelo, é acrescentada a confiança na validade dos resultados de uma avaliação no modelo.

Na área de engenharia é comum utilizar sistemas baseados em modelos matemáticos determinísticos, com dados de entrada definidos por valores fixos. Na realidade, como nos reservatórios de petróleo, é difícil ter essas condições. Algumas propriedades dos campos de petróleo, não podem ser representadas por um único valor, devido a que este não pode expressar o comportamento da propriedade em todo o reservatório. Estas características são chamadas aleatórias ou incertas, e o conhecimento sobre elas é extremadamente limitado porque podem ser medidas em apenas alguns poços exploratórios, portanto no restante do campo, estas propriedades são estimadas através de técnicas geoestatísticas. Assim, num ponto arbitrário do reservatório, não se tem o valor exato, tem-se uma aproximação.

Por outro lado, devido à natureza determinística dos modelos de simulação, é preciso alimentá-los com dados também determinísticos, portanto se devem calcular valores fixos para as propriedades denominadas incertas, utilizando suas distribuições de probabilidade. De esta maneira, é possível reformular o problema determinístico em um problema estocástico, e assim, os simuladores irão propagar as incertezas dos dados de entrada produzindo também saídas incertas.

Na seguinte seção será definido com maior detalhe os parâmetros aleatórios envolvidos nos modelos de simulação de reservatórios.

3.3.1. Parâmetros incertos

A primeira etapa do processo de quantificação de incertezas é descrever os parâmetros que apresentam incerteza mediante suas funções de distribuição de probabilidade. Este é possível parametrizando os dados utilizando um conjunto de variáveis aleatórias independentes, na maioria dos problemas reais, variáveis aleatórias contínuas.

Os tipos de distribuições mais utilizados para a representação dessas variáveis são:

- Normal
- Log-normal
- Uniforme
- Exponencial
- Triangular
- Beta
- Gamma

Em modelos de simulação de reservatórios de petróleo, devido à dificuldade na coleta dos dados, a extração de amostras e o pouco e impreciso conhecimento geológico que tem-se do campo completo, existe uma grande quantidade de parâmetros que apresentam incerteza. Nesta dissertação foram considerados os seguintes, para maior detalhe ver (THOMAS, 2001) e (ROSA, et al., 2006).

- ***Porosidade:***

A porosidade, é uma das mais importantes propriedades das rochas na engenharia de reservatórios, já que ela mede a capacidade de armazenamento de fluidos. A porosidade define-se como a relação entre o volume de vazios de uma rocha e o volume total da mesma. O valor da porosidade para um reservatório como um todo é obtido a partir do tratamento estatístico dos dados de medições realizadas em várias amostras de testemunhos retirados de diversos poços perfurados nesse reservatório. Verifica-se na prática que a porosidade apresenta em geral uma distribuição de probabilidade normal (ROSA, et al., 2006) .

- ***Permeabilidade:***

A permeabilidade de um meio poroso é uma medida de sua capacidade de se deixar atravessar por fluidos. Esta propriedade de reservatórios pode ser medida através de ensaios de laboratório com amostras extraídas da formação ou por meio de testes de pressão realizados em poços perfurados. Na prática observa-se que a permeabilidade segue uma distribuição lognormal (ROSA, et al., 2006).

- **Compressibilidade da rocha:**

Um corpo ao ser submetido a uma compressão, sofrerá uma redução de volume. O quociente entre a redução de volume e o volume inicial do corpo, recebe o nome de variação fracional. A compressibilidade total da rocha, é a variação fracional do volume total da rocha, com a variação unitária da pressão (THOMAS, 2001). Em geral, não é possível correlacionar valores de compressibilidade de rocha, as correlações que podem ser obtidas, fornecem, na melhor das hipóteses, apenas uma ordem de grandeza dos valores de compressibilidade. Valores mais representativos devem, preferencialmente, ser medidos em laboratórios para cada caso de reservatório (ROSA, et al., 2006).

- **Propriedades do Aquífero:**

Quando um reservatório apresenta um aquífero adjacente, a queda de pressão do reservatório decorrente da produção dos fluidos pode provocar a expansão da água do aquífero e o conseqüente influxo de água para o reservatório. O conhecimento do volume de água que flui do aquífero para o reservatório é fundamental para a realização de estudos de previsão do comportamento do reservatório (ROSA, et al., 2006). Neste trabalho é estudado um modelo de simulação de reservatórios baseado no campo PUNQ-S3. O qual, é um modelo de engenharia de reservatório industrial, que tem comunicação com um forte aquífero ao oeste e norte. O raio de este aquífero é considerado um parâmetro incerto que afeta a saída do modelo de simulação.

- **Permeabilidade Relativ:**

A vazão de uma fase em um meio poroso saturado com dois ou mais fluidos é sempre menor que a vazão de uma fase quando somente ela satura o meio poroso. A capacidade de transmissão de um fluido que satura 100% de um médio poroso é definida como *permeabilidade absoluta* (k). No caso em que dois ou mais fluidos saturam o meio poroso, a capacidade de transmissão de um desses fluidos chama-se *permeabilidade efetiva* do meio poroso ao fluido considerado. O quociente entre a permeabilidade efetiva e a permeabilidade absoluta do meio poroso é denominada *permeabilidade relativa* ao fluido (ROSA, et al., 2006). Este valor é afetado por outros fatores do reservatório, por exemplo a molhabilidade, a direção do deslocamento: embebição (se a saturação da fase molhada vai em

aumento) ou drenagem (se a saturação da fase molhada vai diminuindo) e finalmente pelo tamanho dos poros (DESTEFANIS, et al., 2007).

Uma dificuldade na quantificação dos efeitos da molhabilidade sobre as curvas de permeabilidade relativa é a possibilidade de modificação da molhabilidade durante o processo de amostragem. Com isso, os experimentos de laboratório podem não representar o processo de saturação que ocorre no reservatório, se drenagem ou embebição. Por tanto, embora não exista nenhuma expressão teórica geral para as curvas de permeabilidade relativa, tem-se proposto distintas relações empíricas, por exemplo os *modelos de potência* ou *leis de potência* (ROSA, et al., 2006).

Para um sistema água-óleo a curva de permeabilidade relativa à água pode ser representada pela expressão:

$$k_{rw} = k_{rw}^* \left(\frac{S_w - S_{wc}}{1 - S_{wc} - S_{orw}} \right)^{n_w} \quad (3.1)$$

Onde

S_w : Saturação de água.

S_{wc} : Saturação de água irreduzível (mínima)

S_{orw} : Saturação de óleo residual para um sistema água-óleo

Para um sistema gás-óleo, é comum representar a curva de permeabilidade relativa ao gás, em função da saturação de líquidos: ($S_l = \text{saturação água} + \text{saturação óleo}$).

$$k_{rg} = k_{rg}^* \left(\frac{1 - S_l}{1 - S_{wc} - S_{org}} \right)^{n_g} \quad (3.2)$$

Onde

S_w : Saturação de água.

S_{wc} : Saturação de água irreduzível (mínima)

S_{org} : Saturação de óleo residual para um sistema gás-óleo

Note-se que as equações eq. (3.1) e eq. (3.2) têm mais dois parâmetros. Os primeiros, são os pontos terminais ou valores máximos das curvas de permeabilidade relativa (k_{rw}^* e k_{rg}^*), que podem ser determinados a partir de medições feitas em laboratórios, dados de perfis corridos nos poços e correlações.

Os segundos, os expoentes, (n_w e n_g) chamados comumente, expoentes de Corey, podem ser obtidos por meio de um ajuste de histórico de um modelo numérico aos dados de produção do reservatório. Estes valores, fixam a curvatura das permeabilidades relativas, influenciada pela distribuição do tamanho dos poros.

Devido ao pouco conhecimento que tem-se sobre estes quatro parâmetros e à incerteza que eles apresentam pelos erros nas medições ou estimações de ajustes, são considerados nesta dissertação como variáveis aleatórias envolvidas no modelo de simulação de reservatórios de petróleo.

- ***Profundidade dos Contatos dos Fluidos***

A determinação dos contatos de fluidos no interior dos reservatórios petrolíferos é necessária na exploração e produção de petróleo, estas localizações podem ser utilizadas nos projetos de controle de perfurações, por exemplo. A obtenção da localização, em profundidade, dos contatos fluidos dentro do reservatório pode ser alcançada mediante a interpretação completa dos perfis geofísicos ou através dos onerosos e demorados teste de formação. Portanto, os valores da profundidade de contato óleo-gás e óleo-água são considerados como variáveis incertas neste trabalho.

Finalmente, para terminar esta seção, reiteramos o fato que, considerar a incerteza em variáveis cujo comportamento não é conhecido por completo em todas as situações, é fundamental para representar a incerteza nas saídas dos sistemas. Neste trabalho as variáveis antes mencionadas, foram consideradas aleatórias com distribuições de probabilidade normal e log-normal, truncadas para manter as condições geológicas do campo.

Na seguinte seção vamos definir com maior detalhe as técnicas mais utilizadas para a quantificação de incertezas.

3.3.2. Técnicas de Quantificação de Incertezas

Compreender a incerteza que afeta a saída de modelos e sistemas em diferentes áreas das ciências e engenharias, tem sido de interesse para muitos, durante muito tempo. Conseqüentemente tem surgido métodos concebidos para tratar este problema. Os enfoques mais dominantes têm sido aqueles, que tratam

os dados incertos como variáveis aleatórias ou processos estocásticos e reformulam os sistemas, que originalmente eram determinísticos, como sistemas estocásticos.

Existe atualmente um conjunto de métodos populares nas aplicações de engenharia, os quais podem ser classificados genericamente em duas categorias: Os métodos que utilizam amostragem direta e os métodos que não utilizam amostragem direta.

A abordagem por amostragem direta inclui, por exemplo, o tradicional método de simulação de Monte Carlo (FISHMAN, 2013) (LIU, 2008), o qual, apesar de apresentar uma precisão excelente, torna-se inviável para algumas aplicações, em virtude de sua baixa taxa convergência. Por outro lado, para acelerar a convergência, surgiram outras alternativas como: Quase Monte Carlo (FOX, 1999) (HARALD, 1992), *Latin hypercube sampling* (LOH, 1996) (STEIN, 1987) e aproximação por superfícies de resposta (MYERS, et al., 2009). Essas técnicas envolvem amostragem e estimativas e, na maioria dos casos possuem implementações relativamente simples. No contexto de simulação de reservatórios de petróleo, estas técnicas precisam de um número de amostras muito grande, acarretando num tempo computacional elevado e não aceitável para solucionar o problema.

Por outro lado, muitos esforços têm sido alocados em pesquisas para desenvolver metodologias que não utilizem amostragem direta. O método mais popular é o método de perturbação (LIU, et al., 1986), onde os campos aleatórios são expandidos através de series de Taylor em torno de sua média e truncados numa determinada ordem. Outra abordagem é baseada na manipulação de operadores estocásticos, incluindo as expansões de Neumann (YAMAZAKI, et al., 1988), a qual expressa a inversa de um operador estocástico como uma serie de Neumann.

Nesta dissertação, são propostas duas técnicas para resolver o problema de quantificação de incerteza em modelos de simulação de reservatórios. No capítulo 5 são desenvolvidos os fundamentos teóricos para o método denominado Expansões por Caos Polinomial Generalizado, baseado no caos homogêneo de Wiener (WIENER, 1938) e utilizado com sucesso em distintas áreas das ciências e as engenharias. Com caos polinomial, é possível expressar soluções estocásticas como polinômios ortogonais dos parâmetros de entrada aleatórios, e diferentes

tipos de polinômios ortogonais podem ser selecionados para alcançar a melhor convergência. Por outro lado, no capítulo 6 é enunciada uma técnica da área de computação evolucionária, chamada de Programação Genética, baseada num algoritmo de regressão simbólico que permite expressar a saída de um modelo por médio de funções ou operadores. Este método permite construir soluções aproximadas que representem comportamentos não lineares (exp, log, potências) entre os dados de entrada e saída do modelo.

3.3.3. Maldição da Dimensionalidade

O termo *Curse of Dimensionality* refere-se aos fenômenos que surgem quando são analisados e organizados dados de espaços de alta dimensionalidade, os quais não acontecem em problemas de baixa dimensão. Este termo foi introduzido por Richard E. Bellman quando se encontrava analisando problemas de otimização dinâmica (BELLMAN, 1956), para descrever o extraordinariamente rápido que cresce a dificuldade de um problema quando o número de variáveis ou a dimensão aumenta. Uma experiência comum é que o custo de um algoritmo cresça exponencialmente com respeito à dimensão, fazendo que seja inviável utiliza-lo em dimensões moderadas e altas. Esta situação acontece na estimação de integrais multidimensionais mediante a composição de quadraturas uni-variadas. Por este motivo, neste trabalho é proposto o método da quadratura de Smolyak, na seção 4.4.2. Em problemas de modelos de simulação de reservatórios de petróleo, o custo de dimensionalidade, é ainda mais delicado que em problemas simples, devido a que está relacionado com a quantidade de avaliações no modelo, e neste caso, o modelo é um simulador numérico de reservatórios, que já por se mesmo tem um custo computacional grande.

Comentário:

Finalmente, nesta dissertação vão ser desenvolvidos problemas de quantificação de incerteza em modelos de simulação de reservatórios, com até vinte e três parâmetros de entrada incertos, considerados independentes. Primeiro se utilizará expansões por caos polinomial, com a regra de Smolyak e um enfoque de regressão, para aproximar os coeficientes da expansão polinomial, com a menor quantidade de avaliações no simulador.

Por outro lado, é construída uma matriz com os dados de entrada (de acordo a suas distribuições de probabilidade) e suas correspondentes saídas através do modelo de simulação, para gerar um aprendizado no algoritmo de programação genética e obter uma expressão explícita simbólica que aproxime a resposta do problema considerado. As duas técnicas, tanto a de PCE como a de PG, brindam a facilidade e vantagem de utilizar os resultados obtidos, como um modelo substituto ou *proxy* da saída do modelo de simulação e desta maneira realizar estimativas estatísticas, análises de probabilísticos e outros estudos de interesse.

4 Métodos de Integração Numérica

4.1. Introdução

Muitas medidas de interesse para as ciências se expressam através de integrais, e algumas de estas integrais são difíceis de resolver analiticamente. Isto pode ser devido à complexidade da função a integrar, ou ao domínio de integração ou ambos. Além disso, nos casos onde não conhecemos a função a integrar (como no caso de modelos do tipo *black-box*), e somente se tem uma tabela de dados, técnicas de integração como as desenvolvidas neste capítulo, são absolutamente necessárias se deseja-se estimar a integral de alguma maneira.

Se chama integração numérica ao conjunto de técnicas e métodos que se têm desenvolvido para o cálculo aproximado de integrais. A fórmula padrão dos algoritmos de integração está dada pela seguinte soma pondera finita,

$$I = \int_a^b f(x)dx \approx \sum_{i=1}^N \omega_i f(x_i) \quad (4.1)$$

A função $f(x)$ se avalia em N pontos no intervalo $[a, b]$ e os valores $f(x_i)$ se somam com uma ponderação ou peso w_i . Esta soma pode ser exata somente em dois casos, quando $N = \infty$ ou para N finito, se a função integrável é um polinômio. Os algoritmos de integração numérica equivalem a diferentes seleções de pontos de integração ou *nodos* x_i e pesos w_i . Conforme N cresce, aumenta a precisão, mas a melhor aproximação depende em grande maneira da função mesma $f(x)$.

Existem diferentes técnicas de integração numérica para estimar integrais definidas, as mais usuais são as Fórmulas de Newton-Côtes (SÜLI, et al., 2003), que utilizam uma partição igualmente espaçada do intervalo de integração para definir os nodos, e as Regras de Quadraturas que vamos descrever nas próximas seções.

O interesse em técnicas de integração numérica está motivado no fato que, os coeficientes do polinômio gerado por PCE podem ser definidos mediante

integrais multidimensionais. O desempenho da PCE depende do método utilizado para estimar estes valores, portanto, é fundamental investigar e avaliar técnicas de integração numérica que sejam eficientes em termos de precisão e da quantidade de avaliações à função.

Nas próximas seções deste capítulo são enunciadas as técnicas de quadratura Gaussiana e Genz-Keister para o caso de integrais unidimensionais e as quadraturas produto tensor e Smolyak para o caso de integrais multidimensionais.

4.2. Quadraturas Gaussianas Unidimensionais

Pelo contrário das Fórmulas de Newton-Côtes, as regras de quadratura Gaussiana ou de Gauss (SÜLI, et al., 2003) e (HOLTZ, 2010), estão baseadas em determinar um conjunto de pontos do intervalo de integração, de tal maneira que se otimize a exatidão e o custo computacional requerido, no cálculo da integral.

Em alguns casos é útil escrever a fórmula básica de integração da eq. (4.1) de tal forma que seja separada a função de peso ou densidade $W(x)$, da função a integrar, assim,

$$\int_a^b f(x)dx = \int_a^b W(x)g(x)dx \approx \sum_{i=1}^N \omega_i g(x_i) \quad (4.2)$$

Nos métodos de quadraturas Gaussianas, os N pontos e pesos, são escolhidos de maneira tal que o erro de aproximação desapareça se $g(x)$ é um polinômio de grau $(2N - 1)$. Para obter este incrível resultado os pontos x_i não têm uma amostragem uniforme, pelo contrário têm uma distribuição mais complexa no intervalo $[a, b]$. Este fato, faz com que os algoritmos de quadratura de Gauss tenham resultados com uma precisão melhor que dos métodos de Newton-Côtes (Simpson e Trapezoidal).

A integração Gaussiana está baseada no uso de polinômios ortogonais, que definimos na seção 2.3, para aproximar o integrando $g(x)$ no intervalo $[a, b]$.

Os seguintes três teoremas são um resumo da teoria básica de quadraturas Gaussianas, a demonstração pode-se ver em (STROUD, 1974). O interesse em enunciar estes resultados está na necessidade de conhecer quantos pontos são necessários, considerando ao integrando, para obter aproximações precisas.

Teorema 4.1: (STROUD, 1974) Sejam x_1, x_2, \dots, x_N as raízes do polinômio ortogonal $p_N(x)$ (de grau N) com respeito à densidade $\omega(x)$ em $[a, b]$. Suponha que são encontrados os pesos $\omega_1, \omega_2, \dots, \omega_N$ tal que,

$$\int_a^b \omega(x)g(x)dx \approx \sum_{i=1}^N \omega_i g(x_i) \quad (4.3)$$

É exata para todos os polinômios de grau menor ou igual que $N - 1$. Então, a fórmula da eq. (4.3) é exata para todos os polinômios de grau $\leq 2N - 1$.

∇

Este teorema fornece a informação do grau de exatidão polinomial das quadraturas gaussianas, o qual é fundamental na escolha do número de pontos utilizados quando vai-se aproximar uma integral.

O segundo teorema afirma que não é possível encontrar outra fórmula de este tipo com a mesma precisão.

Teorema 4.2: (STROUD, 1974) Se a fórmula

$$\int_a^b \omega(x)g(x)dx \approx \sum_{i=1}^N \omega_i^* g(x_i^*) \quad (4.4)$$

Tem grau de precisão $2N - 1$, então os pontos x_i^* devem ser os zeros do polinômio ortogonal $p_N(x)$ com respeito à densidade $\omega(x)$ em $[a, b]$.

∇

Por último, o terceiro teorema nos indica o signo das raízes.

Teorema 4.3: (STROUD, 1974) Em uma fórmula de quadratura Gaussiana, todos os pesos ω_i são positivos.

∇

Nesta dissertação, nosso interesse se encontra apenas em avaliar integrais de funções peso associadas às famílias de polinômios ortogonais de Legendre, Laguerre e Hermite, polinômios definidos no capítulo anterior. Por tanto, são apresentados alguns exemplos de quadraturas relacionadas com estas famílias de polinômios ortogonais.

4.2.1. Quadratura de Gauss-Hermite

Este tipo de quadratura está construído a partir dos polinômios de Hermite que definimos na seção 2.3, que são ortogonais com respeito à função de peso $\omega(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$ no intervalo $(-\infty, +\infty)$. Originalmente esta regra nos permite obter uma expressão para,

$$\int_{-\infty}^{+\infty} e^{-x^2} \cdot f(x) dx \approx \sum_{i=1}^n \omega_i f(x_i) \quad (4.5)$$

Os n nodos ou pontos de integração, x_1, \dots, x_n estão definidos a partir das raízes dos polinômios de Hermite físicos \hat{H}_{n-1} e os pesos estão definidos por,

$$\omega_i = \frac{2^{n-1} n! \sqrt{\pi}}{n^2 [\hat{H}_{n-1}(x_i)]^2} \quad (4.6)$$

Nesta dissertação vamos calcular constantemente valores esperados (médias) de funções que dependem de variáveis aleatórias, por tanto, é necessário ter uma expressão para a integral que lhe define.

Seja h uma função que depende da variável aleatória y que tem uma distribuição normal, $\mathcal{N}(\mu, \sigma^2)$. Então a média de $h(y)$ está dada por,

$$E[h(y)] = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} h(y) dy \quad (4.7)$$

Devido a que não temos exatamente a integral na equação (4.5), vamos fazer a mudança de variável: $x = \frac{y-\mu}{\sqrt{2}\sigma}$. Por tanto, temos

$$E[h(y)] = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{\pi}} e^{-x^2} h(\sqrt{2}\sigma x + \mu) dx \quad (4.8)$$

Finalmente utilizando a quadratura de Gauss-Hermite com as raízes dos polinômios físicos, podemos expressar a média como,

$$E[h(y)] \approx \frac{1}{\sqrt{\pi}} \sum_{i=1}^n \omega_i h(\sqrt{2}\sigma x_i + \mu) \quad (4.9)$$

Onde $\sqrt{2}x_i$ são as raízes dos polinômios de Hermite probabilísticos. Como foi mencionado anteriormente, com um nível de acurácia n , a quadratura da gauss-hermite têm uma exatidão polinomial de $2n - 1$.

Na próxima seção, é apresentada a quadratura de Gauss-Laguerre que nos permite calcular integrais de funções exponenciais.

4.2.2. Quadratura de Gauss-Laguerre

Como seu nome o indica, a quadratura de Gauss-Laguerre depende das raízes dos polinômios de Laguerre, os quais são ortogonais no intervalo $[0, +\infty)$ com respeito à função de peso $\omega(x) = e^{-x}$.

Esta regra nos fornece da seguinte fórmula de integração numérica,

$$\int_0^{+\infty} e^{-x} f(x) dx \approx \sum_{i=1}^n \omega_i f(x_i) \quad (4.10)$$

As coordenadas x_1, \dots, x_n são os zeros dos polinômios de Laguerre que definimos na seção 2.3, e os pesos ω_i estão são obtidos mediante a seguinte expressão,

$$\omega_i = \frac{(2n!)^2 x_i}{(n+1)^2 [L_{n+1}(x_i)]^2} \quad (4.11)$$

Por último, na seção de quadraturas de Gauss, vamos definir a regra de Legendre que é muito utilizada no caso de integrais que não têm função peso, ou que a função de peso é uma constante. Além disso, são funções definidas num intervalo de integração finito.

4.2.3. Quadratura de Gauss-Legendre

A regra de quadratura de Gauss-Legendre está baseada nos polinômios de Jacobi que vimos na seção 2.3. Os polinômios de Legendre são um caso particular dos polinômios de Jacobi, quando consideramos os parâmetros $\alpha = \beta = 0$. Lembre-se que estes polinômios são ortogonais com respeito à função de peso constante $\omega(x) = 1$ no intervalo $[-1, 1]$, que corresponde à distribuição de probabilidade uniforme.

A ideia fundamental deste método é conseguir uma expressão do tipo,

$$\int_{-1}^1 1 \cdot f(x) dx \approx \sum_{i=-1}^n \omega_i f(x_i) \quad (4.12)$$

Onde os nodos ou pontos de integração x_1, \dots, x_n são pontos no intervalo $[-1, 1]$ de tal maneira que resulte exata a fórmula para o caso onde $f(x)$ é um polinômio de grau menor o igual que $2n - 1$, ver (PEIXOTO, 2008).

Analisemos o caso mais simples, para $n = 2$. Então, temos com um erro de aproximação E ,

$$\int_{-1}^1 f(x)dx = \omega_1 f(x_1) + \omega_2 f(x_2) + E \quad (4.13)$$

Desejamos que a fórmula seja exata para polinômios de grau menor ou igual que $2 \cdot 2 - 1 = 3$. Então, se integramos de maneira exata os monômios: $1, x, x^2$ e x^3 , e tendo em consideração que o erro é zero para estes casos, temos o seguinte sistema de equações:

$$\begin{cases} 2 = \omega_1 + \omega_2 \\ 0 = \omega_1 x_1 + \omega_2 x_2 \\ \frac{2}{3} = \omega_1 x_1^2 + \omega_2 x_2^2 \\ 0 = \omega_1 x_1^3 + \omega_2 x_2^3 \end{cases} \quad (4.14)$$

Com solução,

$$\omega_1 = \omega_2 = 1 \quad ; \quad x_1 = \frac{1}{\sqrt{3}} \quad ; \quad x_2 = \frac{-1}{\sqrt{3}} \quad (4.15)$$

Então temos a fórmula da quadratura de ordem 2,

$$\int_{-1}^1 f(x)dx = f\left(\frac{1}{\sqrt{3}}\right) + f\left(\frac{-1}{\sqrt{3}}\right) + E \quad (4.16)$$

No caso que seja de interesse calcular uma integral com diferentes limites de integração, sempre é possível redefinir a variável na seguinte forma,

$$\int_a^b f(t)dt = \frac{b-a}{2} \cdot \int_{-1}^1 f\left(\frac{(b-a)x + (a+b)}{2}\right) dx \quad (4.17)$$

Agora, este é apenas o caso de ordem dois, de maneira geral, se desejamos estabelecer uma fórmula do tipo da eq. (4.12) com n pontos no intervalo $[-1,1]$, o resultado que é obtido, é o seguinte: Os nodos serão, para cada valor de n , as raízes do correspondente polinômio de Legendre de grau n que definimos na seção 2.3. Por outro lado, os pesos são obtidos pela expressão,

$$\omega_i = \int_{-1}^1 \left(\prod_{j=1, j \neq i}^n \frac{x - x_j}{x_i - x_j} \right) dx \quad (4.18)$$

Estas fórmulas de quadratura têm o nome Gauss-Legendre precisamente porque os polinômios utilizados para o cálculo da integral são os de Legendre.

Na próxima seção vai-se definir uma regra de quadratura que da mesma maneira que a quadratura de Gauss-Hermite, nos permite calcular integrais com uma função de peso gaussiana.

4.3. Quadraturas Genz-Keister

Em contraparte com as quadraturas de Gauss-Hermite, Alan Genz e Bradley Keister, em (GENZ, et al., 1996), construíram as quadraturas de Genz-Keister para integrais de regiões infinitas com funções de peso gaussiano. Os pontos para esta regra são determinados por extensões sucessivas da quadratura de Gauss-Hermite unidimensional de três pontos.

A regra Gauss-Hermite é notoriamente não aninhada, é dizer, utiliza pontos de integração completamente diferentes para cada nível de acurácia. Isto significa que o conjunto de pontos usados para o nível de acurácia i não são reusados para o nível de acurácia $i + 1$. Por tanto o custo computacional é maior par regras não aninhadas. Inicialmente foi estendida a quadratura de Gauss-Legendre no intervalo $[-1,1]$ na quadratura de Gauss-Patterson, descoberta por Patterson em (PATTERSON, 1968), anos depois Genz e Keister também estendem a quadratura Gauss-Hermite a uma regra aninhada. Os autores em (GENZ, et al., 1996), estendem o trabalho de Patterson para $(-\infty, \infty)$, como segui.

Considere a integral unidimensional,

$$G(f) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} f(x) dx \quad (4.19)$$

Denotamos por $Q^{(m)}$ a regra de quadratura para $G(f)$ com grau de exatidão polinomial $2m + 1$. Vamos a denotar por $Q^{(m)}[i_1 + i_2 + \dots + i_k]$ a regra para $G(f)$ com exatidão $2m + 1$ que utiliza $\sum_{j=1}^k i_j$ pontos, e foi construída por extensões sucessivas de regras de baixo grau de exatidão com $\sum_{j=1}^l i_j$ pontos, para $l = 1, 2, \dots, k$.

Inicialmente, considere para $G(f)$ a quadratura de Gauss-Hermite com um ponto $Q^{(0)}[1](f) = f(0)$, com grau de exatidão polinomial 1. Esta regra é estendida acrescentando dois pontos $\pm\lambda_1$ produzindo uma nova regra,

$$Q^{(2)}[1 + 2](f) = \omega_0^{(1)} f(0) + \omega_1^{(1)} f(\lambda_1) \quad (4.20)$$

Os pesos, $\omega_0^{(1)}$ e $\omega_1^{(1)}$, e λ_1 são determinados maximizando o grau de precisão polinomial de $Q^{(3)}$. A solução para isto é, $\lambda_1 = \sqrt{3}$, $\omega_0^{(1)} = 2/3$ e $\omega_1^{(1)} = 1/6$. Agora, note-se que $Q^{(2)}[1 + 2]$, é exatamente a regra de Gauss-Hermite, para $G(f)$ com 3 pontos e grau de precisão polinomial 5. Construindo sobre esta,

pôde-se encontrar outras regras novas. Assim, sucessivamente são calculadas extensões ótimas, completamente aninhadas, para ordens finitos,

- $Q^{(7)}[1 + 2 + 6]$: Regra com 9 pontos e grau de precisão polinomial 15.
- $Q^{(9)}[1 + 2 + 8]$: Regra com 11 pontos e grau de precisão polinomial 19.
- $Q^{(14)}[1 + 2 + 6 + 10]$: Regra com 19 pontos e grau de precisão polinomial 29.
- $Q^{(25)}[1 + 2 + 6 + 10 + 16]$: Regra com 35 pontos e grau de precisão polinomial 51.

Na seguinte Tabela 4.1 mostram-se os nodos e pesos para algumas ordens da quadratura de Genz-Keister.

Tabela 4.1: Nodos e pesos da Quadratura Genz-Keister.

Num. Nodos	Nodos: λ_i	Pesos: ω_i
n=1	0	1
n=3	-1.73205081	0.16666667
	0	0.66666667
	1.73205081	0.16666667
n=7	-4.18495602	0.00069568
	-1.73205081	0.13855327
	-0.74109535	0.13137861
	0	0.45874487
	0.74109535	0.13137861
	1.73205081	0.13855327
	4.18495602	0.00069568
n=9	-4.18495602	9.4269E-05
	-2.86127958	0.00799633
	-1.73205081	0.09485095
	-0.74109535	0.27007433
	0	0.25396825
	0.74109535	0.27007433
	1.73205081	0.09485095
	2.86127958	0.00799633
	4.18495602	9.4269E-05

Com esta regra de quadratura finalizamos as metodologias expostas para estimar integrais uni-variadas. As quadraturas aqui mencionadas serão utilizadas

nas próximas seções, para a construção de técnicas que permitem resolver problemas de integração multidimensional.

4.4. Quadraturas Multivariadas

Na maioria dos problemas que demandam resolver integrais, são problemas multi-variados, onde é preciso generalizar os métodos mencionados nas seções anteriores ou aplicar técnicas diferentes. Nesta seção são enunciadas duas metodologias que podem ser adoptadas para a resolução este tipo de problemas, a primeira delas é a composição de quadraturas uni-variadas. O segundo método proposto é chamado quadratura de Smolyak ou *sparse grid* e está orientado a problemas de alta dimensão onde é necessário ter em consideração o custo computacional envolvido na estimação.

4.4.1. Quadratura Produto Tensor

A composição de quadraturas unidimensionais é a técnica mais intuitiva para o cálculo de integrais multidimensionais. O nome deste método (quadratura produto tensor) deve-se a que a composição coordenada a coordenada de quadraturas uni-variadas conforma um produto tensor completo (URBAN, et al., 2009) (HOLTZ, 2010).

Suponha que estamos interessados em calcular integrais multidimensionais da forma,

$$I_W^d f = \int_{I_1 \times I_2 \times \dots \times I_d} W(z_1, \dots, z_d) f(z_1, \dots, z_d) dz_d \dots dz_1, \quad (4.21)$$

com $\emptyset \neq I_j \subseteq \mathbb{R}$

Onde I_W^d é uma integral d-dimensional, f é integrável no hiper-rectângulo $I_1 \times I_2 \times \dots \times I_d$ e a função de peso $W(z_1, \dots, z_d) = W(z_1) \dots W(z_d)$ é um produto de funções integráveis não negativas em $I_j, \forall j = 1, \dots, d$.

O método direto para calcular esta integral é compor coordenada a coordenada utilizando regras de quadraturas uni-variadas. Seja $(U_k^j)_{j=1}^d$ uma sequência de regras de quadraturas, com k nodos e pontos de integração, tais que $U_k g = I_{W_j}^1 g$ é mantida para todos os polinômios g de grau máximo m_k .

Sejam $(\omega_i^j)_{i=1}^k$ os pesos e $(x_i^j)_{i=1}^k$ os pontos de integração da quadratura U_k^j . Então, podemos aproximar a integral da equação (4.21) por,

$$\begin{aligned} I_W^d f &= \int_{I_1} \int_{I_2} \cdots \int_{I_d} W(z_1) \cdots W(z_d) f(z_1, \dots, z_d) dz_d \cdots dz_1 \\ &\approx \int_{I_1} \int_{I_2} \cdots \int_{I_{d-1}} W(z_1) \cdots W(z_{d-1}) \sum_{i_d=1}^k \omega_{i_d}^d f(z_1, \dots, x_{i_d}^d) dz_{d-1} \cdots dz_1 \\ &\approx \sum_{i_1=1}^k \cdots \sum_{i_d=1}^k \omega_{i_1}^1 \cdots \omega_{i_d}^d f(x_{i_1}^1, \dots, x_{i_d}^d) \end{aligned}$$

Note-se que este último resultado pode-se expressar como,

$$\mathcal{J}_W^d f = \bigotimes_{i=1}^d U_k^{(i)} f + error \quad (4.22)$$

O problema deste enfoque é o número de avaliações na função f que são necessárias para obter a aproximação. No caso de uma integral d-dimensional e utilizando uma sequência de regras de quadratura com k pontos de integração, é preciso fazer k^d avaliações, ou seja, o número de avaliações cresce exponencialmente com respeito à dimensão do problema. Este fenômeno é chamado *curse of dimensionality*, e é a razão pela qual se vai desenvolver a técnica de Smolyak na seguinte subseção.

4.4.2. Quadratura de Smolyak

Nesta subseção é exposta a fundamentação teórica e prática do uso da regra da quadratura de Smolyak (ou *sparse grid*), a qual é utilizada para avaliar integrais de alta dimensão sob regiões de espaços euclidianos. Esta técnica foi descrita pelo matemático russo, Smolyak, em 1963 (SMOLYAK, 1963) e sua construção tem sido aplicada em problemas de integração numérica por distintos autores, fazendo uso da regra do meio ponto (BASZENKI, et al., 1993), da regra trapezoidal (BONK, 1994) e das regras de Gauss (NOVAK, et al., 1999), como base das regras uni-variadas.

Dada uma sequência de regras de quadratura uni-variada, a construção de Smolyak é definida a partir do produto tensor (ver seção 2.4) tomado sobre diferenças consecutivas dessas quadraturas uni-variadas (GERSTNER, et al., 1998). Os pontos de avaliação resultante da quadratura multivariada construída

desta maneira, são distribuídos mais esparsamente que os pontos da quadratura produto tensor tradicional, conseguindo aproximar integrais com uma quantidade de pontos de avaliação menor. Nosso interesse em desenvolver esta técnica é a necessidade de estimar integrais multidimensionais na teoria de caos polinomial, estritamente num tempo e custo computacional baixo. Devido a que nossa área de aplicação é a simulação de reservatórios de petróleo, avaliar volve-se extremadamente caro. Seguidamente vamos definir esta regra de integração baseados na tese doutoral do Kaarnioja (KAARNIOJA, 2013) e o trabalho feito em (HOLTZ, 2010).

Considere a seguinte sequência de regras de quadratura uni-variadas, para $m_k, k \in \mathbb{N}$,

$$U_{m_k} f := \sum_{i=1}^{m_k} \omega_{i,k} f(x_{i,k}) \quad (4.23)$$

onde $\omega_{i,k}$ e $x_{i,k}$ são os m_k pesos e nodos da quadratura U_{m_k} .

Introduzimos o seguinte operador diferença,

$$\begin{aligned} \Delta_k &:= U_{m_k} - U_{m_{k-1}}, \text{ para } k = 1, 2, 3, \dots \\ e \quad U_{m_0} &:= 0 \end{aligned} \quad (4.24)$$

Quando $k \in \mathbb{N}^d$ é um multi-índice, com $k_j > 0$, então

$$\Delta_k := (\Delta_{k_1} \otimes \dots \otimes \Delta_{k_d}) \quad (4.25)$$

Definição 4.1 (Regra de Quadratura de Smolyak). A regra de quadratura de Smolyak de nível ℓ no hiper-retângulo $I_1 \times I_2 \times \dots \times I_d$ para uma integral d -dimensional é o operador,

$$Q_\ell^d = \sum_{\substack{|k|_1 \leq \ell + d - 1 \\ k \in \mathbb{N}^d}} \bigotimes_{i=1}^d \Delta_{k_i} \quad (4.26)$$

Para $k_i > 0$ e $\ell \geq 1$. Onde $|k|_1 = k_1 + \dots + k_d$.

Também é comum encontrar a definição da quadratura de Smolyak com uma ordem $\sigma = \ell + d - 1$.

Observe na definição anterior, que a regra de Smolyak no caso de uma dimensão é a regra de quadratura uni-variada considerada. Assim,

$$Q_\ell^1 = \sum_{i=1}^{\ell+1-1} \Delta_i = (U_{m_1} - U_{m_0}) + (U_{m_2} - U_{m_1}) + (U_{m_3} - U_{m_2}) \dots + (U_{m_\ell} - U_{m_{\ell-1}}) = U_{m_\ell} \quad (4.27)$$

No próximo exemplo vamos desenvolver a construção da quadratura de Smolyak num caso de duas dimensões, exposto em (HOLTZ, 2010).

Exemplo 4.1: (HOLTZ, 2010) Considere uma sequência de três regras de quadratura trapezoidal aberta (QUARTERONI, et al., 2010), com $m_k = 2^k - 1$ nodos. Temos assim, uma sequência de regras com 1, 3 e 7 pontos. Então a quadratura de Smolyak baseada nessa regra de quadratura, com $\ell = 3$ está dada por,

$$\begin{aligned} Q_3^2 &= \sum_{\substack{|k|_1 \leq 3+2-1 \\ k \in \mathbb{N}^2}} \Delta_{(k_1, k_2)} \\ &= \Delta_{(1,1)} + \Delta_{(1,2)} + \Delta_{(1,3)} + \Delta_{(2,1)} + \Delta_{(2,2)} + \Delta_{(3,1)} \\ &= (\Delta_1 \otimes \Delta_1) + (\Delta_1 \otimes \Delta_2) + (\Delta_1 \otimes \Delta_3) + (\Delta_2 \otimes \Delta_1) + (\Delta_2 \otimes \Delta_2) + (\Delta_3 \otimes \Delta_1) \\ &= (U_{m_1} \otimes U_{m_3}) + (U_{m_2} \otimes U_{m_2}) - (U_{m_1} \otimes U_{m_2}) + (U_{m_3} \otimes U_{m_1}) - (U_{m_2} \otimes U_{m_1}) \end{aligned}$$

Note-se que, inicialmente, pelo nível da quadratura de Smolyak, são consideradas as quadraturas uni-variadas com até $m_\ell = 2^\ell - 1 = 7$ nodos ou pontos de integração. Portanto, se fizéramos uma composição de quadraturas uni-variadas com 7 pontos, tenderíamos uma regra $U_{m_3}f \otimes U_{m_3}f$ com $7^2 = 49$ valores para ser avaliados na função. Por outro lado, Smolyak, considere somente alguns dos produtos tensores das regras e consegue aproximar a integral com somente $(1 \times 7) + (3 \times 3) - (1 \times 3) + (7 \times 1) - (3 \times 1) = 17$ pontos. Em Figura 4.1, imagem obtida de (HOLTZ, 2010), pode-se observar este resultado.

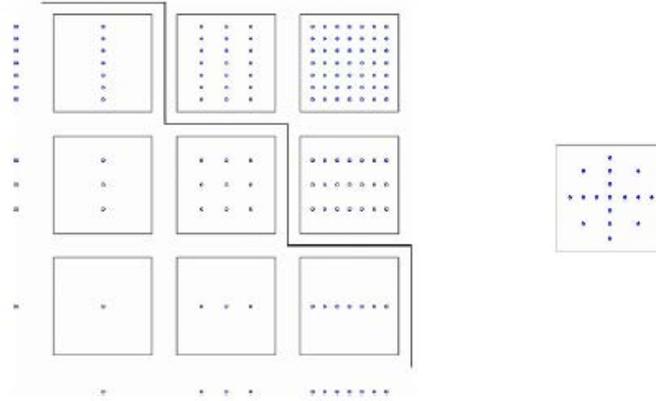


Figura 4.1: Gride completo (esquerda) e o gride de Smolyak (direita), baseados em regras de quadratura trapezoidal de 1, 2 e 7 pontos, tomando de (HOLTZ, 2010).

Uma pergunta óbvia que surge na construção desta quadratura, é sua exatidão polinomial. É dizer, quando podemos ter uma equivalência desta regra com a eq. (4.21), assim

$$I_W^d f = Q_\ell^d \quad (4.28)$$

Sabemos que uma quadratura gaussiana com n pontos de integração, estima com total precisão polinômios de grau máximo $2n - 1$. Por outro lado, foi evidenciado que a regra de Smolyak precisa uma quantidade consideravelmente menor de pontos de avaliação.

Então, no seguinte teorema é mostrado que a característica de precisão nas quadraturas uni-variadas, é trasladada com a mesma capacidade para a regra de Smolyak.

Teorema 4.4: Seja U_i uma sequência de regras gaussianas uni-variadas correspondentes à função de peso W na eq. (4.21). Pode-se afirmar que,

$$I_W^d f = Q_\ell^d \quad \text{for all } f \in \mathbb{P}_{2\ell-1}^d \quad (4.29)$$

Onde \mathbb{P}_k^d é o espaço de polinômios multi-variados,

$$\mathbb{P}_k^d = \left\{ \mathbb{R}^d \ni x \mapsto \sum_{\substack{|\beta|_1 \leq k \\ \beta \in \mathbb{N}^d}} a_\beta x^\beta \in \mathbb{R}; \quad a_\beta \in \mathbb{R}, \forall \beta \in \mathbb{N}^d \right\} \quad (4.30)$$

O teorema anterior nos indica que, o grau de exatidão polinomial para uma regra de Smolyak baseada em gauss, com nível ℓ , é $2\ell - 1$.

Para maior detalhe, Novak e Ritter trabalharam no grau de exatidão polinomial da regra de Smolyak em (NOVAK, et al., 1999). A convergência pode ser estudada em (KAARNIOJA, 2013).

5 Expansões por Caos Polinomial

5.1. Introdução

Um problema usual presente em distintas áreas da mecânica, física, finanças, entre outras, é querer estudar a distribuição de probabilidade de uma certa variável aleatória, que representa a resposta de um modelo (chamemos de modelo, a um conjunto de equações que representam um fenômeno). Tal distribuição de probabilidade, na maioria dos casos, não é conhecida e as respostas do modelo não têm uma representação analítica explícita, devido à complexidade dos fenômenos subjacentes (como os modelos do tipo *black-box*). No entanto, se são conhecidas as densidades de probabilidade dos parâmetros de entrada, é possível realizar uma análise de como é propagada a aleatoriedade através do modelo e quantificar a incerteza na saída. Na Figura 5.1 é apresentado um esquema da propagação de incertezas num modelo estocástico $f(Z) = Y$.

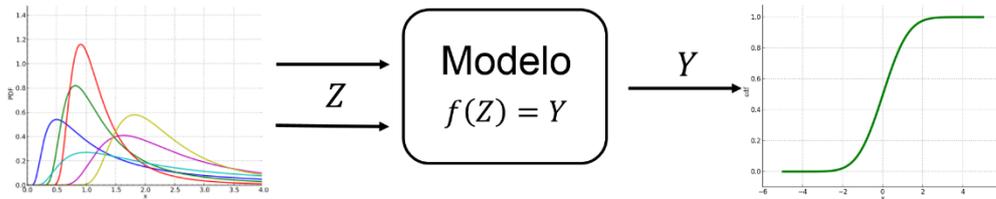


Figura 5.1: Esquema de propagação de incerteza.

O método mais conhecido para tratar este problema é o método de simulação de Monte Carlo. Porém, a grande desvantagem desta técnica encontra-se na sua lenta convergência, i.e., este método tem um custo computacional alto. Portanto, para modelos complexos, como aqueles tratados nesta dissertação, o seu uso é inviável. Por outro lado, Expansões por Caos Polinomial (PCE) é uma técnica de convergência rápida (ERNST, et al., 2012), que aproxima a resposta de um modelo através de uma combinação linear de polinômios, que são ortogonais com respeito à densidade de probabilidade dos parâmetros aleatórios de entrada do modelo.

5.2. Expansões por Caos Polinomial Generalizado

Expansões por Caos Polinomial é uma técnica baseada no conceito de Wiener do caos homogêneo (WIENER, 1938), onde o termo caos refere-se à aleatoriedade nos parâmetros envolvidos em diferentes processos físicos ou mecânicos. A base teórica desta aproximação está nas conclusões de Cameron y Martin (CAMERON, et al., 1947), nas que as expansões de caos polinomial convergem para uma funcional L^2 no espaço de Hilbert. No contexto dos processos estocásticos, isto implica que cada processo estocástico com momento de segunda ordem finito, pode-se representar por uma expansão por caos polinomial (infinita). A forma original da PCE é uma expansão espectral, baseada nos polinômios de Hermite ortogonais em termos de variáveis aleatórias gaussianas e utilizando coeficientes determinísticos. Mas no entanto, ao utilizar esta forma original, a convergência ótima consegue-se somente quando se trabalha com processos estocásticos gaussianos.

Para obter uma convergência ótima para processos estocásticos mais gerais, Xiu e Karniadakis em (XIU, et al., 2002) estenderam esta aproximação a um esquema mais amplo, chamando *expansões por caos polinomial generalizado*. Neste esquema é utilizada a conexão fechada entre as funções de probabilidade de algumas variáveis aleatórias e a função de peso na relação de ortogonalidade de alguns polinômios ortogonais, para representar os processos não gaussianos. Isto significa que, a convergência ótima pode-se conseguir para processos caracterizados por outras distribuições não gaussianas, como distribuição uniforme, gamma e beta. Vamos no resto da seção, definir formalmente este conceito de PCE.

Considere um sistema físico ou mecânico descrito por um modelo f , para nosso caso um modelo de simulação. Suponha que este modelo tem M parâmetros de entrada incertos, representados por um vetor $Z = (Z_1, Z_2, \dots, Z_M)$ de componentes mutuamente independentes e distribuição $F_Z(z_1, \dots, z_M) = P(Z_1 \leq z_1, \dots, Z_M \leq z_M)$.

Para cada $i = 1, \dots, M$, seja $F_{Z_i}(z_i) = P(Z_i \leq z_i)$ a distribuição marginal de Z_i no suporte I_{Z_i} (XIU, 2010). Devido à independência temos que,

- $F_Z(z) = F_{Z_1}(z_1) \cdot F_{Z_2}(z_2) \cdot \dots \cdot F_{Z_M}(z_M)$

$$\blacksquare I_Z = I_{Z_1} \times I_{Z_2} \times \dots \times I_{Z_M}$$

Seja,

$$L^2_{dF_Z}(I_Z) = \left\{ f: I_Z \rightarrow \mathbb{R} \mid \mathbb{E}[f^2(Z)] = \int_{I_Z} f^2(z) dF_Z(z) < \infty \right\} \quad (5.1)$$

O espaço de todas as funções integráveis em média quadrática com norma $\|f\|_{L^2_{dF_Z}} = (\mathbb{E}[f^2])^{1/2}$. Então, se a resposta do modelo $f \in L^2_{dF_Z}(I_Z)$, podemos definir sua representação por caos polinomial como,

$$Y = f(Z) = \sum_{\beta \in \mathbb{N}^M} a_\beta \phi_\beta(Z) \quad (5.2)$$

Onde, a_β são os coeficientes da expansão, $\beta \in \mathbb{N}^M$ são os multi-índices, com $|\beta|_1 = \beta_1 + \beta_2 + \dots + \beta_M$ e $\{\phi_\beta\}_{\beta \in \mathbb{N}^M}$ é a base de polinômios ortogonais multi-variados que satisfazem,

$$\blacksquare \phi_\beta(Z) = \phi_{\beta_1}(Z_1) \cdot \phi_{\beta_2}(Z_2) \cdots \phi_{\beta_M}(Z_M) \quad (5.3)$$

$$\blacksquare \mathbb{E}[\phi_\beta(Z)\phi_\alpha(Z)] = \int \phi_\beta(Z)\phi_\alpha(Z)dF_Z = \gamma_\beta \delta_{\beta\alpha} \quad (5.4)$$

Onde $\gamma_\beta = \mathbb{E}[\phi_{\beta_1}^2(Z_1)] \cdots \mathbb{E}[\phi_{\beta_M}^2(Z_M)]$ são fatores de normalização e $\delta_{\beta\alpha}$ é a delta de Kronecker ($\delta_{\beta\alpha} = 1$ se $\beta = \alpha$ e 0 em outro caso).

A ortogonalidade na equação (5.4) garante que os polinômios podem ser utilizados como uma base de funções para aproximar funções em termos das variáveis aleatórias Z_i .

Como foi mencionado, as expansões por caos polinomial foram utilizadas originalmente para funções dependentes de variáveis aleatórias gaussianas, mas, depois do trabalho de Xiu e Karniadakis existem correspondências entre as distribuições de probabilidade dos parâmetros e sua base de funções ortogonais.

Na Tabela 5.1 são apresentadas algumas destas correspondências.

Tabela 5.1: Correspondência entre o tipo de distribuição e a base de polinômios.

Distribuição de probabilidade de Z	Base de polinômios ortogonais da PCE	Suporte
Gaussiana	Hermite	$(-\infty, \infty)$
Gamma	Laguerre	$[0, \infty)$
Beta	Jacobi	$[a, b]$
Uniforme	Legendre	$[a, b]$

Na prática, a expansão infinita da equação (5.2) é truncada em um número finito de termos e uma ordem finita de expansão. Nesse caso, podemos falar de uma PCE de grau p , onde p é o maior grau dos polinômios que representam as variáveis incertas. A expansão é truncada depois de P termos, onde P é definido por,

$$P = \binom{M+p}{p} = \frac{(M+p)!}{M!p!} \quad (5.5)$$

Assim, podemos aproximar a resposta do modelo por,

$$Y \approx f_p(Z) = \sum_{\substack{|\beta|_1 \leq p \\ \beta \in \mathbb{N}^M}} a_\beta \phi_\beta(Z) = \sum_{k=1}^P a_k \phi_k(Z) \quad (5.6)$$

Onde a correspondência dos β e os k pode ser observada num exemplo de dimensão quatro, na seguinte Tabela 5.2 (XIU, 2010),

Tabela 5.2: Bijeção entre os índices β e k da soma truncada na equação (5.6).

$ \beta _1$	Multi-índice β	Índice associado k
0	(0,0,0,0)	1
1	(1,0,0,0)	2
	(0,1,0,0)	3
	(0,0,1,0)	4
	(0,0,0,1)	5
	(2,0,0,0)	6
2	(1,1,0,0)	7
	(1,0,1,0)	8
	(1,0,0,1)	9
	(0,2,0,0)	10
	(0,1,1,0)	11
	(0,1,0,1)	12
	(0,0,2,0)	13
	(0,0,1,1)	14
	(0,0,0,2)	15
	3	(3,0,0,0)
(2,1,0,0)		17
(2,0,1,0)		18
...

Por outro lado, a quantidade de termos e coeficientes da expansão depende diretamente do número de parâmetros considerados no modelo e do grau da expansão desejada. Esta relação pode ser observada na Tabela 5.3, que mostra a

quantidade de coeficientes da PCE em função do número de parâmetros incertos e do grau maior dos polinômios que representam as variáveis. Note-se que o número de coeficientes cresce rapidamente quando aumenta o número de variáveis ou o grau da expansão desejada.

Tabela 5.3: Número de coeficientes na expansão por caos polinomial.

Grau da PCE (p)	Número de parâmetros incertos no modelo (M)								
	2	3	4	5	10	11	15	25	50
2	6	10	15	21	66	78	136	351	1326
3	10	20	35	56	286	364	816	2376	23426
4	15	35	70	126	1001	1365	3876	23751	316251
5	21	56	126	252	3003	4368	15504	142506	3478761

A maior dificuldade na construção da PCE está no cálculo ou na maioria dos casos, na estimação, dos coeficientes. Várias abordagens têm sido desenvolvidas para obter estes valores e têm sido classificadas em duas categorias, as técnicas *intrusivas* e as *não intrusivas*. No caso, dos métodos intrusivos, é necessário ter um alto grau de conhecimento do modelo, para fazer modificações algébricas no sistema de equações que lhe define, ou alterações no código de computador, um exemplo desta técnica é a projeção de Galerkin. Por outro lado, o método não-intrusivo é um esquema emergente do análise de elementos finitos estocásticos. Esta última técnica está sendo utilizada na análise de modelos cuja resposta está definida por meio de um conjunto de chamadas a um modelo determinístico existente, sem fazer modificações no código de computador (BLATMAN, et al., 2010) (HAGAN, 2013). Neste trabalho vamos desenvolver duas técnicas do tipo não intrusivas, devido a que nosso modelo é um sistema do tipo caixa preta (*black box*).

Nas seguintes seções vamos enunciar duas metodologias não intrusivas para estimar os coeficientes da expansão, a primeira é uma aproximação mediante a projeção espectral ortogonal e a segunda é uma aproximação mediante a regressão linear.

5.3. Aproximação por Projeção Espectral Ortogonal

Os coeficientes da expansão podem ser calculados fazendo um forte uso da ortogonalidade da base de polinômios que representam as variáveis aleatórias.

Para toda função $f \in L^2_{dF_Z}(I_Z)$ podemos definir a *Projeção Ortogonal Espectral de ordem p* , que vamos a denotar por $P_p f$ e vamos definir por (XIU, 2010),

$$P_p f(Z) = \sum_{\substack{|\beta|_1 \leq p \\ \beta \in \mathbb{N}^M}} a_\beta \phi_\beta(Z) \quad (5.7)$$

Onde,

$$a_\beta = \frac{\mathbb{E}[f \phi_\beta]}{\gamma_\beta} = \frac{1}{\gamma_\beta} \int f(z) \phi_\beta(z) dF_Z(z), \quad \forall |\beta|_1 \leq p \quad (5.8)$$

Lembre-se que $\gamma_\beta = \mathbb{E}[\phi_{\beta_1}^2(Z_1)] \cdots \mathbb{E}[\phi_{\beta_M}^2(Z_M)]$.

Note-se que estes coeficientes podem ser obtidos multiplicando a equação (5.6) por $\phi_\beta(Z)$ e tomando a média, devido à ortogonalidade da base de polinômios pode-se obter a expressão (5.8).

Por outro lado, a existência e convergência da projeção, se segue diretamente da teoria de probabilidade clássica. A convergência da projeção no sentido da média quadrática se pode expressar como,

$$\|f - P_p f\|_{L^2_{dF_Z}} \rightarrow 0, \quad \text{quando } p \rightarrow \infty \quad (5.9)$$

Dos resultados mais importantes deste enfoque está o seguinte. Dado um modelo f , se desejamos aproximar esse modelo por um polinômio de grau p , então a projeção espectral ortogonal é o polinômio de grau p que melhor aproxima o modelo f . Assim,

$$\|f - P_p f\|_{L^2_{dF_Z}} = \inf_{g \in \mathbb{P}_p^M} \|f - g\|_{L^2_{dF_Z}} \quad (5.10)$$

Onde \mathbb{P}_N^d é o espaço de todos os polinômios d -dimensionais de $Z \in \mathbb{R}^d$ com grau máximo N .

Note na equação (5.8) que os coeficientes são definidos mediante uma integral multidimensional que depende de f , portanto faz necessário utilizar técnicas de integração numérica para conseguir estimar estes valores. Para casos de baixa dimensão ($M \leq 3$) e modelos simples, estas integrais podem ser

calculadas mediante métodos modestos, mas, quando o número de parâmetros envolvidos no modelo aumenta e o modelo por se mesmo é complexo, é necessário acudir a técnicas de integração mais robustas ou especializadas em integração de alta dimensão. O motivo de isto vem do custo envolvido na avaliação da função e a necessidade de realizar estas aproximações com um tempo e custo computacional baixo.

Nesta dissertação propomos utilizar como método de integração numérica, a Quadratura de Smolyak (enunciado na seção 4.4.2) para a estimação do numerador dos coeficientes a_β na equação (5.8) e a Quadratura de Gauss para o denominador γ_β que não depende do simulador f e além disso, é um produto de integrais unidimensionais.

De esta maneira, a equação (5.8) pode ser reescrita como,

$$a_\beta = \frac{\mathbb{E}[f\phi_\beta]}{\gamma_\beta} \approx \frac{Q_\ell^M}{\prod U_m} \quad (5.11)$$

Onde U_m é a regra de quadratura gaussiana uni-variada, selecionada de acordo à função distribuição de probabilidade dos parâmetros de entrada do modelo f . Esta regra permitirá estimar a integral unidimensional do denominador dos coeficientes da expansão. Note-se que, a quantidade de nodos considerados para aproximar γ_β dependerá do grau máximo dos polinômios da base ortogonal da PCE selecionada.

Por outro lado, o denominador dos coeficientes, depende do modelo de simulação e está definido por meio de uma integral M -dimensional. Por tanto, vai ser aproximado por uma quadratura de Smolyak Q_ℓ^M com nível ℓ . O valor do nível da regra de quadratura depende do grau selecionado da PCE.

No caso que o modelo de simulação considerado tenha uma quantidade baixa de parâmetros, é possível aproximar os coeficientes da PCE utilizando unicamente Quadraturas Gaussianas.

Observe que, para o enfoque da projeção espectral ortogonal, o custo computacional de aproximar um modelo f por meio de uma PCE, está diretamente relacionado com a técnica de integração numérica selecionada para estimar os coeficientes. O custo computacional das técnicas de integração numérica, mencionadas no capítulo 3, depende da quantidade de nodos considerados, é dizer, do número de avaliações na função.

Em Tabela 5.4 e Tabela 5.5 vamos mostrar a relação existente entre o número de avaliações e o nível (acurácia) ou número de nodos considerado em três diferentes regras de quadratura. Supondo que as variáveis aleatórias do modelo, seguem uma distribuição de probabilidade gaussiana, vamos considerar as regras de quadratura Gauss-Hermite (GH), Smolyak baseada em Gauss-Hermite (SGH) e Smolyak baseada em Genz-Keister (SGK).

Tabela 5.4: Correspondência entre o número de avaliações e o nível para as regras GH, SGH e SGK com dimensão 7.

$M = 7$	Número de Avaliações		
Nível: ℓ	Gauss-Hermite	Smolyak baseada em Gauss-Hermite	Smolyak baseada em Genz-Keister
1	1	1	1
2	128	15	15
3	2187	113	99
4	16,384	589	407
5	78,125	2,437	1,317
6	279,936	8,583	3,837
7	823,543	26,769	9,941
8	2,097,969	75,841	22,725
9	4,782,969	198,649	48,401
10	10,000,000	487,175	96,967

Tabela 5.5: Correspondência entre o número de avaliações e o nível para as regras GH, SGH e SGK com dimensão 15.

$M = 15$	Número de Avaliações		
Nível: ℓ	Gauss-Hermite	Smolyak baseada em Gauss-Hermite	Smolyak baseada em Genz-Keister
1	1	1	1
2	32,768	31	31
3	14,348,907	481	451
4	1,073,741,824	5,021	4,151
5	30,517,578,125	39,941	27,701
6	470,184,984,576	259,607	146,477
7	4,747,561,509,943	1,441,297	654,117
8	35,184,372,088,832	7,045,537	2,571,477

Observe a diferença do número de avaliações nas três regras de integração, este valor incrementa conforme aumenta o nível. A quadratura produto tensor com Gauss-Hermite, cresce exponencialmente com respeito à dimensão e o número de nodos ($nodos = \rho^M$), ver seção 4.4.1. Por outro lado, a quadratura de Smolyak construí um conjunto de pontos de integração menor e melhor distribuídos, ver seção 4.4.2. No caso de Genz-Keister, devido a que é uma quadratura aninhada, tem um número ainda menor de pontos para avaliar, se comparado com SGH, sendo assim, a melhor escolha considerando o custo computacional.

5.4. Aproximação por Regressão

Outra alternativa não intrusiva para estimar os coeficientes da PCE na equação (5.6) é obter a melhor aproximação de $Y = f(Z)$ no sentido de mínimos quadrados, para um grau fixo p . Este método tem sido proposto recentemente (BLATMAN, et al., 2008), (BLATMAN, et al., 2010), (SUDRET, et al., 2013), sendo chamado método da regressão.

Considere a seguinte notação vetorial para a equação (5.6),

$$Y \approx f_p(Z) = \sum_{\substack{|\beta|_1 \leq p \\ \beta \in \mathbb{N}^M}} a_\beta \phi_\beta(Z) = \sum_{k=1}^P a_k \phi_k(Z) = a^T \phi(Z) \quad (5.12)$$

Onde,

- a é o vetor coluna de coeficientes da PCE, tamanho $P \times 1$:

$$a = \{a_\beta: |\beta|_1 \leq p\} = (a_1, a_2, \dots, a_P)^T$$

- ϕ é o vetor linha de polinômios ortogonais da base da PCE, tamanho $1 \times P$:

$$\phi = \{\phi_\beta: |\beta|_1 \leq p\} = (\phi_1, \phi_2, \dots, \phi_P)$$

Seja $\mathcal{Z} = (z^{(1)}, z^{(2)}, \dots, z^{(N)})$ um conjunto de N realizações ou amostras para o vetor aleatório de entradas e seja $\mathcal{Y} = (y^{(1)}, y^{(2)}, \dots, y^{(N)})^T$ suas correspondentes saída através do modelo f . À coleção \mathcal{Z} é chamada *Planejamento de Experimentos* (PE). Este conjunto de valores pode ser construído desde uma amostragem aleatório (SUDRET, et al., 2013), mediante a técnica do Hipercubo Latino (CHOI, 2004) ou por meio das raízes dos polinômios ortogonais que foram considerados para a construção da base da PCE (BERVEILLER, et al., 2006)

(SUDRET, 2008). Mais na frente vamos desenvolver esta última técnica como metodologia proposta nesta dissertação.

O problema considerado consiste em encontrar o vetor dos coeficientes \hat{a} que minimizam a soma dos erros quadráticos, é dizer,

$$\hat{a} = \arg \min_a \sum_{i=1}^N (f(z^{(i)}) - a^T \phi(z^{(i)}))^2 \quad (5.13)$$

Esta minimização tem uma resposta exata, para o conjunto de amostras Z ao resolver o seguinte sistema de equações:

$$(\Phi^T \Phi) \hat{a} = \Phi^T \gamma \quad (5.14)$$

Ou equivalentemente,

$$\hat{a} = (\Phi^T \Phi)^{-1} \Phi^T \gamma \quad (5.15)$$

Onde Φ é uma matriz de tamanho $N \times P$ definida por,

$$\Phi_{ij} = (\phi_j(z^{(i)}))_{\substack{i=1,\dots,N \\ j=1,\dots,P}} \quad (5.16)$$

Onde a matriz $\Phi^T \Phi$ é chamada *matriz de informação*.

Observe-se que a matriz Φ está construída a partir do conjunto das realizações Z . Portanto, a estimação dos coeficientes depende diretamente do planejamento de experimentos considerado. Além disso, note-se que, o sistema de equações em (5.14) tem solução se $N \geq P$ (BERVEILLER, et al., 2006), portanto o número de experimentos considerados par estimar os coeficientes da expansão é maior o igual a P .

Surgem então duas dúvidas, a primeira é a maneira mais eficiente de gerar o conjunto de amostras e a segunda é o tamanho desse conjunto. No contexto dos métodos baseados em PCE, tem sido mostrado por Isukapalli e Berveiller em (ISUKAPALLI, 1999) e (BERVEILLER, 2005) que um planejamento de experimentos eficiente pode ser construído desde as raízes dos polinômios ortogonais da base da PCE.

Para o caso de variáveis aleatórias gaussianas, são consideradas as raízes dos polinômios de Hermite e o planejamento é gerado como seguiu (SUDRET, 2008),

- Se p denota o grau máximo dos polinômios da PCE truncada, então considere as $p + 1$ raízes do polinômio de Hermite de grau $p + 1$, denotado por H_{p+1} . Então temos o conjunto $\{r_1, r_2, \dots, r_{p+1}\}$.

- Construía todos os vetores de tamanho $1 \times M$ de combinações destas raízes. Assim, $r = (r_{i_1}, r_{i_2}, \dots, r_{i_M})$ onde $1 \leq i_1 \leq \dots \leq i_M \leq p + 1$. Note que tem-se em total $(p + 1)^M$ vetores de combinações.
- Finalmente, selecione os N vetores do total que tem menor norma euclidiana, ou seja, os vetores que maximizam a função distribuição de probabilidade gaussiana. Portanto, são as amostras com uma probabilidade de incidência maior e assim representam o comportamento mais usual da saída do modelo.

Por outro lado, a quantidade de amostras consideradas tem sido proposta por diferentes autores, Webster seleciona $N = P + 1$ em (WEBSTER, et al., 1996), Isukapalli propoe $N = 2(P + 1)$ em (ISUKAPALLI, 1999) e Berveiller considera $N = (M - 1)P$ ou $N = 2P$ em (BERVEILLER, 2005), mas todas as sugestões são baseadas em experimentos empíricos.

Investigações recentes mostram que o número de amostras requeridos no planejamento de experimentos está relacionado com a invertibilidade da matriz de informação $\Phi^T \Phi$. Sudret em (SUDRET, 2008) propõe o seguinte algoritmo,

- Considere o conjunto de vetores das raízes dos polinômios de Hermite, ordenados de maneira crescente pela norma euclidiana, $\{r^{(1)}, r^{(2)}, \dots, r^{(p+1)^M}\}$.
- Expresse a matriz de informação como uma soma de matrizes quadradas de tamanho $P \times P$ e de posto unitário, assim,

$$(\Phi^T \Phi)_{ij} = \sum_{k=1}^N \phi_i(r^{(k)}) \phi_j(r^{(k)}) = \sum_{k=1}^N a^k \quad (5.17)$$

Onde,

$$a_{ij}^k = \phi_i(r^{(k)}) \phi_j(r^{(k)}) \quad (5.18)$$

- Some o número de matrizes a^k necessárias para que a matriz de informação $\Phi^T \Phi$ tenha posto P . Note-se que é preciso somar pelo menos $N = P$.
- O tamanho do planejamento de experimentos é o número de matrizes que foram necessárias somar, para que a matriz de informação seja invertível.

Observe-se que, este algoritmo consegue minimizar a quantidade de amostras consideradas no planejamento de experimentos. Sudret em (SUDRET, 2008) mostra que esta técnica é mais eficiente em termos de custo computacional que as outras estratégias propostas.

Na próxima seção vamos fazer uso do polinômio construído mediante caos polinomial, para quantificar a incerteza na saída do modelo considerado. Esta análise comumente é feita calculando a média, a variância, alguns percentis ou mais rigorosamente, a distribuição de probabilidade completa. Vamos indicar na seguinte seção como é possível utilizar a PCE para obter estes estatísticos.

5.5. Análise Estatística e Probabilística

Quando a acurácia da PCE calculada para $f(Z)$ é satisfatória, tem-se em efeito uma representação analítica para o modelo f em termos de Z . Esta expressão pode ser utilizada para realizar estudos estatísticos e probabilísticos.

Uma das maiores vantagens da PCE é poder calcular diretamente da expansão polinomial, e com um esforço computacional mínimo a média e a variância.

Seja $f_p(Z)$ a PCE de grau p que aproxima a f i.e. $f_p \approx f$ no sentido da média quadrática. Então a média de f pode ser aproximada por (XIU, 2010),

$$\mu_f \triangleq \mathbb{E}[f(Z)] \approx \mathbb{E}[f_p] = \int \left(\sum_{|\beta|_1 \leq p} a_\beta \phi_\beta(Z) \right) dF_Z(z) = a_0 \quad (5.19)$$

A variância de f pode-se aproximar por,

$$\text{var}(f(Z)) = \mathbb{E}[(f(Z) - \mu_f)^2] \approx \sum_{0 < |\beta|_1 \leq p} \gamma_\beta a_\beta^2 \quad (5.20)$$

Por outro lado, os percentis de f (ver Definição 2.2) ou mais rigorosamente, a distribuição de probabilidade inteira (ver Definição 2.1) podem ser aproximados após uma avaliação numérica sobre f_p . Para a avaliação numérica, deve ser fornecida uma quantidade de amostras que serão avaliadas na expansão. Definimos a continuação o processo para estimar estes valores (KO, et al., 2013).

Seja $\{\xi_1, \xi_2, \dots, \xi_N\}$ um conjunto de N amostras geradas independentemente de acordo à função densidade de Z . A função distribuição de probabilidade de $Y = f(z)$ pode ser estimada utilizando a expansão f_p ,

$$\hat{F}_p(y) = \frac{1}{N} \sum_{k=1}^N \mathbf{1}_{(-\infty, y]}(Y_{k,p}) \quad (5.21)$$

Onde $Y_{k,p} = f_p(\xi_k)$ e $\mathbf{1}_{(-\infty, y]}$ é a função indicadora no intervalo $(-\infty, y]$. A solução $\hat{F}_p(y)$ é chamada estimação empírica de cdf de Y .

O α -percentil de Y pode ser aproximado desde o conjunto ordenado:

$$Y_{(1),p} \leq Y_{(2),p} \leq \dots \leq Y_{(k),p}$$

$$\hat{Y}_{\alpha,p} = Y_{([\alpha N], p)} \quad (5.22)$$

O interesse em saber estes valores, está na necessidade de conhecer a probabilidade com que a saída do modelo estudado vai ficar num intervalo determinado. No caso de modelos de simulação de reservatórios de petróleo, a estimação da *cdf* pode mostrar com que probabilidade a produção acumulada de óleo, por exemplo, vai ser menor a um valor predefinido, num período de tempo. Estas informações são importantes para a tomada de decisões no gerenciamento de reservatórios.

5.6. Tratamento de variáveis com múltiplas distribuições

Na maioria dos problemas reais que envolvem parâmetros aleatórios, a incerteza presente em cada um deles é modelada por diferentes distribuições de probabilidade. Isto faz necessário realizar um tratamento às variáveis, anterior a PCE, para transformar o vetor de parâmetros de entrada em um vetor de variáveis identicamente distribuídas.

Seja $Z = (Z_1, Z_2, \dots, Z_M)$ o vetor de entradas com componentes mutuamente independentes, definimos o seguinte mapeamento um-para-um (BERVEILLER, et al., 2006),

$$\xi_i = \phi^{-1}(F_i(Z_i)) \quad (5.23)$$

Onde, ϕ é a função distribuição de probabilidade (cdf) da gaussiana padrão e $\{F_i(Z_i), i = 1, \dots, M\}$ são as distribuições marginais das Z_i . Desta forma é construído o vetor $\xi = (\xi_1, \xi_2, \dots, \xi_M)$ de entradas independentes gaussianas padrão.

Na próxima seção vamos definir uma medida usualmente utilizada para quantificar o erro da aproximação por PCE.

5.7. Quantificação do Erro de Aproximação

Na prática, é usual calcular o seguinte erro empírico, para quantificar o erro de aproximação da expansão por caos polinomial truncada (BLATMAN, et al., 2010) (SUDRET, et al., 2013),

$$R^2 = 1 - \frac{1/N \sum_{i=1}^N (f(z^{(i)}) - f_p(z^{(i)}))^2}{\widehat{\text{var}}(Y)} \quad (5.24)$$

onde

- $\widehat{\text{var}}(Y)$ é a variância empírica de Y calculada por,

$$\widehat{\text{var}}(Y) = \frac{1}{N-1} \sum_{i=1}^N (f(z^{(i)}) - \bar{f})^2, \quad \bar{f} = \frac{1}{N} \sum_{i=1}^N f(z^{(i)}) \quad (5.25)$$

- $\{z^{(1)}, z^{(2)}, \dots, z^{(N)}\}$ é um conjunto de amostras geradas com a distribuição de probabilidade de Z e é utilizado para avaliar o rendimento da PCE.

A acurácia da aproximação f_p pode ser avaliada desde o significado do *coeficiente de determinação* R^2 . Quando $R^2 = 1$ indica que o modelo conseguiu um ajuste perfeito ao modelo real com resposta Y .

Finalizamos este capítulo reforçando a ideia que PCE é uma técnica que permite quantificar a incerteza, na resposta de um modelo afetado pela aleatoriedade presente nas variáveis de entrada. Na área de petróleo, o problema de quantificação de incertezas consiste na estimação de estatísticas (média, variância, outros momentos) de parâmetros de interesse, calculadas numa etapa de pós-processamentos e a partir dos resultados de um simulador numérico de reservatórios de petróleo, considerando que algumas das variáveis envolvidas nesse modelo de simulação sejam incertas. Uma vez que é aproximada a saída, análises estatísticas de interesse podem ser obtidas com métodos de amostragem na função estimada. Caos polinomial é uma técnica que permite ter algumas estatísticas diretamente y para outros estudos, é possível utilizar a saída aproximada como um modelo substituto (*proxy*).

No próximo capítulo, se enunciará a teoria básica do algoritmo de programação genética, que é outra técnica que permite quantificar incerteza em modelos de simulação de reservatórios de petróleo, e que vamos considerar neste trabalho.

6 Programação Genética

6.1. Introdução

Técnicas de inteligência computacional têm sido utilizadas com muita frequência na indústria de petróleo. Simuladores para a caracterização de reservatórios (como o Petrel da Schlumberger e OpendTect) já possuem há alguns anos, ferramentas de redes neurais (RN) e aplicações com algoritmos genéticos (GA). Trabalhos como (MOHAGHEGH, et al., 1995), (MOHAGHEGH, 2005) e (PACHECO, et al., 2009) formulam propostas de soluções a diferentes áreas na engenharia de petróleo, por meio de algoritmos genéticos, redes neurais e lógica fuzzy. Por outro lado, programação genética (PG), a qual está incluída na família de técnicas de computação evolucionária, vem se destacando cada vez mais em aplicações onde se tenha pouca informação e seja preciso uma expressão matemática explícita para descrever um problema (YU, et al., 2008) (BURKE, et al., 2009).

Na engenharia de reservatórios é preciso quantificar a incerteza em função de todas as variáveis do reservatório (OLIVER, et al., 2008). Entretanto, a relação entre as variáveis do modelo de reservatório (tais como a permeabilidade ou a porosidade) e os dados (tais como as medições do corte de água ou pressão) é com frequência altamente não linear, pelo qual, nestes casos pode ser muito complicado caracterizar a função distribuição de probabilidade (*pdf*) a posteriori.

Como foi mencionado em capítulos anteriores, o método mais comum para resolver este problema é o método de Monte Carlo (MC). Mas sua lenta convergência faz com que seja pouco utilizado em casos onde as avaliações são muito custosas, como em modelos de simulação de reservatórios de petróleo. Por outro lado, PG é uma técnica que vem sendo utilizada com sucesso na estimação da saída de modelos físicos, mecânicos, entre outros. Neste trabalho será utilizado o algoritmo de PG, para encontrar uma expressão explícita, que aproxime a saída do modelo de simulação. Esta saída é influenciada pela incerteza presente nos

parâmetros de entrada. Portanto, a função obtida mediante PG deve representar esta relação e permitir a quantificação da incerteza, por meio do cálculo de estatísticas e probabilidades.

6.2. Programação Genética

O algoritmo de PG é uma coleção de técnicas de computação evolucionária que constrói programas de computador, para solucionar problemas automaticamente sem que o usuário tenha necessidade de conhecer ou especificar a forma da solução (POLI, et al., 2008). Análogo à evolução dos seres vivos, estes programas são submetidos a um processo de evolução com etapas de avaliação, seleção, cruzamentos e mutações. Os indivíduos são estruturas de computador em forma de árvore, compostas por nós e folhas. Os nós representam funções aritméticas básicas (+, -, *) ou operadores mais complexos (cos, exp) e as folhas são os terminais que podem ser variáveis ou constantes.

O algoritmo de PG cria uma população inicial, composta por este tipo de indivíduos. A cada um é atribuído um valor numérico, chamado aptidão, que indica quanto a solução representada por este indivíduo é melhor que as demais soluções da população. A população de soluções candidatas é modificada iterativamente; cada iteração envolve a aplicação de operadores genéticos (seleção, cruzamento e mutação) na espera de evoluir novos e melhores candidatos (KOZA, 1992).

A seguir, será feita uma descrição mais detalhada dos componentes principais do algoritmo de PG.

6.2.1. Representação

Como mencionado antes, em PG, cada indivíduo é representado por uma árvore que equivale a um programa de computador. Este programa tem funções (nós) e terminais (folhas). Os terminais estão situados nas extremidades da árvore e são parâmetros que recebem um valor de entrada. Por outro lado, as funções conectam um ou mais terminais e estão situadas nos vértices internos da árvore. A

seguir, enumeramos as possibilidades para estes dois componentes seguindo Koza (KOZA, 1992).

- Funções (nós):
 - Operações aritméticas: +, -, /, *, %
 - Funções matemáticas: sen, cos, log, exp, sqrt
 - Operações booleanas: and, or, not
 - Operadores condicionais: se-então-senão
 - Operadores iterativos: enquanto (condição) faça
 - Funções recursivas
 - Funções específica do domínio do problema
- Terminais (folhas):
 - Variável: x_1, x_2, \dots, x_M
 - Constante: 0.45, 8, -1,33

Na Figura 6.1, pode-se observar a estrutura em forma de árvore para um indivíduo que tem no conjunto de funções a adição, o produto e a raiz quadrada. Nos terminais, tem as variáveis x e y, e uma constante de 0,12.

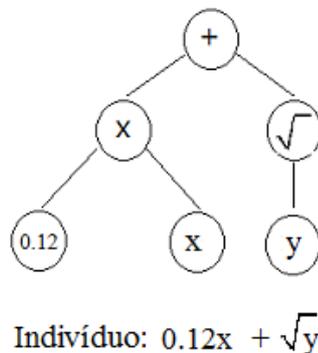


Figura 6.1: Representação em forma de árvore de um indivíduo de PG.

A representação do indivíduo se dá de acordo com a ordem estabelecida na árvore para as funções e os terminais, construindo-se de acima (nó raiz) para baixo (nós componentes), da esquerda à direita.

A combinação de todas as soluções definidas mediante estas funções e terminais, forma o chamado espaço de busca.

6.2.2. Avaliação

A avaliação é o processo aplicado a cada indivíduo para conhecer seu desempenho em relação ao problema considerado. Como mencionado antes, a avaliação de um indivíduo em PG é medida pela sua aptidão, que é representada por um valor numérico. Esta medida de aptidão direciona o processo evolucionário em PG.

A avaliação depende de uma função objetivo, que pode ser medida de várias formas, de maneira implícita ou de maneira explícita. Usualmente está associada a um erro entre a saída do programa e a saída desejada. Por exemplo, é comum utilizar algumas das seguintes funções,

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \quad (6.1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2} \quad (6.2)$$

Onde Y_i representa as saídas reais e \hat{Y}_i as saídas aproximadas.

Neste trabalho foi considerado o erro $RMSE$ para avaliar os indivíduos em cada população do algoritmo de PG. Isto para os conjuntos de treinamento e validação. Para o caso do conjunto de teste, pelo problema considerado, foi calculado o coeficiente de determinação (seção 5.7) para quantificar os erros das soluções finais.

6.2.3. Operadores

Os operadores genéticos em PG têm a função de criar descendentes em cada geração do algoritmo. Seguindo Koza (KOZA, 1992), existem operadores primários e secundários.

- Primários: Reprodução e Cruzamento.
- Secundários: Mutação, Permutação, Edição, Encapsulamento e Decimação.

A reprodução tem o objetivo de preservar na geração seguinte os indivíduos com melhores valores de aptidão. Os indivíduos selecionados são copiados sem alterações e colocados na geração seguinte. Por outro lado, o operador de cruzamento está baseado na troca de material genético entre dois indivíduos, chamados pais. Mediante uma seleção probabilística os pais são selecionados de acordo com a sua aptidão. A troca é feita através de cortes aleatórios nas estruturas individuais. Em Figura 6.2 figura obtida de (POLI, et al., 2008), pode-se observar um exemplo do operador cruzamento.

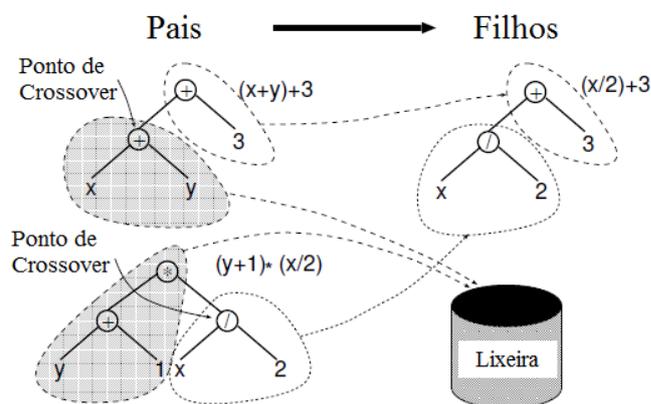


Figura 6.2: Exemplo de Cruzamento. Os indivíduos da esquerda são cópias dos pais, portanto o material genético pode ser utilizado sem realizar alterações nos indivíduos originais.

Da mesma forma em que é gerado o filho da figura anterior, pode também ser gerado um outro filho composto pelas sub-árvores não utilizadas.

No caso dos operadores secundários, o operador mutação opera sobre um indivíduo que é selecionado por meio de sua aptidão. Aleatoriamente é escolhido um ponto de corte (ponto de mutação) na estrutura do indivíduo, e a combinação de funções e terminais a partir desse ponto é substituída por uma nova sub-árvore gerada aleatoriamente. Os demais operadores secundários operam similarmente e podem ser estudados em (KOZA, 1992).

Finalmente, todos os parâmetros de PG devem ser selecionados para iniciar o processo de evolução dos indivíduos: tamanho da população, número de gerações, taxa de cruzamento, mutação, profundidade máxima da árvore, etc. A escolha pode ser realizada por meio de uma sequência de experimentos, nos quais sejam analisados os erros de treino e validação (*RMSE*), assim como a convergência do algoritmo. Neste trabalho, a solução obtida é avaliada

posteriormente num conjunto de teste, com uma função de avaliação diferente à considerado no processo de evolução (R^2).

6.3. Programação Genética Multi-Gene

Programação genética multi-gene (PGMG) é uma variação da PG clássica, na qual, cada indivíduo multi-gene é formado por um ou mais genes e cada um deles é uma árvore da PG tradicional. O número de genes é incrementado em cada indivíduo de maneira tal que seja melhorado o valor de aptidão (por exemplo, reduzir a soma dos erros quadráticos num conjunto de dados).

O modelo geral é uma combinação linear ponderada de cada gene (árvore) e os pesos dos genes. Neste trabalho os pesos foram otimizados mediante uma regressão de mínimos quadrados ordinária. Desta maneira, o resultante modelo pseudo-linear pode capturar o comportamento não linear, presente nos dados do treinamento e validação (SEARSON, et al., 2010).

Matematicamente, o modelo de regressão multi-gene pode ser escrito como,

$$\hat{Y} = d_0 + d_1 \cdot \text{árvore}_1 + d_2 \cdot \text{árvore}_2 + \dots + d_N \cdot \text{árvore}_N \quad (6.3)$$

Onde,

- Cada árvore é uma função de zero ou mais das M variáveis de entrada x_1, x_2, \dots, x_M
- d_0 é o viés
- Os termos d_1, d_2, \dots, d_N são os pesos dos genes
- N é o número de genes, ou seja, a quantidade de árvores que compõe o indivíduo

Na Figura 6.3 pode-se observar um modelo multi-gene com saída (melhor indivíduo) y em função das variáveis de entrada x_1, x_2 e x_3 .

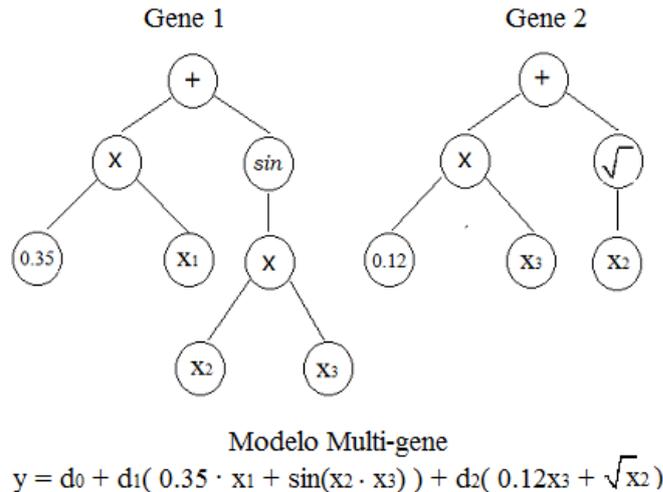


Figura 6.3: Modelo PG Multi-Gene.

Os operadores para PGMG são uma adaptação dos operadores tradicionais de PG. Neste trabalho foi utilizado o pacote de programação genética multi-gene fornecido livremente para *matlab*: *GPTIPS* (KOZA, 1992). Neste programa foram adaptados os operadores para funcionar com indivíduos formados por várias árvores. No caso do cruzamento, pode-se mencionar dois tipos: *cruzamento de dois pontos de alto nível* e *cruzamento de baixo nível*. No primeiro caso, são selecionados dois genes, chamados pontos de cruzamento, nos dois pais. Então, os genes no meio dos pontos de cruzamento são trocados entre os pais para gerar dois novos filhos. No segundo caso, um gene é extraído aleatoriamente de cada pai, se realiza o cruzamento tradicional para genes e os dois genes resultantes são colocados nos pais originais, gerando assim dois filhos. Para o caso da mutação, é selecionado aleatoriamente um gene no pai e a mutação tradicional de PG é efetuada. O gene mudado é substituído no pai original gerando um novo filho (SEARSON, et al., 2010).

De igual maneira que a PG tradicional, os componentes da PGMG, como o tamanho da população, número de gerações, taxa de cruzamento, profundidade máxima da árvore etc., são testados para encontrar o modelo que tenha o erro menor. Nesta dissertação foi utilizado o RMSE (a raiz da média dos erros quadráticos), definido na equação (6.2), como função de avaliação nos conjuntos de treino e validação. No caso do conjunto de teste, foi considerado o coeficiente de determinação para quantificar o ajuste das soluções obtidas.

6.4. Formulação para o Problema de Quantificação de Incerteza

Considere um modelo de simulação de reservatórios de petróleo, cuja saída é a produção acumulada de óleo, gás e água. As respostas do modelo estão em função de um conjunto de parâmetros incertos, como por exemplo: permeabilidade, porosidade, compressibilidade da rocha ou profundidade dos contatos dos fluidos.

Deseja-se estudar a influência dos parâmetros de entrada na resposta do modelo. Primeiramente, deve-se considerar que tais saídas do modelo de simulação são também variáveis incertas, devido à aleatoriedade presente nos parâmetros de entrada. Portanto, estudar a relação entre as entradas e as saídas é analisar a incerteza propagada através do modelo de simulação. Normalmente a análise da incerteza na saída de modelo é feita mediante a estimação de estatísticas e probabilidades. Nesta dissertação estamos interessados num resultado mais rigoroso e, portanto, é aproximada a distribuição de probabilidades inteira.

Sabemos que a relação entre muitas das variáveis de entrada de um modelo de simulação de reservatórios de petróleo e as possíveis saídas (como a produção acumulada), é não linear. Por este motivo, foi desenvolvido e proposto o algoritmo de programação genética multi-gene, para capturar este comportamento.

Para o problema de quantificação de incerteza via PG, os conjuntos de treinamento, validação e teste devem ser construídos seguindo a densidade de probabilidade das entradas do modelo. Desta maneira, o algoritmo de PGMG permitirá encontrar uma expressão matemática explícita, que represente a relação entre os dados de entrada e de saída, mas, tendo em consideração a incerteza presente nas entradas. A expressão simbólica obtida mediante PGMG é aleatória e depende completamente das entradas do modelo e de sua densidade de probabilidade.

Uma vez que seja construída, por meio de PG, uma expressão explícita que aproxime a resposta estudada, esta pode ser utilizada como um *proxy* sob incerteza do modelo de simulação. Desta maneira, é possível realizar análises estatísticas e probabilísticas sem necessidade de utilizar novamente o simulador de reservatórios, diminuindo assim o custo computacional de fazer estas análises.

Por outro lado, para obter os percentis e a distribuição de probabilidade empírica da saída do modelo, pode-se realizar a mesma análise enunciada na seção 5.5. Além disso, assim como na aproximação por caos polinomial, é possível quantificar o erro da estimação por meio do coeficiente de determinação definido na seção 5.7.

Finalmente, o seguinte algoritmo descreve os passos a seguir, de acordo com o mencionado acima:

- i. Determinar a densidade de probabilidade dos parâmetros de entrada
- ii. Construir a matriz de dados, (entradas, saídas) de treinamento e validação para o aprendizado do algoritmo de PGMG. Estas matrizes são construídas aleatoriamente, seguindo a densidade de probabilidade dos parâmetros de entrada.
- iii. Rodar o PGMG com a matriz de dados construída em ii.
- iv. Construir um modelo substituto sob incerteza com a estrutura matemática explícita resultante do PGMG.
- v. Realizar uma análise estatística e probabilística baseada em amostragem, para obter a função distribuição de probabilidade da saída do modelo de simulação.

Na Figura 6.4 pode-se observar este processo.

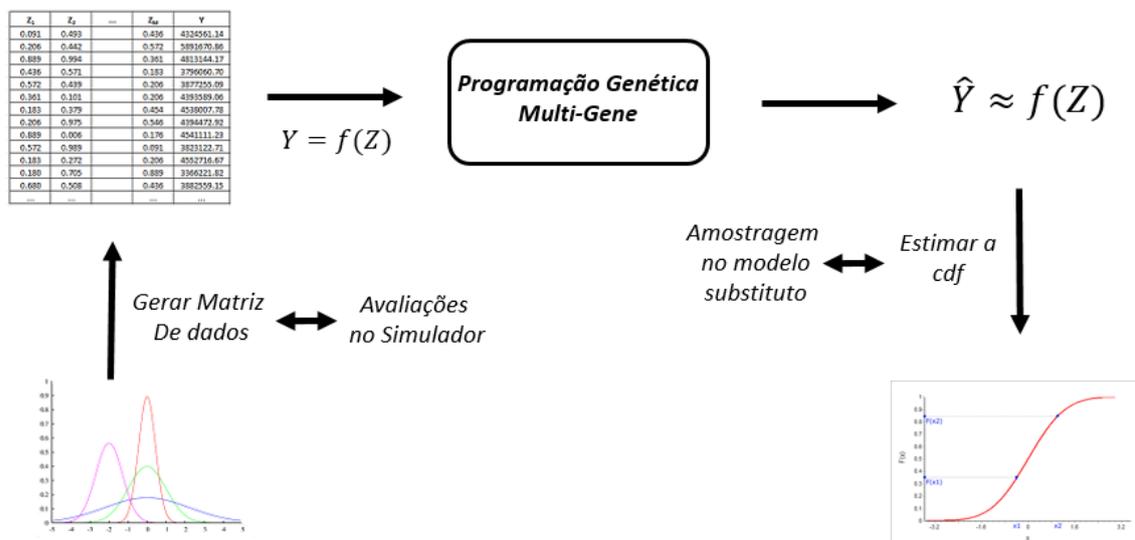


Figura 6.4: PGMG no processo de quantificação de incertezas.

Finalizamos este capítulo reforçando uma das maiores vantagens do algoritmo de PGMG, que é obter uma expressão explícita que pode ser manipulada e utilizada como um modelo substituto sob incerteza da resposta do modelo de simulação (produção acumulada de óleo, gás e água). Uma vez construído o modelo substituto, é possível fazer análises estatísticas e probabilísticas da resposta do simulador, sem necessidade de simular novamente o campo de petróleo. Desta maneira, o custo computacional de realizar estas análises é diminuído consideravelmente. Uma outra vantagem é que, caso se tenha uma ideia do comportamento da resposta do modelo que se deseja estimar, PGMG permite utilizar funções (exp, log, tan) com um comportamento similar e assim orientar a construção das soluções candidatas.

No próximo capítulo, vai-se desenvolver o estudo de casos realizado e a descrição dos modelos de simulação de reservatórios considerados. Inicialmente são expostas as técnicas das quadraturas, com o fim de valorar sua acurácia e custo computacional. Seguidamente, são apresentados os resultados obtidos com PCE e PGMG num modelo de simulação baseado no campo de petróleo PUNQ-S3.

7 Estudo de Casos

7.1. Introdução

Neste capítulo será desenvolvido o estudo de casos realizado e a descrição dos modelos de simulação de reservatórios considerados. Inicialmente serão expostas as técnicas de quadratura para integração numérica, com o fim de valorar sua acurácia e custo computacional. A motivação do primeiro experimento numérico está no fato de que a eficiência das expansões por caos polinomial depende da escolha do método utilizado para aproximar os coeficientes que compõem o polinômio. Seguidamente, será feito um estudo comparativo das técnicas apresentadas nos capítulos anteriores: Expansões por Caos Polinomial via aproximação espectral e aproximação por regressão; e PG. Os métodos vão ser aplicados em três modelos de simulação baseados no caso PUNQ-S3 (IMPERIAL COLLEGE LONDON) e fazendo uso do simulador de reservatórios comercial IMEX, desenvolvido por *Computer Modeling Group* (CMG). Os resultados serão comparados com a técnica de Monte Carlo, para valorar sua precisão e eficiência computacional na quantificação de incertezas em modelos de simulação de reservatórios de petróleo.

7.2. Experimento Numérico

No corpo desta dissertação temos feito ênfase na importância da escolha do método utilizado para calcular os coeficientes da expansão por caos polinomial, quando é considerada esta técnica, para aproximar a saída do modelo estudado. Uma vez a que os coeficientes podem ser expressos por meio de integrais multidimensionais, é possível fazer uso de algum mecanismo de integração numérica para estimar estes valores. No caso de modelos de simulação de reservatórios de petróleo, deve ser ainda mais cuidadosa a seleção do método para

o cálculo dos coeficientes da expansão, já que as avaliações no modelo têm um grande custo computacional inerente ao modelo (campo de petróleo).

Na seção 5.3, propomos utilizar as quadraturas de Smolyak para tratar o problema da dimensionalidade, gerado no cálculo dos coeficientes de expansões com um alto número de variáveis. Esta seção tem a finalidade de mostrar a acurácia desta técnica e o custo computacional, medido em quantidade de avaliações à função, ou seja, número de chamadas ao simulador. Vamos comparar os resultados com a quadratura gaussiana, método usualmente utilizado na literatura (GUEVARA, 2014).

Considere o seguinte problema de integração com dimensão d .

$$\int_{\mathbb{R}^d} (2\pi)^{\frac{-d}{2}} \left(\|x\|^2 + 2x_1^3 + \frac{1}{2}x_d^6 \right) e^{\frac{-1}{2}\|x\|^2} dx, \quad x \in \mathbb{R}^d \quad (7.1)$$

Observe-se que a integral representa a média da função: $f(x) = \|x\|^2 + 2x_1^3 + \frac{1}{2}x_d^6$, onde x é um vetor de variáveis aleatórias normais, com média zero e desvio padrão um. Assim,

$$E[f(x)] = \int_{\mathbb{R}^d} \overbrace{\frac{1}{\sqrt{2\pi}^d} e^{\frac{-1}{2}\|x\|^2}}^{\text{cdf de } x \sim N(0,1)} \cdot \underbrace{\left(\|x\|^2 + 2x_1^3 + \frac{1}{2}x_d^6 \right)}_{f(x)} dx \quad (7.2)$$

Então, temos uma integral d -dimensional, associada a uma função de peso gaussiana $W(x) = \frac{1}{\sqrt{2\pi}^d} e^{\frac{-1}{2}\|x\|^2}$. Portanto, é possível utilizar regras de quadratura de Gauss-Hermite (GH), ou quadraturas de Smolyak baseadas em sequências de Gauss-Hermite (SGH) ou Genz-Keister (SGK) para estimar seu valor.

Para aproximar a integral por meio de quadraturas, deve ser selecionado um nível, valor que determina a quantidade de nodos de integração. A escolha do nível depende do grau da função $f(x)$, pode-se observar que neste caso é um polinômio de grau seis, portanto o nível ($\ell \in \mathbb{N}$) das quadraturas deve satisfazer $2\ell - 1 \geq 6 \Rightarrow \ell \geq 4$.

Para realizar contas numéricas, considere o problema da eq. (7.1) com dimensão $d = 10$. A integral pode ser calculada analiticamente fazendo uso das propriedades da média e mediante algum software de cálculo numérico é possível obter o valor médio $E[f(x)] = 17.5$. Na Tabela 7.1 são mostrados os resultados da estimação para diferentes valores de nível nas regras de quadratura.

Tabela 7.1: Aproximação da integral da eq. (7.1).

Parâmetros	Num. Avaliações/Simulações			Valor estimado da integral		
	Nível	GH	SGH	SGK	GH	SGH
2	1,024	21	21	10.5	10.5	14.5
3	59,049	221	201	14.496	14.5	14.5
4	1,048,576	1,581	1,201	17.5	17.5	17.5
5	9,765,625	8,761	5,301	17.5	17.5	17.5

Pode-se observar o crescimento exponencial da quadratura de Gauss-Hermite ($Num. Simulação_k = nível_k^{10}$) em contrapartida ao crescimento lento das quadraturas com a forma de Smolyak. Por outro lado, as quadraturas conseguem a exatidão polinomial quando o nível é maior o igual a quatro e pode-se notar que Smolyak manteve a acurácia das quadraturas uni-variadas com uma quantidade menor de pontos de integração.

Agora, considere o mesmo problema da eq. (7.1), acrescentando um termo não polinomial. Assim, para $x \in \mathbb{R}^d$, $d = 10$.

$$\int_{\mathbb{R}^d} (2\pi)^{-\frac{d}{2}} \left(\|x\|^2 + 2x_1^3 + \frac{1}{2}x_d^6 + \sum_{k=1}^d \sin(x_k^2) \right) e^{-\frac{1}{2}\|x\|^2} dx \quad (7.3)$$

Neste caso, deseja-se aproximar $E \left[\|x\|^2 + 2x_1^3 + \frac{1}{2}x_d^6 + \sum_{k=1}^d \sin(x_k^2) \right]$. O valor da integral é $E[f(x)] = 21.0158$. A diferença com a situação anterior, é que a função de integração não é um polinômio, portanto, não pode-se afirmar uma equivalência ao aproximar com regras de quadratura. Além disso, a escolha do nível é mais difícil, devido a que não se tem um grau polinomial. Os resultados da estimação deste problema são apresentados na Tabela 7.2, os últimos valores para Gauss-Hermite, não foram avaliados pelo alto custo computacional.

Tabela 7.2: Aproximação da integral da eq. (7.3)

Parâmetros	Num. Avaliações/Simulações			Valor estimado da integral		
	Nível	GH	SGH	SGK	GH	SGH
2	1,024	21	21	18.916	18.915	14.970
3	59,049	221	201	14.967	14.970	14.970
4	1,048,576	1,581	1,201	21.572	21.572	19.249
5	9,765,625	8,761	5,301	21.999	21.999	20.737
6	60,466,176	40,405	19,485	-	19.750	20.737
7	282,475,249	162,025	63,405	-	21.790	20.737
8	1,073,741,824	581,385	185,085	-	20.876	20.737
9	3,486,784,401	1,904,465	490,625	-	20.766	20.840
10	10,000,000,000	5,778,965	1,206,645	-	21.336	21.072

Observe-se que as quadraturas começam a convergir quando o nível é maior que quatro, isto pelo grau do termo polinomial da função $f(x)$. A regra de Genz-Keister consegue com uma quantidade menor de pontos de integração estimar com maior acurácia o valor da média desejada. Isto é devido a que Genz-Keister é uma regra aninhada e o *grid sparse* utiliza só alguns pontos do total do produto tensor.

Desta maneira, temos confiança na utilização das quadraturas da forma de Smolyak, e portanto vamos utilizá-las nos próximos casos de estudo, onde PCE e PG são aplicados em três modelos de simulação de reservatórios baseados no campo PUNQ-S3.

7.3. Modelos de Simulação de Reservatórios de Petróleo

Com o objetivo de quantificar a incerteza nos modelos de simulação de reservatórios de petróleo, foi realizado um procedimento que pode ser resumido nos seguintes passos:

- Definir o modelo e os parâmetros considerados para a análise determinística do sistema.
- Identificar os parâmetros de entrada incertos, determinar seu efeito na saída e modela-los mediante variáveis aleatórias. Ou seja, transformar o modelo determinístico inicial, num modelo estocástico.

- Aproximar a saída do modelo por um polinômio ou uma expressão simbólica, mediante as técnicas de quantificação de incerteza propostas (PCE, PG). Nesta etapa é necessário avaliar no simulador IMEX, para construir um conjunto de respostas do modelo, que será utilizado de acordo às metodologias abordadas.
- Utilizar a expressão obtida, por meio de PCE ou PG, como um substituto sob incerteza da saída do modelo de simulação e realizar cálculos estatísticos e análises probabilísticas para quantificar a incerteza presente na resposta do sistema. Estas análises são feitas mediante amostragem direta na função aproximada.

No software *matlab*, foi implementado um conjunto de scripts, os quais mantêm uma ligação direta com o simulador IMEX. Estas funções no formato de *matlab*, geram, alteram e leem os arquivos de entrada e saída, com extensões *.dat* e *.out* respectivamente, do simulador da CMG e executam os processos necessários para a obtenção das saídas aproximadas mediante PCE e PG (ver formato do arquivo de leitura do programa no apêndice).

Nesta seção vamos desenvolver três casos de estudo, no contexto da quantificação de incertezas em modelos de simulação de reservatórios. Vai-se iniciar descrevendo o campo petrolífero: PUNQ-S3, e seguidamente se expõem os modelos com os parâmetros definidos incertos. Os resultados vão ser comparados com a técnica tradicional de Monte Carlo, com o fim de determinar o desempenho obtido.

7.3.1. Descrição do campo

PUNQ-S3 é um modelo de engenharia de reservatório industrial. Está baseado num campo real que foi operado por *Elf Exploration Production* e desenvolvido por um grupo de companhias de petróleo, universidades e instituições na União Europeia no projeto PUNQ (*Production forecasting with UNcertainty Quantification*). A geologia de PUNQ-S3 é descrita como um modelo de cinco camadas, onde cada camada é caracterizada por uma face única com a largura, distância de separação e direção esperada. Uma descrição completa do problema, junto com os arquivos do modelo podem encontrar-se em (IMPERIAL COLLEGE LONDON).

O objetivo da criação de PUNQ-S3 foi desenvolver um caso de teste para permitir a vários grupos de investigação avaliar sua habilidade para caracterizar a incerteza em predições do desempenho de reservatórios, utilizando informação geológica, *hard data* nos blocos do *grid* dos poços e dados históricos de produção dos primeiros oito anos de produção. Foi solicitado aos participantes do projeto que predissessem a produção acumulada de óleo para 16.5 anos de produção total e caracterizar a incerteza na predição (GAO, et al., 2005). Devido a isto, PUNQ-S3 é um adequado caso de análise na área de quantificação de incertezas e portanto, foi escolhido para este trabalho.

O *grid* de simulação original para o problema PUNQ-S3 consiste de $19 \times 2 \times 85$ blocos, as dimensões de cada bloco nas direções *X* e *Y* são $180 \times 180 \text{ m}^2$ e tem 1761 blocos ativos. O campo é limitado por uma falha ao leste e sul e tem comunicação com um forte aquífero ao oeste e norte, por isto não são considerados poços injetores. O campo original contém 6 poços produtores alocados em torno da zona de contato gás-óleo. Além disso foram definidas 5 posições para poços *in-fill* adicionais (GAO, et al., 2005).

Neste trabalho se vai quantificar a incerteza na saída de um modelo de simulação baseado numa modificação do problema PUNQ-S3 inicial, para um período de 13 anos. Foram considerados os seis poços definidos no caso original, chamados: PRO-1, PRO-4, PRO-5, PRO-11, PRO-12, PRO-15 e foram definidos mais 4, chamados: PRO-23, PRO-24, PRO-29 e PRO-50. Os dez poços foram abertos sequencialmente cada dois meses. Na Figura 7.1 pode-se observar a distribuição destes poços, numa imagem da porosidade do campo PUNQ-S3, obtida do software da CMG. As saídas do simulador tratadas são:

- Produção acumulada de óleo (N_p)
- Produção acumulada de gás (G_p)
- Produção acumulada de água (W_p)

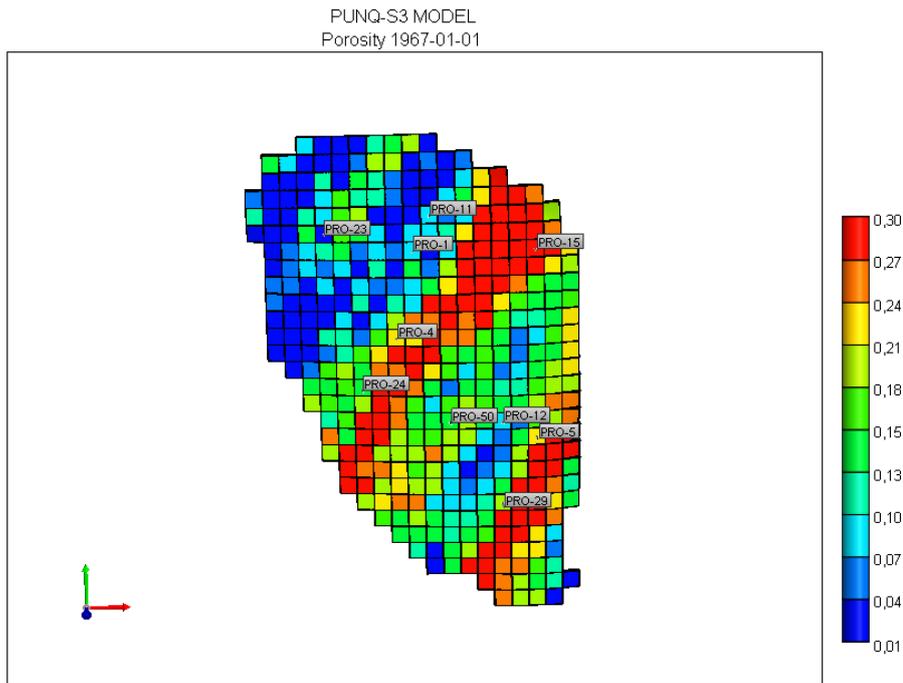


Figura 7.1: Porosidade do campo PUNQ-S3, na data 1967-01-01.

Para o período compreendido entre março 1967 até janeiro 1980, vão ser estudados três modelos baseados neste campo de petróleo, no primeiro caso é exposto um sistema com três parâmetros incertos com distribuição de probabilidade log-normal, no segundo e no terceiro modelo é estudada a propagação da incerteza presente em onze e vinte e três variáveis respectivamente, com distribuição normal. Os parâmetros considerados incertos, são descritos na seção 3.3.1 e nos três casos foi suposta a independência destes parâmetros de entrada no modelo de simulação.

Nesta dissertação se deseja obter com um custo computacional baixo, uma expressão que aproxime eficientemente cada uma destas saídas e seja utilizada como um modelo substituto para o cálculo de estatísticos e análises probabilísticas.

Problemas de reservatórios como este não possuem solução analítica, portanto, para obter um parâmetro de comparação para as soluções obtidas com os métodos de PCE e PG, foi feita uma análise empregando-se a técnica de simulação de Monte Carlo.

7.3.2. Estudo de Caso 1

Inicialmente, no estudo de caso 1, o problema é definido com três parâmetros, que são os valores que afetam as matrizes de porosidade e permeabilidade, estes são chamados de multiplicador da porosidade (*MULTIPOR*), multiplicador da permeabilidade horizontal (*MULTIPERMI*) e multiplicador da permeabilidade vertical (*MULTIPERMK*). Estas características incertas do reservatório vão ser parametrizadas por três variáveis aleatórias que vamos supor independentes, cuja incerteza é modelada com uma distribuição log-normal de média zero e desvio padrão um (pode-se ver seção 3.3.1.).

Uma vez determinada a modelagem estocástica do sistema, foi utilizado o simulador IMEX para gerar um conjunto de saídas, com as quais realizar uma análise paramétrica. Isto com o fim de verificar que as respostas do modelo são variantes com as propriedades que estão-se supondo incertezas. No caso de ser invariantes, não há, na prática, incerteza nas respostas e portanto não tem sentido quantificá-la. Nas seguintes figuras (Figura 7.2, Figura 7.3 e Figura 7.4) podem-se observar as variações na produção acumulada (Np , Gp , Wp) quando são variados cada um dos parâmetros (*MULTIPOR*, *MULTIPERMI* e *MULTIPERMK*) de maneira independente.

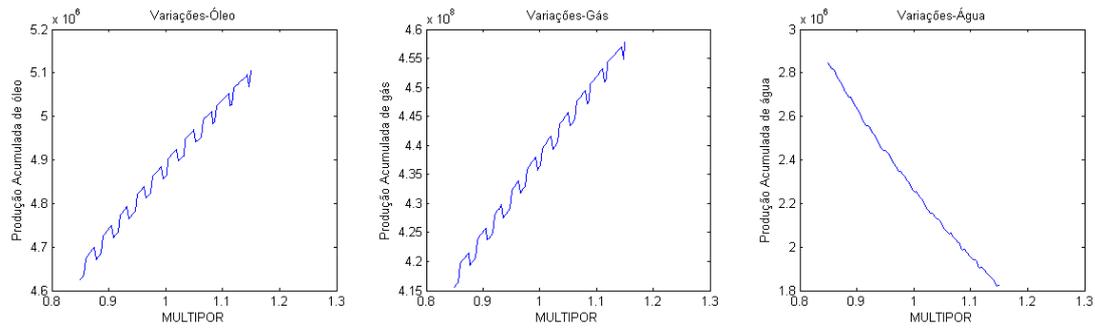


Figura 7.2: Variações na produção acumulada de óleo, gás e água, quando varia o multiplicador da porosidade no intervalo [0.85, 1.15].

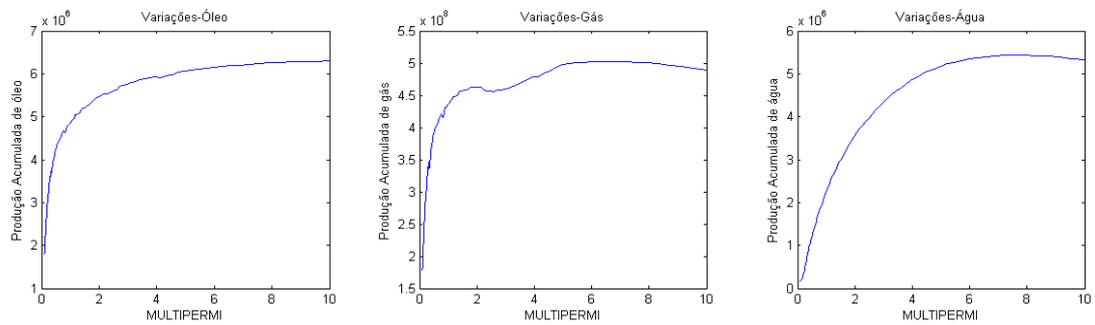


Figura 7.3: Variações na produção acumulada de óleo, gás e água, quando varia o multiplicador da permeabilidade horizontal no intervalo [0.01, 10].

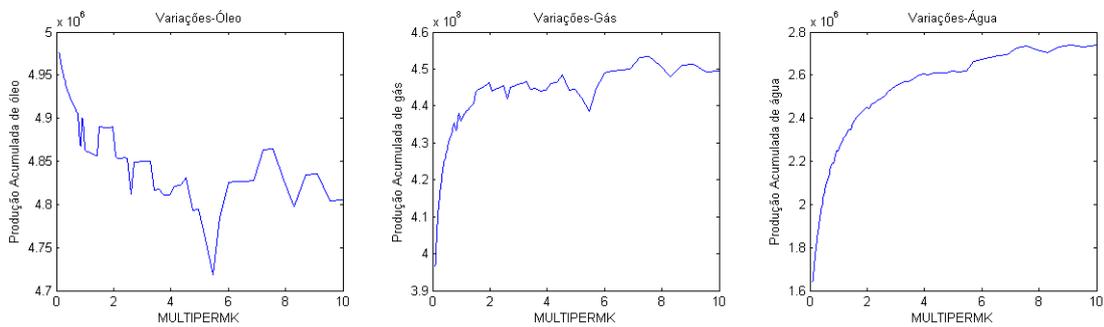


Figura 7.4: Variações na produção acumulada de óleo, gás e água, quando varia o multiplicador da permeabilidade vertical no intervalo [0.01, 10].

Pode-se observar nas figuras anteriores que os três parâmetros afetam a produção acumulada de óleo, gás e água, portanto, tem sentido quantificar a incerteza nestas saídas. Realizadas as parametrizações correspondentes e as análises de efeitos das variáveis sobre as saídas, é possível aplicar as técnicas propostas neste trabalho, para quantificar a incerteza propagada através do simulador.

- **Expansões por caos polinomial**

Na construção da expansão por caos polinomial, a base de funções ortogonais utilizada foi a família de polinômios de Hermite (ver Tabela 5.1), já que uma variável aleatória com distribuição log-normal pode ser expressa mediante uma variável normal com média zero e desvio padrão um, assim,

$$\begin{aligned}
 & Y \text{ Lognormal} \\
 \Leftrightarrow & Y = e^X, \quad \text{com } X \sim N(\mu, \sigma^2) \\
 \Leftrightarrow & Y = f(Z) = e^\mu e^{\sigma Z}, \quad \text{com } Z \sim N(0,1)
 \end{aligned}$$

Caso seja utilizado um enfoque de aproximação por projeção espectral, além da seleção da base de funções, há mais duas escolhas importantes, o grau da expansão (p) e o nível da quadratura de Smolyak (ℓ). O grau da PCE é o máximo grau dos polinômios ortogonais que conformam cada termo da expansão final, sua seleção depende do comportamento que tenha a saída que se deseja aproximar. Por outro lado, o nível da quadratura está relacionado ao número de pontos que serão utilizados para estimar os coeficientes da expansão (número de simulações). A exatidão polinomial da quadratura está em função deste valor e sua escolha depende do argumento das integrais que definem os coeficientes. Pode-se notar na eq. (5.8) que o integrando tem grau aproximadamente $2p$, já que é um produto de um polinômio de grau máximo p e a função f , que está-se supondo tem grau p . Portanto, é necessário um nível igual ou superior a $\ell = p + 1$, para garantir uma exatidão polinomial igual ou superior que $2(p + 1) - 1 = 2p + 1$. Nesta seção foram feitas experiências numéricas variando-se o grau da expansão e o nível da regra de Smolayk, para obter a melhor aproximação da saída com uma quantidade baixa de simulações.

Por outro lado, quando os coeficientes são aproximados mediante um enfoque por regressão é preciso seguir os dois algoritmos descritos na seção 5.4. Desta forma, podem-se obter as amostras e o número necessário delas, que permitem construir um planejamento de experimentos adequado para estimar os coeficientes da expansão. Lembre-se que as amostras do planejamento são geradas por meio das raízes dos polinômios de Hermite de grau $p + 1$ e a quantidade delas (número de simulações) é obtida através da invertibilidade da matriz de informação (ver eq. (5.16)). Portanto, a quantidade de amostras deve ser como mínimo o número de coeficientes da expansão. De igual maneira que no enfoque

por quadraturas, foram gerados vários experimentos mudando o grau da expansão, com o fim de obter a melhor aproximação.

Na Tabela 7.3 pode-se observar os valores do coeficiente de determinação (R^2), gerados com um conjunto de 50,000 amostras (ver eq. (5.24)), avaliadas nas soluções obtidas por PCE, variando o grau e o nível. Este valor permite ter um conhecimento do ajuste da aproximação, com respeito aos valores reais.

Tabela 7.3: Coeficientes de determinação para cada solução obtida mediante expansões por caos polinomial com três dimensões.

	Grau da PCE	Número de Coeficientes	Nível da Quadratura	Número de Simulações	R^2 Óleo	R^2 Gás	R^2 Água
Projeção	2	10	4	39	0.924	0.912	0.881
	2	10	5	93	0.938	0.922	0.899
	3	20	4	39	0.932	0.921	0.913
	3	20	5	93	0.947	0.933	0.940
	4	35	4	39	0.920	0.911	0.897
	4	35	5	93	0.962	0.946	0.950
Regressão	2	10	-	13	0.773	0.736	0.775
	3	20	-	30	0.910	0.865	0.895
	4	35	-	45	0.624	0.153	0.597

Note-se que quando o nível da quadratura aumenta, o ajuste da expansão melhora, isto pela exatidão das regras de quadratura. Além disso, não necessariamente aumentando o grau da expansão, será obtida uma melhor aproximação. Isto depende do comportamento da saída e quanto ela pode ser modelada por um polinômio de grau 2, 3, 4 ou mais. Com respeito ao enfoque da regressão, pode-se dizer que as estimações feitas dependem completamente do planejamento de experimentos gerado. Adicionalmente, deve-se levar em consideração que o polinômio obtido não é necessariamente o polinômio de grau p que melhor aproxima a saída do modelo, aquele mencionado na eq. (5.10).

Finalmente, foram selecionadas as expansões por caos polinomial obtidas com os parâmetros marcados na Tabela 7.3 com 93 e 30 simulações (linhas 7 e 9).

- **Programação Genética Multi-Gene**

No caso do algoritmo de PG, o processo para obter uma solução é menos analítico e mais experimental-evolutivo. Foi executado um algoritmo de PG para cada saída do modelo de simulação considerado. Os parâmetros do algoritmo são testados e selecionados de acordo ao erro alcançado num conjunto de teste. Nesta seção foi abordado o algoritmo proposto e descrito na seção 6.4, lembre-se que, primeiramente se devem gerar os conjuntos de aprendizado: treino e validação seguindo as distribuições de probabilidade dos parâmetros incertos do modelo. Para este problema, foi construída uma matriz de valores distribuídos log-normalmente com média zero e desvio padrão um, o número de amostras foi escolhido empiricamente. As correspondentes saídas foram calculadas utilizando o simulador de reservatórios.

Os operadores selecionados para a construção das soluções candidatas foram: soma, subtração, produto, potência quadrada, função mínimo e função máximo. O operador de mutação utilizado foi aquele descrito na seção 6.3. Para o caso do cruzamento, foram utilizados os cruzamentos de alto e baixo nível (ver seção 6.3), numa porcentagem de 20% e 80% respectivamente. Finalmente, o operador de reprodução copia o indivíduo inteiro à seguinte geração sem modificações.

Uma vez escolhidos os parâmetros adequados para o algoritmo de programação genética multi-gene, foram rodados dez vezes cada algoritmo para garantir que os resultados são consistentes, os erros em cada rodada ficam dentro de um intervalo pequeno e a convergência é mantida. A melhor solução candidata para cada programa genético e os valores dos parâmetros são apresentados na Tabela 7.4. Além do coeficiente de determinação gerado com um conjunto de 50,000 amostras (ver eq. (5.24)).

Tabela 7.4: Valores dos parâmetros e coeficiente de determinação para cada programa genético no caso de três dimensões.

Parâmetros do Algoritmo de PGMG	Oleo	Gás	Água
Tamanho da População	30	30	30
Número de Gerações	850	850	850
Tamanho do Torneio	4	4	4
Elitismo	0.05	0.05	0.05
Taxa de Mutação	0.3	0.3	0.3
Taxa de Cruzamento	0.65	0.65	0.65
Máxima Profundidade dos Genes	3	3	3
Máximo número de Genes	20	20	20
Intervalo de Variação dos nós constantes	[-10,10]	[-10,10]	[-10,10]
Amostras no Treino	40	60	80
Amostras na Validação	20	30	50
Número de Simulações	60	90	130
Coeficiente de Determinação: R^2	0.959	0.955	0.977

Pode-se observar que nos três casos (produção acumulada de óleo, água e gás) são obtidas porcentagens de ajuste acima de 95%, com uma quantidade de simulações abaixo de 130. Com o qual é evidenciado o alto desempenho do algoritmo de PG, num modelo de simulação de reservatórios de baixa dimensão, com parâmetros distribuídos log-normalmente que geram um certo grau de dificuldade nas estimações.

Por outro lado, comparando o desempenho das soluções obtidas mediante PCE e PG, pode-se destacar que para a produção acumulada de óleo e gás, PG consegue ajustar o modelo com maior precisão e realizando uma quantidade menor de simulações. Para o caso da produção acumulada da água, PG precisou mais simulações que PCE para obter um ajuste superior (130 contra 93). Com a finalidade de mostrar o desempenho de PG com uma quantidade menor de amostras no caso da produção de água foi executada uma sequência de experimentos com um total de 93 simulações. As soluções mais representativas foram obtidas com 60 amostras para o treinamento, 33 amostras para a validação e parâmetros equivalentes aos da Tabela 7.4. O coeficiente de determinação médio obtido nos experimentos realizados foi 0.938, este valor apresenta um ajuste inferior à solução obtida mediante PCE com 93 simulações. O número de simulações foi aumentado até obter um ajuste melhor ao alcançado por PCE.

Com o fim de mostrar um exemplo das soluções obtidas mediante o algoritmo de PG, é apresentada na seguinte Figura 7.5 a expressão simbólica que aproxima a produção acumulada de óleo no modelo de simulação considerado neste trabalho. As variáveis z_1, z_2 e z_3 representam os parâmetros *MULTIPOR*, *MULTIPERMI* e *MULTIPERMK* respectivamente.

$$\begin{aligned} \hat{N}p = & 1.24 \cdot 10^8 \cdot z_1 - 3.532 \cdot 10^8 \cdot z_2 + 2.005 \cdot 10^6 \cdot (\max(z_3, z_3))^2 - 4.766 \cdot 10^7 \cdot (z_2 - \\ & 3.971)^2 - 1.428 \cdot 10^7 \cdot \min((z_2)^2, z_1 - z_2) - 1.623 \cdot 10^7 \cdot \max(-3.971 \cdot z_1, \max(z_3, z_1)) - 2.845 \cdot \\ & 10^7 \cdot \max(2 \cdot z_2, (z_2)^2) + 1.345 \cdot 10^7 \cdot z_2 \cdot \max(z_1, z_3) - 4.301 \cdot 10^7 \cdot \max(z_1 - z_2, 5.485 \cdot z_1) + 1.71 \cdot \\ & 10^7 \cdot \max(5.868 \cdot z_1, z_1 - z_3) + 1.3 \cdot 10^7 \cdot z_1 \cdot z_2 + 5.864 \cdot 10^6 \cdot z_1 \cdot z_3 - 7.136 \cdot 10^6 \cdot \max(z_2, z_1) + \\ & 2.993 \cdot 10^6 \cdot \max(z_1, z_3) - 2.993 \cdot 10^6 \cdot \min(z_2, z_3) + 2.027 \cdot 10^7 \cdot \max(2 \cdot z_1, \max(z_1, z_3)) - \\ & 2.8 \cdot 10^6 \cdot \max((z_2)^2, -7.054 \cdot z_3) - 3.681 \cdot 10^6 \cdot z_3 \cdot (z_1)^2 + 1.808 \cdot 10^7 \cdot \max((z_1)^2, z_2) + 6.995 \cdot \\ & 10^7 \cdot \max(z_2, (z_2)^2) - 4.817 \cdot 10^6 \cdot z_1^2 - 1.415 \cdot 10^7 \cdot \max(z_1 - z_2, z_1^2) + 1.19 \cdot 10^9 \end{aligned}$$

Figura 7.5: Solução simbólica de PG para a produção acumulada de óleo.

Por outro lado, independentemente da metodologia empregada para aproximar cada saída do sistema, as soluções podem ser utilizadas como modelos (estocásticos) substitutos sob incerteza para realizar análises estatísticas e probabilísticas. Isto tem a vantagem, de que uma vez obtidas as soluções, não é preciso simular novamente o campo de petróleo para fazer estes estudos, com o qual evita-se um alto custo computacional. Com esta ideia vai-se estimar a distribuição de probabilidade, com a qual é possível obter quaisquer percentis das saídas do modelo. Além disso, com o fim de ter um modelo de comparação, neste problema foi executado o método de Monte Carlo com 300,000 simulações. Se apresenta na Figura 7.6, Figura 7.7 e Figura 7.8 as distribuições de probabilidade empírica para a produção acumulada de óleo, gás e água aproximadas mediante PCE, PG e Monte Carlo.

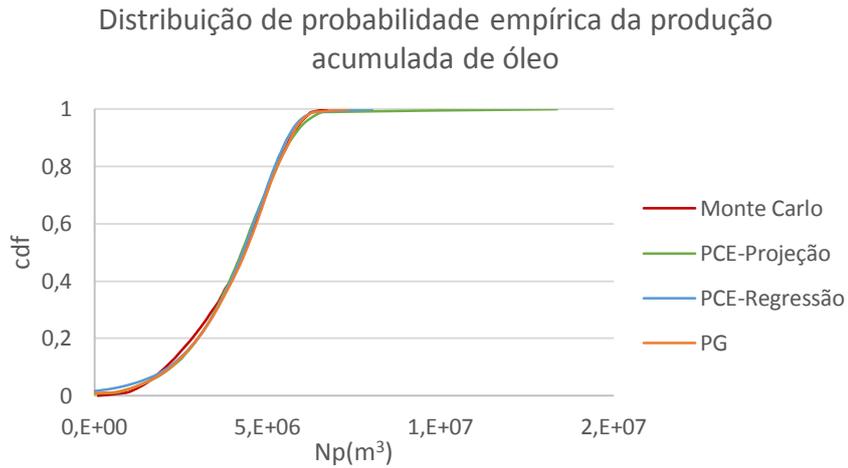


Figura 7.6: Função distribuição de probabilidade empírica da produção acumulada de óleo.

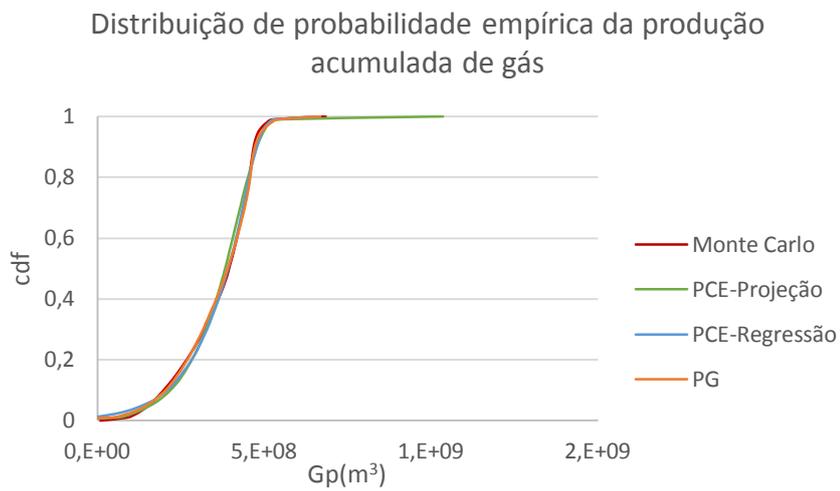


Figura 7.7: Função distribuição de probabilidade empírica da produção acumulada de gás.

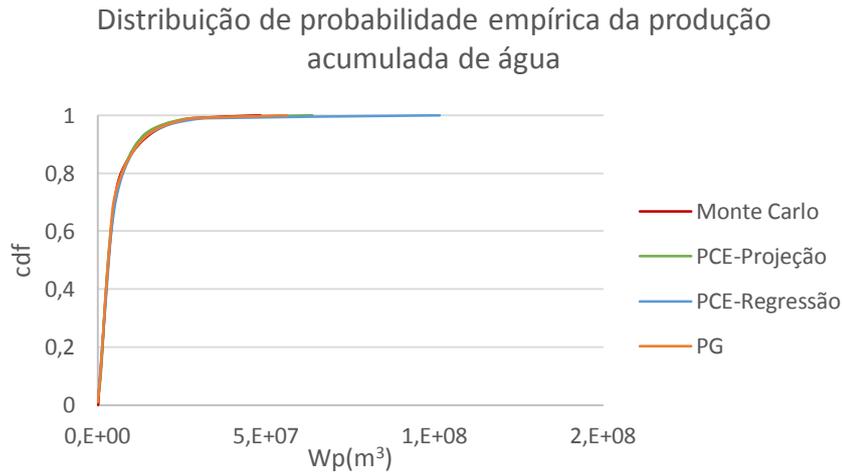


Figura 7.8: Função distribuição de probabilidade empírica da produção acumulada de água.

Na Figura 7.9 se mostram os valores do coeficiente de determinação e o número de simulações para cada aproximação das duas metodologias propostas, com o fim de fazer uma comparação.

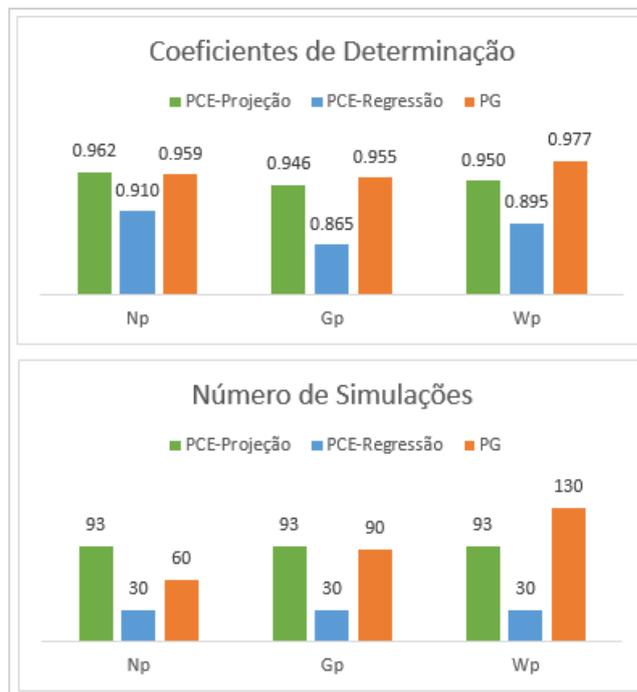


Figura 7.9: Coeficientes de determinação e número de simulações para cada aproximação, modelo de três parâmetros.

As distribuições de probabilidade empírica foram estimadas por meio da formulação descrita na seção 5.5, os percentis que conformam esta curva são da forma da eq. (5.22).

Pode-se observar a proximidade das soluções obtidas por meio de PG e PCE com a solução resultante de Monte Carlo. A principal diferença das estimações radica no número de simulações realizadas para quantificar a incerteza nas saídas do modelo. Este valor busca-se sempre ser o mínimo possível, em especial, para problemas de alta dimensionalidade, como os que são abordados nos próximos estudos de caso.

Finalmente na Tabela 7.5 são apresentadas as médias e as variâncias para cada saída do modelo de simulação. No caso de PCE, estes valores foram estimados utilizando diretamente os coeficientes da expansão, por meio das formulas (5.19) e (5.20). Para o método de PG, foi realizada uma amostragem direta com 100.000 experimentos gerados aleatoriamente seguindo a distribuição de probabilidade dos parâmetros de entrada incertos.

Tabela 7.5: Média e variâncias para cada saída do modelo de 3 parâmetros incertos.

Produção Acumulada de Óleo		
	Média	Variância
Monte Carlo	4,060,102.18	1,791,266,013,125.12
PCE-Projeção	4,077,236.87	1,830,407,624,375.01
PCE-Regressão	4,007,346.46	2,127,007,975,206.21
PG	4,101,232.57	1,829,258,367,026.30
Produção Acumulada de Gás		
	Média	Variância
Monte Carlo	363,106,047.35	11,153,185,807,589,300.00
PCE-Projeção	362,376,519.98	10,749,386,533,295,900.00
PCE-Regressão	361,453,847.82	14,196,314,285,608,500.00
PG	361,563,693.27	12,413,092,352,971,800.00
Produção Acumulada de Água		
	Média	Variância
Monte Carlo	5,019,276.05	32,386,882,380,460.30
PCE-Projeção	4,894,527.96	31,568,285,744,982.30
PCE-Regressão	5,204,298.45	42,969,128,441,697.60
PG	4,922,285.04	32,715,248,581,023.80

Pode-se observar na Tabela 7.5 que as aproximações da média e a variância para cada saída do modelo estão próximas dos valores calculados com o Monte Carlo de 300.000 de simulações. Desta maneira é demonstrado o desempenho das técnicas propostas na estimação de estatísticos. Na Tabela 7.6 é apresentado o resumo dos intervalos de variação nas aproximações destes valores.

Tabela 7.6: Variação das estimações da média e a variância no modelo de 3 parâmetros incertos.

Produção Acumulada de Óleo		
	Média	Variância
PCE-Projeção	0.4%	2.2%
PCE-Regressão	1.3%	18.7%
PG	1.0%	2.1%
Produção Acumulada de Gás		
	Média	Variância
PCE-Projeção	0.2%	3.6%
PCE-Regressão	0.5%	27.3%
PG	0.4%	11.3%
Produção Acumulada de Água		
	Média	Variância
PCE-Projeção	2.5%	2.5%
PCE-Regressão	3.7%	32.7%
PG	1.9%	1.0%

7.3.3. Estudo de Caso 2

A fim de determinar o desempenho das técnicas propostas neste trabalho, na quantificação de incertezas em problemas de média-alta dimensão, foi desenvolvido um modelo de simulação de reservatórios de petróleo com onze variáveis incertas, descritas na seção 3.3.1. e enumeradas a seguir.

1. Multiplicador da porosidade (MULTIPOR)
2. Multiplicador da permeabilidade horizontal (MULTIPERMI)
3. Multiplicador da permeabilidade vertical (MULTIPERMK)
4. Compressibilidade da rocha (CPOR)
5. Raio do aquífero (AQRADIUS)
6. Expoente de Corey da permeabilidade relativa ao gás (NG)
7. Máxima permeabilidade relativa ao gás (KRGMAX)
8. Expoente de Corey da permeabilidade relativa à água (NW)
9. Máxima permeabilidade relativa à água (KRWMAX)
10. Profundidade de contato óleo-água (DELTAWOC)
11. Profundidade de contato óleo-gás (DELTAGOC)

Estas variáveis serão modeladas mediante uma distribuição de probabilidade normal e vamos supor que são mutuamente independentes. Os parâmetros desta modelagem são mostrados na Tabela 7.7.

Tabela 7.7: Parâmetros da modelagem probabilística das variáveis incertas no modelo.

Parâmetro	Valor Base	Média	Desvio Padrão	Mínimo	Máximo
MULTIPOR	1	1	0.05	0.85	1.15
MULTIPERMI	0	0	0.3	-1	1
MULTIPERMK	0	0	0.3	-1	1
CPOR	0.000000450	0.000000450	0.00000050	0.000003	0.000006
AQRADIUS	3000	3000	1000	1	6000
NG	3	3	0.5	1.5	4.5
KRGMAX	0.2	0.2	0.05	0.05	0.35
NW	4	4	0.5	2.5	5.5
KRWMAX	0.585	0.6	0.1	0.3	0.9
DELTAWOC	0	0	1	-3	3
DELTAGOC	0	0	1	-3	3

No caso dos multiplicadores da permeabilidade, os valores utilizados nas simulações são a décima potência, isto pelo comportamento logarítmico que pode-se observar empiricamente nestes parâmetros. Para o caso dos contatos de fluidos, por facilidade foram consideradas as médias zero, mas, são introduzidos nas simulações os valores $2395 + \text{contato óleo-água}$ e $2355 + \text{contato óleo-gás}$.

Por outro lado, da mesma maneira que no estudo de caso 1, é preciso realizar uma análise paramétrica, para determinar se existe influência das variáveis consideradas incertas nas respostas do modelo. Nas Figura 7.10 - Figura 7.17 podem-se observar as variações na produção acumulada de óleo, gás e água, quando são alterados de forma independente cada um dos onze parâmetros. No caso dos multiplicadores da porosidade e da permeabilidade, os gráficos são os apresentados no caso anterior.

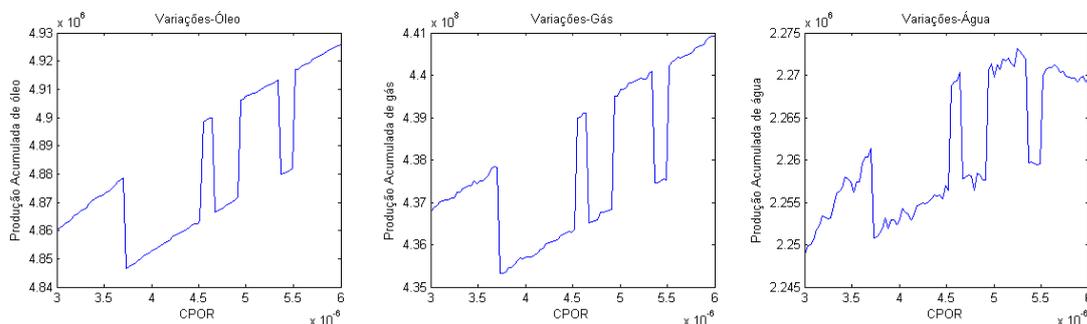


Figura 7.10: Variações na produção acumulada de óleo, gás e água, quando varia a compressibilidade da rocha no intervalo $[0.000003, 0.000006]$.

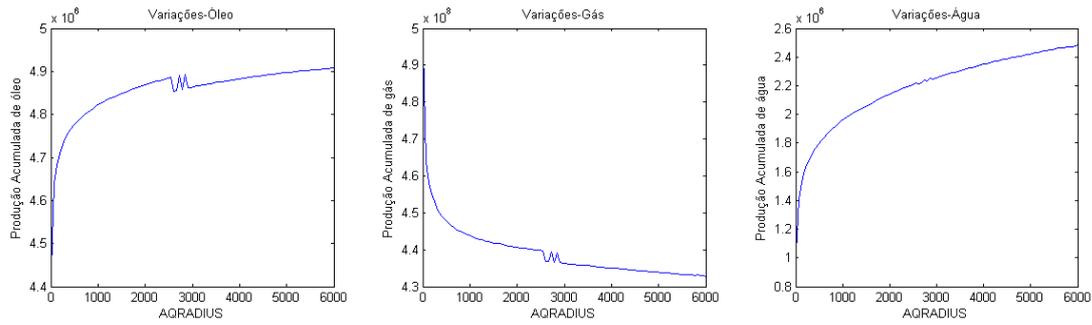


Figura 7.11: Variações na produção acumulada de óleo, gás e água, quando varia o raio do aquífero no intervalo [1, 6000].

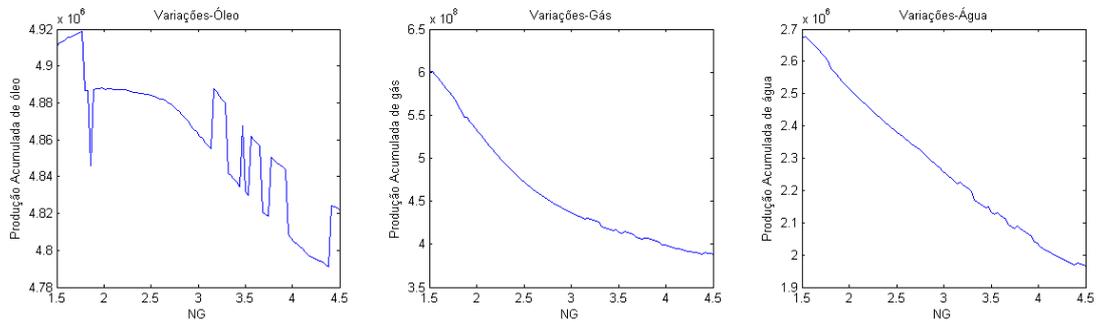


Figura 7.12: Variações na produção acumulada de óleo, gás e água, quando varia o expoente de Corey da permeabilidade relativa ao gás, no intervalo [1.5, 4.5].

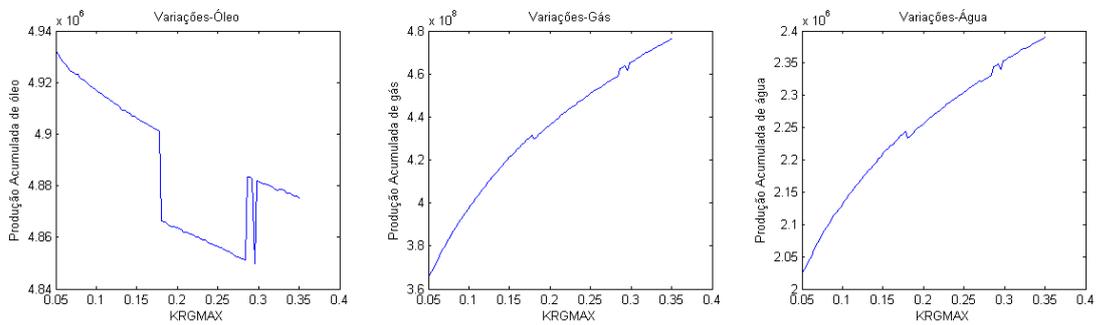


Figura 7.13: Variações na produção acumulada de óleo, gás e água, quando varia o valor máximo da permeabilidade relativa ao gás, no intervalo [0.05, 0.35].

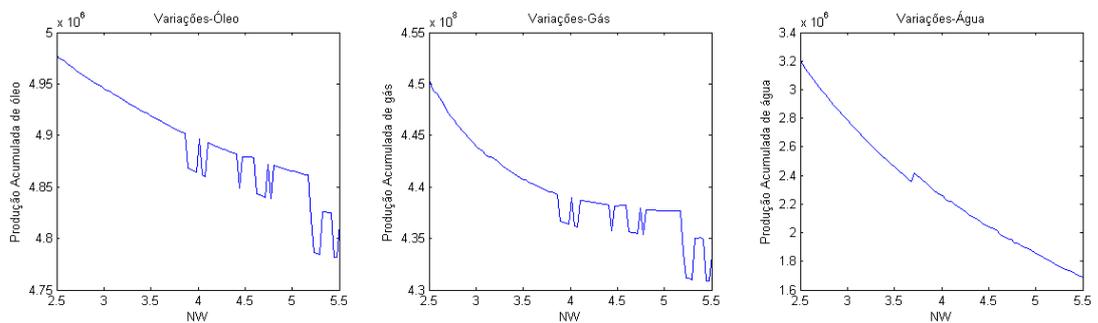


Figura 7.14: Variações na produção acumulada de óleo, gás e água, quando varia o expoente de Corey da permeabilidade relativa à água no intervalo [2.5, 5.5].

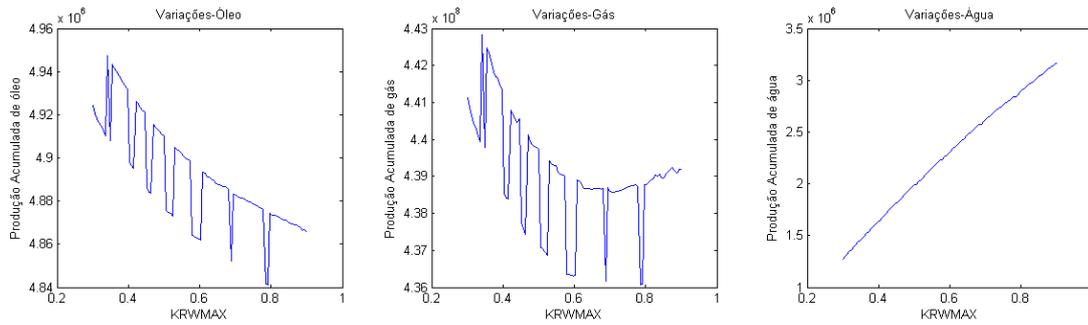


Figura 7.15: Variações na produção acumulada de óleo, gás e água, quando varia o valor máximo da permeabilidade relativa à água no intervalo [0.3, 0.9].

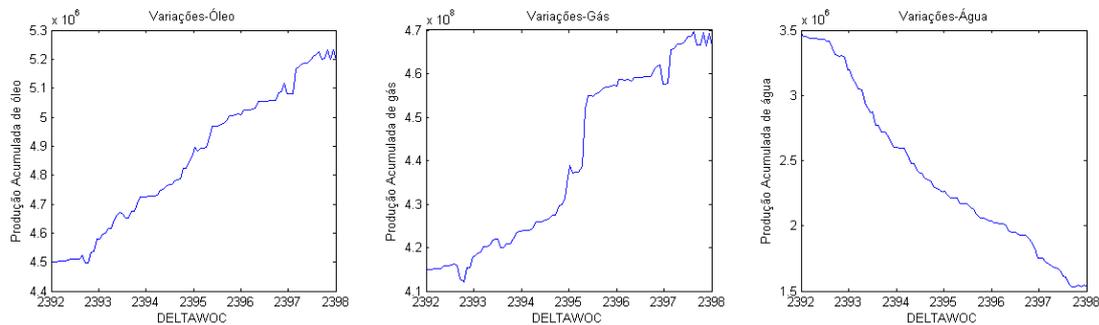


Figura 7.16: Variações na produção acumulada de óleo, gás e água, quando varia a profundidade de contato óleo-água no intervalo [2392, 2398].

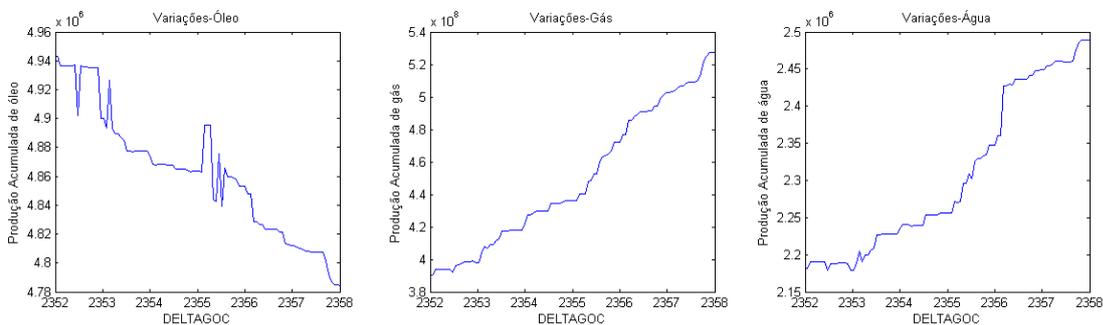


Figura 7.17: Variações na produção acumulada de óleo, gás e água, quando varia a profundidade de contato óleo-gás no intervalo [2352, 2358].

Pode-se observar que em todos os casos, as variáveis incertas afetam as saídas do modelo, portanto, tem sentido propagar a incerteza nelas. É importante mencionar que os parâmetros NG e KRGMAX têm um efeito particular sobre a produção acumulada de óleo, esta saída apresenta-se fortemente irregular e instável frente a variações nestas variáveis. Por outro lado, mudanças nos parâmetros MW e KRWMAX geram comportamentos altamente volúveis nas produções acumuladas de óleo e gás. Na seguinte Figura 7.18 apresenta-se a intensidade e o direcionamento do efeito dos parâmetros aleatórios nas saídas do modelo.

Finalmente, uma vez parametrizadas as entradas aleatórias do modelo e verificado o efeito destas nas saídas, é possível aplicar as técnicas propostas objetivando-se a quantificação da incerteza mediante o uso das soluções obtidas.

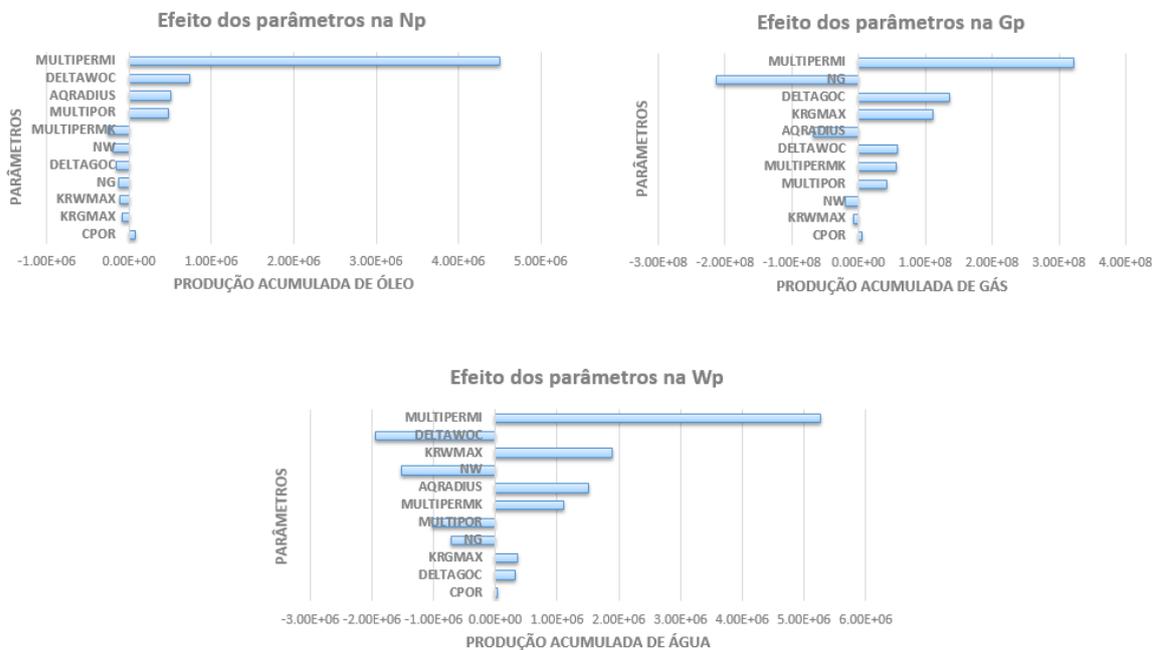


Figura 7.18: Efeito dos parâmetros incertos nas saídas do modelo.

- **Expansões por caos polinomial**

Para o problema considerado, devido a que as entradas são distribuídas normalmente, a expansão foi construída utilizando uma família de polinômios de Hermite, como a base de funções ortogonais (ver Tabela 5.1). Da mesma maneira que no caso de estudo 1, se apresenta o obstáculo de selecionar o grau da expansão e o nível da quadratura utilizada (para o enfoque de projeção). Devido a que é um problema de alta dimensionalidade, deve-se levar em consideração que a quantidade de simulações vai aumentar mais rapidamente que em situações de baixa dimensão, portanto exige-se maior cuidado na escolha destes valores.

No caso do enfoque por regressão, o algoritmo descrito na seção 5.4 fornece as amostras com maior probabilidade de ocorrência e a quantidade mínima de amostras, necessária para que a matriz de informação seja invertível. Este conjunto constitui o planejamento de experimentos que é utilizado para aproximar os coeficientes da expansão.

Nesta seção foram realizados experimentos numéricos variando-se o grau da expansão com níveis de quadratura inclusive nitidamente abaixo do necessário, apenas para levantar o comportamento do método nesta situação extrema. Na Tabela 7.8 são apresentados, para cada solução de PCE, os valores do coeficiente de determinação (R^2), gerado com um conjunto de 50,000 amostras avaliadas nas soluções da PCE (ver eq. (5.24)). Lembre-se que R^2 permite ter um conhecimento do ajuste que foi alcançado pelo modelo de aproximação.

Tabela 7.8: Coeficientes de determinação para cada solução obtida mediante expansões por caos polinomial com onze dimensões.

	Grau da PCE	Número de Coeficientes	Nível da Quadratura	Número de Simulações	R^2 Óleo	R^2 Gás	R^2 Água
Projeção	2	78	3	243	0.855	0.835	0.980
	3	364	3	243	0.826	0.685	0.969
Regressão	2	78	-	133	0.968	0.893	0.980

Pela dimensionalidade deste problema foi considerado somente um nível de quadratura igual a três, devido a que, para o próximo nível ($\ell = 4$) são necessárias 1607 simulações, o qual é inviável para os modelos que estão sendo estudados neste trabalho. No caso da aproximação dos coeficientes por regressão, o planejamento de experimentos com menor quantidade de amostras e com maior probabilidade de ocorrência tem 133 elementos.

Note-se na Tabela 7.8 que as estimações mais precisas foram alcançadas com uma expansão de grau dois, portanto foram selecionadas as soluções para PCE com projeção e com regressão deste grau. Uma observação importante é que as estimações construídas por meio da técnica de regressão têm um ajuste superior às desenvolvidas por meio da projeção espectral, apesar de ser este último método o que garante a obtenção do polinômio mais próximo à função que se deseja estimar. Isto é devido a que, para um problema de alta dimensionalidade como o que está-se abordando neste trabalho, as aproximações dos coeficientes da expansão precisam de uma quantidade maior de avaliações para convergir ao valor certo.

- **Programação Genética Multi- Gene**

Para problemas de alta dimensionalidade, como aquele abordado nesta seção, o algoritmo de PG é executado da mesma maneira que no caso de estudo anterior. Este mecanismo não apresenta dificuldades maiores na construção das expressões simbólicas que aproximam cada saída. Isto devido a que é um algoritmo de aprendizado que utiliza um histórico de dados para treinar-se e conseguir representar a relação entre as informações fornecidas. A principal complicação é enfrentar um espaço de busca maior e obter uma solução possivelmente mais extensa, portanto menos manipulável.

Da mesma maneira que no caso de estudo 1, foi gerada uma matriz de dados correspondentes às entradas e saídas do modelo considerado. As entradas foram números aleatórios distribuídos seguindo a Tabela 7.7 e as saídas correspondentes foram obtidas mediante o simulador IMEX. Esta base de dados foi separada em conjuntos de treino e validação, a quantidade e proporção das amostras em cada conjunto foi escolhida empiricamente. Os operadores de cruzamento, mutação e reprodução foram os mesmos utilizados no caso de estudo 1. As funções usadas para a construção das soluções candidatas foram: soma, subtração, produto, potência quadrada, função mínimo e função máximo. Note-se que as últimas duas funções, permitem a construção de soluções não polinomiais.

Finalmente, uma vez selecionados os parâmetros do algoritmo de programação genética multi-gene, se rodaram trinta vezes cada algoritmo, com o alvo de garantir que os resultados são consistentes, que os erros pertencem a um intervalo pequeno e a convergência é mantida. A melhor solução candidata para cada programa e os valores dos parâmetros são apresentados na seguinte Tabela 7.9. Além do coeficiente de determinação gerado com um conjunto de 50,000 amostras (ver eq. (5.24)).

Tabela 7.9: Valores dos parâmetros e coeficiente de determinação para cada programa genético no caso de onze dimensões.

Parâmetros do Algoritmo de PGMG	Produção Acumulada de Óleo	Produção Acumulada de Gás	Produção Acumulada de Água
Tamanho da População	150	150	250
Número de Gerações	500	500	750
Tamanho do Torneio	4	4	4
Elitismo	0.05	0.05	0.05
Taxa de Mutação	0.3	0.3	0.3
Taxa de Cruzamento	0.65	0.65	0.65
Máxima Profundidade dos Genes	3	3	3
Máximo número de Genes	35	35	35
Intervalo de variação dos nós constantes	[-10,10]	[-10,10]	[-10,10]
Amostras no Treino	80	80	200
Amostras na Validação	20	20	151
Número de Simulações	100	100	351
Coeficiente de Determinação: R^2	0.982	0.915	0.991

Pode-se observar que o algoritmo de PG multi-gene consegue estimar as três saídas do modelo com um ajuste superior a 91% e com um custo computacional, medido em chamadas ao simulador, inferior a 352 avaliações. Estes resultados evidenciam o desempenho do algoritmo de programação genética para problemas de alta dimensionalidade com incerteza nos parâmetros de entrada.

Por outro lado, comparando o desempenho das soluções obtidas mediante PCE e PG, pode-se destacar que para a produção acumulada de óleo e gás, da mesma maneira que no estudo de caso 1, PG consegue ajustar o modelo com maior precisão e realizando uma quantidade menor de simulações. Pelo contrário, na produção acumulada da água, foram realizados múltiplos experimentos aumentando a quantidade de simulações, até obter um ajuste superior ao obtido pela PCE. Entre os experimentos realizados pode-se mencionar o desempenho das soluções obtidas com 243 simulações, 180 amostras para o treinamento e 63 amostras para a validação. O coeficiente de determinação médio obtido nos experimentos realizados foi 0.954, este valor apresenta um ajuste inferior à solução obtida mediante PCE com 243 simulações.

Como foi destacado no caso de estudo 1, uma das principais vantagens das técnicas propostas, é o uso das soluções finais como modelos substitutos (*proxies*) sob incerteza. Portanto, uma vez que são construídas estas expressões, é possível realizar análises estatísticas e probabilísticas evitando simular novamente o campo de petróleo, diminuindo consideravelmente o custo computacional de estes estudos. Fazendo uso desta vantagem, foi estimada a distribuição de probabilidade empírica da produção acumulada de óleo, gás e água no modelo de simulação considerado. Para validar os resultados obtidos, foi desenvolvido o método de simulação de Monte Carlo com 1,000,000 de amostras, tentando conseguir a convergência. As distribuições de probabilidade empírica estimadas com PCE e PG foram comparadas com Monte Carlo e são apresentadas em Figura 7.19, Figura 7.20 e Figura 7.21.

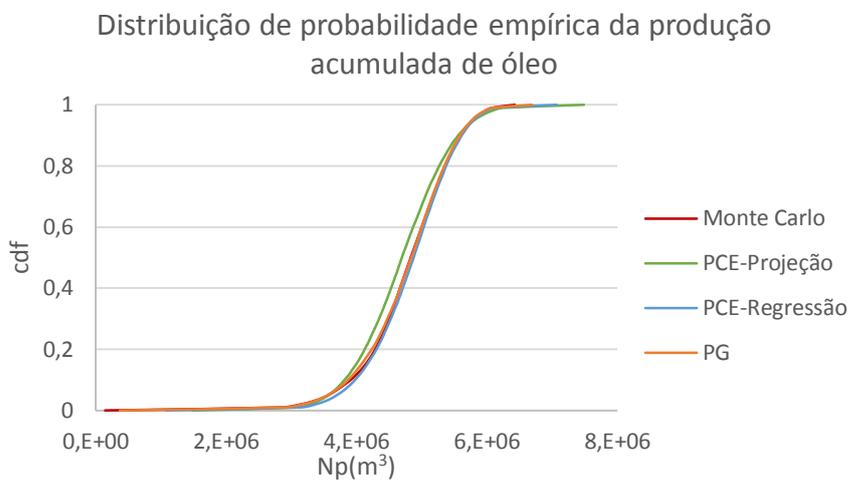


Figura 7.19: Função distribuição de probabilidade empírica da produção acumulada de óleo.

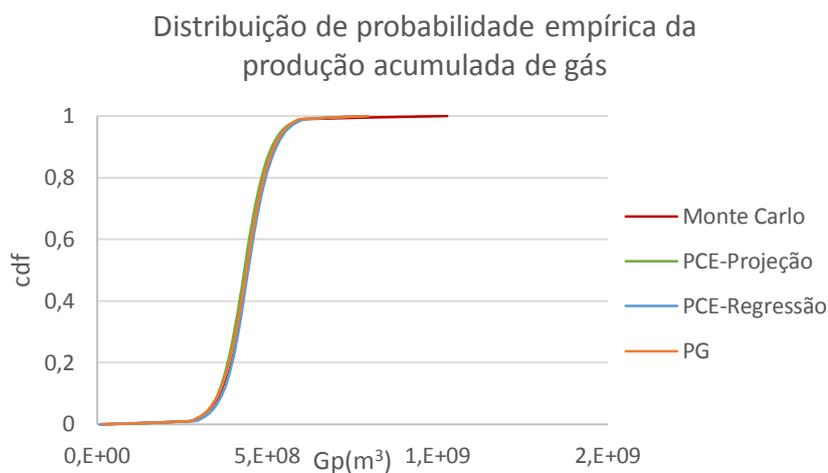


Figura 7.20: Função distribuição de probabilidade empírica da produção acumulada de gás.

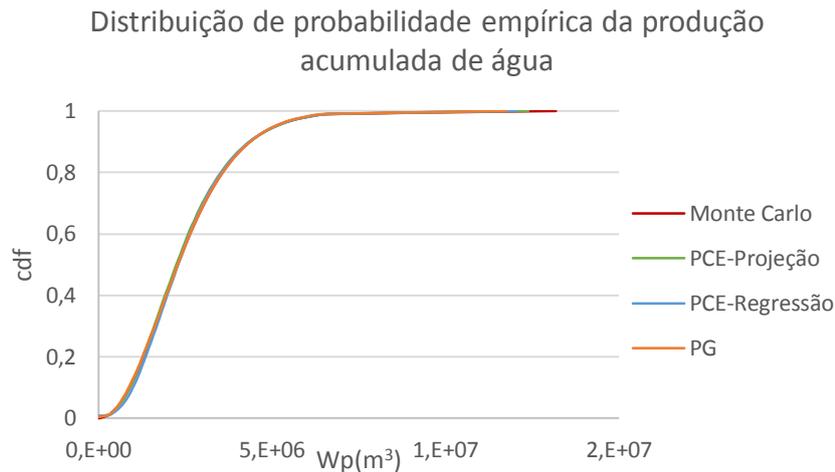


Figura 7.21: Função distribuição de probabilidade empírica da produção acumulada de água.

Observe-se a proximidade das soluções obtidas por meio de PG e PCE com a distribuição calculada a partir do Monte Carlo, com o qual é evidenciada e validada a performance dos métodos propostos para a quantificação de incertezas em modelos de simulação de alta dimensão. Na Figura 7.22 são mostrados os coeficientes de determinação e o número de simulações para cada técnica.

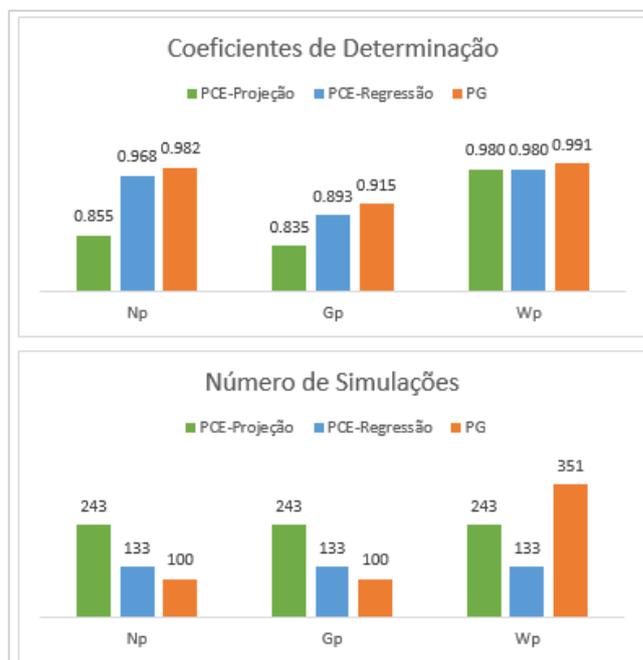


Figura 7.22: Coeficientes de determinação e número de simulações para cada aproximação, modelo de onze parâmetros.

Na Tabela 7.10 são apresentadas as médias e as variâncias para cada saída do modelo de simulação. No caso de PCE, estes valores foram estimados utilizando diretamente os coeficientes da expansão, como no estudo de caso 1. Para o método de PG, foi realizada uma amostragem direta com 100.000 experimentos gerados aleatoriamente seguindo a distribuição de probabilidade dos parâmetros de entrada incertos.

Tabela 7.10: Média e variâncias das saídas do modelo de 11 parâmetros incertos.

Produção Acumulada de Óleo		
	Média	Variância
Monte Carlo	4,784,762.93	457,521,495,659.87
PCE-Projeção	4,696,136.00	468,962,957,578.20
PCE-Regressão	4,826,829.86	417,938,277,724.43
PG	4,774,635.13	467,345,555,372.88
Produção Acumulada de Gás		
	Média	Variância
Monte Carlo	439,324,838.22	4,381,975,348,297,610.00
PCE-Projeção	432,537,796.40	4,184,714,063,800,560.00
PCE-Regressão	444,717,941.82	4,232,855,593,351,630.00
PG	437,510,931.31	4,455,862,080,407,370.00
Produção Acumulada de Água		
	Média	Variância
Monte Carlo	2,488,395.45	1,943,708,422,925.22
PCE-Projeção	2,464,429.77	1,953,262,113,640.13
PCE-Regressão	2,506,720.02	1,867,838,283,826.18
PG	2,480,972.32	1,927,682,189,795.36

Pode-se observar na Tabela 7.10 que as aproximações da média e a variância para cada saída do modelo estão próximos dos valores calculados com o Monte Carlo de 1.000.000 de simulações. Desta maneira é demonstrado o desempenho das técnicas propostas na estimação de estatísticos. Na Tabela 7.11 é apresentado o resumo dos intervalos de variação nas aproximações destes valores.

Tabela 7.11: Variação das estimações da média e da variância.

Produção Acumulada de Óleo		
	Média	Variância
PCE-Projeção	1.9%	2.5%
PCE-Regressão	0.9%	8.7%
PG	0.2%	2.1%
Produção Acumulada de Gás		
	Média	Variância
PCE-Projeção	1.5%	4.5%
PCE-Regressão	1.2%	3.4%
PG	0.4%	1.7%
Produção Acumulada de Água		
	Média	Variância
PCE-Projeção	1.0%	0.5%
PCE-Regressão	0.7%	3.9%
PG	0.3%	0.8%

Finalmente, para o problema estudado mostra-se um desempenho superior, em termos do custo computacional e a porcentagem de ajuste, nas soluções de PG nos casos da produção acumulada de óleo e gás. Na estimação da distribuição de probabilidade da produção acumulada de água, PG precisou uma quantidade de simulações consideravelmente maior que PCE, para conseguir um maior nível de ajuste.

7.3.4. Estudo de Caso 3

No último estudo de caso desenvolvido foram tratados vinte e três parâmetros incertos afetando a produção acumulada de óleo, gás e água no reservatório PUNQ-S3. Este campo de petróleo tem cinco camadas, determinadas no caso de pesquisa original (IMPERIAL COLLEGE LONDON) mediante conhecimento geológico do reservatório real. As camadas 1, 3 e 5 consistem em canais fluviais recobertos por argilito. A camada 2 representa argila marina ou lagunar com alguns depósitos irregulares e a camada 4 está formada por um delta lagunar envolto em argila lagunar. As propriedades do campo, supostas incertas, foram as mesmas utilizadas no estudo de caso 2, com a importante diferença de que os multiplicadores da porosidade, da permeabilidade vertical e a permeabilidade horizontal foram considerados parâmetros diferentes em cada camada do reservatório. Desta maneira, trabalhou-se com 15 multiplicadores (3

multiplicadores para cada uma das cinco camadas) e as restantes 8 variáveis globais do campo.

O principal objetivo da formulação deste novo modelo de simulação está na análise do desempenho das técnicas propostas neste trabalho para um número grande de parâmetros incertos. A descrição das variáveis aleatórias e a análise paramétrica foi exposta na seção 7.3.3, neste caso foram consideradas as mesmas distribuições de probabilidade e suas características.

Seguidamente vai-se desenvolver uma análise da quantificação de incertezas num modelo de simulação de reservatórios de petróleo, que apresenta vinte e três parâmetros de entrada incertos supostos mutuamente independentes.

- **Expansões por caos polinomial**

A expansão foi construída utilizando uma família de polinômios de Hermite, como a base de funções ortogonais, isto devido a que as entradas são distribuídas normalmente. Em casos de dimensão alta, a escolha do grau do polinômio e o nível da quadratura utilizada deve ser realizada com cuidado. Destes valores depende grandemente a quantidade de simulações necessárias para a construção das soluções e portanto, o custo computacional associado à quantificação da incerteza nas saídas do modelo. Na Tabela 7.12 mostra-se o custo computacional medido em chamadas ao simulador, para cada seleção do grau e nível de quadratura, no enfoque da projeção. Lembre-se que o nível da quadratura, que aproxima a integral multidimensional dos coeficientes, deve ser pelo menos o grau da expansão mais um ($\ell \geq p + 1$), para ter a acurácia desejada.

Tabela 7.12: Dependência entre o número de simulações e os parâmetros da PCE com dimensão 23, utilizando o enfoque da projeção.

Grau da PCE	Nível da Quadratura	Número de Simulações
2	3	1,059
2, 3	4	15,319
2, 3, 4	5	161,093

Observe-se que o único caso viável em termos de custo computacional é o mais simples com grau da expansão dois e nível de quadratura três. Desta forma, vai construir-se um polinômio de grau 2 com 300 coeficientes definidos por

integrais de dimensão 23, e cada um deles vai ser aproximado por uma quadratura com nível 3. Devido à alta complexidade do problema considerado, a construção mais simples não conseguiu alcançar bom ajustes na aproximação da distribuição de probabilidade empírica da produção acumulada de óleo, gás e água. O coeficiente de determinação calculado com 100,000 amostras mostrou ajustes inferiores ao 60%. Por causa de isto não será considerada esta metodologia neste estudo de caso.

Por outro lado, no enfoque da regressão a quantidade de simulações está relacionada diretamente com o número de coeficientes da expansão. Na Tabela 7.13 é apresentada esta relação.

Tabela 7.13: Relação entre o número de coeficientes da PCE de dimensão 23 com o grau selecionado.

Grau da PCE	Número de Coeficientes
2	300
3	2,600
4	17,550

Devido a que o planejamento de experimentos é construído com uma quantidade de amostras igual ou superior ao número de coeficientes da PCE, a solução com um custo computacional aceitável é uma expansão de grau 2. A construção de um planejamento otimizado utilizando as raízes dos polinômios de Hermite de grau 3 (como na seção 5.4) requer para este caso de 23 dimensões, a criação de $(p + 1)^M = 4^{23} \approx 7.03 \times 10^{13}$ experimentos e o cálculo da sua norma euclidiana. Isto é inviável computacionalmente, portanto, para este caso foi construído o planejamento de experimentos de maneira aleatória, seguindo a Tabela 7.7 e utilizando a proposta empírica apresentada na seção 5.4 pelo (BERVEILLER, 2005). Os resultados são mostrados em Tabela 7.14, o coeficiente de determinação foi calculado com um conjunto de 100,000 amostras.

Tabela 7.14: Coeficientes de determinação para a solução obtida mediante expansões por caos polinomial com 23 dimensões.

Grau da PCE	Número de Coeficientes	Número de Simulações	R^2 Óleo	R^2 Gás	R^2 Água
2	300	600	0.999	0.974	0.965

Pode-se observar o alto desempenho da técnica, medido no número de chamadas ao simulador e o ajuste da solução, evidenciado no coeficiente de determinação calculado com 100,000 amostras.

- **Programação Genética Multi-Gene**

Apesar da alta dimensão do problema considerado, o algoritmo de programação genética não apresenta dificuldades na construção das expressões simbólicas que aproximam a produção acumulada de óleo, gás e água. O único obstáculo é um espaço de busca maior que pode afetar o processo de evolução do algoritmo.

Da mesma forma que nos casos de estudo 1 e 2, foi gerada uma matriz de dados correspondentes às entradas e saídas do modelo considerado. As entradas foram números aleatórios distribuídos seguindo a Tabela 7.7 e as saídas correspondentes foram obtidas mediante o simulador IMEX. Os operadores de cruzamento, mutação e reprodução foram os mesmos utilizados no caso de estudo 2. As funções usadas para a construção das soluções candidatas foram: soma, subtração, produto, potência quadrada, função mínimo e função máximo.

Uma vez selecionados os parâmetros de PG, se rodou trinta vezes cada algoritmo, garantindo a consistência e a convergência das soluções. A melhor solução candidata para cada programa e os valores dos parâmetros são apresentados na seguinte Tabela 7.15, junto com o coeficiente de determinação gerado com um conjunto de 100,000 amostras.

Tabela 7.15: Valores dos parâmetros e coeficiente de determinação para cada programa genético no caso de vinte e três dimensões.

Parâmetros do Algoritmo de PGMG	Produção Acumulada de Óleo	Produção Acumulada de Gás	Produção Acumulada de Água
Tamanho da População	150	150	250
Número de Gerações	500	500	800
Tamanho do Torneio	4	4	4
Elitismo	0.05	0.05	0.05
Taxa de Mutação	0.3	0.3	0.3
Taxa de Cruzamento	0.65	0.65	0.65
Máxima Profundidade dos Genes	3	3	3
Máximo número de Genes	35	35	35
Intervalo de variação dos nós constantes	[-10,10]	[-10,10]	[-10,10]
Amostras no Treino	300	300	450
Amostras na Validação	70	70	100
Número de Simulações	370	370	550
Coeficiente de Determinação: R^2	0.999	0.985	0.971

O algoritmo de PG multi-gene consegue estimar as três saídas do modelo com um ajuste superior a 97% e com um custo computacional, medido em chamadas ao simulador, inferior a 551 avaliações. Estes resultados evidenciam o desempenho da técnica proposta em problemas de alta dimensionalidade. Pode-se destacar que as 370 simulações, utilizadas no aprendizado para o óleo e o gás, estão contidas nas 550 simulações usadas no caso da água. Por tanto, foram necessárias somente 550 avaliações no simulador para aproximar as três saídas do modelo.

Da mesma maneira que nos estudos de caso anteriores, foram utilizadas as soluções das técnicas como *proxies* sobre incerteza para aproximar a distribuição de probabilidade de cada saída do modelo. Além disso foi desenvolvido o método de Monte Carlo com 1,000,000 de simulações com a finalidade de validar os resultados obtidos. As curvas de distribuição de probabilidade são apresentadas em Figura 7.23, Figura 7.24 e Figura 7.25. Devido ao baixo ajuste no enfoque da projeção com PCE, somente são mostradas as curvas obtidas com o enfoque da regressão e o algoritmo de PG. Isto já que, as curvas imprecisas prejudicam a escala do gráfico e não permitem visualizar claramente as aproximações realizadas com as outras técnicas.

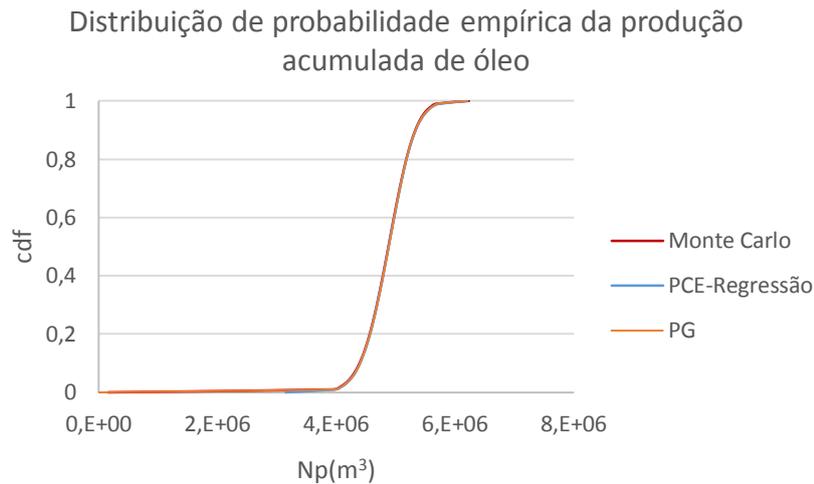


Figura 7.23: Função distribuição de probabilidade empírica da produção acumulada de óleo.

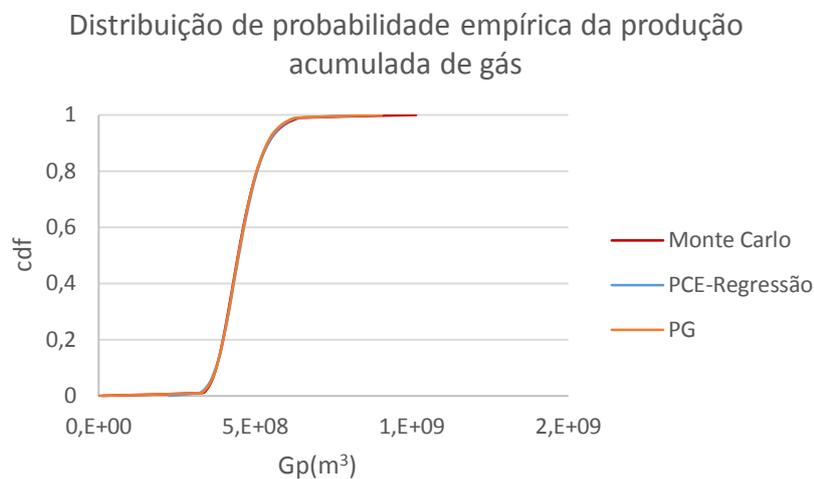


Figura 7.24: Função distribuição de probabilidade empírica da produção acumulada de gás.

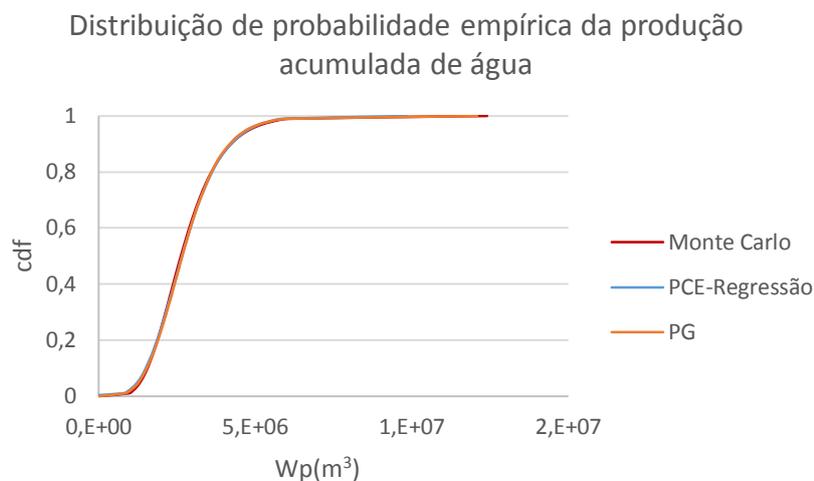


Figura 7.25: Função distribuição de probabilidade empírica da produção acumulada de água.

Observe-se a proximidade das distribuições empíricas, aproximadas utilizando PCE e PG, com a distribuição de probabilidade calculada mediante o Monte Carlo de 1,000,000 simulações. Os erros mínimos nas estimações são evidenciados nas curvas mostradas. Na Figura 7.26 pode-se comparar o ajuste de cada solução com a quantidade de simulações utilizadas.

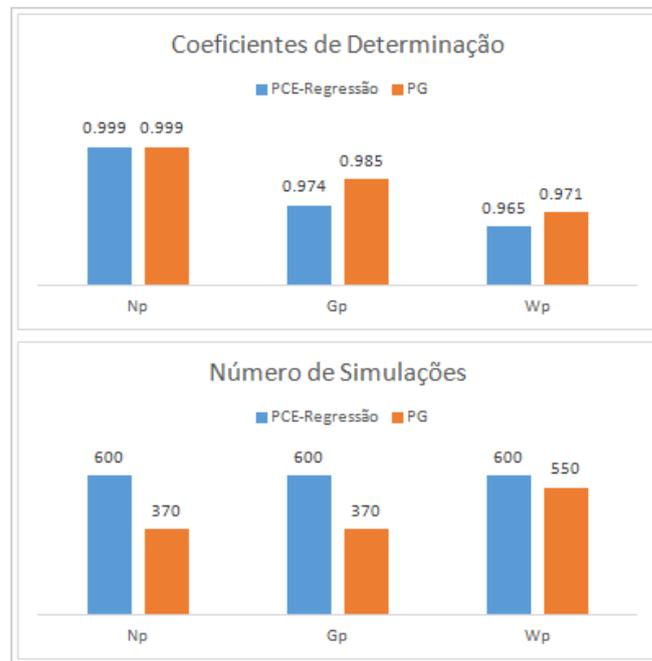


Figura 7.26: Coeficientes de determinação e número de simulações para cada aproximação, modelo de vinte e três parâmetros.

Comparando os resultados de PCE e PG neste estudo de caso, pode-se afirmar que o algoritmo de programação genética consegue para as três saídas um nível de ajuste superior com um número de simulações inferior. Evidenciando um melhor desempenho na quantificação de incertezas neste modelo de simulação de reservatórios.

Finalmente, na Tabela 7.16 são apresentadas as estimações da média e da variância das três saídas consideradas no estudo, produção acumulada de óleo, gás e água. No caso da PCE estes valores foram calculados a partir das equações enunciadas na seção 5.5. Em PG foi realizada uma amostragem direta com 100,000 amostras.

Tabela 7.16: Média e variâncias das saídas do modelo de 23 parâmetros incertos.

Produção Acumulada de Óleo		
	Média	Variância
Monte Carlo	4,866,774.17	135,945,577,114.65
PCE-Regressão	4,872,552.06	135,952,102,173.08
PG	4,871,454.17	138,940,182,109.30
Produção Acumulada de Gás		
	Média	Variância
Monte Carlo	453,008,851.59	4,482,410,396,913,160.00
PCE-Regressão	453,293,155.38	4,473,625,557,509,570.00
PG	452,012,605.42	4,187,462,289,476,150.00
Produção Acumulada de Água		
	Média	Variância
Monte Carlo	2,784,715.84	1,181,817,926,941.62
PCE-Regressão	2,788,035.15	1,192,208,412,288.20
PG	2,790,543.79	1,168,216,475,261.49

Na Tabela 7.17 é apresentado o resumo dos intervalos de variação nas aproximações para a média e a variância.

Tabela 7.17: Variação das estimações da média e da variância.

Produção Acumulada de Óleo		
	Média	Variância
PCE-Regressão	-0.119%	0.005%
PG	-0.096%	-2.203%
Produção Acumulada de Gás		
	Média	Variância
PCE-Regressão	-0.063%	0.196%
PG	0.220%	6.580%
Produção Acumulada de Água		
	Média	Variância
PCE-Regressão	-0.119%	-0.879%
PG	-0.209%	1.151%

As estimações obtidas para a média e a variância apresentam uma precisão superior no método da regressão com PCE que no algoritmo de PG. Isto devido a que, no primeiro caso estes valores são aproximados mediante cálculos diretos a partir dos coeficientes da expansão. Por outro lado, no caso de PG estas aproximações dependem do conjunto de experimentos gerado aleatoriamente, com

o qual é realizada uma amostragem direta na expressão explícita obtida pelo algoritmo evolutivo. Desta maneira, as estimações mudam de acordo com o conjunto de amostras utilizado e a quantidade.

8 Conclusões e Trabalhos Futuros

8.1. Conclusões e Contribuições

De forma a contribuir com o tratamento de incertezas nos modelos de simulação de reservatórios de petróleo, este trabalho desenvolveu e testou algoritmos baseados nas técnicas de PG e caos polinomial. As duas metodologias foram empregadas na quantificação de incertezas conseguindo reduzir o custo computacional demandado neste tipo de análises. Ao contrário de outras técnicas, como o método de simulação de Monte Carlo, as soluções dos algoritmos propostos são funções explícitas, portanto podem ser utilizadas como proxies sob incerteza do modelo de simulação estudado. Com o fim de validar o desempenho de PCE e PG, construíram-se dois modelos de simulação de reservatórios, baseados no campo PUNQ-S3. Os quais consideraram três e onze parâmetros incertos, com distribuições log-normais e normais, respectivamente.

Expansões por caos polinomial conseguiu representar por meio de um polinômio estocástico, a produção acumulada de óleo, gás e água. Os coeficientes destes polinômios foram estimados por meio da quadratura de Smolyak (no enfoque de projeção) e por meio de um planejamento de experimentos (no enfoque de regressão). As expansões obtidas pelas duas metodologias mostraram um desempenho similar e satisfatório, em termos da porcentagem de ajuste. Pode-se destacar que, a abordagem da projeção-Smolyak garante a convergência em média quadrática da expansão polinomial ao polinômio de melhor ajuste, o qual não acontece com o enfoque da regressão. Além disso, a quantidade de simulações com Smolyak é fixa, uma vez selecionado o nível da quadratura e o grau da expansão. No caso da regressão, é preciso um esforço computacional adicional, para calcular o número de amostras do planejamento de experimentos.

Alternativamente, foi pesquisado o algoritmo de programação genética multi-gene, um método menos analítico, que apresenta uma evolução iterativa na construção de suas soluções. Este algoritmo é treinado com uma matriz de dados,

com o objetivo de representar por meio de uma expressão simbólica, a relação existente entre as entradas e as saídas do modelo estudado. A matriz foi construída com entradas, geradas aleatoriamente seguindo uma distribuição de probabilidade definida a priori, e saídas, que são as correspondentes respostas do simulador. O algoritmo de PGMG conseguiu representar a produção acumulada de óleo, gás e água por meio de uma expressão estocástica, não necessariamente polinomial. O desempenho das estimativas foi evidenciado com ajustes superiores a 90%.

As soluções obtidas por PCE e PGMG foram utilizadas como *proxies* sob incerteza, para realizar estatísticas e estudos probabilísticos, que quantifiquem a incerteza nas saídas do modelo de simulação. Foi calculada a distribuição de probabilidade empírica de cada saída do modelo (óleo, gás e água) e comparada com o método de simulação de Monte Carlo. Nos três casos de estudo as soluções foram próximas às obtidas por esta técnica de simulação tradicional, evidenciando a eficiência de PCE e PGMG, em termos da precisão das aproximações e do custo computacional, medido no número de chamadas ao simulador. No entanto, PG alcançou nos três testes níveis de ajuste superiores aos de caos polinomial. Isto possivelmente devido a que PG permite representar as saídas do modelo como uma combinação de funções de qualquer tipo (lógicas, logarítmicas, exponenciais, etc.). Desta maneira, é generalizado o processo de busca da melhor aproximação da saída do modelo.

- **Principais contribuições**

Os principais aportes deste trabalho são enunciados a seguir,

1. Na literatura pesquisada não foram encontradas aplicações do algoritmo de programação genética para a quantificação de incertezas. Ainda menos na abrangência de reservatórios de petróleo. Portanto, torna-se importante a contribuição realizada nesta área.
2. Apesar de ter encontrado aplicações de caos polinomial para a quantificação de incertezas na área de simulação de reservatórios de petróleo, não se encontrou uso dos enfoques de Smolyak e da regressão para o cálculo dos coeficientes da expansão.
3. Na pesquisa bibliográfica realizada não encontrou-se soluções para problemas de alta dimensionalidade como aqueles tratados neste trabalho.

Ainda menos com resultados satisfatórios em termos de custo computacional e precisão.

4. Neste trabalho, as soluções das técnicas propostas foram utilizadas como *proxies* sobre incerteza, num processo posterior ao cálculo de estatísticos básicos, para aproximar a distribuição de probabilidade empírica. Comumente, PCE é só empregado para quantificar a incerteza por meio da estimação da média e a variância, que podem ser obtidas diretamente dos coeficientes da expansão.

8.2. Trabalhos Futuros

A partir desta dissertação destacam-se os seguintes possíveis trabalhos futuros:

- Aproveitar a quantidade pequena de dados necessários no aprendizado do algoritmo de programação genética, para realizar experimentos adicionais utilizando a técnica de validação cruzada, com a finalidade de tentar aumentar os níveis de ajuste.
- Acoplar ao método de caos polinomial à decomposição de Karhunen-Loève, com o fim de melhorar o desempenho. Esta decomposição é particularmente útil para métodos estocásticos baseados em expansões, pois tem por objetivo a redução do espaço probabilístico das entradas estocásticas de ordem alta.
- Aplicação das técnicas propostas em problemas mais complexos, por exemplo em modelos baseados em outros campos, com simulações mais custosas computacionalmente, ou acrescentado variáveis incertas não distribuídas identicamente.
- Utilizar outras técnicas para aproximar as saídas de modelos estocásticos, como por exemplo, algoritmos genéticos para a otimização dos coeficientes de um polinômio ou métodos de superfícies de resposta. Isto com o objetivo de comparar os modelos.
- Implementação (*plug-in*) do trabalho realizado num software de análise de incerteza em modelos de simulação de reservatórios, como por exemplo CMOST, da CMG ou no software OCTOPUS, do ICA.

Referências Bibliográficas

- ADAMSON, G., et al. 1996. *Simulation throughout the life of a reservoir*. s.l. : Oilfield Review, 1996. pp. 16-27. Vol. 2.
- ASKEY, R. e WILSON, J. 1985. *Some basic Hypergeometric Polynomials That Generalize Jacobi Polynomials*. s.l. : Memoirs of the American Mathematical Society, 1985.
- BASZENKI, G. e DELVOS, F. 1993. *Multivariate Boolean midpoint rules*. In: *Numerical Integration IV*. s.l. : Birkhäuser Basel, 1993. pp. 1-11. Vol. 112.
- BAZARGAN, H. e CHRISTIE, M. 2012. *Efficient Polynomial chaos Proxy-based History Matching Quantification for Complex Geological Structures*. s.l. : SPE 163282, 2012.
- BAZARGAN, H., CHRISTIE, M. e TCHELEPI, H. 2013. *Efficient Markov Chain Monte Carlo Sampling Using Polynomial Chaos Expansion*. s.l. : SPE,163663, 2013.
- BELLMAN, R. 1956. *Dynamic programming and Lagrange multipliers*. s.l. : Proceedings of the National Academy of Sciences of the United States of America, 1956. Vol. 42.
- BERVEILLER, M. 2005. *Éléments finis stochastiques: approches intrusive et non intrusive pour des analyses de fiabilité*. s.l. : Diss. Université Blaise Pascal-Clermont-Ferrand II, 2005.
- BERVEILLER, M., SUDRET, B e LEMAIRE, M. 2006. *Stochastic finite element: a non intrusive approach by regression*. s.l. : European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique, 2006. pp. 81-92. Vol. 15.
- BLATMAN, G. e SUDRET, B. 2010. *An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis*. s.l. : Probabilistic Engineering Mechanics, 2010. pp. 183-197. Vol. 25.
- . 2008. *Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach*. s.l. : Comptes Rendus Mécanique, 2008. pp. 518-523. Vol. 336.6.

- BONK, T. 1994. *A new algorithm for multi-dimensional adaptive numerical quadrature*. In: *Adaptive Methods—Algorithms, Theory and Applications*. s.l. : Vieweg+ Teubner Verlag, 1994. pp. 54-68.
- BURKE, E., et al. 2009. *Exploring hyper-heuristic methodologies with genetic programming*. s.l. : In *Computational intelligence*, Springer Berlin Heidelberg, 2009. pp. 177-201.
- CAMERON, R. H. e MARTIN, W. T. 1947. *The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals*. s.l. : *Annals of Mathematics*, 1947. pp. 385-392.
- CHOI, S. 2004. *Polynomial chaos expansion with latin hypercube sampling for estimating response variability*. s.l. : *AIAA journal*, 2004. pp. 1191-1198. Vol. 42.
- DESTEFANIS, M. F. e SAVIOLI, G. B. 2007. *Flujo Agua-Petróleo: Simulación Numérica y Análisis de Sensibilidad*. s.l. : *Mecánica Computacional*, ISSN, 2007. pp. 1666-6070.
- ERNST, O., et al. 2012. *On the convergence of generalized polynomial chaos expansions*. s.l. : *ESAIM: Mathematical Modelling and Numerical Analysis*, 2012. pp. 317-339. Vol. 46.
- ERTEKIN, T., ABOU-KASSEM, J. H. e KING, G. R. 2001. *Basic Applied Reservoir Simulation. , Vol. 7. Richardson SPE*. s.l. : *SPE Textbook Series*, 2001. Vol. 7.
- FALLAH-MEHDIPOUR, E., HADDAD, O. e MARIÑO, M. 2012. *Real-time operation of reservoir system by genetic programming*. s.l. : *Water resources management*, 2012. pp. 4091-4103. Vol. 26.
- FISHMAN, G. 2013. *Monte Carlo: concepts, algorithms, and applications*. s.l. : *Springer Science & Business Media*, 2013.
- FOX, B. 1999. *Strategies for Quasi-Monte Carlo*. s.l. : *Springer Science & Business Media*, 1999.
- GAO, G., ZAFARI, M. e REYNOLDS, A. C. 2005. *Quantifying uncertainty for the PUNQ-S3 problem in a Bayesian setting with RML and EnKF*. s.l. : *SPE Reservoir Simulation Symposium*, Society of Petroleum Engineers, 2005.
- GENZ, A. e KEISTER, B. 1996. *Fully symmetric interpolatory rules for multiple integrals over infinite regions with Gaussian weigh.* s.l. : *Journal of Computational and Applied Mathematics*, 1996. pp. 299-309. Vol. 71(2).

- GERSTNER, T. e GRIEBEL, M. 1998. *Numerical integration using sparse grids*. s.l. : Numerical algorithms, 1998. pp. 209-232. Vol. 18.
- GHANEM, R. e SPANOS, P. 1991. *Stochastic finite elements: A Spectral Approach*. New York : Springer, 1991.
- GUEVARA, K. P. 2014. *Propagação de incertezas via expansões por caos polinomial na simulação de reservatórios de petróleo*. Pontifícia Universidade Católica do Rio de Janeiro : Dissertação de Mestrado, 2014.
- HAGAN, A. 2013. *Polynomial Chaos: A Tutorial and Critique from a Statistician's Perspective*. University of Sheffield-UK : Lecture Note, 2013.
- HARALD, N. 1992. *Random number generation and quasi-Monte Carlo methods*. s.l. : Society for Industrial and Applied Mathematics, 1992.
- HENG LI e SARMA, P. 2011. *A comparative study of the probabilistic-collocation and experimental-design methods for petroleum-reservoir uncertainty quantification*. s.l. : SPE Journal, 2011. pp. 429-439. Vol. 16. 2.
- HOLTZ, M. 2010. *Sparse Grid Quadrature in High Dimensions with Applications in Finance and Insurance*. Berlin Heidelberg : Springer, 2010.
- IMPERIAL COLLEGE LONDON. PUNQ-S3. [Online] <http://www3.imperial.ac.uk/earthscienceandengineering/research/perm/punq-s3model>.
- ISUKAPALLI, S. 1999. *Uncertainty analysis of transport-transformation models*. Rutgers, The State University of New Jersey : Tese de Doutorado, 1999.
- KAARNIOJA, V. 2013. *Smolyak Quadratura*. University of Helsinki : Doctoral dissertation, 2013.
- KAYDANI, H., MOHEBBI, A. e EFTEKHARI, M. 2014. *Permeability estimation in heterogeneous oil reservoirs by multi-gene genetic programming algorithm*. s.l. : Journal of Petroleum Science and Engineering, 2014. pp. 201-206. Vol. 123.
- KO, J. e GARNIER, J. 2013. *Multi-element stochastic spectral projection for high quantile estimation*. s.l. : Journal of Computational Physics, 2013. pp. 87-108. Vol. 243.
- KOZA, J. R. 1992. *Genetic programming: on the programming of computers by means of natural selection*. s.l. : MIT press, 1992. Vol. 1.
- KUBRUSLY, C. 2006. *A Concise Introduction To Tensor Product*. s.l. : Far East Journal of Mathematical Sciences, 2006. pp. 137-174. Vol. 22.

- KUO, F. Y. e SLOAN, I. H. 2005. *Lifting the curse of dimensionality*. s.l. : Notices of the AMS, 2005. pp. 1320-1328. Vol. 52.
- LE MAÎTRE, O. e KNIO, O. 2010. *Introduction: Uncertainty Quantification and Propagation*. s.l. : Springer Netherlands, 2010.
- LIU, J.S. 2008. *Monte Carlo strategies in scientific computing*. s.l. : Springer Science & Business Media, 2008.
- LIU, W.K., BELYTSCHKO, T. e MANI, A. 1986. *Random field finite elements*. s.l. : International Journal for Numerical Methods in Engineering, 1986. pp. 1831-1845. Vol. 23.
- LOH, W.L. 1996. *On Latin hypercube sampling*. s.l. : The annals of statistics, 1996. pp. 2058-2080. Vol. 24.
- MOHAGHEGH, S. e AMERI, S. 1995. *Artificial neural network as a valuable tool for petroleum engineers*. 1995.
- MOHAGHEGH, S. 2005. *Recent developments in application of artificial intelligence in petroleum engineering*. 2005. pp. 86-91.
- MYERS, R.H., MONTGOMERY, D.C. e ANDERSON-COOK, C.M. 2009. *Response surface methodology: process and product optimization using designed experiments*. s.l. : John Wiley & Sons, 2009. Vol. 705.
- NOVAK, E. e RITTER, K. 1999. *Simple cubature formulas with high polynomial exactness*. s.l. : Constructive approximation, 1999. pp. 499-522. Vol. 15.
- OLIVER, D., REYNOLDS, A. e LIU, N. 2008. *Inverse theory for petroleum reservoir characterization and history matching*. s.l. : Cambridge University Press, 2008.
- PACHECO, M.A. e VELLASCO, M. M. 2009. *Intelligent systems in oil field development under uncertainty*. s.l. : Springer, 2009. Vol. 183.
- PATTERSON, T. 1968. *The optimum addition of points to quadrature formulae*. s.l. : Mathematics of Computation, 1968. pp. 847-856. Vol. 22(104).
- PEIXOTO, L. 2008. *Quadratura de Gauss iterativa com base nos polinômios ortogonais clássicos*. s.l. : Programa de Pós-Graduação em Modelagem Matemática e Computacional, 2008. pp. 41-49.
- POLI, R., et al. 2008. *A field guide to genetic programming*. s.l. : Lulu.com, 2008.
- PRANGA, M. e TOMPKINS, M. J. 2012. *Efficient estimation of polynomial chaos proxies using generalized sparse quadrature*. Schlumberger-Doll Research : Technical Paper, 2012.

- QUARTERONI, A., SACCO, R. e SALERI, F. 2010. *Numerical mathematics*. s.l. : Springer Science & Business Media, 2010. Vol. 37.
- ROSA, A., DE SOUZA CARVALHO, R. e XAVIER, J. 2006. *Engenharia de reservatórios de petróleo*. s.l. : Interciência, 2006.
- ROSS, S. 2012. *A first course in probability*. s.l. : Pearson, 2012.
- RYAN, R. 2002. *Introduction To Tensor Products Of Banach Spaces*. s.l. : Springer Science & Business Media, 2002.
- SARMA, P. e XIE, J. 2011. *Efficient and robust uncertainty quantification in reservoir simulation with polynomial chaos expansions and non-intrusive spectral projection*. The Woodlands, TX : In Paper SPE 141963 presented at the SPE Reservoir Simulation Symposium, 2011. pp. 21-23.
- SARMA, P., DURLOFSKY, L. and AZIZ, K. 2005. *Efficient Closed Loop Production Optimization Under Uncertainty*. 2005.
- SARMA, P., et al. 2006. *Efficient real-time reservoir management using asjoint-based optimal control and model updating*. s.l. : Computacional Geosciences, 2006. pp. 3-36. Vol. 10. 1.
- SEARSON, D., LEAHY, D. e WILLIS, M. 2010. *GPTIPS: an open source genetic programming toolbox for multigene symbolic regression*. s.l. : Proceedings of the International multiconference of engineers and computer scientists, 2010. pp. 77-80. Vol. 1.
- SMOLYAK, S. 1963. *Quadrature and interpolation formulas for tensor products of certain classes of functions*. s.l. : Dokl. Akad. Nauk SSSR, 1963. pp. 240-243. Vol. 4.
- SOUZA DE CURSI, J. e SAMPAIO, R. 2012. *Modelagem Estocástica E Quantificação de Incertezas*. s.l. : Notas em Matemática Aplicada, 2012.
- STEIN, M. 1987. *Large sample properties of simulations using Latin hypercube sampling*. s.l. : Technometrics, 1987. pp. 143-151. Vol. 29.
- STROUD, A. 1974. *Numerical quadrature and solution of ordinary differential equations*. s.l. : Springer, 1974.
- SUDRET, B. 2008. *Global sensitivity analysis using polynomial chaos expansions*. s.l. : Reliability Engineering & System Safety, 2008. pp. 964-979. Vol. 93. 7.

- SUDRET, B., BLATMAN, G. e BERVEILLER, M. 2013. *Response Surfaces Based on Polynomial Chaos Expansions*. s.l. : Construction Reliability: Safety, Variability and Sustainability, 2013. pp. 147-167.
- SÜLI, E. e MAYERS, D. 2003. *An Introduction to numerical analysis*. s.l. : Cambridge University Press, 2003.
- SZEGO, Gabor. 1939. *Orthogonal Polynomials*. Rhode Island : American Mathematical Society, 1939.
- THOMAS, J. E. 2001. *Fundamentos de engenharia de petróleo*. s.l. : Interciência, 2001.
- URBAN, K., TANG, S. e ROMETSCH, M. 2009. *Numerical Finance*. s.l. : University of Ulm, lectures notes, 2009.
- WEBSTER, M., TATANG, M. e MCRAE, G. 1996. *Application of the probabilistic collocation method for an uncertainty analysis of a simple ocean model*. 1996.
- WIENER, N. 1938. *The Homogeneous Chaos*. s.l. : American Journal of Mathematics, 1938. pp. 897-936. Vol. 60. 4.
- XIU, D. e KARNIADAKIS, G. 2002. *The Wiener--Askey polynomial chaos for stochastic differential equations*. s.l. : SIAM journal on scientific computing, 2002. pp. 619-644. Vol. 24.
- XIU, D. 2010. *Numerical Methods For Stochastic Computations: A Spectral Method Approach*. s.l. : Princeton University Press, 2010.
- YAMAZAKI, F., SHINOZUKA, M. e DASGUPTA, G. 1988. *Neumann expansion for stochastic finite element analysis*. s.l. : Journal of Engineering Mechanics, 1988.
- YU, T., WILKINSON, D. e CASTELLINI, A. 2008. *Constructing reservoir flow simulator proxies using genetic programming for history matching and production forecast uncertainty analysis*. s.l. : Journal of Artificial Evolution and Applications, 2008. pp. 187-201. Vol. 2008.

Apêndice

O formato do arquivo de texto que é lido pelos *scripts* no *matlab* e modificado para convertê-lo num arquivo *.dat* (compatível como o simulador IMEX) é mostrado a continuação.

```
** Null gridblocks
*INCLUDE 'null.inc'

** Porosity
*INCLUDE 'por.inc'

** Porosity multipliers (MULTIPOR)
MOD
1:19 1:28 1 * <ica>MULTIPOR</ica>
1:19 1:28 2 * <ica>MULTIPOR</ica>
1:19 1:28 3 * <ica>MULTIPOR</ica>
1:19 1:28 4 * <ica>MULTIPOR</ica>
1:19 1:28 5 * <ica>MULTIPOR</ica>

** Horizontal permeability multipliers (MULTIPERMI)
*INCLUDE 'permi.inc'
MOD
1:19 1:28 1 * <ica>MULTIPERMI</ica>
1:19 1:28 2 * <ica>MULTIPERMI</ica>
1:19 1:28 3 * <ica>MULTIPERMI</ica>
1:19 1:28 4 * <ica>MULTIPERMI</ica>
1:19 1:28 5 * <ica>MULTIPERMI</ica>

** Permeability J-dir = I-dir
PERMJ EQUALSI

** Vertical permeability multipliers (MULTIPERMK)
*INCLUDE 'permk.inc'
MOD
1:19 1:28 1 * <ica>MULTIPERMK</ica>
1:19 1:28 2 * <ica>MULTIPERMK</ica>
1:19 1:28 3 * <ica>MULTIPERMK</ica>
1:19 1:28 4 * <ica>MULTIPERMK</ica>
1:19 1:28 5 * <ica>MULTIPERMK</ica>

** Rock compressibility (CPOR)
PRPOR 23500
CPOR <ica>CPOR</ica>
```

** Aquifer

AQUIFER REGION 14 4 5 IDIR
 REGION 15 4 5 JDIR

. ----- *Dados do Aquífero* -----

.
 REGION 12 27 3 IDIR

** Aquifer radius (AQRADIUS)

AQPROP

19.6 0.14 270 <ica>**AQRADIUS**</ica> 0.2638889

AQUIFER REGION 15 1 2 IDIR
 REGION 16 1 2 JDIR

. ----- *Dados do Aquífero* -----

.
 REGION 4 28 1 IDIR

. ----- *Dados do Aquífero* -----

.
 REGION 5 28 1 JDIR

AQPROP

6.0 0.14 270 <ica>**AQRADIUS**</ica> 0.2638889

** Fluid properties

MODEL BLACKOIL

DENSITY OIL 912.0

DENSITY GAS 0.8266

DENSITY WATER 1000.0

PVT BG 1

. ---- *Outros dados próprios do campo PUNQ-S3 e do simulador IMEX* ----

** Relative permeability

ROCKFLUID

KROIL STONE1 SWSG

RPT 1

** Maximum gas relative permeability (KRGMAX)

** Corey exponent of gas-rel perm (NG)

SGT SMOOTH PLAW <ica>**NG**</ica> 4.0 0

** Sg krg krog pc

0.0000 0.0000 0.9000 0.000000

0.8000 <ica>**KRGMAX**</ica> 0.0000 0.000000

** Maximum water relative permeability (KRWMAX)
 ** Corey exponent of gas-rel perm (NW)

SWT SMOOTH PLAW <ica>NW</ica> 2.0 0

** Sw krw krow pc
 0.2000 0.0000 0.9000 0.000000
 0.9000 <ica>KRWMAX</ica> 0.0000 0.000000

KRWEXTRAP ON

** Initialization
 INITIAL

VERTICAL BLOCK_CENTER WATER_OIL_GAS
 REFDEPTH 2355.00
 REFPRES 23446

** Delta oil-water contact (DELTAWOC)
 DWOC <ica>DELTAWOC</ica>

** Delta gas-water contact (DELTAGOC)
 DGOOC <ica>DELTAGOC</ica>

PBT 1
 2175 23446
 2496 23446

NUMERICAL
 RUN

** Simulation start date
 DATE 1967 1 1
 GROUP 'G1' ATTACHTO 'FIELD'
 *INCLUDE 'wells.inc'

** Open a well every 2 month

.
 . ----- *Dados da abertura dos poços* -----
 .

*STOP