



Simone de Carvalho Miyoshi

Modelo Kernel PCA-Genético para Ajuste de Histórico

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica.

Orientadora: Prof. Marley Maria Bernardes Rebuszi Vellasco

Co-orientadora: Prof. Dilza de Mattos Szwarcman

Rio de Janeiro
Setembro de 2010

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.



Simone de Carvalho Miyoshi

Modelo Kernel PCA-Genético para Ajuste de Histórico

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

Prof. Marley Maria Bernardes Rebuszi Vellasco

Orientadora

Departamento de Engenharia Elétrica - PUC-Rio

Prof. Dilza de Mattos Szwarcman

Co-orientadora

Laboratório de Inteligência Computacional - PUC-Rio

Prof. Marco Aurélio Cavalcanti Pacheco

Departamento de Engenharia Elétrica - PUC-Rio

Prof. Adriana da Costa Ferreira Chaves

Sem vínculo

Prof. Ricardo Alexandre Passos Chaves

Petróleo Brasileiro S.A.

Prof. José Eugênio Leal

Coordenador Setorial do Centro

Técnico Científico - PUC-Rio

Rio de Janeiro, 27 de setembro de 2010

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

Simone de Carvalho Miyoshi

Graduou-se em Engenharia Química pela Universidade Federal do Rio de Janeiro / UFRJ em 2008. Concluiu em 2009 o curso de especialização em Engenharia de Processos oferecido Programa de Mobilização da Indústria do Petróleo – PROMINP. Desde 2008, desenvolve atividades de pesquisa e desenvolvimento junto ao Laboratório de Inteligência Computacional – ICA. A sua linha de pesquisa está relacionada a aplicações de inteligência computacional na área de Exploração e Produção de Petróleo, mais especificamente, no problema de ajuste de histórico.

Ficha Catalográfica

Miyoshi, Simone de Carvalho

Modelo Kernel PCA-Genético para Ajuste de Histórico/
Simone de Carvalho Miyoshi; orientadora: Marley Maria Bernardes Rebuszi Vellasco; co-orientadora: Dilza de Mattos Szwarcman. – Rio de Janeiro: PUC, Departamento de Engenharia Elétrica, 2010.

v.1, 110 fl.: 45 il. ; 30,0 cm

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

Incluí referências bibliográficas.

1. Engenharia elétrica – Teses. 2. Ajuste de Histórico. 3. Método Kernel PCA-Genético. 4. Algoritmo Genético. I. Vellasco, Marley Maria Bernardes Rebuszi. II. Szwarcman, Dilza de Mattos. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. IV. Título.

À minha mãe.

Agradecimentos

A Deus.

Às minhas orientadoras Prof. Marley Vellasco e Dra. Dilza de Mattos Szwarcman pelos conhecimentos, sugestões e estímulo fornecidos durante o desenvolvimento do trabalho, sem os quais nada seria possível.

Ao Prof. Marco Aurélio Pacheco por, desde o início, confiar na minha capacidade.

Aos amigos e colegas do Laboratório de Inteligência Computacional (ICA), em pelos conhecimentos e sugestões fornecidas durante o trabalho.

A minha família e amigos, o meu alicerce.

Ao CNPq e à PUC-Rio, pelos auxílios concedidos, sem os quais este trabalho não poderia ser realizado.

Resumo

Miyoshi, Simone de Carvalho; Vellasco, Marley Maria Bernardes Rebuzzi; Szwarcman, Dilza de Mattos. **Modelo Kernel PCA-Genético para Ajuste de Histórico**. Rio de Janeiro, 2010. 110 p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Esse estudo propõe um novo método para ajuste de histórico de parâmetros de permeabilidade, baseado em Kernel PCA (KPCA) e em Algoritmo Genético. O ajuste de histórico é um problema inverso, de alta dimensionalidade, onde alguns parâmetros do modelo do reservatório são ajustados com a finalidade de que a simulação de produção seja capaz de oferecer uma previsão mais realista da produção de um determinado campo de petróleo. Apesar de já existirem várias metodologias para esse problema, a maioria delas não leva em consideração restrições geoestatísticas. Além disso, vários desses estudos utilizam métodos de otimização baseados em gradientes que tendem a ficar restritos a mínimos locais. O objetivo desse estudo é desenvolver uma nova metodologia que seja capaz de honrar a estatística de múltiplos pontos; não assuma uma distribuição multigaussiana; faça uso de métodos de otimização que procurem por uma solução global e encontre parâmetros de permeabilidade realísticos. O KPCA é uma análise de componentes principais efetuada em um espaço mapeado implicitamente através de uma função kernel. Este método foi escolhido devido a sua capacidade de: extrair parte dos momentos de estatística de ordem superior; comprimir e reconstruir dados; e de extrair características não lineares. Já o Algoritmo Genético foi selecionado por ser um método de otimização robusto que busca uma solução global. O algoritmo proposto foi testado em um campo sintético e apresentou bons resultados, ou seja, uma aproximação suave dos valores do campo de permeabilidade preservando a forma de canais de permeabilidade.

Palavras-chave

Engenharia elétrica; ajuste de histórico; método Kernel PCA-Genético; Algoritmo Genético.

Abstract

Miyoshi, Simone de Carvalho; Vellasco, Marley Maria Bernardes Rebuzzi (Advisor); Szwarcman, Dilza de Mattos (Advisor). **Modelo Kernel PCA-Genético para Ajuste de Histórico**. Rio de Janeiro, 2010. 110p. MSc. Dissertation – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

This study proposes a new approach for history matching based on Kernel PCA (KPCA) and Genetic Algorithm to adjust the reservoir permeability field. The history matching is an inverse problem, with huge dimensionality, where some parameters of the reservoir model are adjusted so that the simulation of the mathematical models could be able to make a realistic prediction of the future production of an Oil & Gas field. Although there are several methodologies in literature for history matching, most of them don't take into account geostatistical restrictions. Besides, several studies make use of optimization methods based on gradients that can be trapped in local minima. The objective of this study is to develop a new methodology that: honor the multipoint statistics; do not assume multi-gaussian distribution; make use of optimization methods that search for global solutions and find realistic permeability parameters. Kernel PCA is a principal component analysis in a high dimension space called feature space that is mapped implicitly by a kernel function. Kernel PCA was chosen due to its ability to: capture part of statistics moments of high order; compress and accurately reconstruct data; and extract non-linear characteristics. Moreover, the Genetic Algorithm was selected because it is a robust optimization method that searches for local minima. The proposed algorithm was tested in a synthetic field and had good results, that is, a good smooth approximation of the values of the permeability field, preserving the shape of the permeability channels.

Keywords

Electrical engineering; history matching; Genetic-Kernel PCA model; Genetic Algorithm.

Sumário

1 Introdução	16
1.1. Motivação	17
1.2. Objetivos	19
1.3. Descrição do Trabalho	19
1.4. Organização do Trabalho	21
2 Caracterização de Reservatórios	23
2.1. Acúmulo de Hidrocarbonetos em Reservatórios	23
2.2. Propriedades Macroscópicas de Estruturas Porosas	24
2.2.1. Porosidade	25
2.2.2. Permeabilidade	26
2.3. Caracterização de Reservatórios	28
2.3.1. Fontes de Dados para a Caracterização de Reservatórios	29
2.3.1.1. Dados Sísmicos de Reflexão	30
2.3.1.2. Perfilagem	32
2.3.1.3. Testemunhos	32
2.3.1.4. Teste de formação	33
2.3.2. Métodos Geoestatísticos	34
2.3.2.1. Métodos geoestatísticos convencionais	34
2.3.2.2. Algoritmo de Simulação Geoestatística de Múltiplos Pontos - Snesim	35
3 Ajuste de Histórico	38
3.1. O Problema	38
3.2. Abordagens de Ajuste de Histórico	40
3.2.1. Abordagens por Zoneamento do Reservatório	40
3.2.2. Abordagens utilizando Inversão Bayesiana	41
3.2.2.1. Ajuste de Histórico por Inferência Bayesiana	42
3.2.2.2. Métodos no Contexto Bayesiano baseados em Simulação Seqüencial	46

3.2.3. Outras Metodologias	48
3.2.3.1. Deformação Gradual	48
3.2.3.2. Recozimento Simulado/ Algoritmo Genético	49
3.2.3.3. Programação Genética	52
3.3. Discussão das principais técnicas utilizadas	52
4 Estado da Arte do Kernel PCA	54
4.1. Análise de Componentes Principais (PCA)	54
4.2. Mapeamento Não Linear através de uma Função Kernel	56
4.3. O Kernel PCA (KPCA)	58
4.3.1. Algoritmo do Kernel PCA (KPCA)	58
4.4. <i>Pre-Image</i>	61
4.4.1. Métodos de <i>Pre-Image</i>	63
4.4.1.1. Método do Ponto Fixo	63
4.4.1.2. Método de Regularização do Espaço de Entrada	65
4.4.1.3. Escalonamento Multidimensional (<i>Multidimensional Scaling</i>)	66
4.4.1.4. Método de Aprendizagem	68
4.4.1.5. Discussão de cada método de <i>Pre-Image</i>	69
5 Modelo Kernel PCA Genético para Ajuste de Histórico	70
5.1. Descrição do Modelo Kernel PCA-Genético para Ajuste de Histórico	71
5.1.1. Simulação Geoestatística de Múltiplos Pontos	73
5.1.2. Cálculo do KPCA	74
5.1.3. Algoritmo Genético	74
5.1.3.1. Conjunto de Soluções Iniciais	75
5.1.3.2. Módulo de Avaliação	75
5.1.3.3. Seleção	78
5.1.3.4. Recombinação	79
5.1.3.5. Critério de Convergência	82
5.2. Discussão da metodologia proposta	82
6 Estudos de Caso	83

6.1. Primeira Etapa: Testes em Remoção de Ruído com realizações de permeabilidade obtidas a partir da estatística de múltiplos pontos	83
6.1.1. Resultados da Aplicação de Ruído	84
6.1.2. Discussão dos Resultados da Aplicação de Ruído	87
6.2. Aplicação de Ajuste de Histórico utilizando o modelo Kernel PCA-Genético	88
6.2.1. Etapa A	91
6.2.2. Etapa B	93
7 Conclusões & Trabalhos Futuros	99
8 Referências	101
Anexos	108
A1. Mercer Kernels	108
A2. Considerações sobre a existência de <i>Pre-Image</i>	108

Lista de figuras

Figura 1. Acúmulo de hidrocarbonetos em um reservatório	24
Figura 2. Ilustração de meios porosos.	25
Figura 3. Imagem de Microscopia 3D em um espaço poroso de um arenito (<i>Brent Sandstone</i>).	26
Figura 4. Diagrama esquemático do escoamento em meios porosos	27
Figura 5. Efeito do tamanho do grão na permeabilidade	28
Figura 6. Escala das diferentes fontes de dados	29
Figura 7. Diagrama esquemático da sísmica de reflexão no mar.	31
Figura 8. Ilustração do Perfilagem de um poço de petróleo	32
Figura 9. Três diferentes heterogeneidades geológicas e três variogramas similares.	35
Figura 10. Diagrama do Problema de Ajuste de Histórico	39
Figura 11. Ilustração da Solução apresentada por Jacquard & Jain (1965)	41
Figura 12. Exemplo do operador <i>crossover</i> proposto	50
Figura 13. Reservatório subdividido em regiões (contorno em preto) e em sub-regiões (áreas coloridas)	51
Figura 14. Exemplo de PCA	55
Figura 15. Diagrama esquemático de um mapeamento implícito	56
Figura 16. Exemplo de mapeamento de espaço de um kernel.	57
Figura 17. Esquema do KPCA.	58
Figura 18. Diagrama esquemático do problema de <i>Pre-Image</i> .	62
Figura 19. Algoritmo do Modelo Kernel PCA-Genético.	72
Figura 20. Diagrama esquemático de um indivíduo com n genes	75
Figura 21. Diagrama esquemático do cruzamento de um ponto	80
Figura 22. Imagem de treinamento utilizada.	84
Figura 23. Exemplo de campos de permeabilidades obtidos.	84
Figura 24. Campo de permeabilidade escolhido e c/ ruído (a direita).	85
Figura 25. Resultado da remoção de ruído utilizando-se kernel RBF.	85
Figura 26. O comportamento da energia acumulada obtida para a	

função kernel RBF em relação ao parâmetro sigma da função kernel RBF.	86
Figura 27. Relação entre o número de autovetores e a energia acumulada para a função kernel RBF.	86
Figura 28. Resultado da remoção de ruído com um kernel sigmóide.	87
Figura 29. Resultado da remoção de ruído com um kernel polinomial.	87
Figura 30. Etapas dos testes do modelo Kernel PCA-Genético.	88
Figura 31. Campo sintético utilizado.	89
Figura 32. Comparação entre as curvas de produção de água e óleo reais e as aproximadas pelo ajuste de histórico obtido pelo modelo KPCA-Genético proposto.	92
Figura 33. Comparação entre as curvas de produção de gás real e as aproximadas pelo ajuste de histórico obtido pelo modelo KPCA-Genético proposto.	92
Figura 34. Campo de permeabilidade reconstruído (esquerda) e campo de permeabilidade real (direita).	92
Figura 35. Campo de permeabilidade ajustado pelo modelo Kernel-PCA Genético a direita e Campo de permeabilidade real a esquerda.	95
Figura 36. Campo de permeabilidade ajustado pela metodologia de Zoneamento.	95
Figura 37. Curvas de produção do poço 3 simulada pelos modelos ajustados pelo método proposto (KPCA-Genético) e pelo método de Zoneamento.	96
Figura 38. Curvas de Produção do poço 5 simulada pelos modelos ajustados pelo método proposto (KPCA-Genético) e pelo método de Zoneamento.	96
Figura 39. Curvas de Produção do poço 8 simulada pelos modelos ajustados pelo método proposto (KPCA-Genético) e pelo método de Zoneamento.	96
Figura 40. Curvas de Produção do poço 10 simulada pelos modelos ajustados pelo método proposto (KPCA-Genético) e pelo método de Zoneamento.	97
Figura 41. Curvas de Produção do poço 13 simulada pelos modelos	

ajustados pelo método proposto (KPCA-Genético) e pelo método de Zoneamento.	97
Figura 42. Curvas de Produção do poço 16 simulada pelos modelos ajustados pelo método proposto (KPCA-Genético) e pelo método de Zoneamento.	97
Figura 43. Ilustração de Funções sobrejetoras.	109
Figura 44. Ilustração de Funções Bijetoras.	109
Figura 45. Ilustração de Funções não-sobrejetoras.	110

Lista de tabelas

Tabela 1. Configuração Kernels	90
Tabela 2. Energia Acumulada para cada kernel	91
Tabela 3. MAPE Obtido pelo Ajuste de Histórico na Etapa A	91
Tabela 4. MAPE Obtido pelo Ajuste de Histórico na Etapa B	94
Tabela 5. Relação entre o tipo de função e a sua <i>Pre-image</i>	110

“Engenheiros Químicos não são pessoas gentis; eles gostam de altas temperaturas e pressões elevadas”. Autor desconhecido.

1 Introdução

Uma previsão realista do comportamento de um reservatório de petróleo e/ou gás é de extrema importância para o gerenciamento de um campo (Maschio, Vidal e Schiozer, 2008). Essa previsão pode ser realizada através de uma simulação numérica de um modelo de reservatório utilizando equações de escoamento em meios porosos.

Segundo Deutsch (2002), um modelo de reservatório que é consistente com os dados disponíveis é uma importante fonte de informação para o plano de desenvolvimento e para o gerenciamento de um campo de petróleo ou gás.

Na etapa de desenvolvimento do campo, um modelo de reservatório adequado fornece estimativas confiáveis de volumes de hidrocarbonetos no reservatório. Além disso, subsidia uma otimização do plano de desenvolvimento, de forma a permitir encontrar um número ótimo de poços, suas localizações e também os tipos de poços a serem perfurados (Deutsch, 2002). Na etapa de gerenciamento de um campo, um modelo de reservatório realista pode, por exemplo, auxiliar no ajuste da melhor estratégia de injeção (Deutsch, 2002).

Contudo, para tal, é necessária uma caracterização realista das propriedades dos fluidos e das rochas que compõem o reservatório. Tipicamente, a construção de um modelo de reservatório é elaborada em dois estágios. Em um estágio inicial, a caracterização do reservatório é realizada utilizando-se dados estáticos como perfil de poço, dados sísmicos, etc. (Maschio, Vidal e Schiozer, 2008). Em uma segunda etapa, é realizado o ajuste de histórico ou ajuste aos dados dinâmicos.

O processo de ajuste de histórico é um problema inverso, onde alguns parâmetros do modelo do reservatório são ajustados de forma que a simulação de produção seja capaz de reproduzir os dados reais de produção de um determinado campo de petróleo (Rosa, Carvalho e Xavier, 2006). Esses dados de produção possuem informações importantes, dado que eles são uma medida direta da resposta do reservatório (Wen, Capilla, *et al.*, 1999).

Cada etapa da caracterização do reservatório tem a sua importância, pois cada uma tem uma abrangência diferente. Os dados de perfil de poço possuem uma alta definição, mas são dados pontuais. Já os dados de produção possuem uma área de cobertura maior, mas difundem as informações sobre as heterogeneidades locais de um reservatório (Tureyen e Caers, 2002).

Com isso, o processo de ajuste de histórico necessita integrar todos os dados disponíveis para reduzir as incertezas sobre as previsões de produção do reservatório. O modelo ajustado deve não somente reproduzir os dados de produção através da simulação numérica, mas também ser consistente com o conhecimento geológico do reservatório (Roggero e Hu, 1998).

1.1. Motivação

O ajuste de histórico é um problema desafiador dado os seguintes fatores:

- Existe uma relação não linear entre os parâmetros de entrada e as curvas de produção;
- Há a necessidade de se integrar diversas fontes de dados, que podem, inclusive ser conflitantes (Deutsch, 2002);
- A quantidade de parâmetros de entrada é muito maior do que a de parâmetros de saída.

Esse é um exemplo de um problema inverso onde múltiplas configurações de parâmetros de entrada podem resultar na mesma saída (Kelkar e Perez, 2002), ou seja, um problema tipicamente mal condicionado.

A indústria petrolífera, de uma maneira geral, tem efetuado o ajuste de histórico pelo método de tentativa e erro, multiplicando os parâmetros do reservatório por fatores diversos em diferentes regiões (Maschio, Vidal e Schiozer, 2008). Os modelos de reservatório construídos pela indústria, após o ajuste de histórico, frequentemente não são realísticos de um ponto de vista geológico ou as mudanças são realizadas sem qualquer consideração geológica (Hoffman, Wen e Strebelle, 2005).

O problema do ajuste de histórico não é novo. Os primeiros trabalhos encontrados na literatura sobre o tema são de Jacquard (1964) e Jacquard e Jain (1965).

A partir daí novas metodologias foram sendo desenvolvidas com o objetivo de resolver o problema de ajuste de histórico, tais como: Modelos Adjuntos (Chen, Gavalas, *et al.*, 1974); Expansão de Karhunen-Loève (Gavalas, Shah e Seinfeld, 1976); Redes Neurais (Ouenes, Richardson e Weiss, 1995) e (Ramgulam, Ertekin e Flemings, 2007); *Simulated Annealing* (Datta-Gupta, Vasco e Long, 1995); Cadeia de Markov (BONET-CUNHA, OLIVER, *et al.*, 1996); Algoritmos baseados em linhas de corrente (Wang e Kovscek, 2000); Método de Deformação Gradual (Hu, Blanc e Noetinger, 2001); *Ensemble Kalman Filter* (Nævdal, Johnsen e Aanonsen, 2003); *Probability Perturbation Method* (Caers e Hoffman, 2006); Kernel PCA (KPCA) (Sarma, 2006); e Programação Genética (Yu, Wilkinson e Castellini, 2008).

Vários desses estudos utilizam métodos de otimização baseados em gradientes que tendem a ficar restritos a mínimos locais, característica que é ainda mais crítica quando se tem um problema com um grande número de parâmetros.

Além disso, muitos desses métodos não consideram as restrições geoestatísticas e/ou não são baseados em modelos geoestatísticos (Wen, Capilla, *et al.*, 1999). Ou ainda, muitos métodos assumem distribuição de probabilidade multigaussiana dos parâmetros ou relação linear entre os dados de produção e os parâmetros do modelo (Wen, Capilla, *et al.*, 1999).

A simulação multigaussiana está limitada à reprodução do modelo de um variograma ou de uma função de covariância, ambos inferidos a partir de momentos de ordem dois. Dependendo do tipo de reservatório, a estatística de dois pontos, contudo, pode não ser suficiente para reproduzir a continuidade espacial de padrões interconectados característicos de reservatórios de petróleo (Strebelle e Journel, 2001).

Daí surge a necessidade de se desenvolver um método de ajuste de parâmetros geológicos de forma a mantê-los compatíveis com os modelos geoestatísticos. Para isso, é importante não assumir uma distribuição multigaussiana dos parâmetros, utilizar métodos de otimização robustos, e criar modelos ajustados cuja simulação seja capaz de prever mais fielmente o comportamento futuro do reservatório.

1.2. Objetivos

O objetivo deste trabalho é desenvolver uma nova metodologia para ajuste de parâmetros geofísicos, tais como a permeabilidade, considerando os seguintes requisitos:

- Honrar os modelos geoestatísticos de múltiplos pontos durante o processo de ajuste dos dados;
- Não assumir uma distribuição multigaussiana dos parâmetros do modelo;
- Utilizar métodos de otimização que busquem mínimos globais; e
- Ser capaz de efetuar uma previsão realista do comportamento futuro do reservatório.

Desse modo, este trabalho tem como objetivo contribuir para que se disponibilize uma metodologia de ajuste histórico onde os parâmetros geológicos ajustados sejam mais realistas e consistentes, obtendo-se assim uma simulação de produção capaz de prever mais precisamente o comportamento futuro do reservatório.

1.3. Descrição do Trabalho

Este trabalho foi estruturado nas seguintes etapas:

- Estudo do problema de ajuste de histórico e identificação das principais metodologias desenvolvidas;
- Investigação do Kernel PCA como tecnologia de referência para resolver o problema de ajuste de histórico;
- Estudo teórico do Kernel PCA e outra técnica associada chamada de *Pre-Image*;
- Proposição de um novo modelo baseado em Kernel PCA e Algoritmo Genético;
- Testes preliminares do Kernel PCA e *Pre-Image* em uma aplicação tradicional de remoção de ruído;
- Implementação do novo modelo e avaliação dos resultados a partir de um estudo de caso.

O estudo do problema de ajuste de histórico consistiu no levantamento bibliográfico das principais abordagens já propostas para a resolução do mesmo, destacando-se os pontos fortes e fracos de cada metodologia. Além disso, também foram identificadas algumas técnicas para a simulação geoestatística utilizando múltiplos pontos, de forma a reproduzir a continuidade espacial da geometria de litofácies e obter modelos de campo realistas. Dentre elas, destaca-se o modelo desenvolvido por Strebelle e Journel (2001).

Dentre as metodologias identificadas para ajuste de histórico, destaca-se a apresentada por Sarma, Durlofsky, *et al.*, (2007), a qual utiliza o Kernel PCA acoplado à inversão bayesiana e considerando a distribuição de probabilidade dos parâmetros como multigaussiana. Além disso, a metodologia se baseia em métodos para otimização baseados em gradientes.

Apesar dessas restrições, o Kernel PCA foi identificado como uma tecnologia promissora para o ajuste de histórico, devido a sua capacidade de capturar parte da estatística de ordem superior (Kim, Franz e Schölkopf, 2005), a capacidade de extrair características não lineares, e a habilidade de redução de dimensionalidade do problema. Assim, o Kernel PCA foi tomado como a tecnologia de referência para esse estudo.

Em seguida, efetuou-se o estudo teórico do Kernel PCA e do *Pre-Image*, identificando-se os principais trabalhos sobre esses temas.

Foi proposto então um novo modelo para resolução do problema de ajuste de histórico baseado em KPCA e em otimização por Algoritmo Genético. O modelo proposto não faz qualquer consideração a priori sobre a densidade de probabilidade do parâmetro geológico a ser ajustado, preserva a estatística de múltiplos pontos do modelo geológico e utiliza Algoritmo Genético na busca por uma solução global para o problema de ajuste de histórico.

Foram feitos testes iniciais com a aplicação de ruído para análise da técnica de Kernel PCA para a reconstrução de campos de permeabilidade que possuíam padrões sinuosos obtidos através de uma imagem de treinamento, obtendo-se resultados promissores que levaram o uso na aplicação seguinte.

Em seguida, foi realizada a implementação computacional do novo modelo e foram avaliados os resultados em um estudo de caso de ajuste de histórico. Foram obtidos bons resultados em termos de ajuste da curva de produção e em

termos de encontrar um conjunto de dados que representa bem o campo de permeabilidade considerado real.

1.4. Organização do Trabalho

Esta dissertação inclui seis capítulos adicionais.

No Capítulo II, introduzem-se importantes conceitos sobre a caracterização de reservatórios. O principal objetivo desse capítulo é descrever conceitos considerados necessários para a compreensão da fase inicial de caracterização de propriedades de reservatórios. As seções abrangem uma breve descrição de algumas propriedades macroscópicas de meios porosos e introduz algumas técnicas para a caracterização de reservatórios de petróleo: sísmica de reflexão, testemunhagem, perfilagem e teste de formação. O capítulo também apresenta alguns conceitos sobre métodos geoestatísticos tradicionais e sobre geoestatística de múltiplos pontos.

No Capítulo III, aborda-se todo o processo de ajuste de histórico. O principal objetivo desse capítulo é descrever conceitos considerados necessários para a compreensão das questões envolvidas no ajuste de histórico e as principais metodologias existentes para solução do problema.

No Capítulo IV, descreve-se o estado da arte das técnicas de Kernel PCA e de *Pre-Image*. O objetivo desse capítulo é apresentar os conceitos necessários para o entendimento dessas técnicas. As seções desse capítulo descrevem a Análise de Componentes Principais (PCA) tradicional; o mapeamento não linear através de uma função Kernel; o algoritmo do Kernel PCA (KPCA), bem como apresenta vários métodos recentemente desenvolvidos para solução do problema de *Pre-Image*.

No Capítulo V, apresenta-se em detalhes o novo modelo proposto para ajuste de histórico baseado em kernel PCA e Algoritmo Genético.

No Capítulo VI, apresentam-se os resultados obtidos nos estudos de caso realizados neste trabalho. Esse capítulo é dividido em três seções conforme o desenvolvimento do trabalho. Na primeira são apresentados testes para aplicação de remoção de ruído; na segunda seção são descritos os testes para resolução do *Pre-Image*; e na terceira seção são apresentados testes do novo modelo proposto.

Por fim, no Capítulo VII, são apresentadas as conclusões do trabalho e possíveis trabalhos futuros.

2 Caracterização de Reservatórios

Este capítulo visa introduzir os conhecimentos relacionados à caracterização de reservatórios considerados necessários para o completo entendimento do trabalho. Inicialmente, apresenta-se uma pequena introdução sobre o acúmulo de hidrocarbonetos em reservatórios. A seção seguinte faz uma breve descrição de algumas propriedades macroscópicas de meios porosos, tais como permeabilidade e porosidade. Por fim, são descritas algumas etapas da caracterização, como o perfil de poço, a sísmica de reflexão, testemunhagens, teste de produção e alguns conceitos de geoestatística de múltiplos pontos.

2.1. Acúmulo de Hidrocarbonetos em Reservatórios

O petróleo consiste de uma mistura de diversos tipos de hidrocarbonetos com diferentes pesos moleculares e propriedades termodinâmicas.

A teoria mais aceita hoje em dia associa a origem do petróleo à matéria orgânica depositada. Essa matéria orgânica teria se misturado com lama e teria sido empurrada para debaixo de camadas pesadas de sedimento, o que resultaria em altos níveis de calor e pressão (*diagenesis*). Essas condições de calor e pressão teriam causado alterações na matéria orgânica, transformando-a em um material ceroso denominado *kerogen* e, posteriormente, nos hidrocarbonetos líquidos e gasosos por um processo denominado *catagenesis* (Thomas, 2001).

Para a formação de reservatórios de petróleo, três características básicas precisam estar presentes:

- Rocha geradora rica em hidrocarbonetos;
- Rocha reservatório contendo espaços vazios para armazenar os hidrocarbonetos (porosidade) e conectividade entre os espaços porosos de forma a permitir o escoamento (permeabilidade);

- Rocha Capeadora que trapeia (sela) quantidades de hidrocarbonetos ou outro mecanismo para prevenir uma migração do óleo gerado pelas rochas geradoras.

A figura 1 ilustra as três condições citadas para formação de reservatórios de petróleo.

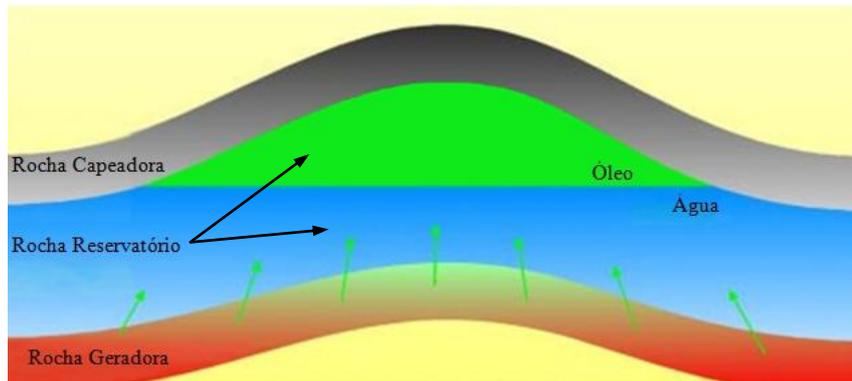


Figura 1. Acúmulo de hidrocarbonetos em um reservatório

O petróleo, por possuir uma densidade média inferior a das rochas que constituem o subsolo, tende a migrar das rochas geradoras para a superfície. Se, no caminho, o óleo encontrar uma camada impermeável (armadilha) que impeça a sua migração e uma estrutura porosa e permeável (reservatório) que faça o seu confinamento, acaba se formando um reservatório de petróleo (Thomas, 2001). Já dentro dos reservatórios, os fluidos irão se organizar de acordo com as suas respectivas densidades.

Como mencionado anteriormente, as propriedades macroscópicas porosidade e permeabilidade das rochas são requisitos fundamentais para a formação de reservatórios de petróleo. Na seção seguinte, detalham-se essas propriedades.

2.2. Propriedades Macroscópicas de Estruturas Porosas

Propriedades macroscópicas de estruturas porosas, como a permeabilidade e a porosidade, são influenciadas apenas pela estrutura do poro, ou seja, pela estrutura microscópica, definida pela geometria, pelo tamanho do poro, etc. Essas

propriedades macroscópicas, utilizadas em modelos de engenharia, representam um comportamento médio de uma amostra contendo muitos poros (Dullien, 1992).

Os parâmetros macroscópicos mais importantes são a porosidade, a permeabilidade, a área específica de superfície, o fator de resistividade de formação e a pressão capilar (Dullien, 1992).

Com isso, simuladores de reservatório utilizam-se de modelos matemáticos de escoamento em meios porosos para descrever o fenômeno físico do fluxo de massa e/ ou energia, a partir de parâmetros macroscópicos .

2.2.1. Porosidade

Porosidade (ϕ) é definida como a fração de espaços vazios em uma rocha em relação ao volume total da rocha (Rosa, Carvalho e Xavier, 2006). A figura 2 ilustra a disposição de espaços em meios porosos.

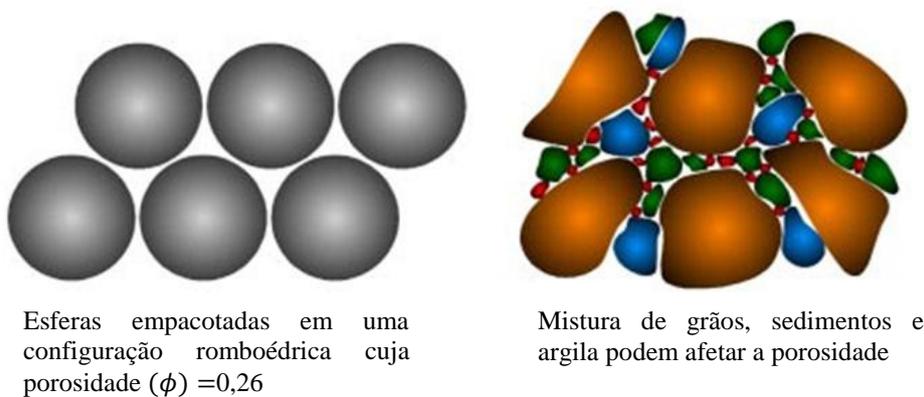


Figura 2. Ilustração de meios porosos.

Fonte: Slumberger (1998).

Como se observa na figura 2, no caso das esferas empacotadas com configuração romboédrica (a esquerda), o volume de espaços vazios entre os grãos é maior do que na configuração espacial apresentada na mistura de grãos (a direita). Com isso, a porosidade da configuração com as esferas empacotadas é maior do que a configuração da mistura de grãos, conforme mostra a figura acima.

Em um meio poroso, alguns poros podem estar interconectados; outros, não. Com isso, apenas a porosidade efetiva, ou seja, a relação entre os espaços vazios interconectados de uma rocha e o volume total da mesma, é de interesse. Isso

ocorre porque os poros isolados não são capazes de contribuir no escoamento (Deutsch, 2002).

A porosidade é uma propriedade importante, pois o volume total de óleo no reservatório é diretamente proporcional a ela.

Na figura 3, mostra-se um exemplo de espaço poroso de uma rocha de arenito (Brent *sandstone*), cuja porosidade efetiva é de 0,1602. A formação está transparente e o espaço poroso está em cinza.

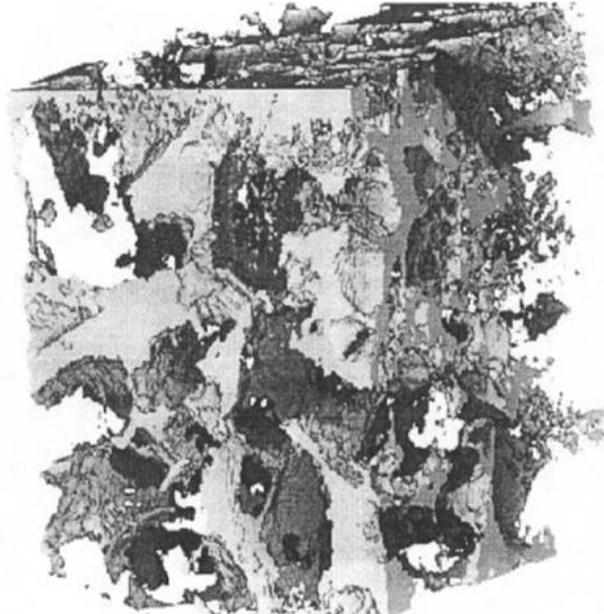


Figura 3. Imagem de Microscopia 3D em um espaço poroso de um arenito (Brent *Sandstone*).

A porosidade pode ser medida, em laboratório, a partir de uma amostra através de métodos destrutivos como a moendra, ou também através de ensaios não destrutivos como a gravimetria, injeção de gás, petrografia e processamento digital de imagens (Dullien, 1992).

2.2.2. Permeabilidade

Permeabilidade é a medida da facilidade que a rocha permite o escoamento do fluido. A permeabilidade também depende do sentido do escoamento do fluido (Deutsch, 2002).

A permeabilidade é definida pela lei de Darcy em um fluxo suficientemente lento, permanente e unidirecional (Dullien, 1992), conforme ilustra a equação (1) e a figura 4.

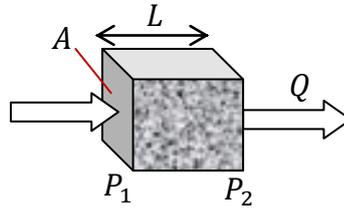


Figura 4. Diagrama esquemático do escoamento em meios porosos

$$k = \frac{Q \mu L}{A (P_1 - P_2)} \quad (1)$$

Onde:

Q é a vazão volumétrica do fluido (cm^3/s)

μ é a viscosidade do fluido (cP)

L é o comprimento do meio poroso na direção do fluxo (cm)

A é a área seccional da amostra (cm^2)

$P_1 - P_2$ é a diferença de pressão hidrostática (atm)

A unidade de permeabilidade é o Darcy, conforme mostra a equação (2).

$$1 \text{ darcy} = \frac{1 \left(\frac{cm^3}{s} \right) 1 (cP) 1 cm}{1 (cm^2) 1 (atm)} \quad (2)$$

Portanto, um meio poroso possui a permeabilidade igual a 1 darcy se a diferença de pressão de 1 atm produzir uma taxa de escoamento de 1 cm^3/s de um fluido com uma viscosidade igual a 1 cP através de um cubo com 1 cm de lado.

Como a porosidade, a permeabilidade pode ser medida em um laboratório a partir de uma amostra, mas não há um instrumento capaz de medir a permeabilidade diretamente no campo. Entretanto, ela pode ser calculada através de equações diferenciais ao se submeter o reservatório a uma condição dinâmica e

monitorar as respostas de pressão e temperatura (Dullien, 1992), a esse processo dá-se o nome de teste de formação (*well-test*).

Enquanto o tamanho do grão tem um efeito insignificante na porosidade da rocha, esse parâmetro tem um efeito predominante na permeabilidade. Isso ocorre porque se lida com o escoamento e com o atrito do fluido contra a superfície da área dos grãos de rocha. Cada grão de rocha está envolvido por uma superfície molhada, onde a velocidade do fluido é sempre zero por definição. Então, forma-se um atrito entre a camada de velocidade igual a zero e onde o fluido está escoando. Dessa forma, para um mesmo fluido e para meios de mesma porosidade, ocorre um atrito maior quando o fluido passa por um meio empacotado com grânulos finos do que com grãos mais largos (Rosa, Carvalho e Xavier, 2006).

Com isso, a permeabilidade do meio com grãos maiores é maior do que a permeabilidade do meio com grãos menores, considerando que ambos os meios possuem a mesma porosidade, conforme mostra a figura 5.

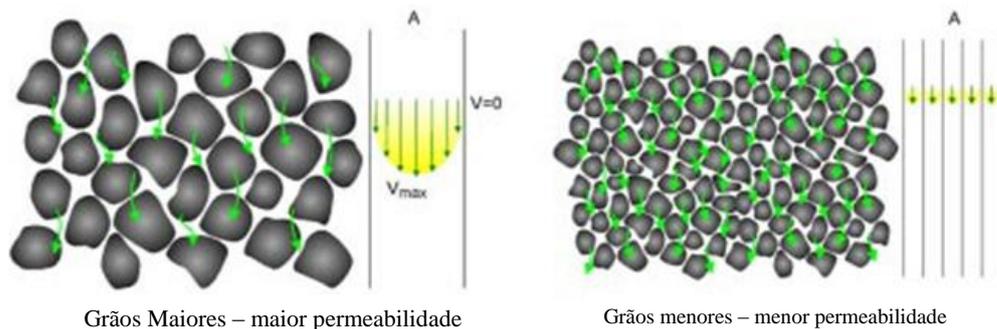


Figura 5. Efeito do tamanho do grão na permeabilidade

A seguir são descritas algumas etapas para caracterização de reservatórios.

2.3. Caracterização de Reservatórios

Idealmente, se o reservatório for homogêneo, a medida de uma propriedade em qualquer local permite uma descrição completa do reservatório. Entretanto, sabe-se que os reservatórios possuem características heterogêneas, ou seja, as propriedades de um reservatório variam em função do espaço (Kelkar e Perez, 2002). Com isso, a caracterização de um reservatório torna-se um problema complexo.

A construção de um modelo de reservatório é elaborada em dois estágios. Em um estágio inicial, a caracterização do reservatório é realizada utilizando-se dados estáticos como perfil de poço, dados sísmicos, etc. (Maschio, Vidal e Schiozer, 2008). Em uma segunda etapa, é realizado o ajuste de histórico ou ajuste aos dados dinâmicos.

Um reservatório pode ser caracterizado de uma forma direta (*hard data*) ou de forma indireta (*soft data*) e/ou a partir da inferência geoestatística de propriedades do reservatório (Deutsch, 2002). As principais fontes de dados para medidas, sejam indiretas ou diretas, são descritas a seguir.

2.3.1. Fontes de Dados para a Caracterização de Reservatórios

Sondagens, amostragens e testemunhagens são procedimentos que efetuam medidas de forma direta, ou seja, possibilitam o contato direto (visual ou tátil) com o material que compõe o reservatório, e também o desenvolvimento de uma série de ensaios laboratoriais nas amostras coletadas, cujos resultados irão subsidiar uma precisa caracterização do tipo ou da natureza do material coletado (Souza, 2006).

Por formas indiretas de investigação entendem-se aquelas que viabilizam a caracterização dos materiais constituintes do reservatório à distância e, portanto, não possibilitam o contato direto do pesquisador com o reservatório (Souza, 2006). Entre as formas indiretas estão: coletas de dados sísmicos, perfil de poço e testes de formação (Deutsch, 2002).

Cada etapa da caracterização do reservatório, seja direta ou indireta, possui uma escala de dados e abrangência diferentes, conforme figura 6.

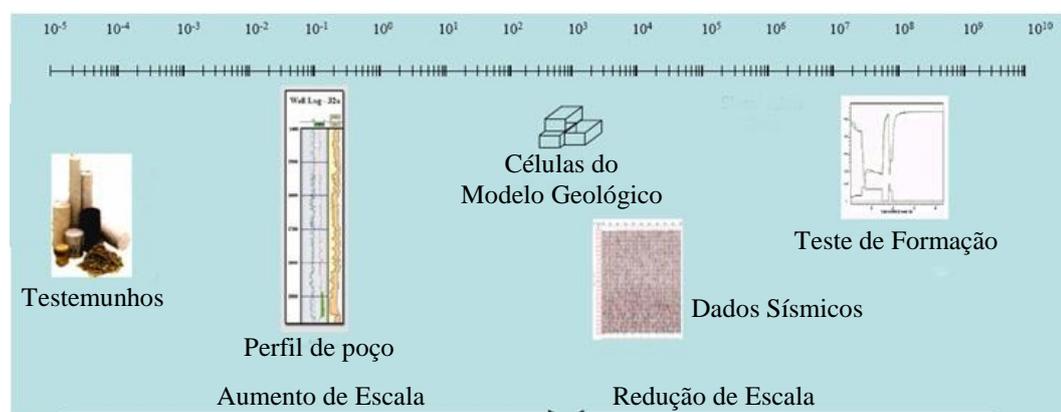


Figura 6. Escala das diferentes fontes de dados

Observa-se pela figura 6, que os dados de perfil de poço (*well-log*), testemunhos (*core data*), dados sísmicos e testes de formação (*well-test*) apresentam escalas diferentes. Dessa forma, para se efetuar a integração desses dados é necessário aumentar a escala (*upscaling*) dos dados de perfil de poço e de testemunhos e uma redução da escala (*downscaling*) dos dados de produção e dados sísmicos.

Quanto à abrangência, os dados do perfil de poço e de testemunhos possuem uma alta definição, mas são dados pontuais. Já os dados de produção possuem uma área de cobertura maior e difundem características heterogêneas de um reservatório (Tureyen e Caers, 2002).

Nas sub-seções a seguir são descritas as principais fontes de dados para a caracterização de propriedades de reservatórios.

2.3.1.1. Dados Sísmicos de Reflexão

Os métodos sísmicos são baseados nos princípios de propagação das ondas acústicas. A sísmica de reflexão lida com o fenômeno da reflexão do som nas interfaces entre os meios com distintas propriedades físicas, ou seja, com contrastes de impedância acústica. Teoricamente, corpos extremamente homogêneos não refletem as ondas acústicas, até que estas encontrem obstáculos ou interfaces acústicas (outro tipo de rocha, por exemplo) (Souza, 2006).

Basicamente, a utilização dos métodos sísmicos de reflexão em áreas marítimas se dá a partir da emissão de pulsos sonoros por meio de fontes acústicas rebocadas na superfície da água, ou muito próximos dela. Em alguns métodos, a fonte sísmica é posicionada na coluna d'água, a certa distância da superfície de fundo, conforme mostra a figura 7.

As frentes de ondas emitidas pela fonte acústica se propagam através da água e dos estratos sedimentares subjacentes em velocidades que variam de acordo com as características físicas de cada um dos meios geológicos atravessados pelo sinal (Souza, 2006).

Quando ocorrem contrastes de velocidade de propagação e de densidade, ou seja, contrastes de impedância acústica entre dois estratos, uma parte da energia das frentes de onda é refletida na interface entre os meios e retorna à superfície, e

a outra penetra nas camadas sedimentares, retornando aos sensores (hidrofonos) subsequentemente (Souza, 2006).

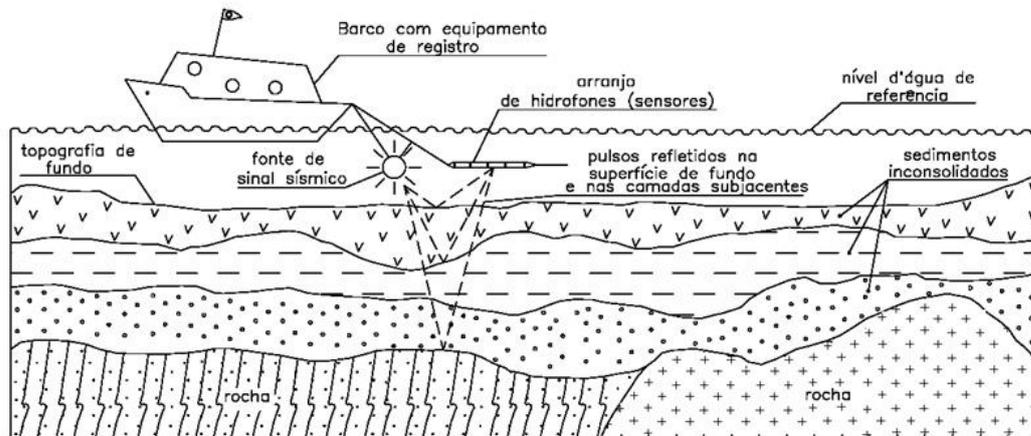


Figura 7. Diagrama esquemático da sísmica de reflexão no mar.

Fonte: Souza L. A. (2006)

Os sinais recebidos pelos sensores são conduzidos ao sistema de processamento e gravação, para finalmente serem impressos na forma de uma seção, em papel eletrossensível, ou, no caso dos equipamentos mais modernos, gravada digitalmente (Souza, 2006).

Após essa etapa de coleta de dados, são realizadas etapas consecutivas de interpretação e de avaliação qualitativa para que sejam inferidas propriedades de interesse para a exploração ou exploração¹ de um reservatório. Devido a essas etapas adicionais, mais incertezas são associadas às propriedades levantadas por processo sísmico (Kelkar e Perez, 2002).

Da sísmica de reflexão derivam a interpretação estrutural, superfície de *grids* e localização de falhas. Dela também derivam informações de atributos como variações em larga escala de proporções de fácies² e porosidade (Deutsch, 2002).

¹ O termo exploração, em geologia, relaciona-se à fase de prospecção: busca e reconhecimento da ocorrência dos recursos naturais, e estudos para determinar se as reservas são economicamente viáveis. A exploração é a retirada do recurso com máquinas adequadas, para fins de beneficiamento, transformação e utilização.

² Refere-se à caracterização de um tipo de rocha ou de uma associação de rochas (litotipo), dado qualquer aspecto genético, composicional, químico ou mineralógico, morfológico, estrutural ou textural distintivo para fins de referência em um estudo geológico.

2.3.1.2. Perfilagem

Perfilagem geofísica é o registro de medidas geofísicas (perfis elétricos, acústicos, radioativos, mecânicos, térmicos, etc.) realizadas por sensores de perfilagem deslocados em um poço (Girão, 2006). A figura 8 a seguir ilustra o registro da perfilagem.

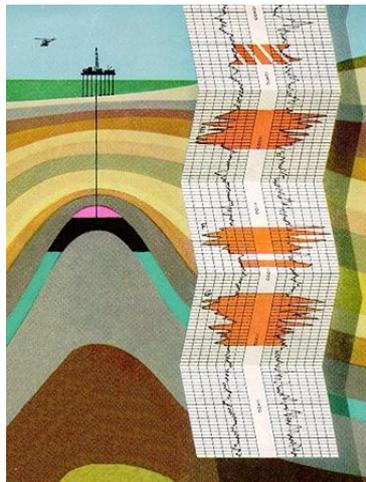


Figura 8. Ilustração do Perfilagem de um poço de petróleo

Fonte: Departamento de Energia dos Estados Unidos (2010)

A interpretação das informações obtidas na perfilagem geofísica permite inferir propriedades da formação. Isso é possível porque as rochas podem ser identificadas em função de suas propriedades elétricas (condutividade elétrica, polarização induzida, constante dielétrica ou potencial eletroquímico natural), acústicas (velocidade de propagação ou tempo de trânsito de ondas elásticas compressoriais ou cisalhantes), radioativas (radioatividade natural ou induzida), mecânicas, térmicas, etc. (Girão, 2006).

A perfilagem permite uma informação precisa das superfícies estratigráficas e das falhas, como também uma inferência das medidas de propriedades tais como tipo de fácies, porosidades e, por vezes, permeabilidade (Deutsch, 2002).

2.3.1.3. Testemunhos

As medidas de permeabilidade e de porosidade, feitas através de testemunhos, são medidas diretas das propriedades de reservatórios. Por isso, esses dados são menos incertos (Kelkar e Perez, 2002).

Contudo, para a realização de tais medidas, é necessário que seja perfurado um poço para a obtenção de tais testemunhos, o que requer um custo de investimento, se o poço for apenas para investigação das propriedades do reservatório. Assim, as propriedades inferidas através de testemunhos são mais precisas, porém esparsas.

2.3.1.4. Teste de formação

Os testes de formação podem ser realizados em diferentes estágios de perfuração, completação e produção. O objetivo de cada teste pode ir ser da simples identificação dos fluidos produzidos à determinação de características complexas do reservatório. Em termos gerais, os testes podem ser agrupados em testes de produtividade ou testes para caracterização de reservatórios (Slumberger, 1998).

Os testes de produtividade são conduzidos para:

- Identificar os fluidos produzidos e determinar as suas respectivas razões volumétricas de produção;
- Medir a pressão e a temperatura do reservatório;
- Obter amostras para a análise de PVT (Pressão, Volume e Temperatura);
- Determinar a produtividade do poço (*well deliverability*);
- Avaliar a eficiência da completação;
- Caracterizar os danos do poço.

Já os testes para caracterização possuem o objetivo de:

- Avaliar os parâmetros do reservatório tais como permeabilidade, espessura, *channels widths*, padrões conectados de escoamento e barreiras (Deutsch, 2002);
- Caracterizar as heterogeneidades do reservatório;
- Determinar a extensão do reservatório, espessura e a sua geometria.
- Determinar a comunicação hidráulica entre poços.

Esses últimos são possivelmente os dados mais valiosos sob o ponto de vista de um engenheiro de petróleo. Na perspectiva de um engenheiro, a construção da

caracterização de um reservatório não tem valor se o modelo de reservatório não pode reproduzir os dados de produção (Kelkar e Perez, 2002).

2.3.2. Métodos Geoestatísticos

2.3.2.1. Métodos geoestatísticos convencionais

O variograma é a medida de correlação espacial mais utilizada para modelagem de grids de porosidade e permeabilidade. A seleção criteriosa dos parâmetros do variograma é crítica para se obter um variograma de amostras limpo e interpretável (Deutsch, 2002).

O variograma é definido como o valor esperado de:

$$\gamma_{z,z}(h) = \frac{1}{2} E\{[Z(u) - Z(u + h)]^2\} \quad (3)$$

Onde:

u é o vetor de coordenadas;

$Z(u)$ é uma variável aleatória definida na região de estudo;

h é o vetor de distância (lag) que separa os dados;

$Z(u + h)$ versão lagged da variável aleatória.

O crossvariograma estendido para duas variáveis aleatórias $Z(u)$ e $Y(u)$ é definido pelo valor esperado conforme expressão abaixo.

$$\gamma_{z,y}(h) = \frac{1}{2} E\{[Z(u) - Z(u + h)][Y(u) - Y(u + h)]\} \quad (4)$$

O variograma e o crossvariograma são momentos de ordem dois, ou seja, estatística de “dois pontos”, que geralmente são inferidas a partir de dados de obtidos através de medidas precisas e esparsas de alguns poucos poços.

O variograma realiza uma descrição adequada das heterogeneidades dentro de uma única fácies. Contudo, o variograma não é capaz de capturar as formas complexas curvilíneas (como canais, estratificação oblíqua) e também não é capaz de descrever grandes conectividades dentro de um reservatório (Caers e Zhang, 2004).

2.3.2.2.

Algoritmo de Simulação Geoestatística de Múltiplos Pontos - Snesim

A geoestatística de múltiplos pontos se baseia no conceito de imagens de treinamento. Essas imagens são um banco de dados de padrões geológicos, a partir da qual a estatística de múltiplos pontos pode ser inferida (Caers e Zhang, 2004).

Segundo Strebelle e Journel (2001), essa imagem de treinamento deve refletir as heterogeneidades da sub-superfície como a presença de canais interconectados, que são característicos de reservatórios de petróleo. A imagem de treinamento não precisa representar nenhuma informação localizada precisa, mas sim necessita refletir o conceito obtido pela análise sísmica, pelo perfil de poço e pela comparação entre reservatórios análogos, ou seja, precisa refletir o conceito estrutural construído pelos dados estáticos.

Caers e Zhang (2002) comparam três imagens de treinamento e os seus respectivos variogramas, conforme mostra figura 9.

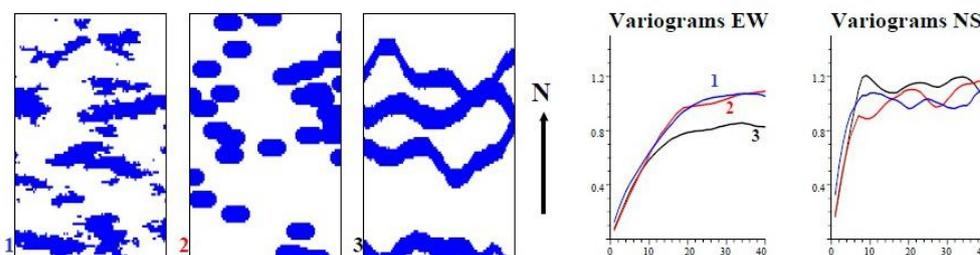


Figura 9. Três diferentes heterogeneidades geológicas e três variogramas similares.

Fonte: CAERS, ZHANG (2002).

A figura 9 mostra três possíveis candidatas para uma imagem de treinamento; a primeira com uma textura estratificada; a segunda, um conjunto de formas elípticas; e a terceira, um conjunto de formas fluviais. Como o variograma é baseado em momentos de ordem 2, ou seja, “estatística de dois pontos”, ele não é capaz de captar essas heterogeneidades dos padrões.

Daí, a necessidade de outros métodos geoestatísticos de forma a melhor caracterizar estruturas heterogêneas as presentes em reservatórios de petróleo.

A abordagem geoestatística de múltiplos pontos, ou seja, estatística baseada em momentos de ordem superior, proposta por Strebelle e Journel (2001), utiliza uma imagem de treinamento, para estimar a probabilidade condicional de uma

variável aleatória, dadas outras variáveis vizinhas. Com isso, a estatística de múltiplos pontos expressa a variabilidade conjunta de mais de dois locais ao mesmo tempo a partir de dados de uma imagem de treinamento.

A partir da inferência da geoestatística de múltiplos pontos a partir de imagens de treinamento, Strebelle & Journel (2001) propuseram o algoritmo *Snesim* para realizar uma simulação geoestatística que gera realizações de forma que as mesmas considerem a geoestatística de múltiplos pontos.

O *Snesim* é um algoritmo de simulação seqüencial. Este método se baseia na idéia de simular cada nó do grid de fácies, ou de uma propriedade geológica, seqüencialmente por um caminho aleatório, onde a simulação de nós a serem simulados é restringida pela realização dos nós já simulados. O algoritmo funciona da seguinte forma:

Inicialmente é construído um grid 3D ou 2D no qual cada nó não possui valor de propriedade. Em seguida, são preenchidos os valores da propriedade onde existe um poço ou onde a propriedade é conhecida. Em seguida, efetua-se um ciclo com quatro passos:

- No primeiro passo, escolhe-se um nó sem propriedade através de um caminho aleatório.
- No segundo passo, procura-se pelos nós mais próximos que contenham dados de poços e pelos nós vizinhos já simulados.
- No terceiro passo, baseados nesses nós selecionados no segundo passo, constrói-se um modelo de probabilidade a partir de uma imagem de treinamento considerando esses múltiplos pontos selecionados.
- No quarto passo, através do modelo probabilístico calculado no passo anterior, efetua-se uma realização para o nó. Obtida a realização o ciclo se perpetua até que todos nós que não contém valor de propriedade seja visitado.

O algoritmo do *Snesim* está disponível em um programa chamado SGeMs - Stanford Geostatistical Modeling Software (Remy, Boucher e Wu, 2009). Com isso, através do *Snesim* é possível obter realizações de um campo de permeabilidade, por exemplo, utilizando estatística de múltiplos pontos a partir do

conceito geológico construído nas etapas iniciais de caracterização do reservatório.

No próximo capítulo, será descrito mais detalhadamente o tema do ajuste de histórico, destacando-se algumas das metodologias já desenvolvidas.

3

Ajuste de Histórico

O capítulo anterior fez uma breve introdução a algumas técnicas de caracterização de reservatórios como sísmica, perfilagem, testemunhos e teste de formação, bem como apresentou algumas técnicas de modelagem geoestatística. Conforme já mencionado, essa aquisição de dados é fundamental para a caracterização das propriedades de um reservatório e para a construção de modelos de reservatórios. Contudo, esses modelos necessitam de uma etapa consecutiva de ajuste para refletir os dados históricos de produção.

Deste modo, esse capítulo visa caracterizar o problema do ajuste de histórico que representa a segunda etapa da construção do modelo de reservatório.

3.1.

O Problema

O problema de ajuste de histórico é na verdade um problema inverso, onde existe um conjunto de dados observáveis (dados de produção, dados de pressão de fundo de poço, etc.) e deseja-se obter um conjunto de parâmetros não observáveis (parâmetros do modelo matemático do reservatório), de forma que a simulação do modelo matemático seja capaz de reproduzir os dados reais de produção e efetuar uma previsão do comportamento futuro de um reservatório.

O ajuste dos modelos é uma tarefa recorrente para o engenheiro de reservatório por representar uma parte importante da integração das atividades de caracterização de reservatórios e previsão de produção de petróleo (Schiozer, Sousa e Maschio, 2009).

O objetivo final do ajuste de histórico é obter uma previsão mais realista a produção de petróleo para subsidiar o processo de tomada de decisões durante o plano de desenvolvimento ou no gerenciamento de um campo de óleo e/ou gás conforme esquematizado na figura 10.



Figura 10. Diagrama do Problema de Ajuste de Histórico

No processo de tomada de decisões, um modelo de reservatório que é consistente com os dados disponíveis é uma importante fonte de informação para o plano de desenvolvimento ou para o gerenciamento de um campo de petróleo ou gás (Deutsch, 2002), como por exemplo:

- Na etapa de desenvolvimento do campo, um modelo de reservatório adequado fornece estimativas confiáveis de volumes de hidrocarbonetos no reservatório. Além disso, subsidia uma otimização do plano de desenvolvimento, de forma a permitir encontrar um número ótimo de poços, a sua localização e também o tipo de poços a serem perfurados (Deutsch, 2002).
- Na etapa de gerenciamento de um campo, um modelo de reservatório realista pode, por exemplo, auxiliar no ajuste da melhor estratégia de injeção (Deutsch, 2002).

Contudo, para isso, é necessária uma caracterização realista das propriedades dos fluidos e das rochas que compõem o reservatório. Dessa forma, o processo de ajuste de histórico necessita de integrar todos os dados disponíveis obtidos nas etapas anteriores de caracterização para reduzir as incertezas sobre as previsões de produção do reservatório. O modelo ajustado deve não somente reproduzir os dados de produção através da simulação numérica, mas também ser consistente com o conhecimento geológico do reservatório (Roggero e Hu, 1998).

O ajuste de histórico é um problema desafiador, de grande importância econômica e de natureza técnica complexa das quais se destacam:

- A grande dimensionalidade;
- As múltiplas soluções (*Ill posed*);
- A necessidade de atender à caracterização geológica realizada a priori com base em dados estáticos;
- A necessidade de preservar características não lineares presentes inerentes aos reservatórios;
- O alto custo computacional relativo à simulação de modelos de reservatório.

Devido à grande discretização das malhas de simulação de reservatórios, um *grid* de permeabilidade, por exemplo, pode ter mais do que 10^5 parâmetros. Além disso, como o número de parâmetros é muito maior do que a quantidade de dados de produção, o problema do ajuste de histórico é um problema tipicamente mal condicionado (*Ill posed*). Ou seja, pode existir uma variedade de conjuntos de parâmetros de modelos de reservatório que geram a mesma curva de produção (Kelkar e Perez, 2002).

O problema do ajuste de histórico não é novo. Os primeiros trabalhos encontrados na literatura sobre o tema são de Jacquard (1964) e Jacquard e Jain (1965). A partir daí, novas metodologias foram sendo desenvolvidas com o objetivo de resolver o problema de ajuste de histórico. As principais abordagens são descritas a seguir.

3.2. Abordagens de Ajuste de Histórico

As principais abordagens de ajuste de histórico são apresentadas neste trabalho de acordo com a natureza das considerações iniciais, estando subdivididas em: abordagens por zoneamento do reservatório, abordagens utilizando a inversão bayesiana e outras metodologias.

3.2.1. Abordagens por Zoneamento do Reservatório

A mais simples parametrização é feita pelo método de zoneamento, onde o reservatório é subdividido em pequenas regiões nas quais as propriedades são assumidas uniformes, o que transforma o ajuste de histórico em um problema bem

condicionado dada a redução do número de variáveis. Jacquard & Jain (1965), Jahns (1966), Ramgulum, Ertekin & Flemings (2007), entre outros, utilizaram esse método para reduzir o número de parâmetros para ajuste de histórico automático.

Jacquard & Jain (1965) dividiram o reservatório em regiões cujos parâmetros de permeabilidade eram considerados homogêneos, conforme mostra a figura 11, e utilizaram métodos baseados em gradientes para cálculo dos parâmetros.

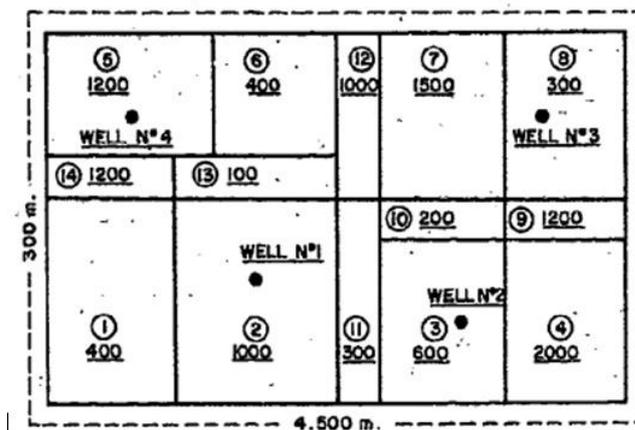


Figura 11. Ilustração da Solução apresentada por Jacquard & Jain (1965)

Ramgulum, Ertekin, & Flemings (2007) utilizaram Rede Neurais para aproximar propriedades como porosidade, permeabilidade e altura do reservatório, utilizando reservatórios com regiões homogêneas predefinidas.

Ainda, no contexto da abordagem por Zoneamento, a indústria petrolífera de uma maneira geral tem efetuado o ajuste de histórico, pelo método de tentativa e erro, multiplicando por regiões os parâmetros do reservatório (Maschio, Vidal e Schiozer, 2008). Segundo Hoffman, Wen, & Strebelle (2005), os modelos de reservatório construídos pela indústria freqüentemente não são realísticos de um ponto de vista geológico ou as mudanças são realizadas sem qualquer consideração geológica.

3.2.2. Abordagens utilizando Inversão Bayesiana

Foram desenvolvidas muitas metodologias para ajuste de histórico através da abordagem da Inversão Bayesiana. Em muitas dessas abordagens, considera-se que a densidade de probabilidade de cada variável aleatória é independente e é

uma gaussiana. Com isso, a densidade de probabilidade do conjunto de variáveis aleatórias é multigaussiana. Gavalas, Shah e Seinfeld (1976) foram os precursores do uso da teoria da Inversão Bayesiana para a resolução de problemas de ajuste de histórico. Uma breve descrição desse trabalho é apresentada a seguir.

3.2.2.1. Ajuste de Histórico por Inferência Bayesiana

Esse trabalho, desenvolvido por Gavalas, Shah e Seinfeld (1976), se baseia no teorema de Bayes, onde é estimada a densidade de probabilidade condicional para as variáveis do modelo (m) dadas as observações (d_{obs}), conforme equação (5).

$$f(m|d_{obs}) = \frac{f(d_{obs}|m)f(m)}{f(d_{obs})} \quad (5)$$

Onde:

$f(m)$ é a densidade de probabilidade dos parâmetros do modelo a priori, ou seja, antes de incorporar as observações;

$f(d_{obs}|m)$ é a densidade de probabilidade das observações (d_{obs}) dados os parâmetros do modelo (m) essa função também é chamada de verossimilhança dos parâmetros do modelo (m) dadas as observações (d_{obs}).

Como $f(d_{obs})$ é uma constante, então, define-se $f(m|d_{obs})$ conforme a equação:

$$f(m|d_{obs}) = a f(d_{obs}|m)f(m)$$

Onde:

a representa uma constante de normalização; (6)

$f(m)$ é a densidade de probabilidade dos parâmetros do modelo a priori, ou seja, antes de incorporar as observações;

$f(d_{obs}|m)$ é a densidade de probabilidade das observações (d_{obs}) dados os parâmetros do modelo (m) essa função também é chamada de verossimilhança dos parâmetros do modelo (m) dadas as observações (d_{obs}).

Para se estimar a densidade de probabilidade condicional $f(m|d_{obs})$, neste trabalho, são efetuadas as seguintes considerações:

- É considerada uma relação linear entre os parâmetros do modelo (m) e os dados de produção (d), conforme equação (7).

$$d = Gm \quad (7)$$

- A densidade de probabilidade dos parâmetros do modelo (a priori) também são multigaussianas com média igual a μ e matriz de covariância igual a C_M , conforme apresentado pela equação (8).

$$f(m) = \frac{1}{(2\pi)^{\frac{N_m}{2}} \det(C_M)} \exp\left(-\frac{1}{2}(m - \mu)^T C_M^{-1} (m - \mu)\right) \quad (8)$$

Onde:

N_m é o número de parâmetros do modelo a serem ajustados;

m são os parâmetros do modelo de reservatório a serem ajustados;

μ é o valor esperado dos parâmetros do modelo (m);

C_M é a matriz de covariância dos parâmetros do modelo, calculada com base no variograma obtido a partir dos dados de poços.

- As medidas dos dados de produção podem ser quantificadas e que não há erros na modelagem. Com isso, a relação entre as observações (d_{obs}) e os parâmetros do modelo (m) é dada pela equação (9).

$$d_{obs} = Gm + \epsilon \quad (9)$$

Onde:

ϵ representa os erros de medida dos dados.

- As medidas do erro possuem uma distribuição multigaussiana com média igual a zero e matriz de covariância C_D .

Dada as considerações acima, a densidade de probabilidade das observações dados os parâmetros do modelo $f(d_{obs}|m)$, também chamada de verossimilhança, é também uma distribuição multigaussiana e é dada pela equação (10).

$$f(d_{obs}|m) = \frac{1}{(2\pi)^{\frac{N_d}{2}} \det(C_D)} \exp\left(-\frac{1}{2}(d_{obs} - Gm)^T C_D^{-1} (d_{obs} - Gm)\right) \quad (10)$$

Onde:

N_d é o número de observações dos dados;

m são os parâmetros do modelo de reservatório a serem ajustados;

d_{obs} são os dados observados;

C_D é a matriz de covariância dos erros de medida dos dados.

Substituindo a equação (8) e (10) em (6), tem-se a expressão apresentada pela equação (11) e (12).

$$f(m|d_{obs}) = a' \exp(-O) \quad (11)$$

$$O = (d_{obs} - Gm)^T C_D^{-1} (d_{obs} - Gm) + (m - \mu)^T C_M^{-1} (m - \mu) \quad (12)$$

Onde:

a' representa uma constante de normalização;

m são os parâmetros do modelo de reservatório a serem ajustados;

d_{obs} são os dados observados;

C_D é a matriz de covariância dos erros de medida dos dados;

C_M é a matriz de covariância a priori dos parâmetros do modelo.

O objetivo é encontrar o estimador de máxima a posteriori (MAP), ou seja, o conjunto de parâmetros m que maximiza a densidade de probabilidade condicional $f(m|d_{obs})$, dada pela equação (11). Essa maximização corresponde a minimizar a função quadrática O representada na equação (12).

Gavalas, Shah e Seinfeld (1976) não sugerem que os parâmetros do modelo m sejam otimizados diretamente. Eles sugerem que a matriz de covariância seja decomposta espectralmente e que os parâmetros do modelo sejam reconstruídos através de um somatório das projeções dos parâmetros nos autovetores multiplicadas aos autovetores, como na expansão de Karhunen-Loève (Karhunen, 1947; Loeve, 1977), apesar dessa expansão não ser citada diretamente. Essa parametrização permite uma redução do número de parâmetros a serem ajustados.

Com isso, Gavalas, Shah e Seinfeld (1976) sugerem que o parâmetro a ser minimizado seja a projeção dos parâmetros a serem ajustados nos autovetores utilizando métodos de otimização baseados em gradientes.

Muitos outros trabalhos, dentro da mesma linha de abordagem, foram desenvolvidos entre eles: Gomez-Hernandez *et al.* (1997), Zhang *et al.* (2002) aplicaram o problema inverso em uma aplicação similar realizando o ajuste de histórico.

Gomez-Hernandez *et al.* (1997) desenvolveram o método *Sequential Self-Calibration Method* (SSC), sugerindo um algoritmo de minimização baseado em métodos gradientes onde a função objetivo era minimizada através do cálculo de coeficientes de sensibilidade. Contudo, o cálculo de tais coeficientes é computacionalmente custoso.

Zhang *et al.* (2002) propuseram, dentro do contexto de Inversão Bayesiana, que a função objetivo fosse minimizada através dos métodos de otimização quasi-Newton BFGS e L-BFGS.

Outros métodos dentro de um contexto bayesiano também foram propostos, como:

- *Monte Carlo Markov Chain* (MCMC);
 - *Ensemble Kalman Filter* (EnKF);
 - Kernel PCA (KPCA);
 - Probability Perturbation Method.
- } Baseados em Simulação
Sequencial / preservar a
estatísticas de Múltiplos
pontos

Moosegard (1995) propôs o uso de amostragens Monte Carlo mais eficientes para aplicações em ajuste de histórico. Posteriormente, Oliver, Cunha e Reynolds (1997) propuseram a amostragem de soluções de Monte Carlo em Cadeias de Markov (MCMC) para a aplicação em ajuste de histórico.

Em 2003, Nævdal, Johnsen e Aanonsen (2003) propuseram uma metodologia para o ajuste de histórico utilizando o Ensemble Kalman Filter (EnKF). O EnKF é uma implementação Monte Carlo do problema de atualização bayesiana. Dada uma função de densidade de probabilidade dos estados e um sistema e da verossimilhança aos dados, o teorema de Bayes é utilizado para obter a função de densidade de probabilidade a posteriori, cujo procedimento é chamado de atualização bayesiana.

No EnKF, a atualização bayesiana é combinada com um modelo que avança no tempo, incorporando novos dados de tempos em tempos. O EnKF representa uma distribuição dos estados do sistema utilizando um conjunto de vetores de

estado, chamados de *ensemble*, e atualizam a matriz de covariância através de amostras computadas a partir do *ensemble* (Nævdal, Johnsen e Aanonsen, 2003).

A seguir são apresentadas abordagens que não utilizam a inversão Bayesiana tradicional e considera a estatística de múltiplos pontos durante o processo de ajuste de dados.

3.2.2.2.

Métodos no Contexto Bayesiano baseados em Simulação Seqüencial

3.2.2.2.1.

Kernel PCA

Ainda no contexto da inversão bayesiana, Sarma (2006) propôs parametrizar o campo de permeabilidade através de Kernel PCA (Schölkopf, Knirsch, *et al.*, 1998) e a expressão da parametrização à inversão bayesiana.

Sarma (2006) sugere que realizações do campo de permeabilidade sejam obtidas através do algoritmo de simulação seqüencial *Snesim*, que gera realizações de acordo com a estatística de múltiplos pontos inferida a partir de uma imagem de treinamento.

Em seguida, Sarma parametriza o campo de permeabilidade $y_F(\xi)$ no espaço das características (F) através do Kernel PCA (KPCA), onde ξ é uma variável aleatória não correlacionada.

Este espaço é mapeado de forma implícita através de uma função kernel. Esse mapeamento existe se a função kernel for contínua e positiva definida dentro de um determinado intervalo e essa função kernel corresponderá ao produto escalar no espaço das características. Após esse mapeamento implícito, é necessário que se efetue um mapeamento do espaço das características para o espaço das entradas, que é chamado de *Pre-Image*. No capítulo a seguir será realizado um detalhamento desse mapeamento implícito, do Kernel PCA e do *Pre-Image*.

Sarma sugere que a *Pre-Image* do campo de permeabilidade mapeado no espaço das características $y_{RN}(\xi)$ seja embutida na expressão da inversão bayesiana como na equação (13).

$$\min_{\xi} \left[S = \{y_{RN}(\xi) - y_{prior}(\xi)\}^T C_M^{-1} \{y_{RN}(\xi) - y_{prior}(\xi)\} + \right. \quad (13)$$

$$\left. d\xi - dprior\xi TI d\xi, x - dprior\xi, x \right]$$

Onde:

y_{RN} é o campo aleatório reconstruído no espaço das entradas. É

função da variável aleatória independente (ξ);

$y_{prior}(\xi)$ é o campo aleatório obtido a priori;

C_M é a matriz de covariância do campo aleatório y ;

d_{prior} é o valor do parâmetro de produção observado;

d é o valor do parâmetro de produção calculado;

x são os estados dinâmicos, a serem definidos dentro da estratégia de controle do campo.

O modelo proposto por Sarma (2006) leva em consideração que a densidade de probabilidade do campo de permeabilidade é multigaussiana e a relação entre os parâmetros de entrada e saída são lineares, como na equação (13).

A distribuição multigaussiana é o único caso em que a decomposição é conhecida analiticamente. A verossimilhança, a distribuição a posteriori e a priori são todas multigaussianas somente se a relação entre os parâmetros do modelo e os de saída for uma função linear ou puder ser suficientemente linearizada (Caers e Hoffman, 2006).

Entretanto, de forma geral, a relação entre os dados de produção e os parâmetros do modelo são não lineares nos reservatórios de petróleo. Com isso, dado a relação entre os dados de entrada e os de saída não são lineares, minimizando-se S não se garante que o estimador de máxima a posteriori seja y (Sarma, 2006).

Além disso, Sarma (2007) sugere utilizar métodos de gradientes para minimização da função objetivo proposta por ele. Isso restringe o espaço de busca, já que o conceito é se distanciar minimamente do chute inicial, ou seja, o que ficar preso a mínimos locais.

3.2.2.2.1.1.**Perturbação Probabilística (*Probability Perturbation*)**

Caers e Hoffman (2006) propõe este novo método de ajuste de histórico que garante que a geologia é honrada durante o ajuste de dados e permite que as realizações sejam encontradas de forma eficiente.

A técnica *Probability Perturbation* é também uma técnica de parametrização. Ao invés de parametrizar um conjunto de parâmetros como no Kernel PCA, a *Probability Perturbation* é baseada na parametrização das leis das probabilidades e na forma analítica (tau-model) dada por (Journel, 2002) para aproximar as probabilidades condicionais (Hu, 2008). A densidade de probabilidade condicional é obtida através da expressão (14):

$$P(A|D) = (1 - r)u + rP(A), \quad r \in [0,1] \quad (14)$$

Onde: A é um determinado evento do parâmetro desconhecido;

D é um evento observado;

u é a realização inicial do parâmetro desconhecido;

r é o parâmetro a ser otimizado.

A partir dessa probabilidade condicional, serão obtidas realizações tal que o parâmetro r seja otimizado de forma que o erro entre a produção simulada e a produção observada seja mínimo.

3.2.3.**Outras Metodologias****3.2.3.1.****Deformação Gradual**

Uma metodologia similar à *Probability Perturbation* é a Deformação Gradual, na verdade, a *Probability Perturbation* é inspirada na Deformação Gradual.

O método de Deformação Gradual foi proposto por Roggero e Hu (1998). Este método permite que o modelo inicial do reservatório seja modificado de forma contínua e gradual preservando o modelo geoestatístico (gaussiano ou não).

Primeiramente, são consideradas realizações que honrem os dados do perfilagem e aplica-se uma transformada para cada realização, obtendo-se uma parametrização linear das realizações anteriores, como na equação (15).

$$Z_{it} = \cos(\rho\pi) Z_{it-2} + \text{sen}(\rho\pi) Z_{it-1} \quad (15)$$

Onde:

Z é um conjunto de parâmetros do modelo de reservatório

it é o número de iterações

ρ é um valor a ser otimizado.

Em seguida, o parâmetro ρ é otimizado minimizando-se uma função objetivo descrita pela equação (16) a qual relaciona o somatório dos erros quadráticos entre a produção simulada e a produção observada.

$$F = 0,5 \sum_{i=1}^{n_{\text{obs}}} w_i (d_{\text{sim}}^i(Z_{it}) - d_{\text{obs}}^i)^2 \quad (16)$$

Onde:

d_{sim} é a produção simulada

d_{obs} é a produção observada

w são os coeficientes de peso

n_{obs} é a o número total de medidas

Z_{it} é um conjunto de parâmetros do modelo de reservatório

Dado o ρ obtido pelo processo de otimização, o Z_{it} substituirá uma das realizações iniciais e sucessivas iterações serão obtidas até que se obtenha um conjunto de parâmetros que seja considerado satisfatório.

Um método semelhante ao proposto por Roggero e Hu (1998) foi proposto por Hu, Blanc e Noetinger (2001), que utiliza o método de deformação gradual a partir de realizações obtidas com simulação sequencial, mostrando como o método pode ser adaptado para atender a estatística de múltiplos pontos.

3.2.3.2. Recozimento Simulado/ Algoritmo Genético

Algumas das tentativas para condicionar os campos de permeabilidade aos dados de produção são baseadas em algoritmos estocásticos de otimização como recozimento simulado (*Simulated Annealing*) e Algoritmo Genético.

O *Simulated Annealing* é um método probabilístico proposto por Kirkpatrick, Gellett e Vecchi (1983) e Cerny (1985) para encontrar um mínimo global de uma função de custo que pode ter vários mínimos locais. Esse método é inspirado em processos físicos nos quais um sólido é resfriado lentamente até que

a sua estrutura esteja completamente imobilizada, o que ocorre na configuração de energia mínima.

No contexto do ajuste de histórico, Ouenes, Brrfort, *et al.* (1993) sugeriram utilizar *Simulated Annealing* para otimização de alguns parâmetros do modelo de reservatório, contudo, não foi imposta qualquer restrição geoestatística. Sagar, Kelkar e Thompson (1993) e Sen, Gupta, *et al.* (1992) propuseram que fosse ajustado um modelo de reservatório de forma que a estatística de dois pontos obtida através do variograma fosse preservada.

Já, o Algoritmo Genético é um modelo de otimização inspirado na teoria da evolução, que buscam uma solução global (Michalewicz, 1996). Nesse algoritmo de otimização, cada cromossomo ou indivíduo representa uma solução para o problema. O processo evolutivo ocorre na população de indivíduos e corresponde a uma busca sobre o espaço de soluções potenciais. Cada busca tem dois objetivos: explorar as melhores soluções e explorar todo espaço de busca. O Algoritmo Genético busca um equilíbrio entre a prospecção e a exploração do espaço de busca.

Romero e Carter (2001) propuseram efetuar o ajuste de histórico através da otimização com Algoritmo Genético. Nesse trabalho, foram propostos o uso de mutação *jump* e *creep* e uma generalização do *k-point* crossover. Essa generalização consistiu da utilização de uma máscara para definir de qual pai um gene será copiado durante o processo de *crossover*. A figura 12 ilustra o *crossover* proposto para um caso bidimensional.

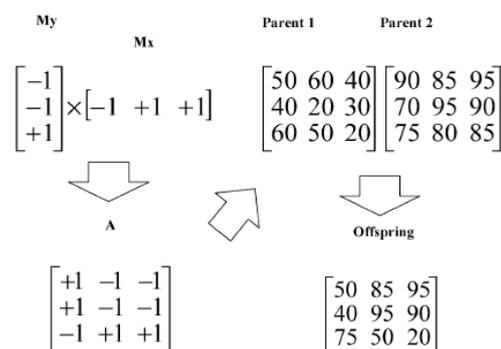


Figura 12. Exemplo do operador *crossover* proposto

Schulze-Riegert, Axmann, *et al.* (2002) aplicaram um Algoritmo Genético para efetuar o ajuste de histórico de um reservatório. Esse trabalho dividiu o reservatório em regiões, que por sua vez eram divididas sub-regiões conforme mostra a figura 13.

Nesse trabalho, foi otimizado um parâmetro que multiplicava a permeabilidade de cada região do reservatório. Schulze-Riegert, Axmann, *et al.* (2002) também sugeriram utilizar métodos de computação em paralelo para avaliar a função objetivo.

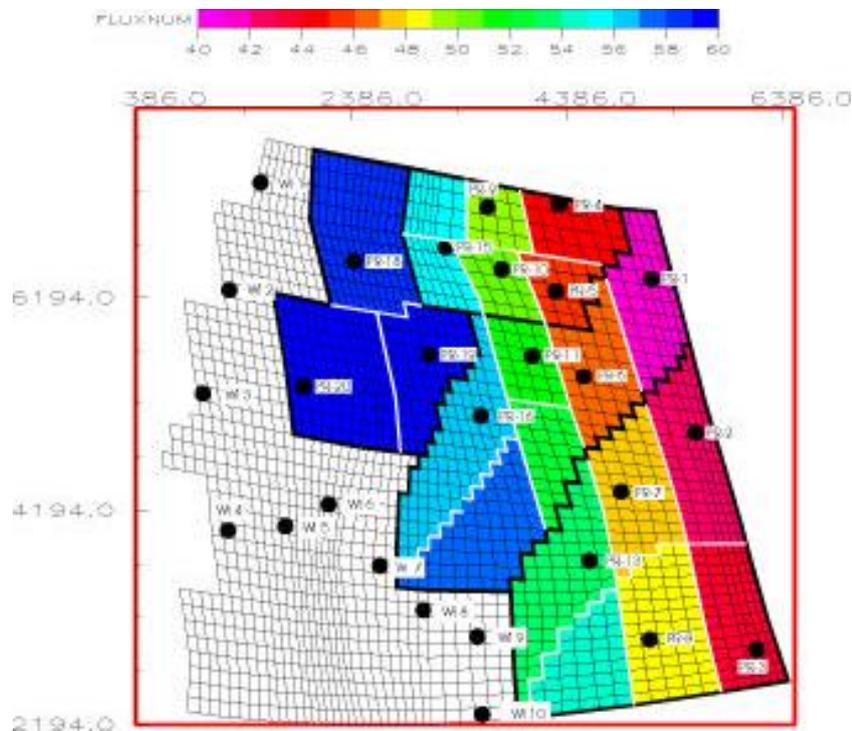


Figura 13. Reservatório subdividido em regiões (contorno em preto) e em sub-regiões (áreas coloridas)

Maschio, Vidal e Schiozer (2008) propôs um algoritmo para ajuste de histórico utilizando Algoritmo Genético de forma a obedecer as restrições geoestatísticas. Inicialmente algumas realizações de litofácies são obtidas a partir dos dados obtidos da perfilagem, por exemplo, utilizando-se um simulador de Simulação Sequencial. Em seguida, são obtidas realizações de parâmetros do reservatório considerando a densidade de probabilidade de cada parâmetro de cada litofÁCIE uma gaussiana. Durante o processo evolutivo, 5 parâmetros correspondentes a simulação geoestatística são evoluídos de forma que o erro quadrático entre a curva de produção simulada e a curva de produção observada seja mínima.

3.2.3.3. Programação Genética

Yu, Wilkinson e Castellini (2008) propuseram um algoritmo baseado em programação genética para selecionar possíveis realizações de modelos de reservatório as quais se ajustam à curva de produção. Em seguida, efetuam uma análise de incerteza dos cenários futuros de produção.

3.3. Discussão das principais técnicas utilizadas

Apesar de existirem muitas metodologias que se propõem a resolver o problema de ajuste de histórico, a maioria delas não leva em consideração restrições geoestatísticas ou ainda, que haja relações lineares entre os dados de produção e os parâmetros do modelo de reservatório (Wen, Capilla, *et al.*, 1999).

A maioria dos modelos baseados na abordagem de zoneamento citados podem gerar modelos de reservatórios inconsistentes.

Já em relação aos modelos tradicionais que envolvem a Inversão Bayesiana. Esses se baseiam na distribuição a priori e na verossimilhança e posteriormente a amostragem para efetuar a distribuição a posteriori (Caers e Hoffman, 2006). A distribuição multigaussiana é o único caso em que a decomposição é conhecida analiticamente. A verossimilhança, a distribuição a posteriori e a priori são todas multigaussianas somente se a relação entre os parâmetros do modelo e os de saída for uma função linear ou puder ser suficientemente linearizada (Caers e Hoffman, 2006). Entretanto, de forma geral, a relação entre os dados de produção e os parâmetros do modelo são não lineares nos reservatórios de petróleo.

Os modelos tradicionais que se baseiam na Inversão Bayesiana requerem uma matriz de covariância a priori, que geralmente é estimada a partir de observações diretas dos parâmetros do modelo. Essas medidas são de amostras retiradas de poços em reservatórios. Como o número de poços exploratórios é limitado devido ao alto custo financeiro dos mesmos, esses dados são tipicamente esparsos.

Outros métodos baseados em simulação seqüencial, como o KPCA e *Perturbation Method* vêm sendo desenvolvidos, indicando uma nova tendência para o ajuste de histórico, onde é preservada a estatística de múltiplos pontos.

Esses métodos possuem a vantagem de gerar realizações mais realistas durante o processo de ajuste, dependendo da imagem de treinamento.

Modelos de ajuste de histórico baseados em técnicas de otimização robustas como *Simulated Annealing* e Algoritmo Genético foram pouco investigados e a maioria das abordagens não leva em consideração restrições geoestatísticas durante o processo de ajuste dos dados.

Além disso, as poucas metodologias desenvolvidas com essas técnicas, que consideram a geoestatística durante o processo de ajuste, utilizam apenas momentos de ordem dois obtidos de dados de poços como restrições geoestatísticas.

Dessa forma, esses algoritmos robustos de otimização possuem um grande potencial para auxiliar no problema de ajuste de histórico dado que esse é um problema de otimização complexo. Contudo, há a necessidade do desenvolvimento de metodologias para preservar às restrições geoestatísticas durante o processo de ajuste de dados, principalmente, que levem em consideração a estatística de múltiplos pontos.

4

Estado da Arte do Kernel PCA

Conforme apresentado no Capítulo 1, o Kernel PCA é uma técnica promissora para auxiliar no processo de ajuste de histórico, devido a sua capacidade de extrair parte da estatística de múltiplos pontos (Kim, Franz e Schölkopf, 2005). Essa característica é altamente desejável no contexto do ajuste de histórico.

O Kernel PCA (KPCA) é uma versão não-linear da Análise de Componentes Principais. Esse método é calculado implicitamente através de uma função Kernel para um espaço de maior dimensionalidade. O KPCA é capaz de extrair características não lineares a partir de um conjunto de dados (Schölkopf, Knirsch, *et al.*, 1998).

Este capítulo, portanto, apresenta os fundamentos da técnica Kernel PCA (KPCA), de forma a facilitar a compreensão do modelo proposto neste trabalho. As seções seguintes apresentam a Análise de Componentes Principais convencional, como o mapeamento implícito para um espaço de maior dimensionalidade é efetuado, os detalhes sobre Kernel PCA (KPCA) e, por fim, os conceitos sobre o problema de *Pre-Image*.

4.1.

Análise de Componentes Principais (PCA)

O objetivo principal do PCA é a redução de dimensionalidade de um conjunto de dados, cujas variáveis estão inter-relacionadas, e manter o máximo de variância presente no conjunto de dados (Jolliffe, 2005).

Esta redução é alcançada pela transformação do conjunto de coordenadas originais para um novo conjunto de coordenadas, as componentes principais, que não são correlacionadas e que estão ordenadas de forma que um subconjunto reduzido de componentes preserve o máximo de variância presente nas variáveis originais.

No algoritmo de PCA, a matriz de covariância (C) do conjunto de dados $\{x_1, \dots, x_l\} \subset R^N$ com média igual a zero, é decomposta em auto-vetores (V) e auto-valores (λ), conforme equação (17):

$$\lambda V = CV \quad (17)$$

Os autovetores obtidos são as componentes principais. Como a matriz de covariância é simétrica, os autovetores são ortogonais entre si e, portanto, não são correlacionados. Quanto maior o autovalor relacionado ao autovetor, maior é a variância dos dados na direção do autovetor. Portanto, para se preservar a máxima variância possível, os autovetores são ordenados conforme os seus respectivos autovalores. Na figura 14, observa-se um exemplo de PCA, onde os vetores em vermelho são as componentes principais.

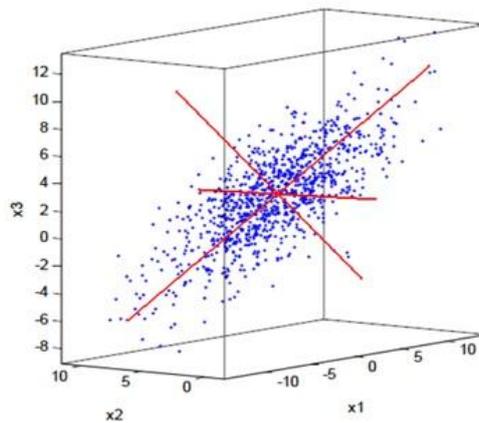


Figura 14. Exemplo de PCA

Em aplicações de redução de ruído e de redução de dimensionalidade, um dado x pode ser reconstruído através de uma expansão, que relaciona os autovetores (V_j) e a projeção (β_j) do dado x nos autovetores, para $j = 1, \dots, n < N$.

$$x \cong \sum_{j=1}^n V_j \beta_j + \bar{x} \quad (18)$$

Onde:

V_j é o autovetor k dos dados originais,

β_j é a projeção de x em V_j .

4.2. Mapeamento Não Linear através de uma Função Kernel

Segundo Schölkopf, Smola e Müller (1998), o teorema Mercer de análise funcional implica que se k é uma função real contínua tal que o operador integral é positivo, existe um mapeamento ϕ para um espaço F onde k atua como produto escalar.

$$\int_{X \times X} k(z, x) f(z) f(x) dz dx > 0, \forall f \neq 0 \tag{19}$$

$$k(z, x) \text{ converge para } \langle \phi(z), \phi(x) \rangle \tag{20}$$

$$\phi: R^N \rightarrow F, \quad x \rightarrow \phi(x), \quad x' \rightarrow \phi(x') \tag{21}$$

As funções k que obedecem ao teorema de Mercer são chamadas de kernels de Mercer. As funções kernels contínuas positivas definidas dentro de um determinado intervalo do espaço de entrada também podem ser representadas como produto escalar em um espaço F (Schölkopf e Smola, 2002). Este trabalho também demonstra que as funções kernels de Mercer são positivas definidas.

Dessa forma, a função kernel precisa ser contínua e positiva definida dentro de um intervalo do espaço de entrada para que seja capaz de efetuar um mapeamento implícito do espaço de entrada para um outro espaço, sendo que o produto escalar neste outro espaço é dado por essa função.

Uma representação desse mapeamento pode ser observada através da figura 15.

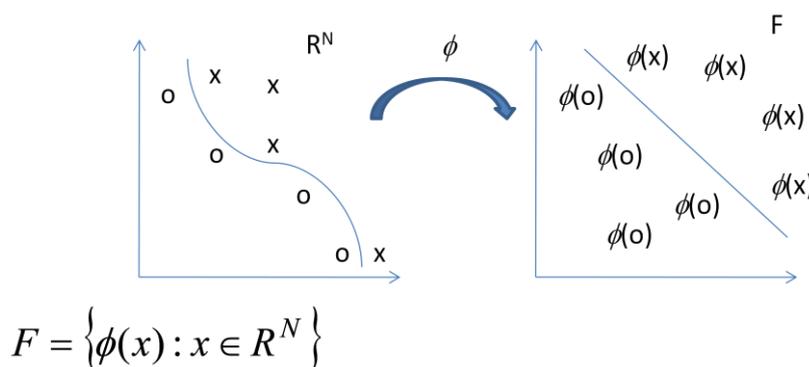


Figura 15. Diagrama esquemático de um mapeamento implícito

Alguns tipos de funções kernel têm sido freqüentemente adotados, tais como:

- Função kernel polinomiais de ordem d:

$$k(x, x') = (x \cdot x')^d \quad (22)$$

- Função kernel gaussiana de funções radiais (RBF):

$$k(x, x') = \exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right) \quad (23)$$

- Função kernel sigmóide:

$$k(x, x') = \tanh(\kappa(x \cdot x') + \theta) \quad (24)$$

A escolha da função kernel determina o tipo de mapeamento implícito ϕ que será realizado para o espaço das características F (Schölkopf, Smola e Müller, 1998).

Como exemplo, observa-se o mapeamento realizado pela função $\phi = (x^2, y^2, \sqrt{2}xy)^T$ corresponde ao mapeamento da função kernel polinomial de ordem 2 na figura 16 abaixo.

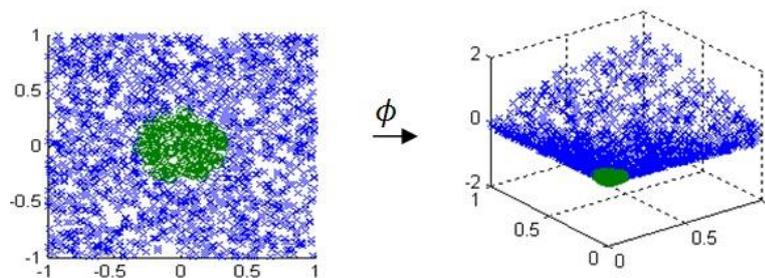


Figura 16. Exemplo de mapeamento de espaço de um kernel.

Dedução de que o mapeamento $\phi = (x^2, y^2, \sqrt{2}xy)^T$ corresponde ao mapeamento da função kernel polinomial de ordem 2:

Dados os pontos x e $x' \in R^2$, $x = (x_1, y_1)^T$ e $x' = (x_2, y_2)^T$, a função kernel polinomial de ordem 2 é dada por:

$$k(x, x') = (x \cdot x')^2 = (x_1x_2 + y_1y_2)^2 \quad (25)$$

$$k(x, x') = x_1^2x_2^2 + y_1^2y_2^2 + 2x_1y_1x_2y_2 \quad (26)$$

Para o mapeamento dado, $\phi(x) = (x_1^2, y_1^2, \sqrt{2}x_1y_1)^T$ e $\phi(x') = (x_2^2, y_2^2, \sqrt{2}x_2y_2)^T$. Então,

$$\phi(x) \cdot \phi(x') = x_1^2x_2^2 + y_1^2y_2^2 + 2x_1y_1x_2y_2 \quad (27)$$

$$\phi(x) \cdot \phi(x') = (x \cdot x')^2 \quad (28)$$

4.3. O Kernel PCA (KPCA)

O KPCA é uma Análise de Componentes Principais não-linear. Essa característica se deve ao fato que o KPCA é uma análise de componentes principais realizada em um espaço de maior dimensionalidade. Este espaço é mapeado implicitamente de forma não linear através da função Kernel. Daí, a capacidade que o KPCA possui de extrair características não-lineares (Schölkopf e Smola, 2002).

Conforme mencionado, outra característica importante do KPCA é a capacidade de capturar parte da estatística de ordem superior de uma série de dados (Kim, Franz e Schölkopf, 2005). Além disso, o KPCA possibilita uma compactação da representação de um conjunto de dados, uma vez que possibilita a redução do número de variáveis.

4.3.1. Algoritmo do Kernel PCA (KPCA)

Schölkopf, Smola e Müller (1997) propõem que se efetue uma análise de componentes principais no espaço F , chamando-a de Kernel PCA (KPCA). Para um conjunto de dados $\{x_1, \dots, x_L\} \subset R^N$, o KPCA equivale à análise de componentes principais desses dados mapeados no espaço F , através de uma função kernel.

A idéia básica do Kernel PCA é utilizar uma função kernel não linear de forma a computar implicitamente o PCA em um possível espaço de maior dimensionalidade F , o qual é não linearmente relacionado ao espaço de entrada (Schölkopf, Smola e Müller, 1997), conforme na figura 17.

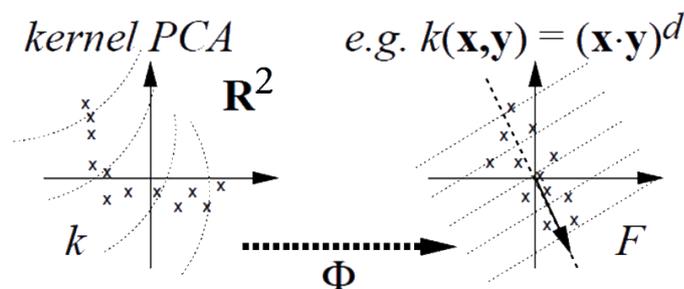


Figura 17. Esquema do KPCA.

Fonte: Schölkopf, Smola e Müller (1997).

As componentes principais no KPCA, entretanto, não são obtidas diretamente da matriz de covariância, como na análise de componentes principais (PCA) ou expansão de Expansão de Karhunen-Loève.

Como não se dispõe dos dados mapeados no espaço das características F , não é possível calcular diretamente a matriz de covariância (\bar{C}) dos dados mapeados em F , como na equação (30). Portanto, também **não** é possível decompor a matriz de covariância em autovetores (v) e autovalores (λ) de forma direta.

$$\lambda_j v_j = \bar{C} v_j, \text{ para todo } j = 1, \dots, L. \quad (29)$$

$$\text{Onde } \bar{C} = \frac{1}{L} \sum_{i=1}^L \phi(x_i) \phi(x_i)^T \quad (30)$$

Substituindo (30) em (29), tem-se que:

$$\lambda_j v_j = \frac{1}{L} \sum_{i=1}^L \phi(x_i) \phi(x_i)^T v_j, \text{ para todo } k = 1, \dots, L. \quad (31)$$

Então:

$$\lambda_j v_j = \frac{1}{L} \sum_{i=1}^L \phi(x_i) \langle \phi(x_i), v_j \rangle, \text{ para todo } k = 1, \dots, L. \quad (32)$$

Como λ_j é uma constante e $\langle \phi(x_i), v_j \rangle$ também é uma constante, então, percebe-se que todas as soluções de v_j estão no span dos dados de treinamento mapeados no espaço das características F , ou seja, existem $\alpha_1^k, \dots, \alpha_L^k$ coeficientes para todo $j = 1, \dots, L$ tal que:

$$v_j = \sum_{i=1}^L \alpha_i^k \phi(x_i) \quad (33)$$

Segundo Schölkopf, Smola e Müller (1998), isso implica que pode-se considerar o sistema equivalente abaixo:

$$\lambda_j (\phi(x_j) \cdot v_j) = (\phi(x_j) \cdot \bar{C} v_j), \text{ para todo } j = 1, \dots, L. \quad (34)$$

Substituindo (30) e (33) em (34), tem-se que:

$$L\lambda\alpha = K\alpha \quad (35)$$

Onde: K é a matriz Kernel³ ou matriz Gram é dada por:

$$K_{ij} = (\phi(x_i) \cdot \phi(x_j)) = k(x_i, x_j) \quad (36)$$

Com isso, é possível decompor a matriz covariância dos dados em F , com média zero, de forma implícita através da matriz Kernel. Schölkopf sugere que a

³ Uma função kernel contínua positiva definida que gera uma matriz Kernel (K) ou Gram matriz positiva definida. Note que uma matriz simétrica é positiva definida se e somente se os seus autovalores são positivos.

matriz Kernel seja decomposta em autovetores (α) e autovalores (λ), como em (35).

Como, a média dos dados (x_i) mapeados no espaço das características F é dada por:

$$\bar{\phi} = \frac{1}{l} \sum_{i=1}^L \phi(x_i) \quad (37)$$

Os dados mapeados (x_i) com média igual a zero são dados por:

$$\tilde{\phi}(x_i) = \phi(x_i) - 1_{L \times 1} \bar{\phi} \quad (38)$$

Onde: $1_{L \times 1} = [1, 1, \dots, 1]^T$ cujas dimensões é $L \times 1$

Dado que os dados precisam estar centralizados (ou seja possuírem média igual a zero) no espaço das características, a matriz \tilde{K} *centralizada* é dada por:

$$\tilde{K}_{ij} = (\tilde{\phi}(x_i) \cdot \tilde{\phi}(x_j)) \quad (39)$$

Substituindo-se (38) em (39), tem-se:

$$\tilde{K}_{ij} = ((\phi(x_i) - 1_{L \times 1} \bar{\phi}) \cdot (\phi(x_i) - 1_{L \times 1} \bar{\phi})) \quad (40)$$

Então,

$$\tilde{K}_{ij} = \phi(x_i)^T \phi(x_i) - \phi(x_i)^T 1_{L \times 1} \bar{\phi} - \bar{\phi}^T 1_{1 \times L} \phi(x_i) + \bar{\phi}^T 1_{1 \times L} 1_{L \times 1} \bar{\phi} \quad (41)$$

Na forma matricial, a equação (41) equivale a equação (42) (Schölkopf, Mika, *et al.*, 1998):

$$\tilde{K} = K - \frac{1}{L} 1_{L \times L} K - \frac{1}{L} K 1_{L \times L} + \frac{1}{L^2} 1_{L \times L} K 1_{L \times L} \quad (42)$$

Com isso, a matriz \tilde{K} , cujos dados no espaço das características possuem média zero, passa a ser decomposta pela expressão (Schölkopf, Smola e Müller, 1998):

$$\tilde{K} \alpha_i = \lambda_i \alpha_i \quad , \text{ Onde } i \in [1, L] \quad (43)$$

Dessa forma, os autovetores da matriz de covariância são dados por:

$$v_j = \sum_{i=1}^L \alpha_i^k \tilde{\phi}(x_i) \quad (44)$$

Como está se realizando uma análise de componentes principais no domínio F , para um padrão x , é possível obter a expressão de $\phi(x)$ reconstruída através de uma expansão com os n primeiros autovetores (v_j), como no PCA convencional.

$$\phi(x) \cong \sum_{j=1}^n v_j b_j + 1_{L \times 1} \bar{\phi} \quad (45)$$

Onde: b_k é a projeção de $\phi(x)$ em v_k .

Esta projeção de $\phi(x)$ em v_k pode ser calculada pela expressão:

$$b_j = \sum_{i=1}^L \alpha_i^j k(x_i, x) \quad (46)$$

Durante a etapa de reconstrução, pode-se determinar o quanto de variância pode ser preservada, em relação à variância total, ao se truncar a expansão nos n primeiros termos, como em (45). Essa medida é definida como energia acumulada (E) como em (47):

$$E = \frac{\sum_{j=1}^n \lambda_j}{\sum_{j=1}^N \lambda_j} \quad (47)$$

A expressão de reconstrução obtida pela equação (45) ainda pode ser reduzida em uma expressão mais simples. Substituindo-se (44) em (45), obtém-se:

$$\phi(x) \cong \sum_{j=1}^n b_j \sum_{i=1}^L \alpha_i^j \tilde{\phi}(x_i) + 1_{L \times 1} \bar{\phi} \quad (48)$$

Rearranjado-se (48), tem-se:

$$\phi(x) \cong \sum_{i=1}^L \sum_{j=1}^n b_j \alpha_i^j \tilde{\phi}(x_i) + 1_{L \times 1} \bar{\phi} \quad (49)$$

Definindo-se:

$$\gamma_i = \sum_{j=1}^n b_j \alpha_i^j \quad (50)$$

E substituindo-se (50) em (49), tem-se a expressão de $\phi(x)$ reconstruída:

$$\phi(x) \cong \sum_{i=1}^L \gamma_i \tilde{\phi}(x_i) + 1_{L \times 1} \bar{\phi} \quad (51)$$

Na equação de reconstrução, como $\phi(x)$ é função de $\tilde{\phi}(x_i)$ em (54), não é possível calcular essa expansão diretamente, o que recai no chamado problema de *Pre-Image*.

4.4. *Pre-Image*

O problema de encontrar padrões no espaço de entrada de forma a reconstruir dados no espaço das características denomina-se *Pre-Image* ou

imagem inversa. A solução do *Pre-Image* é a chave para a redução de ruído e reconstrução não linear a partir do KPCA.

Tipicamente, a *Pre-Image* exata que resolva o problema de reconstrução de dados do espaço das características para o espaço inicial não existe, contudo, pode-se buscar por uma solução aproximada (Mika, Schölkopf, *et al.*, 1999). Mesmo assim não é um problema trivial já que a dimensionalidade do espaço das características F pode ser infinito (Kwok e Tsang, 2004). Considerações sobre a existência ou não de uma *Pre-Image* estão dispostas no anexo A2.

A formulação inicial do problema para encontrar uma *Pre-Image* aproximada foi desenvolvida por (Schölkopf, Mika, *et al.*, 1998). Eles sugeriram que fosse realizada uma minimização para encontrar $z \in R^N$ tal que o quadrado da distância euclidiana entre $\phi(z)$ e $\phi(x)$ seja mínimo, conforme mostra a figura 18.

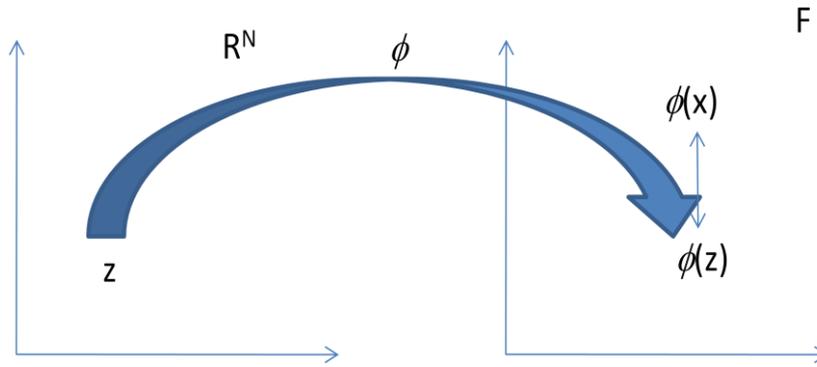


Figura 18. Diagrama esquemático do problema de *Pre-Image*.

Na figura 18, $\phi(x)$ é o elemento no espaço F que se deseja obter a *Pre-Image* e $\phi(z)$ é o mapeamento direto da preimage obtida. Ou seja, deve-se encontrar $z \in R^N$ tal que o quadrado da distância euclidiana (ρ) seja mínimo, conforme mostra a equação (53).

$$z = \arg \min_z \rho(z) \tag{52}$$

Como, o quadrado da distância euclidiana é dado por:

$$\rho(z) = \|\phi(x) - \phi(z)\|^2 \tag{53}$$

Substituindo-se (51) em (53), tem-se que:

$$\rho(z) = \phi(z)' \phi(z) - 2 \left(\sum_{i=1}^L \gamma_i \tilde{\phi}(x_i)' \phi(z) + \frac{1}{L} \sum_{i=1}^L \phi(x_i)' \phi(z) \right) + \phi(x)' \phi(x) \tag{54}$$

Substituindo (38) em (54), tem-se que:

$$\rho(z) = \tag{55}$$

$$\phi(z)' \phi(z) - 2 \left(\sum_{i=1}^L \gamma_i \left(\phi(x_i) - \frac{1}{L} \sum_{j=1}^L \phi(x_j) \right)' \phi(z) + \frac{1}{L} \sum_{i=1}^L \phi(x_i)' \phi(z) \right) + \phi(x)' \phi(x)$$

Então,

$$\rho(z) = k(z, z) - 2 \left(\sum_{i=1}^L \gamma_i \left(k(x_i, z) - \frac{1}{L} \sum_{j=1}^L k(x_j, z) \right) + \frac{1}{L} \sum_{i=1}^L k(x_i, z) \right) + k(x, x) \quad (56)$$

Que equivale a:

$$\rho(z) = k(z, z) - 2 \left(\sum_{i=1}^L \gamma_i k(x_i, z) + \frac{1}{L} \sum_{i=1}^L k(x_i, z) \left(1 - \sum_{j=1}^L \gamma_j \right) \right) + k(x, x) \quad (57)$$

Colocando o termo $k(x_i, z)$ em evidência, tem-se:

$$\rho(z) = k(z, z) - 2 \left(\sum_{i=1}^L k(x_i, z) \tilde{\gamma}_i \right) + k(x, x) \quad (58)$$

$$\text{Onde: } \tilde{\gamma}_i = \gamma_i + \frac{1}{L} \left(1 - \sum_{j=1}^L \gamma_j \right) \quad (59)$$

Alguns métodos foram desenvolvidos para realizar tal minimização dada pela equação (52), que estão descritos a seguir.

4.4.1. Métodos de *Pre-Image*

Os métodos encontrados na literatura para encontrar uma *Pre-Image* aproximada são descritos a seguir:

- Método do Ponto Fixo (Mika, Schölkopf, *et al.*, 1999);
- Método da Regularização do Espaço de Entrada (Abrahamsen e Hansen, 2009);
- Multidimensional Scaling (Kwok e Tsang, 2004); e
- Método da Aprendizagem (Bakir, Weston e Schölkopf, 2004).

Em seguida é realizada uma discussão sobre cada método.

4.4.1.1. Método do Ponto Fixo

O Método do Ponto fixo foi proposto por (Mika, Schölkopf, *et al.*, 1999) para resolver o problema do *Pre-Image*. O método consiste em utilizar otimização baseada em gradiente decrescente para encontrar um ponto de mínimo da distância ao quadrado (ρ) utilizando-se um algoritmo de ponto fixo. Contudo, este

método pode sofrer de instabilidades numéricas (Kim, Franz e Schölkopf, 2005). Além disso, também pode ficar preso em mínimos locais.

Mika, Schölkopf, Smola, Müller, Scholz, & Rätsch (1999) sugerem derivar a função $\rho(z)$ em relação a z , igualar essa derivada a zero e obter o valor de z por iteração do ponto fixo. Este é um método iterativo, que nem sempre converge e a solução final também depende da solução inicial.

Derivando-se a equação (58) e igualando-a a zero, tem-se que:

$$\frac{\partial \rho(z)}{\partial z} = \frac{\partial k(z,z)}{\partial z} - 2 \left(\sum_{i=1}^L \tilde{\gamma}_i \frac{\partial (k(x_i, z))}{\partial z} \right) = 0 \quad (60)$$

Deste modo, para cada tipo de função kernel pode ser obtida uma expressão utilizando-se do método iterativo de ponto fixo.

- Para a função kernel RBF:

$$z_{t+1} = \frac{\sum_{i=1}^L \tilde{\gamma}_i \exp(-\|z_t - x_i\|^2/c) x_i}{\sum_{i=1}^L \tilde{\gamma}_i \exp(-\|z_t - x_i\|^2/c)} \quad (61)$$

- Para a função kernel sigmóide:

$$z_{t+1} = \frac{\sum_{i=1}^L \tilde{\gamma}_i \operatorname{sech}^2(\kappa(x_i \cdot z_t) + \theta) x_i}{\operatorname{sech}^2(\kappa(z_t \cdot z_t) + \theta)} \quad (62)$$

- Para a função kernel polinomial:

$$z_{t+1} = \frac{\sum_{i=1}^L \tilde{\gamma}_i (x_i \cdot z_t)^{d-1} x_i}{(z_t \cdot z_t)^{d-1}} \quad (63)$$

Contudo, apenas para a função kernel RBF, essa formulação apresenta um resultado numericamente estável. Para a função kernel polinomial, essa formulação atinge o número máximo de iterações sem apresentar um resultado satisfatório. A formulação para a função kernel sigmóide nem sempre converge, dependendo da posição inicial e da formulação de ponto fixo.

Alguns outros pesquisadores sugeriram pequenas alterações no método de obtenção de *Pre-Images* através de Ponto Fixo no que se refere ao chute inicial.

Kim, Franz, & Schölkopf (2005) sugerem que o chute inicial (z_0) seja o elemento do conjunto de dados que esteja mais próximo do ponto desejado. Este ponto é obtido através da medida da distância entre o ponto desejado $\phi(x)$ e dos seus vizinhos $\phi(x_i)$ através da expressão (58).

Teixeira, Tome, Stadthanner, & E.W. (2008) sugerem que o chute inicial seja a média aritmética entre os S vizinhos mais próximos, conforme equação a seguir:

$$z_0 = \frac{1}{S} Q j_s \quad (64)$$

Onde: $Q = [x_1, x_2, \dots, x_s]$

$$j_s = [1, 1, \dots, 1]^T \text{ de dimensão } (S \times 1) \begin{matrix} \phi: R^N \rightarrow F, x \rightarrow \phi(x) \\ y \rightarrow \phi(y) \end{matrix}$$

Por fim, em 2009, foi proposta uma regularização da distância de forma a obter uma solução da *Pre-Image* mais estável do que pelo método de ponto fixo. Este método foi chamado de Método de Regularização do Espaço de Entrada (*Input Space Regularization Method*), descrito na seção a seguir.

4.4.1.2. Método de Regularização do Espaço de Entrada

Esse método, originalmente chamado de *Input Space Regularization Method*, foi desenvolvido por Abrahamsen e Hansen (2009). Ele se baseia em procurar uma solução do problema de otimização não linear que seja próximo ao padrão que apresenta ruído. Com isso, o termo da distância conforme equação (58) é alterado com um termo de regularização conforme equação (65).

$$\rho_1(z) = \|\phi(x) - \phi(z)\|^2 + \lambda \|z - x_0\|^2 \quad (65)$$

Onde:

λ é o parâmetro de regularização não negativo e x_0 é o conjunto de dados com ruído.

Expandindo os termos, tem-se que:

$$\rho_1(z) = k(z, z) - 2(\sum_{i=1}^L k(x_i, z) \tilde{y}_i) + \text{cte} + \lambda(z^T z - 2z^T x_0 + x_0^T x_0) \quad (66)$$

Desprezando-se os termos constantes, tem-se a equação (67) que é o termo que se deseja minimizar:

$$\rho_1(z) = k(z, z) - 2(\sum_{i=1}^L k(x_i, z) \tilde{y}_i) + \lambda(z^T z - 2z^T x_0) \quad (67)$$

Ignorando-se o primeiro termo, transforma-se o problema em um problema de maximização, com a seguinte função objetivo:

$$\rho_2(z) = 2(\sum_{i=1}^L k(x_i, z) \tilde{y}_i) - \lambda(z^T z - 2z^T x_0) \quad (68)$$

Em seguida, deriva-se essa última expressão, iguala-se a mesma a zero e obtêm os termos da iteração por ponto fixo para cada tipo de kernel.

Para o Kernel RBF, Abrahamsen e Hansen (2009) sugere que seja utilizada a seguinte expressão:

$$z_{t+1} = \frac{\frac{2}{c} \sum_{i=1}^L \tilde{\gamma}_i \exp\left(-\frac{1}{c} \|z_t - x_i\|^2\right) x_i + \lambda x_0}{\frac{2}{c} \sum_{i=1}^L \tilde{\gamma}_i \exp\left(-\frac{1}{c} \|z_t - x_i\|^2\right) + \lambda} \quad \begin{array}{l} \phi: R^N \rightarrow F, x \rightarrow \phi(x) \\ y \rightarrow \phi(y) \end{array} \quad (69)$$

Outra formulação para *Pre-image* disponível na literatura é a *Multidimensional Scaling* formulada por (Kwok e Tsang, 2004).

4.4.1.3.

Escalonamento Multidimensional (*Multidimensional Scaling*)

Kwok e Tsang (2004) foram os precursores do método *Multidimensional Scaling Based Technique* (MDS). Eles correlacionaram as distâncias no espaço das características com as do espaço das entradas. O método MDS utiliza a distância entre os pontos vizinhos para construir uma *Pre-Image* aproximada (Bakir, Weston e Schölkopf, 2004). O método MDS é não iterativo, ao contrário dos métodos de *Pre-Image* utilizando pontos fixos.

Para exemplificar o método, considera-se aqui o caso de um kernel RBF que possui a seguinte função:

$$k(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \quad (70)$$

O quadrado da distância entre um ponto $\phi(x)$ e um ponto $\phi(y)$ no espaço das características (d_F^2) para o kernel RBF é:

$$d_F^2(x, y) = \|\phi(x) - \phi(y)\|^2 = 2 - 2\exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \quad (71)$$

Então, para o kernel RBF, a relação entre o quadrado das distâncias no espaço das características e o quadrado das distâncias no espaço de entrada é dado por:

$$d_F^2(x, y) = 2 - 2\exp\left(-\frac{d_Z^2(x, y)}{2\sigma^2}\right) \quad (72)$$

Colocando a expressão (71) em termos do quadrado distância no espaço de entrada (d_Z^2), tem-se:

$$d_Z^2(x, y) = \|x - y\|^2 = -2\sigma^2 \ln\left(1 - \frac{1}{2} \|\phi(x) - \phi(y)\|^2\right) \quad (73)$$

Com isso é possível relacionar, para o kernel RBF, a distância de dois pontos no espaço das características à distância no espaço de entrada (d_Z).

Dado que $\phi(x)$ é o ponto ao qual se deseja obter a *Pre-Image* e que pode ser expresso por uma expansão conforme equação (51), o quadrado da distância para

o kernel RBF entre o ponto $\phi(x)$ e um ponto vizinho $\phi(x_i)$ pode ser obtida pela expressão:

$$\rho_i = \|\phi(x_i) - \phi(x)\|^2 = 2 - 2(\sum_{i=1}^L k(x_i, x_i) \tilde{\gamma}_i) \quad (74)$$

$$\text{Onde: } \tilde{\gamma}_i = \gamma_i + \frac{1}{L}(1 - \sum_{i=1}^L \gamma_i)$$

Substituindo a expressão (74) na equação (73), obtêm-se o quadrado da distância ($dist_i^2$) no espaço das entradas entre a *Pre-Image* exata e um ponto vizinho (x_i).

$$dist_i^2 = -2\sigma^2 \ln(\sum_{i=1}^L k(x_i, x_i) \tilde{\gamma}_i) \quad (75)$$

Assim pode ser construído um vetor dos quadrados das distâncias entre o ponto de interesse no espaço de entradas e L pontos vizinhos.

$$dist^2 = [dist_1^2, \dots, dist_L^2] \quad (76)$$

Com isso, o problema de *Pre-Image* na abordagem MDS se baseia em encontrar:

$$z^* = \arg \min_{z^* \in \mathfrak{R}} \sum_{i=1}^L (\|x_i - z^*\|^2 - dist_i^2)^2 \quad (77)$$

Ou seja, encontrar a *Pre-Image* (z^*) tal que o erro quadrático entre $dist_i^2$ e o quadrado da distância entre a *Pre-Image* e um ponto vizinho x_i seja mínimo.

A partir daí é feita uma minimização dessa função, obtendo-se uma expressão que não envolve métodos iterativos.

São escolhidos $n < L$ dos i -ézimos vizinhos mais próximos do ponto de interesse em \mathfrak{R}^N , obtidos de acordo com a distância d_i^2 . Em seguida, calcula-se uma média (\bar{x}) de cada variável dos n vizinhos.

$$\bar{x} = \sum_{i=1}^n x_i \quad (78)$$

De cada vizinho é retirada a média calculada pela equação (78).

$$\tilde{x}_i = x_i - \bar{x} \quad (79)$$

Esses vizinhos mais próximos sem a média (\tilde{x}_i) serão elementos do conjunto X .

$$X = \{\tilde{x}_1, \dots, \tilde{x}_n\} \in \mathfrak{R}^{N \times n} \quad (80)$$

Pela solução sugerida por Kwok e Tsang (2004), a *Pre-Image* aproximada é obtida pela expressão (81):

$$z^* = \frac{1}{2}(XX^T)^{-1}X(d_0^2 - dist^2) + \bar{x} \quad (81)$$

Onde:

$$d_0^2 = [x_1^T x_1, \dots, x_n^T x_n]$$

$dist^2 = [dist_1^2, \dots, dist_n^2]$ é o vetor do quadrado das distâncias.

4.4.1.4.

Método de Aprendizagem

Bakir, Weston e Schölkopf (2004) propuseram estimar uma função Γ através da aprendizagem de um conjunto de dados de treinamento $\{v_k^T \phi(x_i), x_i\}_{i=1}^L$, ou seja, as projeções dos dados no espaço das características e os dados originais. Sugere-se a obtenção de um mapeamento $\hat{\Gamma}: R^V \rightarrow R^N$, onde R^V é o subespaço cujas coordenadas são os v auto-vetores computados pelo Kernel PCA no espaço das características, tal que $\Gamma(v_k^T \phi(x_i)) \approx x_i$. Bakir, Weston e Schölkopf (2004) sugerem que sejam encontrados um conjunto de pesos ($w_j \in R^V$) dado que $j=1, \dots, N$; tal que: $\Gamma_j(\psi) = w_j^T \psi$.

Com isso, Bakir, Weston e Schölkopf (2004) propõe que seja resolvido o seguinte problema de minimização:

$$w_j = \arg \min_{w_j} \sum_{i=1}^N (x_i^j - \Gamma_j(x_i | w_j))^2 + \lambda \|w_j\|^2 \quad (82)$$

Onde:

λ é um termo de regularização

A solução da minimização desse problema pode ser obtida através da abordagem de *Ridge Regression Learning Algorithm in Dual Variables* (Sauders, Gammernan e Vovk, 1998), sendo adaptado para esse novo problema. Essa abordagem é utilizada para regressão utilizando funções *kernel*.

Com isso, a expressão do *Pre-Image* é obtida dada pela expressão:

$$z^* = X' (\lambda I + K)^{-1} k' (v_j^T \phi(x_i), v_j^T \phi(x)) \quad (83)$$

Onde:

$$K_{ij} = [k'(v_j^T \phi(x_i), v_j^T \phi(x_j))]$$

k' é a função kernel produto escalar que visa aproximar o mapeamento reverso.

$$X = \{x_1, \dots, x_N\}$$

Contudo ainda é necessário encontrar uma boa sintonia para a função kernel (k').

4.4.1.5. Discussão de cada método de *Pre-Image*

Apresenta-se aqui uma análise de cada um dos métodos discutidos nas seções anteriores.

O Método do Ponto Fixo proposto por Mika, Schoölkopf, *et al.* (1999) é um método Iterativo e rápido. Para funções Kernel Sigmóides e Polinomiais pode apresentar instabilidade e pode não convergir. Além disso, o método pode ser sensível aos pontos de partida (chute inicial) . Algumas modificações quanto ao ponto de partida são apresentadas por Kim, Franz e Schölkopf (2005) e por Teixeira, Tomé, *et al.* (2008).

O *Multidimensional Scaling* proposto por Kwok e Tsang (2004) é uma solução não iterativa, porém depende da seleção dos vizinhos mais próximos da *Pre-Image*.

O Método da Aprendizagem proposto por Bakir, Weston e Schölkopf (2004) não é um método iterativo; contudo, depende da sintonia do kernel que aproxima o mapeamento reverso.

O *Input Space Regularization Method* desenvolvido por Abrahamsen e Hansen (2009) é uma variação do método do ponto fixo. É iterativo e mais preciso do que o método de Ponto Fixo tradicional.

5

Modelo Kernel PCA Genético para Ajuste de Histórico

Conforme descrito na seção 3.2.2.2.1, em um estudo anterior, Sarma, Durlofsky, *et al.* (2007) parametrizaram o campo de permeabilidade através do Kernel PCA, acoplando o campo a uma expressão da inversão bayesiana. Sarma, Durlofsky, *et al.* (2007) também consideravam a densidade de probabilidade do campo de permeabilidade como uma função multigaussiana e que a relação entre os parâmetros do modelo e os dados de produção era linear. Além disso, utilizavam métodos de gradientes para minimização da função objetivo proposta por ele.

Porém, nem sempre a distribuição de probabilidade dos parâmetros é conhecida. Além disso, no contexto da inversão bayesiana, a verossimilhança, as distribuições a posteriori e a priori são todas multigaussianas se e somente se a relação entre os parâmetros do modelo e os parâmetros observáveis for linear (Caers e Hoffman, 2006). Com isso, de acordo com Sarma (2006) não se pode garantir que a função de máxima verossimilhança seja de fato a expressão que está sendo otimizada no seu trabalho.

Contudo, o Kernel PCA (KPCA) é capaz de capturar parte da estatística de ordem superior de uma série de dados (Kim, Franz e Schölkopf, 2005) e, por consequência, é capaz de preservar a estatística de múltiplos pontos de uma série de realizações. Outra característica importante é o fato do KPCA permitir uma redução de dimensionalidade. Esses atributos mencionados conferem ao KPCA um grande potencial de auxiliar no processo de ajuste de histórico.

Já o Algoritmo Genético é um algoritmo de otimização, inspirado pela teoria da evolução, que busca uma solução global. A base teórica do esse algoritmo foi introduzida por Holland (1975) e sua primeira implementação e adaptação do algoritmo para otimização de parâmetros foi realizada por De Jong (1975). O algoritmo genético também foi estudado por inúmeros outros pesquisadores como Goldberg (1989), Whitley (1989) e Bäck (1992). Desde a sua proposição, esse algoritmo vem comprovando a sua robustez e efetividade em

uma grande variedade de aplicações práticas bem como em problemas de pesquisa como problemas de agendamento de recursos, controle dinâmico, teoria dos jogos, análise de dados, problemas de modelagem ótima de muitos produtos técnicos incluindo problemas de otimização complexos que envolvem restrições (Bäck, 1992). Além disso, o modelo computacional do Algoritmo Genético pode ser facilmente paralelizável o que pode ser uma vantagem para o uso de sistemas computacionais distribuídos. Esses fatores tornam essa técnica particularmente interessante para a resolução de problemas de ajuste de histórico.

Neste trabalho, outra metodologia, inspirada no trabalho de Sarma, é proposta nesta dissertação. Nessa nova metodologia não se faz qualquer consideração a priori sobre a densidade de probabilidade da permeabilidade, preserva-se a estatística de múltiplos pontos do modelo geológico e utiliza-se do Algoritmo Genético na busca por uma solução global para o problema de ajuste de histórico.

Nesse novo algoritmo, o campo de permeabilidade é reconstruído por uma expansão em um espaço de maior dimensionalidade e os seus termos são ajustados através do Algoritmo Genético, obedecendo às restrições geoestatísticas de múltiplos pontos.

5.1. Descrição do Modelo Kernel PCA-Genético para Ajuste de Histórico

Uma nova metodologia, chamada aqui de *Modelo Kernel PCA-Genético*, é proposta para resolver o problema de ajuste de histórico em reservatórios de óleo e gás. Propõe-se que sejam otimizadas, através de Algoritmos Genéticos, as projeções do campo de permeabilidade no espaço das características $\{b_1, \dots, b_n\}$.

Esse novo modelo une a robustez do Algoritmo Genético e a característica de capturar parte dos momentos de ordem superior do Kernel PCA. O algoritmo desse novo modelo está esquematizado na figura 19.

Como no Algoritmo Genético, cada cromossomo ou indivíduo representa uma solução para o problema. No modelo Kernel PCA-Genético, cada solução do problema é um conjunto de projeções do campo de permeabilidade no espaço das características $\{b_1, \dots, b_n\}$ (Miyoshi, Szwarcman e Vellasco, 2010). Esse

conjunto de projeções é reconstruído para o espaço original através do *Pre-Image* e do Kernel PCA.

Nesse modelo, cada conjunto de novas soluções representa uma nova população. Essas projeções serão evoluídas por um Algoritmo Genético de forma que se encontre uma solução cujo campo de permeabilidade associado a elas permita que simulação do modelo de reservatório atualizado seja capaz de reproduzir os dados reais de produção e efetuar uma previsão do comportamento futuro de um reservatório.

Esse processo evolucionário das projeções no espaço das características permite que a estatística de múltiplos pontos seja preservada. Durante o processo evolucionário, alguns indivíduos são selecionados de acordo com a sua aptidão e recombinados gerando novos indivíduos. Estes novos indivíduos substituem os indivíduos menos aptos formando uma nova geração de indivíduos. Em seguida, os novos indivíduos são avaliados.

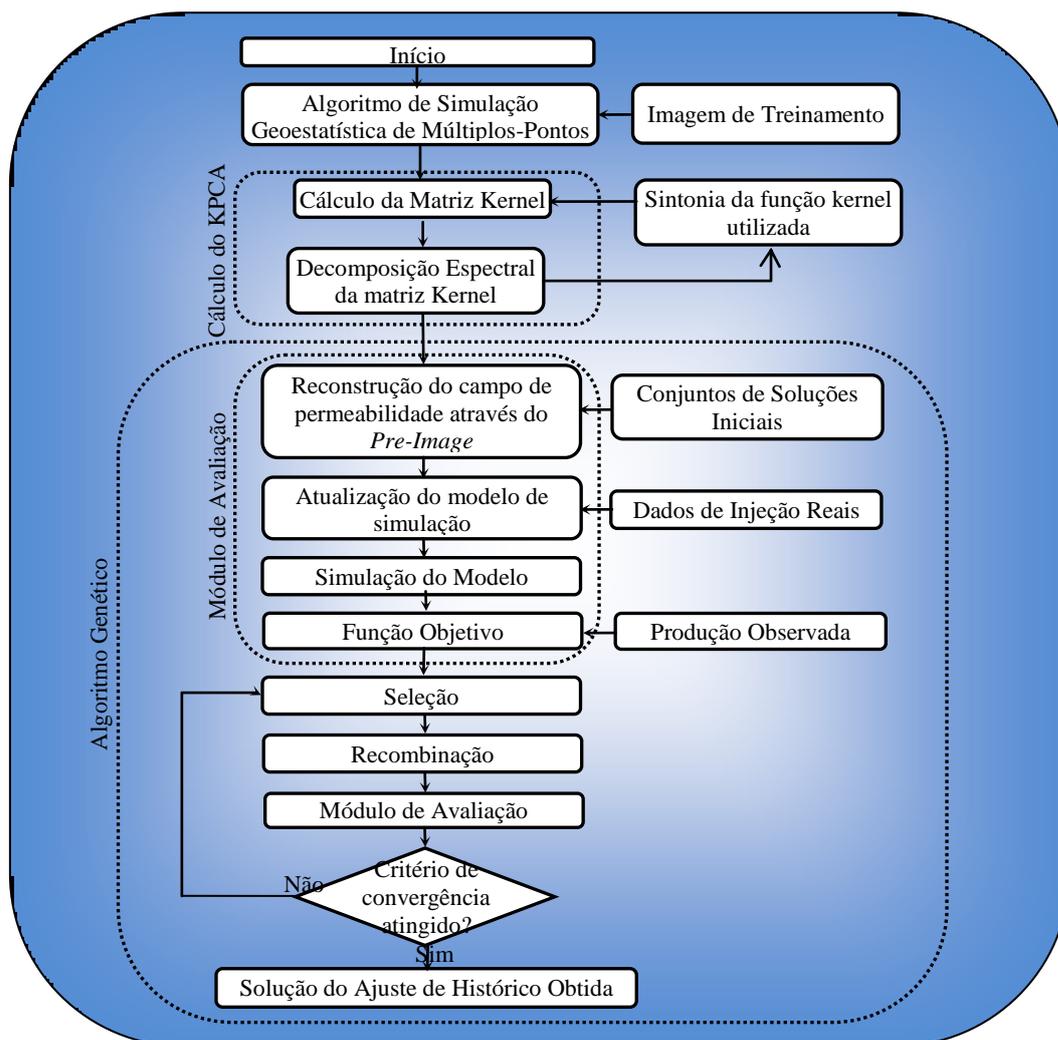


Figura 19. Algoritmo do Modelo Kernel PCA-Genético.

No módulo de avaliação, o campo de permeabilidade é reconstruído no espaço de entrada pelo *Pre-Image* a partir dos autovetores e autovalores calculados pela decomposição da matriz Kernel no algoritmo do KPCA. Em seguida, o modelo de reservatório é atualizado e após, é simulado. A função objetivo sugerida é erro percentual entre a produção simulada e a produção observada.

Os indivíduos melhor avaliados possuem uma probabilidade maior de serem selecionados para reprodução genética que permite uma busca sobre o espaço de soluções potenciais. Cada busca tem dois objetivos: explorar as melhores soluções (cruzamento), explorar todo espaço de busca (mutação).

O processo evolucionário de seleção, recombinação e substituição dos indivíduos menos aptos por novos indivíduos formando um novo conjunto de soluções, que serão avaliados é realizado até que o critério de convergência seja atingido, obtendo-se a solução do ajuste. As seções subsequentes apresentam detalhadamente as etapas referentes a esse novo algoritmo.

5.1.1. Simulação Geoestatística de Múltiplos Pontos

Primeiramente deve ser efetuada a Simulação Geoestatística de Múltiplos Pontos. Nessa etapa, utiliza-se um algoritmo de Simulação Geoestatística de Múltiplos Pontos chamado *Snesim*, disponível no software SGeMs desenvolvido pelo laboratório SCRF na Universidade de Stanford (Remy, Boucher e Wu, 2009).

O *Snesim* é um algoritmo de simulação seqüencial que infere a estatística de múltiplos pontos a partir de uma imagem de treinamento, a qual deve refletir o conceito geológico e estrutural concebido nas etapas anteriores do ajuste de histórico. A partir dessa estatística de múltiplos pontos são obtidas realizações de cada parâmetro do campo de permeabilidade, e ao final do processo obtêm-se uma realização do campo de permeabilidade. O detalhamento do funcionamento do *Snesim* foi apresentado na seção 2.3.2.2.

Dessa forma, realizações de campos de permeabilidade são obtidas de acordo com essa estatística de múltiplos pontos inferida. Com isso, obtêm-se um conjunto de dados $\{x_1, \dots, x_L\} \subset R^N$, onde L é o número de realizações e N é a quantidade de parâmetros a serem ajustados.

5.1.2. Cálculo do KPCA

Na etapa do cálculo do KPCA, obtêm-se a matriz Kernel conforme apresentado na seção 4.3.1 e, em seguida, efetua-se a decomposição espectral da mesma, obtendo-se os seus respectivos autovetores e autovalores, conforme equação (43).

Um importante aspecto é a sintonia da função kernel utilizada. A sintonia dos parâmetros da função kernel deve ser efetuada de forma que a energia obtida na decomposição da matriz Kernel, dada pela equação (47), seja a máxima possível.

Com isso, deve-se utilizar algum método de otimização para auxiliar no processo de sintonia, tal como Algoritmo Genético. Nessa sintonia, o uso de Algoritmos Genéticos é sugerido devido a sua robustez e capacidade de explorar o espaço de busca de forma inteligente. Outra vantagem do uso do Algoritmo Genético deve-se ao desconhecimento da relação entre a energia acumulada e os parâmetros da função kernel.

A partir da decomposição espectral da matriz Kernel, é possível obter uma expressão da permeabilidade mapeada no espaço das características dada pela equação (49). Em seguida, os termos dessa expansão da permeabilidade serão otimizados através de um Algoritmo Genético (GA).

5.1.3. Algoritmo Genético

No Algoritmo Genético, um conjunto de soluções iniciais representa uma população inicial. Essa população é composta por indivíduos ou cromossomas, que são soluções em potencial para o problema. Por sua vez, cada indivíduo é composto por um conjunto de genes, que representam as variáveis do problema.

No modelo aqui proposto, cada indivíduo é um conjunto de projeções a ser evoluído pelo Algoritmo Genético. Se a solução for composta por n projeções, então, cada indivíduo possui n genes com representação real, como na figura 20.

$$\underline{b_1} \quad \underline{b_2} \quad \underline{b_3} \quad \underline{b_4} \quad \dots \quad \underline{b_n}$$

Figura 20. Diagrama esquemático de um indivíduo com n genes

Isto é, neste modelo, cada indivíduo ou cromossoma é constituído por um conjunto de genes: $\{b_1, \dots, b_n\}$, onde n é o número total de projeções e cada gene possui uma representação real.

Segundo Michalewicz (1996), o desempenho do uso da representação real é muito superior em relação à representação binária no que se refere ao resultado da otimização e em termos de tempo computacional.

5.1.3.1. Conjunto de Soluções Iniciais

Propõe-se que o conjunto de soluções iniciais seja composto por indivíduos gerados de forma aleatória para se evitar que o algoritmo fique preso em soluções sub-ótimas. A esse conjunto de soluções iniciais dá-se o nome de população inicial.

Em seguida, cada conjunto de soluções é avaliado de acordo com a sua função objetivo através do Módulo de Avaliação.

5.1.3.2. Módulo de Avaliação

O módulo de avaliação avalia a qualidade dos indivíduos do algoritmo genético. Como o indivíduo é na verdade parte da informação do campo de permeabilidade compactada através do Kernel PCA, é necessário que seja efetuada uma descompactação dessa informação, para que, em seguida, seja avaliado o quão bem o campo de permeabilidade descompactado permite que a produção dada pelo modelo de reservatório se aproxima da curva de produção real.

Dessa forma, esse módulo de avaliação é composto pelas seguintes etapas: reconstrução do campo de permeabilidade através do *Pre-Image*, atualização do modelo de simulação, simulação do modelo e cálculo de uma função objetivo.

5.1.3.2.1.

Reconstrução do campo de permeabilidade através do *Pre-Image*

Como a expansão do campo de permeabilidade está sendo reconstruída no espaço das características, então é necessário que seja realizado um mapeamento dessa expansão dada pela equação (44) para o espaço de entrada. Esse mapeamento é denominado *Pre-Image*, conforme foi descrito na seção 4.4.1 onde também vários métodos estão descritos.

Nesse modelo, o método de *Pre-Image* utilizado é o Método do Ponto Fixo proposto por Mika, Schölkopf, *et al.* (1999) descrito na seção 4.4.1.1.

Utiliza-se como ponto inicial do processo iterativo o elemento do conjunto de dados que se encontra mais próximo do ponto reconstruído pela expansão das projeções no espaço das características dada pela equação (44), conforme proposto por Kim, Franz, & Schölkopf (2005).

Primeiramente, calcula-se a medida da distância euclidiana entre o ponto desejado $\phi(x)$ e dos seus vizinhos $\phi(x_i)$, que é dada através da expressão (58).

Depois, ordenam-se as distâncias euclidianas calculadas e se encontra o elemento do conjunto de dados que se encontra mais próximo da expansão reconstruída de $\phi(x)$.

Em seguida, prossegue-se o processo iterativo dado pelas equações (61) a (63) até que o processo iterativo atinja a sua convergência, utilizando como ponto de partida o $\phi(x_i)$ que se encontra mais próximo do ponto desejado.

Através desse procedimento, para cada conjunto de projeções $\{b_1, \dots, b_n\}$, os campos de permeabilidade correspondentes a reconstrução dessas projeções no espaço das características são aproximados através desse processo de *Pre-Image*, descrito na seção 4.4.1.

Em outras palavras, é aproximado um mapa de permeabilidade no espaço de entrada cujo mapeamento direto se aproxima do mapa reconstruído através das projeções no espaço das características.

Na próxima etapa é realizada a atualização do modelo de simulação.

5.1.3.2.2.**Atualização do Modelo de Simulação e Simulação do modelo**

É efetuada a atualização do modelo de simulação com cada novo campo de permeabilidade e com os dados de injeção. Cada modelo de reservatório ajustado é então simulado e é obtida uma produção simulada correspondente. Em seguida, calcula-se a função objetivo.

5.1.3.2.3.**Função Objetivo**

A função objetivo deve informar de forma quantitativa ao algoritmo o quão boa foi a aproximação das curvas de produção simuladas das curvas de produção reais, dessa forma, deve avaliar o erro entre a produção simulada e a produção real dentro da janela de tempo de ajuste de histórico. A função objetivo também deve considerar a produção por poço e por produto seja água, óleo ou gás e a pressão de fundo de poços, pois a mesma é conhecida e pode também ser ajustada.

Uma das possíveis funções objetivo seria o erro percentual médio entre as produções simuladas e a produção real por poço e por produto.

Em seguida deve-se calcular a aptidão de cada indivíduo. A aptidão é calculada a partir da normalização linear do ranqueamento de cada indivíduo na população, como a seguir:

Primeiramente, ordena-se todos indivíduos da população de acordo com os respectivos valores da função de avaliação de forma decrescente. O indivíduo menos apto, ou seja, o indivíduo com maior função de avaliação recebe ranqueamento (*ranque*) igual a um, o segundo indivíduo menos apto recebe ranqueamento (*ranque*) igual a dois, e assim sucessivamente.

Em seguida, a aptidão é calculada a partir da equação:

$$A_i = \min + \left(\frac{\max - \min}{n_{pop} - 1} \right) (\text{ranque} - 1)$$

Onde:

max e min são parâmetros da normalização;

n_{pop} é o número total de indivíduos da população;

ranque é o ranqueamento do indivíduo na população;

A_i é a aptidão do indivíduo *i*.

(84)

O cálculo da aptidão a partir do ranqueamento é importante para controlar a pressão seletiva evitando-se uma convergência prematura (Whitley, 1989).

Aumentando-se a pressão seletiva, a busca se foca nos indivíduos mais aptos, o que pode levar ao problema de superindivíduos. Isto é, indivíduos que possuem uma aptidão tão melhor que o restante da população, fazendo que os seus descendentes dominem a população rapidamente, ocasionando perda de diversidade ou uma convergência prematura.

Reduzindo-se a pressão seletiva aumenta-se a exploração efetuando-se uma busca em uma população mais diversa (Whitley, 1989).

Após a etapa de avaliação, os indivíduos são selecionados de acordo com a sua aptidão e recombinados.

5.1.3.3. Seleção

Inicialmente, os indivíduos da população atual (pais) que possuem as melhores aptidões são clonados para a população da próxima geração (filhos). Essa quantidade de indivíduos clonados é calculada segundo a política chamada de estado estacionário, dada pela equação (85).

$$Q_{indSS} = npop \cdot (1 - GAP) \quad (85)$$

Onde:

Q_{indSS} é a quantidade de indivíduos clonados para a próxima geração;

$npop$ é total de indivíduos da população;

GAP é a fração dos indivíduos da população a serem trocados a cada geração.

Em seguida, o restante da nova população (filhos) é preenchida com os indivíduos obtidos pelos processos de seleção pela roleta e recombinação, conforme sugerido em Michalewicz (1996).

A roleta seleciona os indivíduos pais aleatoriamente, proporcionando maiores chances de reprodução aos mais aptos. A roleta é realizada da seguinte maneira (Michalewicz, 1996):

- Calcula-se a probabilidade de cada indivíduo ser selecionado através da equação (86):

$$p_i = \frac{A_i}{\sum_{i=1}^{npop} A_i} \quad (86)$$

Onde:

p_i é a probabilidade do indivíduo i ser selecionado;

A_i é o valor da aptidão do indivíduo i ;

$npop$ é o número total de indivíduos.

- Calcula-se a probabilidade cumulativa (q_i) para cada indivíduo pela equação (87):

$$q_i = \sum_{j=1}^i p_j \quad (87)$$

Onde:

p_i é a probabilidade do indivíduo i ser selecionado;

q_i é a probabilidade cumulativa do indivíduo i .

- Em seguida é gerado um número aleatório r no intervalo de zero a um. Se $r < q_1$, seleciona-se o primeiro cromossomo; ou então, seleciona-se o i -ésimo cromossomo tal que $q_{i-1} < r < q_i$.

Desta forma, os indivíduos melhor avaliados têm uma probabilidade maior de serem selecionados.

Depois da etapa de seleção, os indivíduos pais selecionados são recombinados através dos operadores genéticos.

5.1.3.4. Recombinação

São realizadas recombinações através dos operadores genéticos: o cruzamento e a mutação. O cruzamento é um acelerador do processo de busca e tira proveito das soluções mais promissoras e a mutação é um operador exploratório e dispersa a população pelo espaço de busca.

5.1.3.4.1. Cruzamento

Para se escolher quais indivíduos serão selecionados para o cruzamento, gera-se um número aleatório r entre zero e um. Se r for menor do que a

probabilidade de cruzamento (pc) aplica-se o operador nos cruzamento nos indivíduos selecionados dois a dois (Michalewicz, 1996).

Neste modelo, são utilizados dois tipos de cruzamento: o cruzamento de um ponto de corte e o cruzamento aritmético.

- Cruzamento de um ponto de corte: efetua o cruzamento dos genes de dois indivíduos genitores em uma posição aleatoriamente escolhida, criando dois possíveis descendentes. O ponto de cruzamento é gerado aleatoriamente no intervalo entre 1 e $n - 1$, onde n é o número de projeções a serem otimizadas.

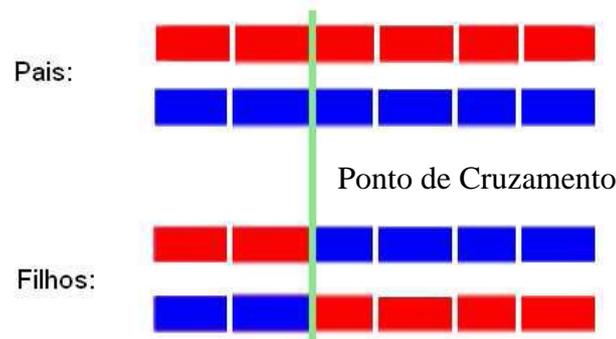


Figura 21. Diagrama esquemático do cruzamento de um ponto

- Cruzamento Aritmético. Neste tipo de cruzamento é realizada uma combinação linear de dois indivíduos dadas pelas equações (88) e (89).

$$c1' = r \cdot c1 + (1 - r) \cdot c2 \quad (88)$$

$$c2' = r \cdot c2 + (1 - r) \cdot c1 \quad (89)$$

Onde:

$c1'$ e $c2'$ são indivíduos filhos;

$c1$ e $c2$ são indivíduos pais;

r é um número aleatório.

5.1.3.4.2. Mutaç o

O operador mutaç o tem por objetivo introduzir uma diversidade nos indivíduos da populaç o, de forma a evitar que o mesmo fique preso em m nimos ou m ximos locais. Entretanto, se a taxa de mutaç o for muito elevada o

algoritmo genético corre o risco de perder antigos parâmetros ótimos e executar uma busca puramente aleatória, o que não é desejável, pois a velocidade de convergência para a solução tendera a diminuir (Michalewicz, 1996).

Para determinar em qual gene ocorrerá mutação, gera-se um número aleatório r variando de zero a um para todos os genes dos indivíduos da nova população, se r referente ao gene for menor do que a probabilidade de mutação (pm), seleciona-se esse gene para a mutação (Michalewicz, 1996).

Descreve-se a seguir dois tipos de mutação: mutação uniforme e mutação não-uniforme.

- Na mutação uniforme, o resultado da mutação é um valor aleatório entre o valor do limite inferior e superior do gene.
- A mutação não-uniforme, sorteia-se um bit aleatório 0 ou 1, em seguida, o valor final do gene é dado pela equação (90) e (91):

$$b'_i = b_i + (LS - b_i) \cdot (1 - rm^{(1-\frac{t}{T})^b}) \quad \text{se o dígito do número aleatório for zero} \quad (90)$$

$$b'_i = b_i - (b_i - LI) \cdot (1 - rm^{(1-\frac{t}{T})^c}) \quad \text{se o dígito do número aleatório for um} \quad (91)$$

Onde:

b_i é o valor da projeção i ;

b'_i é o valor da projeção i após a mutação;

LS é o limite superior da projeção b_i ;

LI é o limite inferior da projeção b_i ;

rm é um número aleatório de zero a um;

t é o número de gerações;

T é o número total de gerações;

c é um parâmetro do sistema.

Após a etapa de recombinação através dos operadores genéticos, os indivíduos filhos gerados nessas etapas de seleção e recombinação irão fazer parte da parte da próxima geração.

Essa nova geração de indivíduos será avaliada e reinicia-se o processo evolutivo até que o critério de convergência seja atingido.

5.1.3.5. Critério de Convergência

O critério de convergência proposto que finaliza o processo de otimização do Algoritmo Genético é na verdade um critério de parada dado pelo número de gerações máximo do Algoritmo Genético ou pelo número máximo de avaliações realizadas no Algoritmo Genético.

5.2. Discussão da metodologia proposta

A principal vantagem desse algoritmo é sem dúvida o fato de preservar a estatística de múltiplos pontos durante o processo de ajuste. Isso possibilita que sejam obtidos campos de permeabilidade mais realistas do que os obtidos pelos métodos tradicionais citados na abordagem por zoneamento e por inversão bayesiana.

Outra vantagem em relação aos outros métodos empregados na literatura é a não consideração de uma função de densidade de probabilidade a priori. Conforme detalhado no capítulo 3, tipicamente essas funções de densidade de probabilidade são inferidas a partir de dados esparsos de poços e com isso não necessariamente reflete o conceito estrutural e geológico do reservatório.

Esse método também é mais robusto que os métodos tradicionalmente empregados baseados em gradientes. Esses métodos são sensíveis às condições iniciais. Já o método proposto é baseado em um método de otimização mais robusto, o Algoritmo Genético, que busca uma solução global.

Contudo, o método proposto possui algumas desvantagens. Como um número maior de avaliações até a convergência é requerido pelos algoritmos genéticos, um número maior de simulações e um maior custo computacional são necessários durante o processo de ajuste.

Além disso, a *Pre-Image* exata nem sempre existe; na verdade se procura uma *Pre-Image* aproximada, conforme discutido na seção 4.4.1. Contudo, o método de ponto fixo sugerido por Mika, Schölkopf (1999) utilizando as funções kernel gaussianas sempre converge (2002).

6 Estudos de Caso

Alguns estudos de caso foram realizados para verificar a eficácia do KPCA em aplicações de redução de ruído, reprodução de padrões e para o teste do algoritmo proposto em ajuste de histórico.

Na primeira etapa, foi realizado um teste da aplicação de KPCA para redução de ruído em campos de permeabilidade obtidas através do algoritmo *Snesim*, que infere a estatística de múltiplos pontos de uma imagem de treinamento.

Na segunda etapa, efetuou-se um teste do Modelo Kernel PCA-Genético para Ajuste de Histórico proposto no ajuste de histórico de um campo sintético utilizando-se todos os dados históricos de produção em uma fase e utilizando apenas os 5 primeiros anos dos dados históricos de produção em outra fase.

6.1. **Primeira Etapa: Testes em Remoção de Ruído com realizações de permeabilidade obtidas a partir da estatística de múltiplos pontos**

Para obtenção das realizações de permeabilidade, foi utilizado o programa de modelagem geoestatística 3D chamado S-GeMS, desenvolvido pelo laboratório SCRF da Universidade de Stanford.

Dentre as implementações geoestatísticas disponíveis no programa, utilizou-se o algoritmo *Snesim_std* para a obtenção das realizações do campo de permeabilidade a partir da imagem de treinamento (*training image*) mostrada na figura 22.

Ao total, foram obtidas 2000 realizações de campo de permeabilidade (45x45). A figura 23 apresenta alguns exemplos de campos de permeabilidade obtidos.

Uma das realizações obtidas foi escolhida e a ela adicionou-se ruído gaussiano com média igual a zero e desvio padrão de 0.7. Em seguida, foi implementado o algoritmo para redução de ruído KPCA em Matlab®.

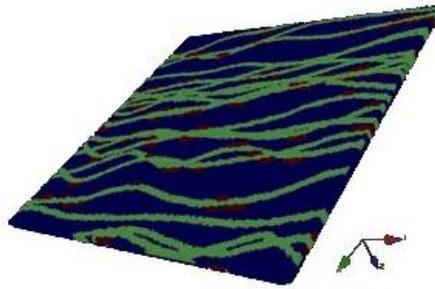


Figura 22. Imagem de treinamento utilizada.

Fonte: Remy (2009).

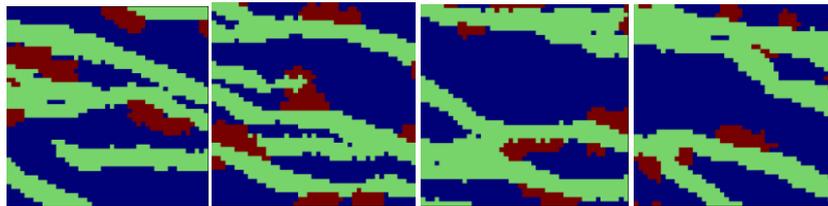


Figura 23. Exemplo de campos de permeabilidades obtidos.

O algoritmo faz a decomposição da matriz Kernel. Em seguida, calcula-se um conjunto de projeções $\{b_1, \dots, b_j\}$ de um padrão ruidoso nos j primeiros autovetores da matriz Kernel, conforme equação (46). A redução do número de auto-vetores na aplicação de remoção de ruído é particularmente útil, pois se elimina a parte dos dados que apresenta alta frequência, já que algumas componentes principais de direção de menor variância dos dados são descartadas.

Em seguida, dado o conjunto de projeções calculado $\{b_1, \dots, b_j\}$ e dos autovetores da matriz Kernel, é possível reconstruir o padrão ruidoso no espaço das características, a partir de uma expansão dada pela equação (51).

Para a aplicação de redução de ruído, utilizando KPCA, foram testados três tipos de kernels utilizando o mesmo número de autovetores em cada teste (194 autovetores). Foram realizados testes para os kernels polinomiais, sigmóides e RBF.

6.1.1. Resultados da Aplicação de Ruído

O campo de permeabilidade escolhido dentre os gerados pelo SGeMs. A e esse mesmo campo após a inserção de ruído são mostrados na figura 24.

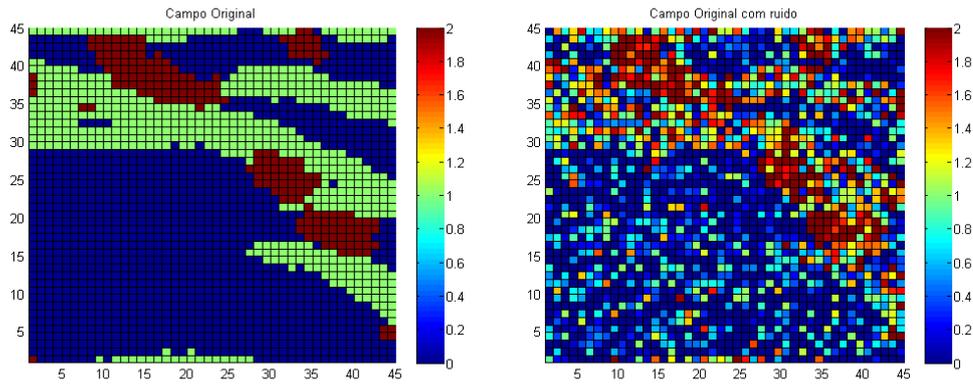


Figura 24. Campo de permeabilidade escolhido e c/ ruído (a direita).

Os seguintes resultados foram obtidos utilizando o KPCA para redução de ruído utilizando-se o mesmo número de autovetores (194).

Para o kernel RBF com $\sigma=20, 50, 90$ e 300 obteve-se 32,33%, 71,67%, 80,09% e 83,82% de energia acumulada respectivamente, com o campo de permeabilidade mostrados na figura 25.

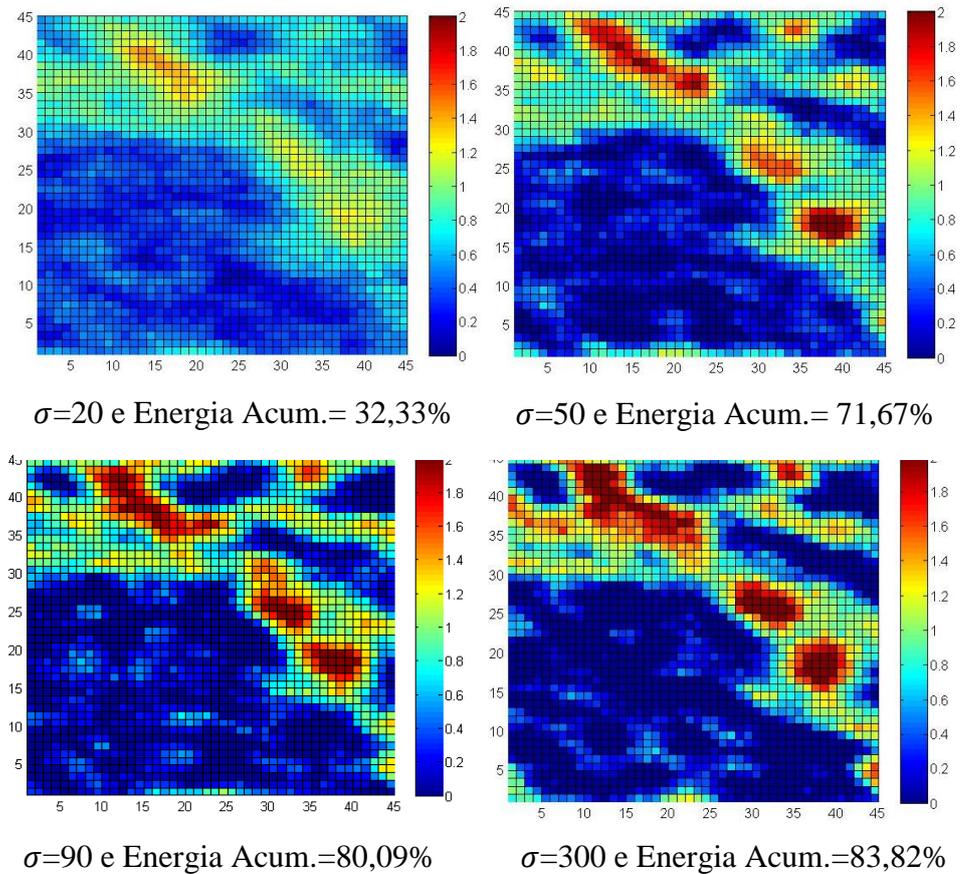


Figura 25. Resultado da remoção de ruído utilizando-se kernel RBF.

Pela figura 25 é possível observar que a qualidade da redução de ruído obtida está relacionada a energia acumulada: para maiores energias acumuladas, obtiveram melhores resultados. Ainda, a energia acumulada obtida depende da sintonia da função kernel como podemos observar na figura 26.

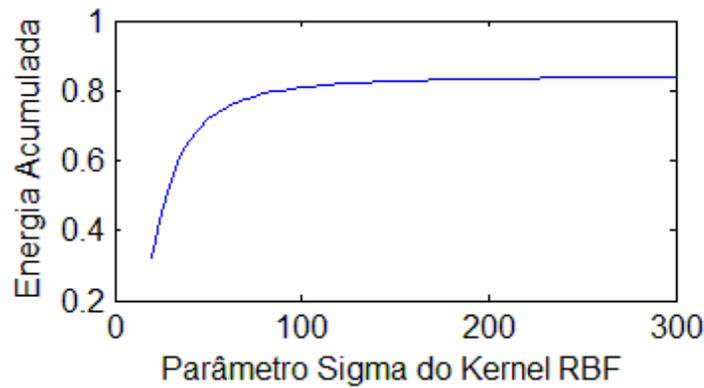


Figura 26. O comportamento da energia acumulada obtida para a função kernel RBF em relação ao parâmetro sigma da função kernel RBF.

A figura 27 mostra a energia acumulada em relação aos números de autovetores para o kernel rbf com σ igual a 90. Como pode-se observar na figura 27, a configuração de 194 autovetores apresenta 80% de energia acumulada.

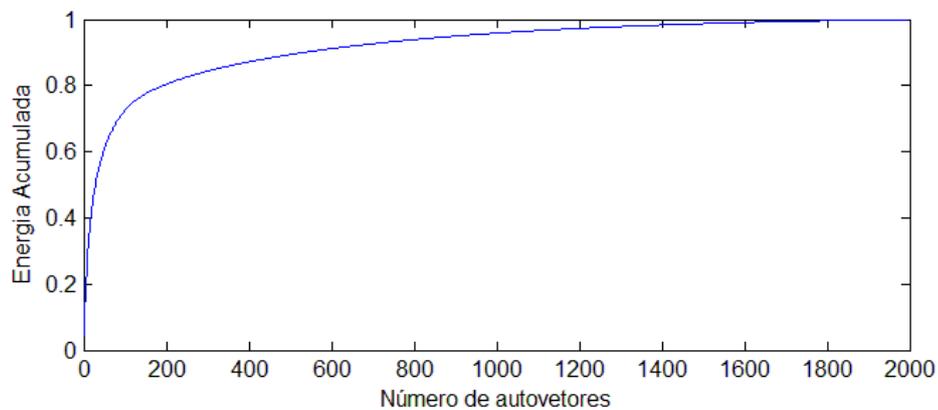


Figura 27. Relação entre o número de autovetores e a energia acumulada para a função kernel RBF.

Para o kernel sigmóide com κ igual a $1.0e-4$, $2.4691e-4$ e $1e-6$ e $\theta=0$, obteve-se 84,40%, 85,88% e 84,20% respectivamente e os campos de permeabilidade apresentados na figura 28.

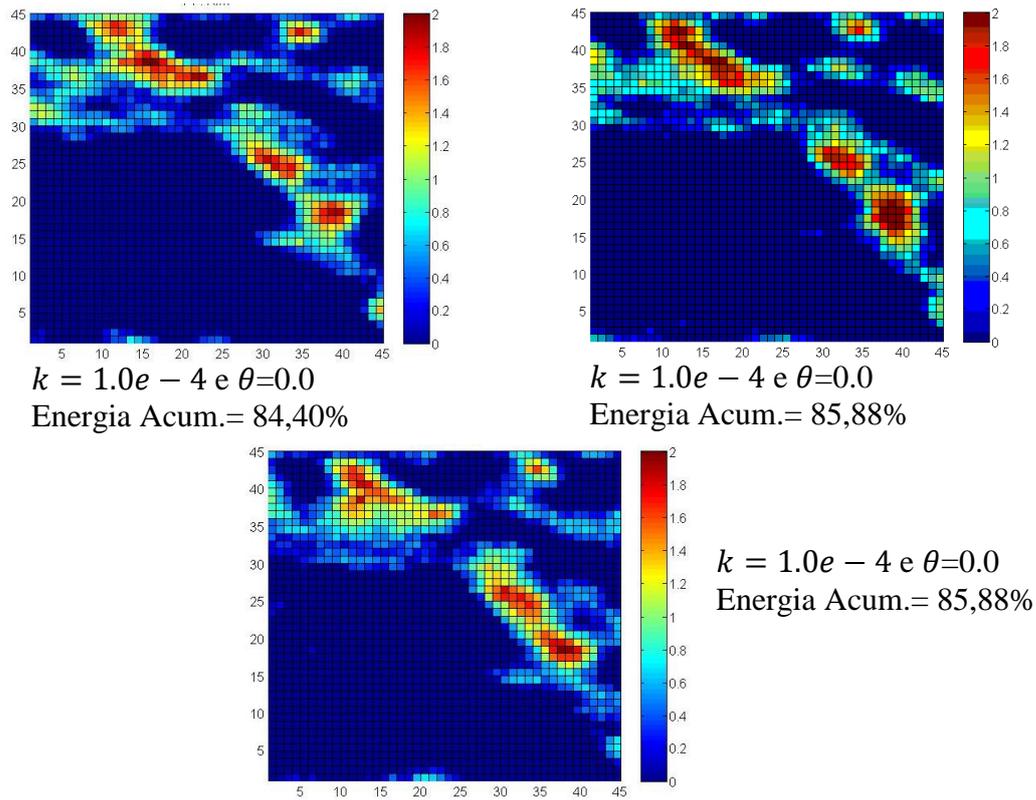


Figura 28. Resultado da remoção de ruído com um kernel sigmóide.

Por fim, para o kernel polinomial de ordem 2, obteve-se 30% de energia acumulada e o campo de permeabilidade da figura 29.

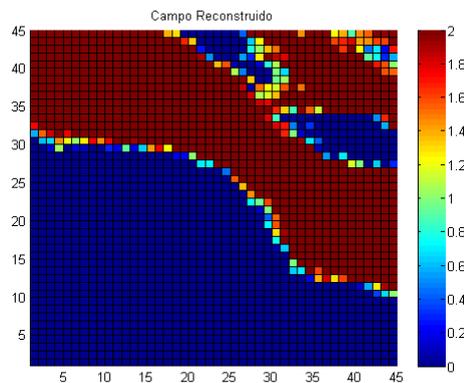


Figura 29. Resultado da remoção de ruído com um kernel polinomial.

6.1.2. Discussão dos Resultados da Aplicação de Ruído

As funções kernels sigmóides e RBF obtiveram um bom resultado em termos de redução de ruído. Entretanto, o kernel polinomial de ordem dois não foi capaz de remover o ruído de forma eficiente 1.

Foi mostrado que a eficiência da remoção de ruído através do Kernel PCA depende do tipo de função kernel utilizada, bem como da sintonia da mesma. Foi mostrado também que boas reduções de ruído estavam relacionadas a maiores níveis de energia.

Para esse caso, o KPCA mostrou-se ser uma técnica eficiente para remoção de ruído.

6.2. Aplicação de Ajuste de Histórico utilizando o modelo Kernel PCA-Genético

O teste do modelo Kernel PCA-Genético na aplicação de ajuste de histórico foi realizado em duas etapas, conforme mostra a figura 30.

- Etapa A: Universo de Produção: 20 anos /7201 dias (ajuste).
- Etapa B: Universo de Produção: 5 anos / 1842 dias (ajuste) e 15 anos/ 5349 dias (teste).

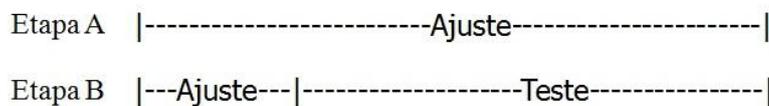
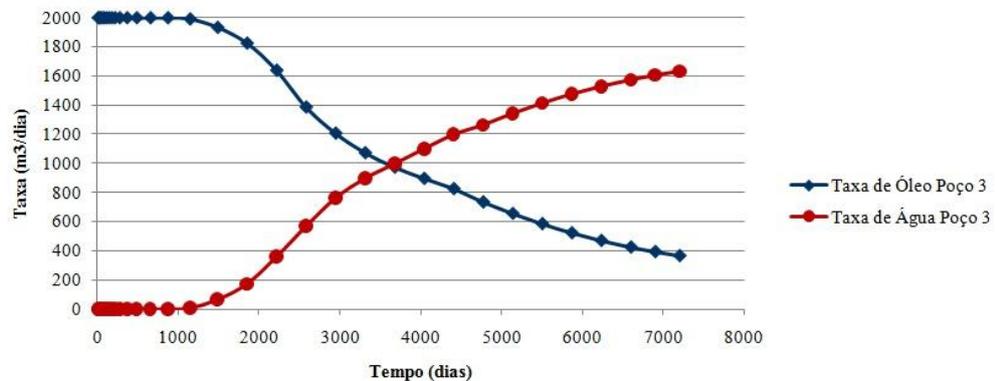


Figura 30. Etapas dos testes do modelo Kernel PCA-Genético.

Para os dois testes, foi utilizado um campo sintético quadrado que possui um *grid* de 30x30x1 e contém 6 poços injetores e 6 poços produtores. Esse campo está apresentado na figura 31.

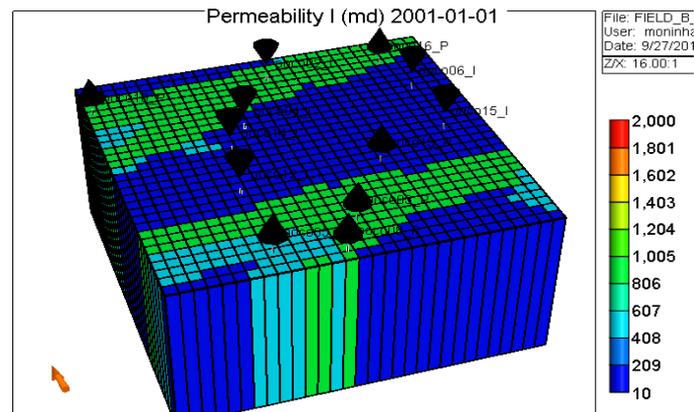


Figura 31. Campo sintético utilizado.

Durante o processo de ajuste, foi considerado que apenas as informações sobre o parâmetro da permeabilidade eram desconhecidas.

Para a implementação do algoritmo do modelo Kernel PCA-Genético, esquematizado pela figura 19, foram obtidas 800 realizações de permeabilidade, a partir de uma imagem de treinamento, através do algoritmo *Snesim* disponível no programa SGeMS, (Remy, Boucher e Wu, 2009).

A imagem de treinamento utilizada foi a mesma do teste de redução de ruído que foi apresentada na figura 22. Essa imagem de treinamento possui 3 litofácies, indicando regiões de alta permeabilidade (em verde), regiões de baixa permeabilidade (azul) e regiões de permeabilidade intermediária (vermelha).

Dentre as realizações obtidas, uma delas foi escolhida para ser o campo de permeabilidade real e o restante foi utilizado para gerar o KPCA. O campo de permeabilidade real foi inserido no modelo de reservatório e este foi simulado, obtendo-se dados da produção considerada real ou produção observada.

Conforme o algoritmo do modelo Kernel PCA-Genético mostrado na figura 19, o Algoritmo Genético evoluiu um conjunto de projeções do campo de permeabilidade em um espaço de maior dimensionalidade. Cada indivíduo do Algoritmo Genético constitui uma projeção no espaço das características.

Para implementação do Algoritmo Genético, foi utilizada a biblioteca GACOM elaborada pelo Laboratório de Inteligência Computacional da PUC-Rio.

Campos de permeabilidade foram reconstruídos, a partir das projeções no espaço de características, pela metodologia *Pre-Image* de ponto fixo sugerida por Mika, Schölkopf, *et al.* (1999). Os algoritmos do KPCA e do *Pre-Image* foram implementados no software Matlab[®].

O modelo de reservatório foi atualizado com cada um dos campos de permeabilidade reconstruídos. Em seguida, os modelos ajustados foram simulados, através do *software* IMEX[®] da CMG[®], obtendo-se a produção simulada de água, óleo e gás por poço para cada modelo. Foi então calculada a função objetivo de cada modelo ajustado, considerando o erro percentual médio absoluto (MAPE) da produção de água, óleo e gás por poço e por período de tempo segundo a equação (92).

$$MAPE = \frac{1}{3 * nt * np} \sum_{t=1}^{nt} \sum_{ip\text{ço}=1}^{np} \sum_{prod=1}^3 \left| \frac{prod.simul.t,ip\text{ço},prod - prod.observ.t,ip\text{ço},prod}{prod.observ.t,ip\text{ço},prod} \right| \quad (92)$$

Onde: t são os períodos de tempo; nt é o período de tempo final simulado; $ip\text{ço}$ refere-se ao número de cada poço, np é o número total de poços, $prod = 1$ refere-se ao óleo, $prod = 2$ refere-se ao água, $prod = 3$ refere-se ao gás produzido.

Por fim, utilizou-se como critério de convergência um critério de parada, baseado no número de avaliações.

Para se garantir a convergência do algoritmo, foi realizada um segundo processo de otimização, onde o resultado obtido na primeira otimização foi inserido como ponto de partida dessa segunda otimização. Por sua vez essa, primeira otimização foi efetuada a partir de indivíduos gerados aleatoriamente.

Foram feitos alguns testes do modelo proposto com dois tipos de kernel: sigmóide e RBF Gaussiana. A configuração de cada função kernel utilizada foi obtida através de vários testes, de forma que represente a maior energia acumulada. A configuração de cada kernel obtida está indicada na Tabela 1.

Tabela 1. Configuração Kernels

Tipo de Kernel	Kernel $k(\mathbf{x}, \mathbf{y})$	Parâmetro	Valor
RBF Gaussiana	$\exp\left(-\frac{\ \mathbf{x} - \mathbf{y}\ ^2}{2\sigma^2}\right)$	σ	7.0148e+004
Sigmóide 1	$\tanh(\kappa(x \cdot y) + \theta)$	κ	2.7222e-4
		θ	0.00
Sigmóide 2	$\tanh(\kappa(x \cdot y) + \theta)$	κ	2.7222e-4
		θ	-0.08

A Tabela 2 mostra as energias acumuladas obtidas com cada função kernel sugerida, de acordo com o número de autovetores definido.

Tabela 2. Energia Acumulada para cada kernel

Tipo de Kernel	Número de Autovetores	Energia Acumulada
RBF Gaussiana	9	54,95%
Sigmóide 1	9	54,99%
Sigmóide 2	9	54,79%
RBF Gaussiana	10	57,28%
Sigmóide 1	10	57,32%
Sigmóide 2	10	57,11%
RBF Gaussiana	50	83,32%
Sigmóide 1	50	83,37%
Sigmóide 2	50	83,09%

6.2.1. Etapa A

Nesta etapa foi efetuado o teste do algoritmo considerando todos os dados históricos, ou seja, o universo de produção de 20 anos. Foram utilizados testes com as funções kernel do tipo sigmóide na configuração 1 e do tipo gaussiano RBF na configuração dada pela Tabela 1, foram feitos testes com 10 e 50 autovetores.

Os resultados obtidos na etapa A são apresentados na Tabela 3 onde apresenta-se a média dos erros percentuais médios absolutos (MAPE), entre a produção simulada e a produção observada de água, óleo e gás, para cada poço, obtida através do modelo Kernel PCA-Genético no teste A.

Tabela 3. MAPE Obtido pelo Ajuste de Histórico na Etapa A

MAPE	Tipo de Kernel	
	Sigmóide 1	RBF
Número de	10	17,629%
Autovetores	50	23,728 %

Para o melhor caso deste teste (17,628% de erro), são apresentadas na figura 32 e na figura 33 as comparações entre as curvas de produção de água, óleo e gás reais e as aproximadas pelo ajuste de histórico proposto para o poço 3 do campo.

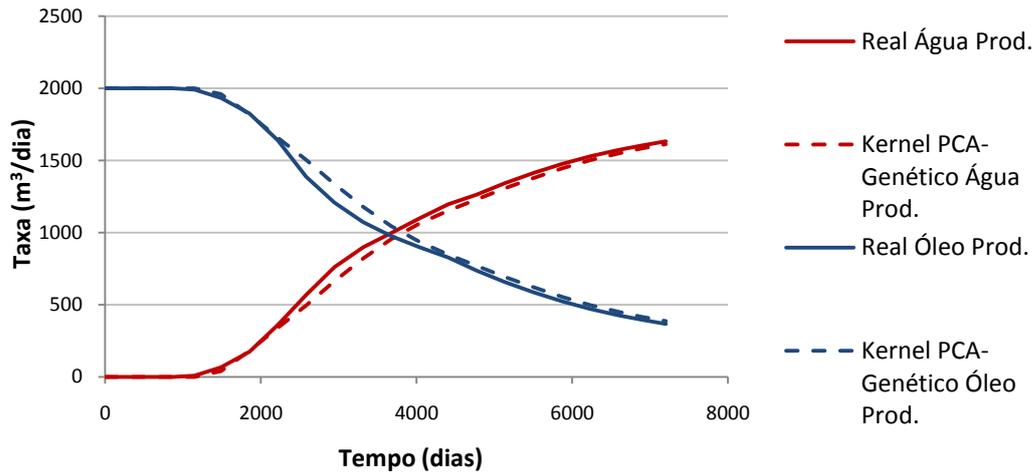


Figura 32. Comparação entre as curvas de produção de água e óleo reais e as aproximadas pelo ajuste de histórico obtido pelo modelo KPCA-Genético proposto.

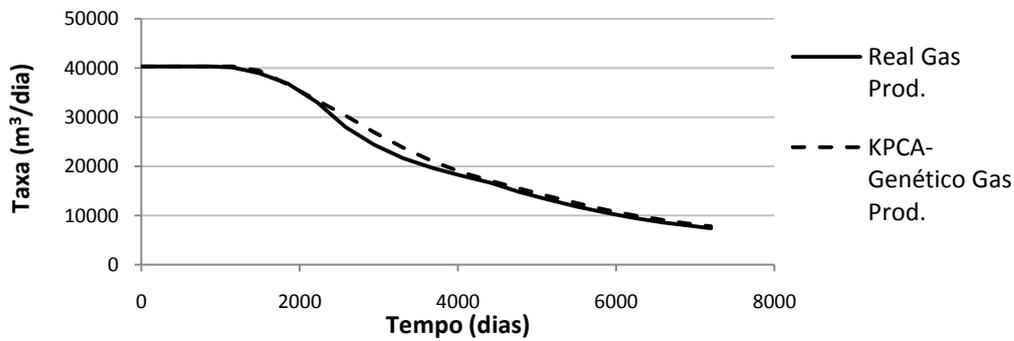


Figura 33. Comparação entre as curvas de produção de gás real e as aproximada pelo ajuste de histórico obtido pelo modelo KPCA-Genético proposto.

A comparação entre os campos de permeabilidade reconstruídos através do modelo Kernel PCA-Genético proposto e do campo considerado real é apresentada na figura 34. A área em vermelho representa formações de permeabilidade alta (1000 md - millidarcies), em verde as de permeabilidade média (500 md) e em azul as de permeabilidade baixa (10 md).

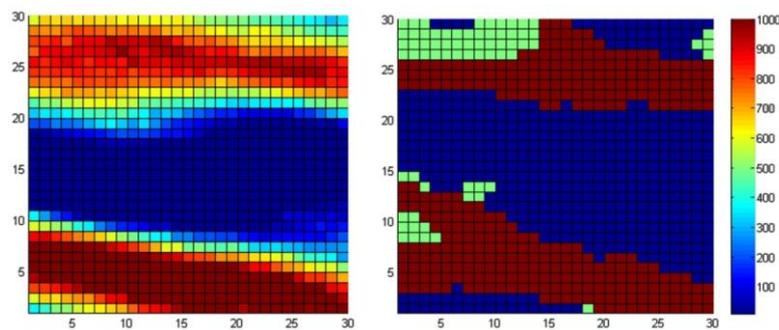


Figura 34. Campo de permeabilidade reconstruído (esquerda) e campo de permeabilidade real (direita).

Através da figura 34, pode-se observar que os canais de baixa e alta permeabilidade foram preservados durante o processo de ajuste. O campo de permeabilidade ajustado reconstruiu as áreas de permeabilidade média de forma suave. Na realidade, em reservatórios reais, as mudanças nas propriedades não ocorrem abruptamente como no campo de permeabilidade considerado real. Com isso, o campo de permeabilidade reconstruído é um caso em que a permeabilidade é realística e o campo de permeabilidade considerado real é um caso onde o campo de permeabilidade não era realístico. Dessa forma, a suavização demonstrada, no processo de ajuste, é uma característica muito positiva dessa metodologia.

6.2.2. Etapa B

Nesse teste, o ajuste de histórico foi efetuado utilizando-se o algoritmo Kernel PCA-Genético considerando apenas os dados históricos de 1842 dias (~5 anos). Os demais dados históricos referentes aos 5349 dias restantes foram utilizados para teste onde se verificou se o algoritmo conseguiu ajustar a permeabilidade de forma que o modelo fosse capaz de efetuar uma boa previsão futura.

Também foi realizada uma comparação com do Modelo Kernel PCA-Genético proposto com o método de zoneamento, no qual as regiões do reservatório foram multiplicadas por região. Foram efetuados testes dividindo-se igualmente a malha em 9 e 10 regiões. No zoneamento, o algoritmo de otimização utilizado também foi o algoritmo genético com as mesmas configurações do algoritmo utilizado no modelo Kernel PCA Genético.

As configurações do Algoritmo Genético utilizada nesse teste está descrita a seguir:

- Número de Avaliações: 600
- Tamanho da População: 50
- Taxa Inicial de Crossover: 0,7
- Taxa Final de Crossover: 0,45
- Taxa Inicial de Mutação: 0,2
- Taxa Final de Mutação: 0,2

- GAP: 0,9
- Número de Experimentos: 10
- Next Run Seed Rate: 0,10

Foram realizados testes com as funções kernel do tipo sigmóide na configuração 1 e na configuração 2 e do tipo gaussiano RBF na configuração dada pela Tabela 1, foram feitos testes com 9 e 10 autovetores.

Os erros (MAPE) obtidos através do processo de ajuste de histórico são apresentados na tabela 4. Pela tabela 4, pode-se observar que o erro percentual médio absoluto (MAPE) na etapa de teste (após 1842 dias) foi maior do que o MAPE obtido durante o processo de ajuste (até 1842 dias), o que já era esperado.

Também é verificado que o erro percentual médio absoluto (MAPE) obtido após 1842 dias pelo modelo Kernel PCA-Genético foi, na maioria dos casos, menor do que o MAPE obtido pela metodologia de zoneamento.

Para o melhor caso, otimizando-se 10 parâmetros através do modelo Kernel PCA-Genético, utilizando-se a função kernel sigmóide na configuração 1, obteve-se um MAPE de 10,61% na etapa de ajuste e de 13,81% na etapa de teste. O resultado obtido foi bem superior aos 35,09% e 25,54% de MAPE na etapa de teste obtidos pela metodologia de zoneamento.

Tabela 4. MAPE Obtido pelo Ajuste de Histórico na Etapa B

MAPE	Qde de Parâmetros Tipo de Kernel	10 Parâmetros				9 Parâmetros			
		Sigmóide 1	RBF	Sigmóide 2	Zoneamento (5X2)	Sigmóide 1	RBF	Sigmóide 2	Zoneamento (3X3)
Parcial ate 1842	Oil Rate SC	5.44%	9.61%	5.41%	10.49%	11.30%	11.36%	4.87%	8.02%
	Gas Rate SC	5.44%	9.61%	5.41%	10.49%	11.30%	11.36%	4.87%	8.02%
	Water Rate SC	20.96%	31.69%	22.05%	31.92%	22.48%	45.57%	23.52%	32.23%
	Media	10.61%	16.97%	10.96%	17.64%	15.02%	22.76%	11.09%	16.09%
Apos 1842	Oil Rate SC	17.34%	32.69%	17.75%	44.59%	27.20%	31.72%	16.99%	31.30%
	Gas Rate SC	17.34%	32.69%	17.75%	44.59%	27.20%	31.72%	16.99%	31.30%
	Water Rate SC	6.75%	13.28%	6.99%	16.09%	12.75%	14.54%	6.94%	14.02%
	Media	13.81%	26.22%	14.16%	35.09%	22.38%	25.99%	13.64%	25.54%
Total	Oil Rate SC	10.85%	20.10%	11.02%	25.99%	18.5%	20.6%	10.4%	18.60%
	Gas Rate SC	10.85%	20.10%	11.02%	25.99%	18.5%	20.6%	10.4%	18.60%
	Water Rate SC	14.50%	23.32%	15.21%	24.73%	18.1%	31.5%	16.0%	23.95%
	Media	12.07%	21.18%	12.41%	25.57%	18.4%	24.2%	12.2%	20.38%

Esse resultado pode ter ocorrido devido ao fato de que o campo de permeabilidade ajustado pela metodologia de zoneamento não é realístico. A permeabilidade pode até ser ajustada aos dados históricos através do método de

zoneamento, mas a curva de produção correspondente não será suficientemente precisa durante a etapa de previsão de produção.

O melhor caso alcançado utilizando-se a metodologia Kernel PCA-Genético para Ajuste de Histórico, considerando o universo de produção de 5 anos para a etapa de ajuste e sem fixar os valores de permeabilidade dos locais onde há poços durante a etapa de simulação geostatística foi obtido pela configuração Sigmóde 1.

O campo de permeabilidade obtido para essa configuração é apresentado pela figura 35 (à direita) e o campo de permeabilidade obtido pela metodologia de Zoneamento é apresentado através da figura 36. As figuras 37-42 apresentam as curvas da produção reais, as curvas de produção referente esse melhor caso e as curvas de produção obtidas pelo método de zoneamento.

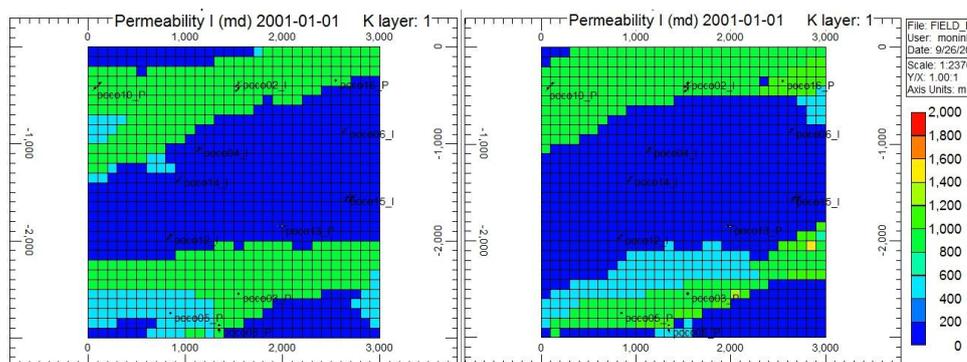


Figura 35. Campo de permeabilidade ajustado pelo modelo Kernel-PCA Genético a direita e Campo de permeabilidade real a esquerda.

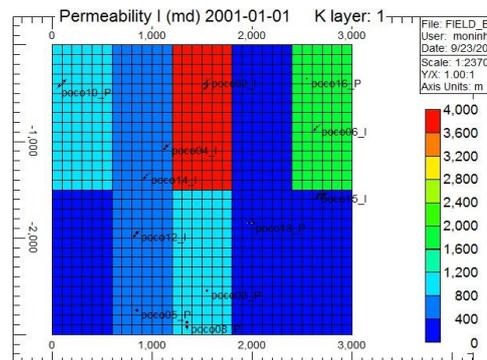


Figura 36. Campo de permeabilidade ajustado pela metodologia de Zoneamento.

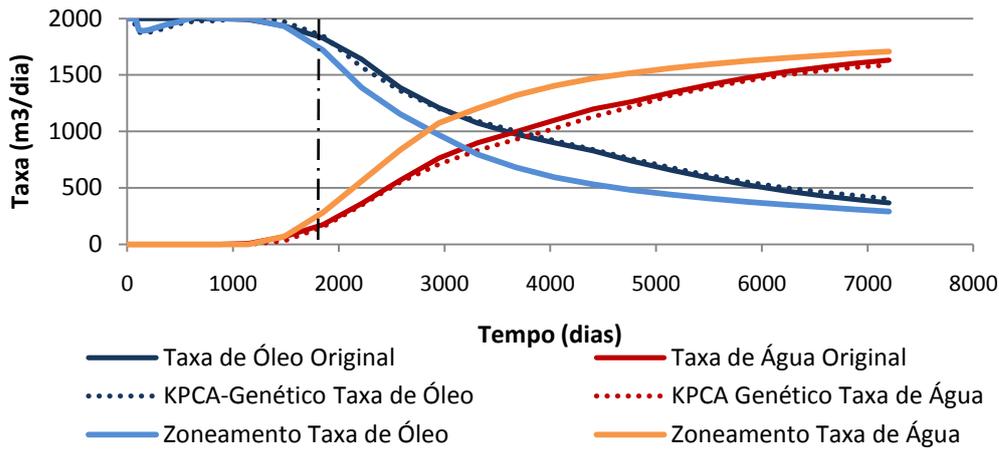


Figura 37. Curvas de produção do poço 3 simulada pelos modelos ajustados pelo método proposto (KPCA-Genético) e pelo método de Zoneamento.

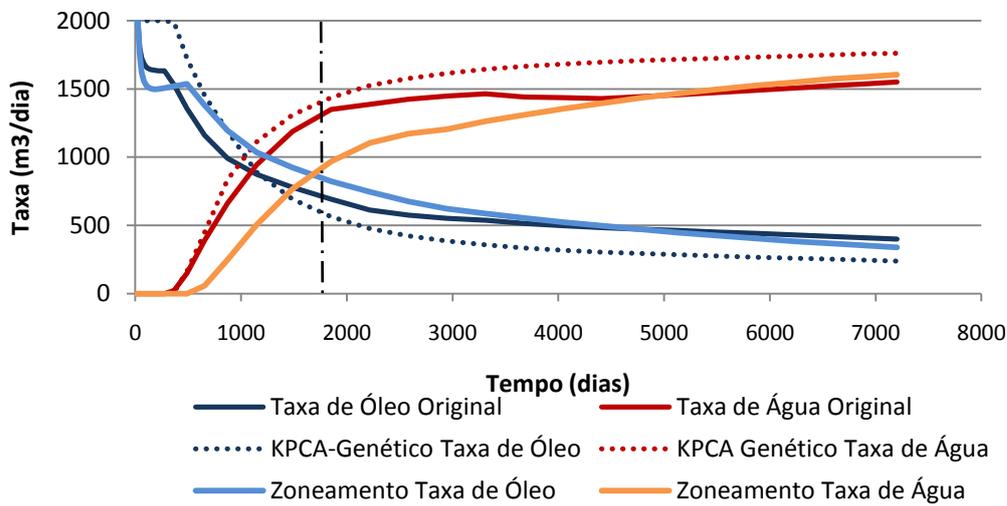


Figura 38. Curvas de Produção do poço 5 simulada pelos modelos ajustados pelo método proposto (KPCA-Genético) e pelo método de Zoneamento.

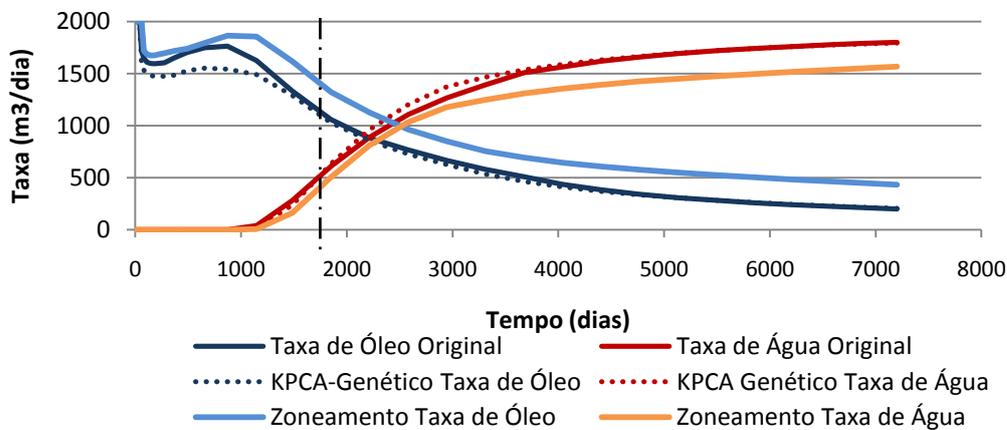


Figura 39. Curvas de Produção do poço 8 simulada pelos modelos ajustados pelo método proposto (KPCA-Genético) e pelo método de Zoneamento.

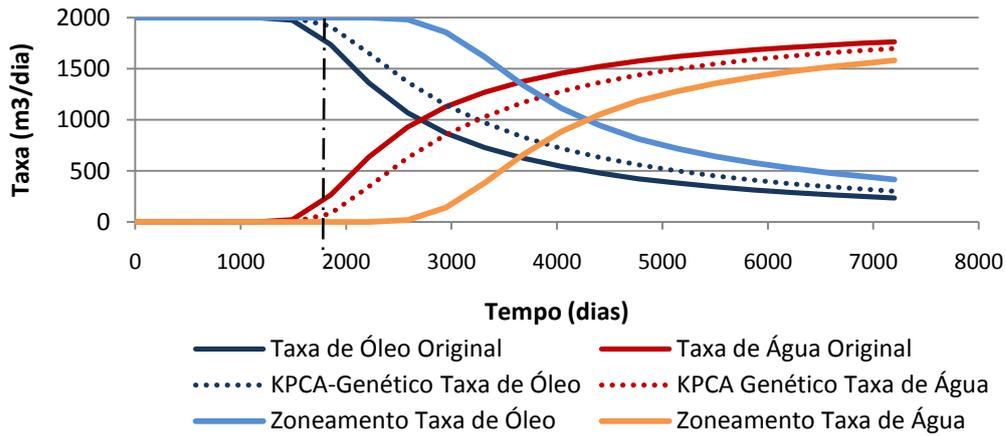


Figura 40. Curvas de Produção do poço 10 simulada pelos modelos ajustados pelo método proposto (KPCA-Genético) e pelo método de Zoneamento.

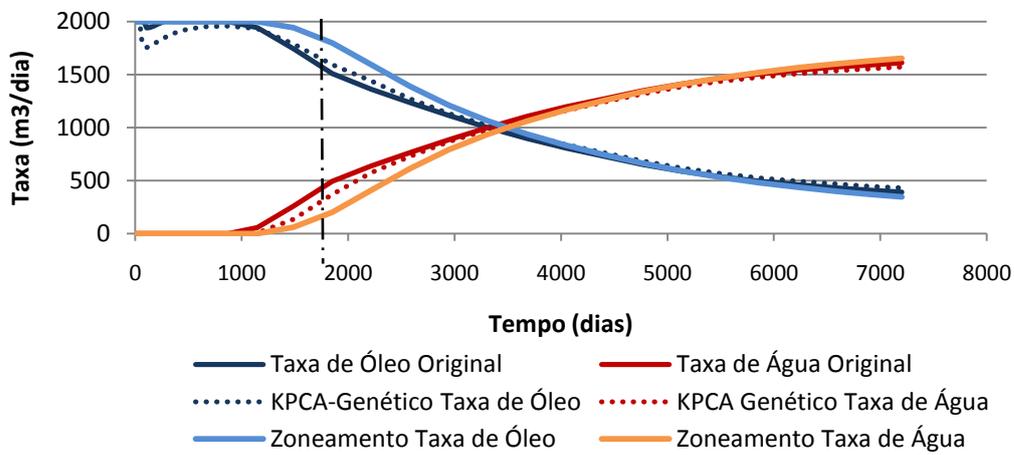


Figura 41. Curvas de Produção do poço 13 simulada pelos modelos ajustados pelo método proposto (KPCA-Genético) e pelo método de Zoneamento.

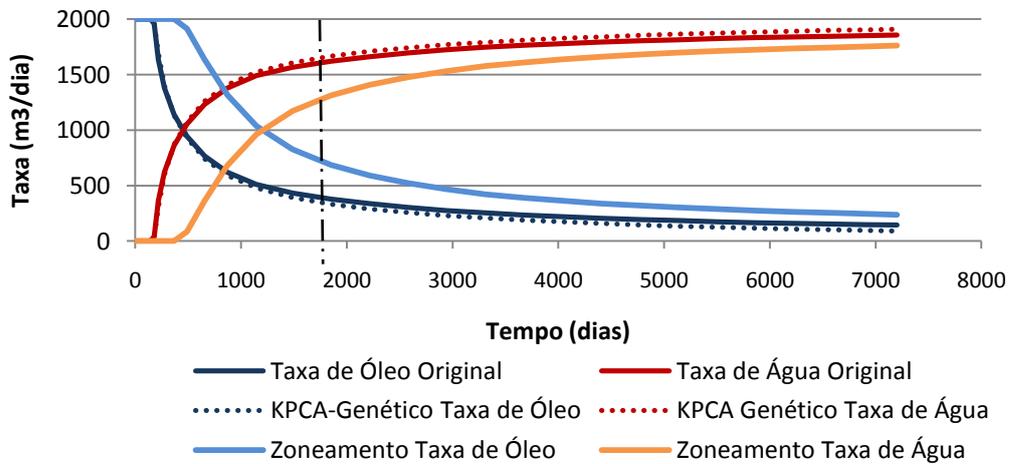


Figura 42. Curvas de Produção do poço 16 simulada pelos modelos ajustados pelo método proposto (KPCA-Genético) e pelo método de Zoneamento.

Através das figuras 37-42, pode-se observar que o resultado obtido para o ajuste de histórico através do modelo Kernel PCA-Genético proposto, utilizando apenas 1842 dias durante o ajuste, obteve uma excelente aproximação das curvas de produção reais de água e óleo para os poços 3, 8,13 ,16.

Já no poço 5, o modelo Kernel PCA-Genético, obteve um erro percentual médio absoluto (MAPE) de 17,73% e o modelo de zoneamento obteve um MAPE de 18,56%. Já para o poço. E no poço 10, o modelo Kernel PCA-Genético obteve um MAPE de 28,87% e o método de zoneamento obteve MAPE de 31,05%. Ou seja, para a produção de todos os poços o modelo proposto apresentou um resultado superior do que o obtido através do método de zoneamento. Contudo, para os poços 5 e 10, o modelo proposto ainda não apresentou resultados tão significativos quanto os obtidos por esse modelo nos demais poços.

Com isso, neste capítulo foram apresentados testes da aplicação do Kernel PCA em redução de ruído e dois testes da proposta de um novo algoritmo para ajuste de histórico utilizando-se da técnica KPCA e algoritmos genéticos. O modelo proposto obteve um bom desempenho para o ajuste de histórico do campo escolhido. Os melhores resultados nos testes do modelo proposto foram obtidos utilizando-se funções kernel sigmóides.

7

Conclusões & Trabalhos Futuros

Este trabalho propôs uma nova metodologia para efetuar ajuste de histórico de reservatórios de óleo e gás que não assume que a densidade de probabilidade da permeabilidade seja multigaussiana e que preserva a estatística de múltiplos pontos do modelo geológico. Além disso, a otimização tende a não ficar presa a mínimos locais.

O Modelo Kernel PCA-Genético para Ajuste de Histórico apresentado é baseado em Kernel PCA, otimização com Algoritmo Genético e simulação geoestatística de múltiplos pontos.

Nos testes realizados, primeiramente, o Kernel PCA foi testado para o caso da remoção de ruídos, onde foi possível estudar e verificar quais os melhores tipos de kernel para representar os campos de permeabilidade. Nos testes realizados, as funções kernel RBF e sigmóide apresentaram um melhor resultado do que o kernel polinomial de ordem dois. Nesses testes de remoção de ruído, também foram observadas a importância da boa sintonia da função kernel.

Em seguida, a metodologia foi testada em um campo sintético de óleo e gás em dois casos. No primeiro caso foi efetuado o ajuste de histórico da permeabilidade considerando o universo de 20 anos. Já no segundo caso, foram ajustados apenas os dados históricos dos 5 primeiros anos, os 15 anos restantes foram utilizados para efetuar um teste da metodologia. Através desse teste pode-se verificar se o modelo ajustado é capaz de efetuar uma boa previsão futura do comportamento do reservatório.

Como resultados, foram obtidos bons resultados em ambos os testes. Obteve-se uma boa aproximação dos valores da permeabilidade em relação ao campo de permeabilidade considerado real. Preservou-se a forma dos canais de permeabilidade e gerando uma solução suave sem mudanças abruptas.

No segundo teste, observou-se que para o melhor resultado obtido pelo método Kernel PCA-Genético, as curvas de produção de água e óleo conseguiram efetuar uma boa previsão do comportamento futuro da produção de óleo e gás

obtendo-se um erro absoluto percentual de 13,81%, cuja previsão foi superior a efetuada pelo modelo ajustado pelo método de zoneamento que apresentou um erro absoluto percentual de 35,09% para o ajuste de 10 parâmetros.

Como trabalhos futuros, pretende-se:

- Usar métodos multigaussianos para comparação;
- Testar o algoritmo para o ajuste de pressão;
- Investigar metodologias que efetuem a simulação probabilística de múltiplos pontos durante o processo de ajuste;
- Investigar a utilização dos dados dos poços durante a etapa de simulação geoestatística de múltiplos pontos;
- Investigar uma abordagem probabilística no espaço das características de forma a garantir a estatística de múltiplos pontos e de forma a obter cenários baseados em modelos probabilísticos.

8

Referências

ABRAHAMSEN, T. J.; HANSEN, L. K. **Input Space Regularization Stabilizes Pre-images for Kernel PCA De-noising**. IEEE Workshop on Machine Learning for Signal Processing. Grenoble: IEEE. 2009. p. 1-6.

BÄCK, T. Evolutionary Algorithms. **ACM SIGBIO Newsletter.**, New York, NY., XXII, n. 2, Junho 1992. 26-31.

BAKIR, G. H.; WESTON, J.; SCHÖLKOPF, B. **Learning to find pre-images**. Advances in Neural Information Processing Systems. Vancouver: the MIT Press. 2004. p. 449-456.

BONET-CUNHA, L. et al. **A Hybrid Markov Chain Monte Carlo Method for Generating Permeability Fields Conditioned to Multiwell Pressure Data and Prior Information**. SPE Annual Technical Conference and Exhibition. Denver: SPE. 1996. p. 1-16.

CAERS, J.; HOFFMAN, T. The Probability Perturbation Method: A New Look at Bayesian Inverse Modeling. **Mathematical Geology**, v. 38, n. 1, p. 81-100, Janeiro 2006.

CAERS, J.; ZHANG, T. Multiple-point geostatistics: a quantitative vehicle for intergration geologic analogs into multiple reservoir model. In: GRAMMER, G. M.; HARRIS, P.; EBERLI, G. P. **Integration of Outcrop and Modern Analog in Reservoir Modeling**. 1. ed. Tulsa: AAPG memoir 80, v. 80, 2004. p. 383-394.

CERNY, V. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. **Journal of optimization theory and applications**, v. 45, n. 1, p. 41-55, 1985.

CHEN, W. H. et al. A new algorithm for automatic history matching. **SPE Journal**, 14, n. 6, December 1974. 593-608.

DATTA-GUPTA, A.; VASCO, D. W.; LONG, J. C. S. **Detailed Characterization of a Fractured Limestone Formation Using Stochastic**

Inverse Approaches. Symposium on Imported Oil Recovery. Tulsa: SPE 27744. 1995. p. 133-140.

DE JONG, K. **An Analysis of the behavior of a class of genetic adaptative systems.** University of Michigan. Ann Arbor, p. 256. 1975.

DEPARTAMENTO DE ENERGIA DOS ESTADOS UNIDOS. Sobre o Departamento de Energia dos Estados Unidos. **Site do escritório de Física Nuclear/ Departamento de Energia dos Estados Unidos**, 2010. Disponível em: <<http://www.sc.doe.gov/np/benefits/index.shtml>>. Acesso em: 11 Agosto 2010.

DEUTSCH, C. V. **Geostatistical Reservoir Modeling.** 1st Edition. ed. Alberta: Oxford University Press, v. I, 2002.

DULLIEN, F. A. L. **Porous Media.** 2nd Edition. ed. California: Academic Press, v. I, 1992.

GAVALAS, G. R.; SHAH, P. C.; SEINFELD, J. H. Reservoir history matching by Bayesian estimation. **SPE Journal**, 16, n. 6, December 1976. 337-350.

GIRÃO. **Site da Hydrolog, Serviços de Perfilagem Ltda.**, 2006. Disponível em: <<http://geraldogirao.com/apostilas.htm>>. Acesso em: 19 Maio 2010.

GOLDBERG, D. E. **Genetic Algorithms in search, optimization and machine learning.** 1a Edição. ed. Boston, MA: Addison Wesley, v. I, 1989.

GOMEZ-HERNANDEZ, J. J.; SAHUQUILLO, A.; CAPILLA, J. E. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data. **Journal of Hydrology**, Amsterdam, v. 203, n. 1, p. 162-174, 1997.

HOFFMAN, B. T.; WEN, X.-H.; STREBELLE, S. **Geologically Consistent History Matching of a Deepwater Turbidite Reservoir.** SPE Annual Technical Conference and Exhibition. Dallas: SPE. 2005. p. 1-12.

HOLLAND, J. **Adaptation in natural and artificial systems.** 1. ed. Ann Arbor: The University of Michigan Press, v. I, 1975.

HU, L. Extended Probability Perturbation Method for Calibrating Stochastic Reservoir Models. **Math Geosci**, v. 40, p. 875–885, March 2008.

HU, L. Y.; BLANC, G.; NOETINGER, B. Gradual Deformation and Iterative Calibration. **Mathematical Geology**, 33, n. 4, 2001. 475-489.

JACQUARD, P. Théorie de l'interprétation des mesures de pression. **Revue de L'Institut Français du Pétrole**, v. XIX, n. 3, p. 297-334, 1964.

JACQUARD, P.; JAIN, C. Permeability Distribution From Field Pressure Data. **Society of Petroleum Engineers Journal**, Houston, v. 5, n. 4, p. 281-294, December 1965. ISSN SPE 1307.

JAHNS, H. O. **A Rapid Method for Obtaining a Two-Dimensional Reservoir Description from well Pressure Response Data**. SPE 41st Annual Fall Meeting. Dallas: SPE. 1966. p. 315-327.

JOLLIFFE, I. T. **Principal component analysis**. 2nd edition. ed. New York: Springer verlag, 2005.

JOURNEL, A. G. Combining Knowledge from Diverse Sources: an Alternative to Tradicional Data Independence Hypothesis. **Mathematical Geology**, v. 34, n. 6, p. 573-596, Julho 2002.

KARHUNEN, K. Uber Lineare Methoden in der Wahrscheinlichkeitsrechnung. **Annales Academiae Scientiarum Fennicae, Series AI: Mathematica-Physic**, v. XXXVII, n. 1, p. 3-79, 1947.

KELKAR, M.; PEREZ, G. **Applied Geostatistics for Reservoir Characterization**. 1. ed. Houston: Society of Petroleum, v. I, 2002.

KIM, K. I.; FRANZ, M. O.; SCHÖLKOPF, B. Iterative Kernel Principal Component Analysis for Image Modeling. **Pattern Analysis and Machine Intelligence**, Los Alamitos, v. 27, n. 9, p. 1351-1365, September 2005. ISSN 0162-8828.

KIRKPATRICK, S.; GELATT JR, C. D.; VECCHI, M. P. Optimization by simulated annealing. **Science**, New York, v. 220, n. 4598, p. 671-680, May 1983.

KWOK, J. T. Y.; TSANG, I. W. H. The pre-image problem in kernel methods. **IEEE Transactions on Neural Networks**, Beijing, v. 15, n. 6, p. 1517-1525, November 2004. ISSN 1045-9227.

LOEVE, M. **Probability Theory**. 4. ed. Berlin: Springer-Verlag, v. I, 1977.

MASCHIO, C.; VIDAL, A. C.; SCHIOZER, D. J. A framework to integrate history matching and geostatistical modeling using genetic algorithm and direct search methods. **Journal of Petroleum Science and Engineering**, n. 63, p. 34-42, 2008.

MICHALEWICZ, Z. **Genetic Algorithms+ Data Structures = Evolution Programs**. 3. ed. New York: Springer-Verlag, 1996. ISBN 978-3-540-60676-5.

MIKA, S. et al. Kernel PCA and De-Noising in Feature Spaces. **Advances in neural information processing systems**, v. 11, n. 1, p. 536-542, 1999.

MIYOSHI, S. C.; SZWARCMANN, D. M.; VELLASCO, M. M. B. R. **A new Approach for History Matching of Oil and Gas Reservoir**. Proceedings of the WCCI 2001 IEEE World Congress on Computational Intelligence. Barcelona: IEEE. 2010. p. 295-302.

MOOSEGARD, K. A. T. A. Monte Carlo sampling of solutions to inverse problems. **Journal of Geophysical Research**, v. 100, n. 12, p. 431-447, 1995.

NÆVDAL, G.; JOHNSEN, L. M.; AANONSEN, S. I. **Reservoir Monitoring and Continuous Model Updating Using Ensemble Kalman Filter**. SPE Annual Technical Conference and Exhibition. Denver: SPE. 2003. p. 66-74.

OLIVER, D. S.; CUNHA, L. B.; REYNOLDS, A. C. Markov Chain Monte Carlo Methods for Conditioning Permeability Field to Pressure Data. **Mathematical Geology**, v. 29, n. 1, p. 61-91, 1997.

OUENES, A. et al. **A new algorithm for automatic history matching: Application of simulated annealing method (SAM) to reservoir inverse modeling**. Petroleum Engineering Society. Houston, p. 1-30. 1993. (SPE 26297).

OUENES, A.; RICHARDSON, S.; WEISS, W. W. **Fractured Reservoir Characterization and Performance Forecasting Using Geomechanics and Artificial Intelligence**. SPE Annual Technical Conference & Exhibition. Dallas: SPE. 1995. p. 425-436.

RAMGULAM, A.; ERTEKIN, T.; FLEMINGS, P. B. **Utilization of Artificial Neural Networks in Optimization of History Matching**. Latin American and Caribbean Petroleum Engineering Conference. Buenos Aires: SPE. 2007. p. 1-15.

REMY, N.; BOUCHER, A.; WU, J. **Applied Geostatistics with SGeMS. A User's Guide**. 1. ed. Cambridge: Cambridge University Press., 2009. ISBN 978-0-521-51414-9.

ROGGERO, F.; HU, L. Y. **Gradual Deformation of Continuous Geostatistical Models for History Matching**. SPE Annual Technical Conference and Exhibition. New Orleans: SPE. 1998. p. 1-16.

ROMERO, C. E.; CARTER, J. N. Using genetic algorithms for reservoir characterisation. **Journal of Petroleum Science and Engineering**, London, 2001. 113-123.

ROSA, A. J.; CARVALHO, R. D. S.; XAVIER, J. A. D. **Engenharia de Reservatórios de Petróleo**. 1. ed. Rio de Janeiro: Interciência, 2006.

SAGAR, R. K.; KELKAR, B. G.; THOMPSON, L. G. **Reservoir description by integration of well test data and spatial statistics**. Proc. 68th Ann. Tech. Conf. Exhibition. Dallas: SPE. 1993. p. 475-489.

SARMA, P. **Efficient Closed-Loop Optimal Control of Petroleum Reservoirs under Uncertainty**. Thesis (PhD). Stanford University. Palo Alto, California, p. 221. 2006.

SARMA, P. et al. **A New Approach to Automatic History Matching Using Kernel PCA**. SPE Reservoir Simulation Symposium. Houston: SPE. 2007. p. 1-19.

SAUDERS, C.; GAMMERMAN, A.; VOVK, V. **Ridge Regression Learning Algorithm in Dual Variables**. Proceedings of the Fifteenth International Conference on Machine Learning. Madison: Morgan Kaufmann Publishers Inc. 1998. p. 515-521.

SCHIOZER, D. J.; SOUSA, S. H. G.; MASCHIO, C. Ajuste de histórico de produção assistido. **Boletim Técnico da Petrobras**, Rio de Janeiro, v. 3, n. 1, p. 63-82, 2009.

SCHÖLKOPF, B. et al. **Fast Approximation of Support Vector Kernel Expansions, and an Interpretation of Clustering as Approximation in Feature Spaces**. Pattern Recognition Symposium. London: Springer-Verlag. 1998. p. 125-132.

SCHÖLKOPF, B. et al. **Kernel PCA Pattern Reconstruction via Approximate Pre-Images**. Proceedings of the 8th International Conference on Artificial Neural Networks, Perspectives in Neural Computing. London: Springer Verlag. 1998. p. 147-152.

SCHÖLKOPF, B.; SMOLA, A. J. **Learning with kernels: Support vector machines, regularization, optimization, and beyond**. 1. ed. Cambridge: the MIT Press, 2002.

SCHÖLKOPF, B.; SMOLA, A.; MÜLLER, K.-R. **Kernel Principal Component Analysis**. 7th International Conference Lausanne Artificial Neural Networks — ICANN'97. Switzerland: Springer. 1997. p. 583-588.

SCHÖLKOPF, B.; SMOLA, A.; MÜLLER, K.-R. Nonlinear component analysis as a kernel eigenvalue problem. **Neural Computation**, 10, n. 5, 1998. 1299–1319.

SCHULZE-RIEGERT, R. W. et al. Evolutionary Algorithms Applied to History Matching of Complex Reservoirs. **SPE Reservoir Evaluation & Engineering**, April 2002. 163-173.

SEN, M. K. et al. **Stochastic reservoir modeling using simulated annealing and genetic algorithm**. Proceedings of 67th Annual Technical Conference and Exhibition of the SPE. Washington: SPE. 1992. p. 939-950.

SLUMBERGER. **Introduction to Well Testing**. Slumberger. Bath, p. 309. 1998.

SOUZA, L. A. P. D. **Revisão Crítica da Aplicabilidade dos Métodos Geofísicos na Investigação de Áreas Submersas Rasas**. Tese (Doutorado) - Instituto Oceanográfico. Universidade de São Paulo, p. 311. 2006.

STREBELLE, S. B.; JOURNAL, A. G. **Reservoir Modeling Using Multiple-Point Statistics**. SPE Annual Technical Conference and Exhibition. New Orleans: SPE. 2001. p. 1-11.

TEIXEIRA, A. R. et al. KPCA denoising and the pre-image problem revisited. **Digital Signal Processing**, Aveiro, v. 18, n. 4, p. 568-580, July 2008.

THOMAS, J. E. **Fundamentos da Engenharia de Petróleo**. Rio de Janeiro: Interciência, 2001.

TUREYEN, O. I.; CAERS, J. A **Geostatistical Approach to History Matching Pressure and Flow Data on Non-Uniform Grids**. Proceeding of 8th European Conference on the Mathematics of Oil Recovery. Freiberg: SPE. 2002. p. 1-8.

WANG, Y.; KOVSCEK, A. R. Streamline Approach for History Matching Production Data. **SPE Journal**, v. 5, n. 4, p. 353-362, December 2000.

WEN, X. H. et al. A program to create permeability fields that honor single-phase flow rate and pressure data. **Computers & Geosciences**, v. 25, n. 3, p. 217-230, April 1999.

WHITLEY, D. **The GENITOR Algorithm and Selection Pressure: Why Rank-Based Allocation of Reproductive Trials is Best**. Proceedings of the third international conference on Genetic algorithms. Fairfax: J. David Schaffer. 1989. p. 116-121.

YU, T.; WILKINSON, D.; CASTELLINI, A. Constructing Reservoir Flow Simulator Proxies Using Genetic Programming for History Matching and Production Forecast Uncertainty Analysis. **Journal of Artificial Evolution and Applications**, New York, v. 2008, p. 1-13, January 2008.

ZHANG, F.; REYNOLDS, A. C. **Optimization algorithms for automatic history matching of production data**. Proceedings of 8th European Conference on the Mathematics of Oil Recovery. Freiberg: SPE. 2002. p. 1-10.

Anexos

A1. Mercer Kernels

O teorema de Mercer de análise funcional condiciona os requisitos aos quais é possível construir um mapeamento ϕ a partir da decomposição espectral da função kernel (k) (Schölkopf, Smola e Müller, 1998).

Se k é uma função kernel contínua de um operador integral $\mathcal{K}: L^2 \rightarrow L^2$, $(\mathcal{K}f)(y) = \int k(x, y)f(x)dx$, o qual é positivo, ou seja:

$$\int \int f(x)k(x, y)f(y) dx dy \geq 0, \text{ para todo } f \in L^2, \quad (93)$$

Então, k pode ser expandida em uma série convergente,

$$k(x, y) = \sum_{i=1}^{\infty} \lambda_i \phi_i(x)\phi_i(y), \text{ para } \lambda_i \geq 0 \quad (94)$$

Neste caso, ϕ é um mapeamento para um espaço F tal que k atue como produto escalar, como na equação (95) e (96).

$$\phi: x \rightarrow (\sqrt{\lambda_1}\phi_1(x), \sqrt{\lambda_2}\phi_2(x), \dots) \quad (95)$$

$$\phi(x) \cdot \phi(y) = k(x, y) \quad (96)$$

A2. Considerações sobre a existência de *Pre-Image*

Matematicamente, seja $\phi: R^N \rightarrow F$ uma função não necessariamente inversível. A *Pre-Image* de qualquer elemento $y \in F$, é o subconjunto de todos os elementos de R^N que são mapeados para y .

$$f^{-1}(y) = \{x \in X : f(x) = y\}.$$

Se $\phi: R^N \rightarrow F$ for uma função sobrejetora, ou seja todo elemento do codomínio F é mapeado por algum elemento do domínio R^N , então, existe pelo menos uma *Pre-Image* de y para todo $y \in F$, entretanto a *Pre-Image* pode não ser única.

$$\forall y \in B, \exists x \in A \text{ such that } y = f(x).$$

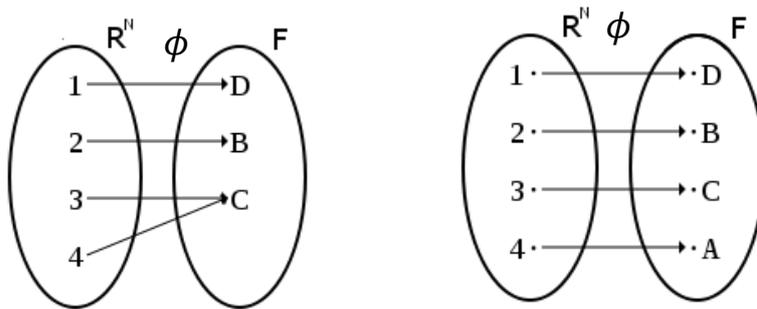


Figura 43. Ilustração de Funções sobrejetoras.

Se o mapeamento ϕ for uma função bijetora (sobrejetora e injetora) do conjunto R^N para o conjunto F , ou seja, todo elemento do codomínio F é mapeado por apenas um elemento do domínio R^N , existe uma única *Pre-Image* para todo $y \in F$.

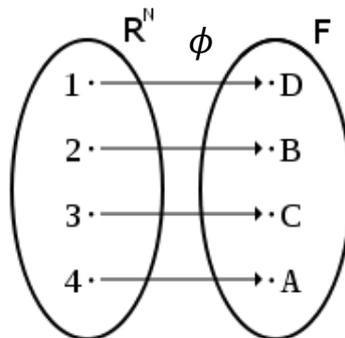


Figura 44. Ilustração de Funções Bijetoras.

Se $\phi: R^N \rightarrow F$ não for sobrejetora, nem todos os elementos de F são a imagem de algum $x \in R^N$. Com isso, a *Pre-Image* de qualquer elemento em F , neste caso, pode ser única, não ser única, ou pode até não existir.

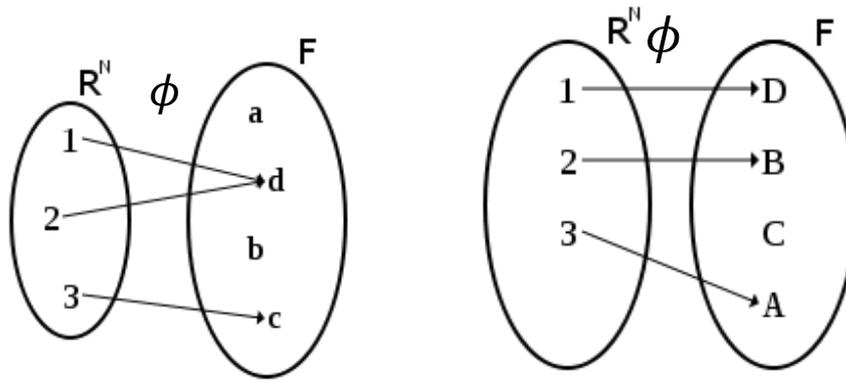


Figura 45. Ilustração de Funções não-sobrejetoras.

Para muitas funções kernel, a dimensão de F tende a ser muito maior que a dimensão de R^N e a função ϕ mapeada implicitamente não é sobrejetora (Abrahamsen e Hansen, 2009).

Com isso, dependendo do tipo de função mapeada implicitamente (sobrejetora, bijetora, não sobrejetora) a preimage de qualquer $y \in F$ pode ser única, pode não ser única ou até pode não existir, conforme resumido na tabela 5:

Tabela 5. Relação entre o tipo de função e a sua *Pre-image*

Tipo de Função $\phi: R^N \rightarrow F$	<i>Pre-Image</i>
Sobrejetora	Qualquer $y \in F$ apresenta pelo menos uma <i>Pre-Image</i>
Bijetora	Qualquer $y \in F$ apresenta apenas uma <i>Pre-Image</i>
Não sobrejetora	Qualquer $y \in F$ apresenta uma ou mais <i>Pre-Image</i> , como também pode não existir uma <i>Pre-Image</i> exata.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)