

Universidade Federal de Santa Catarina
Departamento de Informática e Estatística
Programa de Pós-Graduação em Ciência da Computação

Fernando Luis Bordignon

MODELO HÍBRIDO BAYESIANO GEOESTATÍSTICO PARA SOLUÇÃO DE PROBLEMAS
INVERSOS COM MODELAGEM DE INCERTEZA

Proposta de projeto para Doutorado sanduíche em
parceria com o grupo CERENA-IST da Universi-
dade de Lisboa.

Orientador: Mauro Roisenberg
Coorientador: Amílcar Soares

Florianópolis
2014

Resumo

O processo de caracterização de reservatórios de hidrocarbonetos consiste na determinação tridimensional e quantitativa da estrutura e das propriedades petrofísicas das rochas (porosidade, densidade, saturação de fluídos e etc) de um campo. A utilização de técnicas de inversão sísmica neste processo tem grande importância, pois possibilita a obtenção das propriedades elástica e acústica do reservatório. Inversão sísmica é um problema complexo, não linear, mal determinado e ambíguo, o que leva à necessidade de alta intervenção de especialistas para restringir o resultado. Neste processo também é importante a modelagem da incerteza dos resultados obtidos, fornecendo informações sobre os riscos envolvidos nos processos de perfuração e na produção de óleo e gás. Na literatura recente, métodos foram propostos para tentar solucionar o problema da inversão sísmica com modelagem de incerteza integrada. O uso de ferramental probabilístico baseado no teorema de Bayes e utilizando restrições a priori pode ser aplicado, em conjunto com o amostrador de Gibbs, para gerar realizações estocásticas do resultado. Outra proposta destacada na literatura é inversão Geoestatística, que também gera realizações à um custo maior do que a anterior, mas consegue modelar melhor a continuidade espacial dos fenômenos. Este documento descreve o projeto para a criação de um novo modelo integrando o método Bayesiano à inversão Geoestatística, afim melhorar a convergência do segundo mantendo sua capacidade de modelar continuidade espacial.

Palavras chave: Inversão Sísmica; Modelagem de Incerteza; Inversão Geoestatística; Inversão Bayesiana.

Sumário

1	Introdução	1
1.1	Organização do Texto	5
2	Modelos de Inversão Sísmica com Modelagem de Incerteza	6
2.1	Inversão Sísmica Linear e Não Linear	7
2.2	Modelagem de Incerteza	8
2.3	Resumo	12
3	Modelo Híbrido Bayesiano Geoestatístico	13
3.1	Resultados Preliminares	13
3.2	Plano de trabalho	14
	Bibliografia	17

Capítulo 1

Introdução

Teorias físicas permitem a realização de previsões do comportamento de um sistema, ou seja, dada a descrição completa de um sistema físico, é possível prever o resultado de algumas medidas. Esse processo de prever o resultado de medidas é chamado de problema direto (*forward problem*). O problema de inversão (*inverse problem*) consiste em utilizar o resultados de medidas efetuadas para inferir os valores de parâmetros que caracterizam o sistema (Tarantola, 2005). O problema direto é considerado determinístico. Dado o comportamento do sistema, é possível prever o resultado da medida. Contrariamente, o problema inverso muitas vezes não é determinístico.

Por exemplo: considere uma configuração de lâmpadas e suas potências luminosas dentro de uma sala de aula fechada. É possível determinar a luminosidade em um certo ponto utilizando o modelo direto, que opera sobre os parâmetros do sistema e resulta na luminosidade em um ponto. O problema inverso, neste caso, consiste em determinar a quantidade de lâmpadas e suas potências, dada a luminosidade medida em um ou mais pontos. Claramente este problema é ambíguo, pois existem várias configurações de lâmpadas que produzem a mesma luminosidade. Ainda utilizando o mesmo exemplo, é possível inserir conhecimentos a-priori, como os tipos de lâmpadas que comumente são utilizadas em salas de aula, para auxiliar na inversão. Num sistema real, existem ruídos que influenciam na medida. No exemplo em questão, a sala pode sofrer entrada de luz externa e alterar a medida num certo ponto.

No contexto da indústria de óleo e gás, o problema de inversão sísmica é importante para o processo de caracterização de reservatórios de hidrocarbonetos, que consiste na determinação tridimensional e quantitativa da estrutura e das propriedades petrofísicas (porosidade, densidade, etc) das rochas da subsuperfície. O método de aquisição sísmica de reflexão utiliza pulsos sísmicos de uma fonte artificial controlada e monitora a resposta em função do tempo. Neste

sistema, cada interface entre dois tipos de rochas diferentes gera reflexão e refração do pulso sísmico, como demonstrado na Figura 1.1.

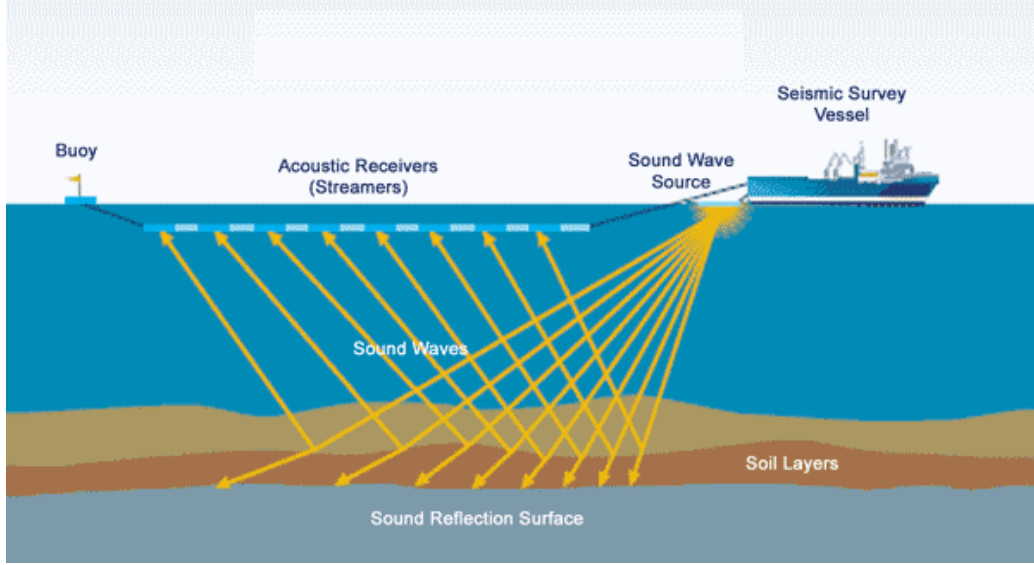


Figura 1.1: Método de sísmica de reflexão

Para inversão acústica, o modelo mais utilizado para representar a resposta do pulso sísmico atravessando as interfaces entre rochas é o convolucional. Nele é assumido um modelo em camadas para a subsuperfície e ângulo de reflexão de 90° . A medida efetuada é representada pelo resultado da convolução do pulso sísmico, também chamado de *wavelet*, com as refletividades das interfaces. Os coeficientes de reflexão com ângulo de reflexão normal são modelados por (Sen, 2006, p. 69):

$$r(t) = \frac{z(t + \delta t) - z(t)}{z(t + \delta t) + z(t)} \quad (1.1)$$

onde $z(t)$ é a impedância acústica no tempo t definida por $z(t) = \rho(t)v(t)$, onde $\rho(t)$ é a densidade da rocha e $v(t)$ a velocidade de propagação da onda acústica. Utilizando os coeficientes de reflexão, modela-se a resposta do sistema $d(t)$ aplicando a convolução da *wavelet* s com os coeficientes de refletividade:

$$d(t) = \int_{-\infty}^{\infty} s(\tau)r(t - \tau)d\tau + e_d(t) \quad (1.2)$$

onde é assumida a presença de um ruído aleatório $e_d(t)$ e d é o chamado dado sísmico. Cada sequência de dados d representa um ponto no plano xy e suas posições (t) são coordenadas em tempo que se relacionam com a profundidade.

O objetivo da inversão sísmica acústica é determinar os valores de impedância acústica das camadas de rocha. Esse é um problema não linear, pois as equações que determinam o modelo direto são não lineares. Também é considerado mal posto, pois é ambíguo, ou seja, várias combinações de camadas e suas impedâncias podem gerar o mesmo resultado no dado sísmico. Essas características geram a necessidade de alta intervenção de especialistas para restringir o resultado, ações também chamadas de regularização do problema. Quando poços são perfurados, são utilizadas outras ferramentas para observar dados mais próximos aos reais. Os principais métodos de inversão contam com a inserção desses dados de poços.

No processo de inversão é importante a modelagem da incerteza dos resultados obtidos, ou seja, obter soluções equivalentes respeitando os dados medidos e as informações a priori. Essas soluções devem representar bem a região onde o erro é menor que a tolerância desejada, chamada de região de equivalência (Tompkins; Fernández Martínez e Fernández Muñiz, 2013). Existem várias razões para essa região de equivalência existir, incluindo erros nas medidas, cobertura dos dados sobre as localizações desejadas, limitação de banda do pulso sísmico, suposições sobre o modelo físico (e.g. isotropia, modelo de camadas, homogeneidade, etc.) e aproximações matemáticas do modelo direto. A modelagem de incerteza ajuda na avaliação dos riscos envolvidos nos processos de perfuração e produção de óleo e gás.

As médias e variâncias locais das soluções não são suficientes para caracterizar a incerteza na inversão, pois os modelos de impedância geralmente são utilizados como entrada de processos como a simulação de fluxo. Esse tipo de simulação é essencialmente uma função não linear, ou seja, a média das simulações de fluxo não é necessariamente igual a simulação da média das impedâncias: $F(\bar{\mathbf{m}}) \neq \overline{F(\mathbf{m})}$, onde \mathbf{m} representa as soluções equiprováveis, $F(\cdot)$ a função não linear da simulação de fluxo e $\overline{(\cdot)}$ a operação de média. Amostragem, de uma forma geral, possibilita análises mais sofisticadas da distribuição posterior (Hansen; Cordua e Mosegaard, 2012).

Na literatura recente, métodos foram propostos para tentar solucionar o problema da inversão sísmica com modelagem de incerteza integrada. Modelos envolvendo técnicas de inteligência computacional foram propostos para tentar resolver o problema (Sen e Stoffa, 1991; Mallick, 1995; Ma, 2002; Artun e Mohaghegh, 2011; Fernández Martínez et al., 2012; Sambridge, 2013). Apesar de terem oferecido bons resultados, esses modelos baseados em otimização não capturam de forma ideal o conhecimento do especialista. A incerteza resultante desses métodos é relacionada com o nível em que foi explorada a superfície de erro, ao invés da incerteza relacionada aos dados e conhecimentos a priori inseridos.

O uso de ferramenta probabilístico baseado no teorema de Bayes pode ser aplicado, em conjunto com o amostrador de Gibbs, para gerar realizações estocásticas da distribuição a posteriori (Figueiredo et al., 2013). Assim, pode-se computar estatísticas do conjunto de soluções, obtido à alto custo computacional. Para atacar a maldição da dimensionalidade e diminuir o custo da amostragem de soluções, foram propostos métodos que utilizam redução dimensional (Tompkins et al., 2011). Neste caso, aplicando análise de componentes principais (PCA) é possível reduzir a dimensão do problema e utilizar um esquema de amostragem determinística, ao custo de perda na resolução espacial das amostras.

A área de Geoestatística teve início a partir do uso de ferramentas estatísticas para tratar dados espaciais, suas continuidades e lidar com incertezas presentes nos problemas da Geologia (Isaaks e Srivastava, 1989, p. 3). Antes de sua criação, um atributo de rocha, por exemplo, era tratado como só uma variável aleatória. Os valores desse atributo nos pontos espaciais de uma área eram tratados como amostras de uma mesma distribuição. A Geoestatística passou a assumir a característica espacial dos dados, onde cada posição da área é tratada como uma variável aleatória possuindo correlações com suas vizinhas.

Soares; Diet e Guerreiro (2007) propõem um método de inversão, chamado *Global Stochastic Inversion (GSI)*, baseado em Simulação Sequencial Direta (SSD), uma importante ferramenta da Geoestatística. Na SSD são utilizados modelos de continuidade direcional dos dados (variogramas direcionais) para simular pontos não observados da área, respeitando assim a continuidade espacial, a variabilidade e o histograma dos dados de poços. O processo de inversão faz uso de SSD para encontrar um conjunto de soluções que tenha alta correlação com o dado sísmico. Apesar da vantagem de se modelar a continuidade espacial em todas as direções, o método foi concebido para trabalhar em áreas pequenas, pois assume a estacionariedade das médias. Além disso, não converge em problemas mais avançados como a inversão para dados de porosidade.

O tema sugerido no presente documento consiste em propor um novo modelo para integrar a atualização Bayesiana no processo de inversão GSI. Resultados prévios indicam que utilizar o método determinístico de inversão Bayesiana (Buland e Omre, 2003 a), como base para iniciar a Simulação Sequencial Direta, obtem maior correlação com a sísmica na primeira iteração da inversão GSI. Com o novo modelo espera-se viabilizar a amostragem de áreas maiores do que a amostragem Bayesiana no mesmo tempo computacional, além de viabilizar a aplicação em inversão sísmica mais complexa como inverter para outras propriedades de rocha. A modelagem da continuidade espacial é outra vantagem de se integrar as propostas. No modelo de amostragem Bayesiana atual é inviável computacionalmente modelar a continuidade em direções que

não sejam a vertical, pois para tanto é necessário inserir as covariâncias entre todos os pontos a serem invertidos.

1.1 Organização do Texto

Este documento está organizado da seguinte forma. Após esta breve introdução, o Capítulo 2 apresenta o estado da arte em modelos de inversão sísmica com modelagem de incerteza. O Capítulo 3 trata do tema do projeto para a integração de atualização Bayesiana em inversão Geoestatística, a ser desenvolvido em parceria com o CERENA - Centro de Recursos Naturais do Instituto Técnico da Universidade de Lisboa, sob orientação do professor Dr. Amílcar Soares, especialista em Geoestatística e inversão sísmica.

Capítulo 2

Modelos de Inversão Sísmica com Modelagem de Incerteza

Numa configuração de experimentos físicos se tem o espaço do modelo e o espaço das medidas. No contexto da sísmica de reflexão, os dados sísmicos d são representados no espaço das medidas e a propriedade de impedância acústica m das rochas é representada no espaço do modelo.

A princípio parece descomplicada a utilização de equações físicas que descrevem o sistema de forma inversa, fazendo o mapeamento do espaço das medidas ao espaço do modelo. No entanto, esses métodos de inversão direta sofrem de instabilidades devido à ruído e características do problema (Sen, 2006, p. 50). Outra opção é utilizar tentativa e erro para ajustar os parâmetros até conseguir uma resposta semelhante aos dados. Formalmente isto é automatizado utilizando otimização. Para tanto, é preciso definir uma função de custo, ou função objetivo, que mede o ajuste dos dados produzidos pelos parâmetros do modelo (dado sintético) ao dado medido.

O objetivo da inversão, no entanto, vai além de encontrar os parâmetros que melhor se ajustam aos dados. Quando os dados são ruidosos, o modelo direto não é exato e não existem dados suficientes, a inversão tem solução não única, ou seja, vários modelos ajustam aos dados de forma equivalente. Consequentemente é importante modelar a incerteza envolvida no processo, indicando qual a variabilidade dos modelos que se ajustam bem aos dados.

A relação entre o modelo e os dados (modelo direto) é dada por:

$$d = g(m) + e \tag{2.1}$$

onde $g(\cdot)$ é uma função não linear, e assume-se que um ruído e está presente. Em teoria o ruído é uma interferência aleatória que não se tem controle, na prática se considera ruído tudo que

não é explicado pela função g , e.g. imprecisões no modelo físico e problemas com filtragem e processamento dos dados.

2.1 Inversão Sísmica Linear e Não Linear

O modelo mais utilizado para aproximar a função g , no caso da inversão sísmica, é o modelo convolucional. No caso discreto a convolução é dada pelo produto acumulado de cada amostra do vetor de refletividades por todas as amostras da *wavelet*. Portanto pode ser representado por uma operação matricial:

$$\mathbf{d} = \mathbf{G}\mathbf{r} \quad (2.2)$$

onde \mathbf{G} é uma matriz convolucional construída utilizando as amostras da *wavelet* e \mathbf{r} o vetor das refletividades definido por:

$$r(t) = \frac{z(t + \delta t) - z(t)}{z(t + \delta t) + z(t)} \quad (2.3)$$

o que torna não linear a relação entre impedância \mathbf{z} e o dado sísmico. Uma aproximação válida quando valores de refletividades não ultrapassam 0.3 é:

$$r(t) = \frac{1}{2}\Delta \ln(z(t)) \quad (2.4)$$

Utilizando estas aproximações para o modelo direto, a alternativa mais objetiva é inverter a matriz \mathbf{G} para obter \mathbf{z} . Neste caso temos os seguintes problemas: existência; unicidade; estabilidade; e robustez (Sen, 2006, p. 56-57). Utilizando a formulação de mínimos quadrados também é possível resolver sistemas sobredeterminados, solucionando problemas com a melhor estimativa possível no sentido de minimizar o erro quadrático (norma L_2). Apesar de ser uma solução mais geral, ainda é utilizado somente um critério de ajuste aos dados, o que não possibilita a inserção de conhecimento a priori. É possível regularizar o método de mínimos quadrados, mas ainda não se tem muita liberdade para inserir conhecimentos a priori e de outras fontes (Clapp; Clapp e Biondi, 2005).

Quando não é possível o uso da aproximação da Equação 2.4, o problema deve ser abordado utilizando métodos de otimização não linear. Com isso os erros devido às aproximações do modelo forward diminuem, mas a otimização se torna mais custosa. Como a relação entre os dados e os parâmetros é não linear, a função objetivo a ser minimizada irá possuir mínimos

locais, tornando necessário o uso de métodos de otimização global. Esta prática está bem documentada na literatura de inversão, como o uso de *simulated annealing* (Ma, 2002), de algoritmos genéticos (Mallick, 1995) e enxame de partículas (Zhe e Hanming, 2013).

Outra forma de inversão presente na literatura é a elástica (Azevedo et al., 2013; Buland e Omre, 2003 a). Nesse tipo de inversão os dados sísmicos estão em um nível diferente de processamento onde são empilhados em ângulos de reflexão. Com isso é possível inverter para V_p (velocidade primária/compressional), V_s (velocidade secundária/cisalhante) e ρ (densidade), ao invés de somente impedância acústica. A Equação 2.5 modela a relação da refletividade c_{pp} com V_p (α), V_s (β) e ρ para cada ângulo disponível.

$$c_{pp}(\theta) = a_\alpha(\theta) \frac{\Delta\alpha}{\bar{\alpha}} + a_\beta(\theta) \frac{\Delta\beta}{\bar{\beta}} + a_\rho(\theta) \frac{\Delta\rho}{\bar{\rho}} \quad (2.5)$$

onde:

$$a_\alpha(\theta) = \frac{1}{2}(1 + \tan^2 \theta), \quad (2.6)$$

$$a_\beta(\theta) = -4 \frac{\bar{\beta}^2}{\bar{\alpha}^2} \sin^2 \theta, \quad (2.7)$$

$$a_\rho(\theta) = \frac{1}{2} \left(1 - 4 \frac{\bar{\beta}^2}{\bar{\alpha}^2} \sin^2 \theta \right). \quad (2.8)$$

Adicionalmente $\bar{\alpha}$, $\bar{\beta}$ e $\bar{\rho}$ são as respectivas médias sobre a interface; $\Delta\alpha$, $\Delta\beta$ e $\Delta\rho$ são os contrastes e θ o ângulo de reflexão. Essas propriedades são importantes, pois a velocidade cisalhante (V_s) tem forte correlação com a presença de líquidos, o que pode indicar interfaces rocha/óleo e água/óleo. As metodologias para resolução do problema são semelhantes, mas neste caso aumentando a dimensão dos dados e números de parâmetros a serem estimados (Buland e Omre, 2003 a).

2.2 Modelagem de Incerteza

Os métodos tratados na seção anterior não modelam a incerteza envolvendo os resultados. Novas propostas surgiram recentemente para tentar encontrar modelos $\mathbf{m}_f \in \mathbf{M} \subset \mathbf{R}^N$, onde \mathbf{M} é o conjunto de modelos admissíveis que atendem o conhecimento a priori e são consistentes com o domínio do problema (Tompkins et al., 2011). Para o problema da incerteza, o objetivo é encontrar o conjunto de modelos pertencentes a \mathbf{M} que se ajustam aos dados observados com

uma certa tolerância (tol):

$$\|\mathbf{G}(\mathbf{m}) - \mathbf{d}\|_p \leq tol \quad (2.9)$$

onde $\|\cdot\|_p$ é uma norma L_p de erro escolhida. Na prática esse parâmetro cria um vale plano na função de erro (Fernández Martínez; Fernández Muñiz e Tompkins, 2012), como ilustrado em uma dimensão na Figura 2.1.

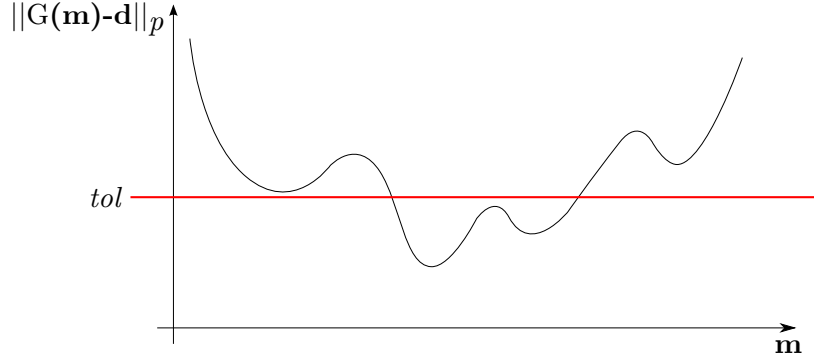


Figura 2.1: Efeito da tolerância na função de erro

A justificativa para o uso desse nível de tolerância é evitar o ajuste dos dados ao ruído. No caso do erro atingir o mínimo, o modelo está tentando explicar o ruído presente nos dados medidos. Apesar da Figura 2.1 ser uma simplificação, o conceito é ilustrado, pois ao adicionar a tolerância um conjunto de modelos passa a ser aceitável. Sem a tolerância um só modelo seria eleito como melhor e não teríamos a estimativa da incerteza. Este parâmetro é definido empiricamente e deve refletir a incerteza e o ruído presente nos dados.

Em Buland e Omre (2003 a) uma técnica de inversão é proposta inserida num *framework* Bayesiano. O objetivo é obter a distribuição posterior das propriedades das rochas, ou o máximo a posteriori. O algoritmo é baseado no modelo convolucional e na aproximação da Equação 2.4. Utilizando esse arcabouço probabilístico, a solução é representada por uma distribuição posterior Gaussiana com expressões explícitas para o valor esperado e para a covariância, ou seja, é calculado o máximo a posteriori (MAP) da distribuição. Como as expressões são analíticas, não são necessárias iterações de ajuste do modelo, tornando o método eficiente. A vantagem de ser eficiente torna o método útil em casos de uso reais, possibilitando o especialista alterar parâmetros e avaliar o resultado em tempo real numa pequena área. Satisfeito com a parametrização, o método é aplicado a todo o campo. Além de retornar a matriz de covariância indicando a incerteza presente no resultado, não é necessário definir a tolerância explicitamente,

mas é preciso definir a matriz de covariância a priori do resultado esperado, ou seja, é preciso ter conhecimento, mesmo que de forma grosseira, das correlações espaciais e variâncias que se espera do resultado. Ao final, a covariância posterior é calculada.

Uma desvantagem importante é a dificuldade em inserir modelos de continuidade mais abrangentes, pois para tanto é necessário incluir as covariâncias entre os pontos a serem estimados. Atualmente é inserida somente covariância entre amostras no mesmo traço, ou seja, somente na direção vertical. Inserir covariância em outras direções aumenta o número de elementos da matriz de covariância de forma quadrática, tornando o processo custoso.

Utilizando uma abordagem estocástica, Figueiredo et al. (2013) leva em consideração outras fontes de incerteza no processo. Com isso o problema continua linearizado, mas é preciso utilizar um esquema de amostragem baseado em simulação *Markov-chain Monte Carlo* (MCMC) via algoritmo de Gibbs, pois não é possível obter as distribuições posteriores analiticamente. Com realizações da distribuição posterior, é possível calcular as estatísticas de interesse e modelar a incerteza envolvida. A desvantagem é o alto custo computacional para gerar amostras dessas distribuições, principalmente com o aumento da dimensão do problema. Um estudo recente utiliza MCMC para realizar amostragem da propriedades de porosidade e permeabilidade das rochas (Zunino et al., 2014). Estas propriedades são geralmente estimadas em uma etapa posterior à estimativa de propriedades primárias, como a impedância. Portanto o método gera interesse ao realizar inversão de forma mais avançada no fluxo de trabalho, além de gerar estimativas da incerteza. Apesar de sofrer com o alto custo computacional da amostragem o autor relata bons resultados.

A estimativa de *wavelets* também é objeto de estudo na literatura. O problema consiste em determinar o pulso sísmico que gerou o sinal medido, utilizando os dados de poços como referência. *Wavelets* podem ser consideradas o elo entre os dados sísmicos e as propriedades de rocha. Como podem haver erros de conversão da escala de profundidade dos poços para a escala de tempo dos dados sísmicos, é necessário avaliar a incerteza na estimativa da *wavelet*. Por isso, um esquema de amostragem semelhante, baseado em MCMC, foi utilizado em Buland e Omre (2003 b).

Esforços na área de inteligência computacional foram inicialmente empregados utilizando métodos que possuem componentes estocásticas, mas foram desenvolvidos originalmente para exploração, ou seja, encontrar rapidamente o mínimo global da função objetivo (Sen e Stoffa, 1995). Otimização por enxame de partículas é um exemplo. O método foi modificado para permanecer em sua fase de exploração, favorecendo soluções que ficam abaixo de certa tolerância.

Realizando, portanto, o que é chamado de *importance sampling* (Martínez et al., 2010). Os autores dessa metodologia defendem que a função de erro é um bom *proxy* para a distribuição posterior.

Trabalhos anteriores tentaram sem êxito utilizar várias execuções de algoritmos genéticos (GA) para estimar a incerteza (Sen e Stoffa, 1995, p. 152). A hipótese de trabalho era a que ao executar o algoritmo várias vezes, seriam encontrados vários mínimos locais que representariam a incerteza. Mas como demonstrado em Fernández Martínez; Fernández Muñoz e Tompkins (2012), a topografia da função de erro em inversão sísmica possui grandes vales alongados de mínimo global. Com isso os métodos devem ser modificados para capturar essa característica, a simples execução múltipla não gera boa estimativa da incerteza.

O trabalho de Tompkins et al. (2011) trata de inversão de dados eletromagnéticos, outro tipo de metodologia utilizada na indústria para imageamento de subsuperfície. Neste problema o modelo direto é mais complexo e não linear. Apesar de não utilizar dados sísmicos, a metodologia proposta tem o objetivos pertinentes à inversão sísmica. A hipótese é que não é necessário amostrar exaustivamente a distribuição posterior para modelar a incerteza, e sim utilizar amostras representativas. A proposta utiliza redução dimensional com análise de componentes principais (PCA) sobre a matriz de covariância. Esta matriz, por sua vez, é calculada a partir de um conjunto de soluções ou aproximada utilizando informação do gradiente da função objetivo após a execução de um método de inversão não linear. Restrições a priori são mapeadas para o espaço reduzido e a técnica de Smolyak (1963) é utilizada para amostrar de forma determinística e hierárquica este espaço. Retornando ao espaço original, é utilizado um critério de rejeição para eliminar soluções acima de certa tolerância. Os passos de amostragem e rejeição são repetidos, refinando o *grid* de Smolyak, até que haja convergência de estatísticas das amostras aceitas.

Ao reduzir a dimensão do problema se perde resolução espacial das amostras a serem geradas, pois serão considerados somente parte dos primeiros autovetores, ordenados de forma decrescente em relação aos seus autovalores, da matriz de covariância. Ou seja, menores variabilidades são desconsideradas (Jolliffe, 2002, p. 3). Essa perda é justificada por serem desprezíveis em certos casos e pela diminuição do custo computacional ao trabalhar no espaço reduzido. Além do custo de computar os autovetores e autovalores, são adicionados os custos de computar o *grid* de Smolyak e do mapeamento de restrições do espaço original ao espaço reduzido. Esses custos extras se tornam proibitivos quando a dimensão do problema cresce e não é mais possível reduzir a dimensão à centenas de variáveis. Com isso alternativas surgiram para tentar

expandir o limite do tamanho do problema, em detrimento da resolução das amostras utilizando decomposição em valores singulares (Tompkins et al., 2011) e utilizando outro esquema de amostragem determinística (Tompkins, 2012). Apesar dos avanços não se sabe se é possível aplicar a metodologia aos problemas de larga escala ($> 10^5$ variáveis) (Tompkins; Fernández Martínez e Fernández Muñiz, 2013).

A metodologia de Scheidt e Caers (2009) também utiliza redução dimensional, mas baseada em *multi-dimensional scaling* (MDS). Técnica que visa preservar distâncias do espaço original no espaço reduzido. Característica que possibilita o uso de diferentes métricas para modelar distâncias entre modelos. Análise de agrupamento de dados (*clustering analysis*) é utilizada para definir modelos equivalentes no espaço reduzido, selecionando somente poucos modelos para simulação de fluxo de petróleo. Esse fluxo de trabalho é possível pois os autores modelam a incerteza já com o objetivo de verificar sua influência na simulação de fluxo.

Soares (2001) propõe o uso de Simulação Sequencial Direta (SSD) utilizando dados de poços para uma amostragem sequencial da impedância baseada na variância da Krigagem. O autor avança no sentido de simular diretamente da distribuição global sem precisar fazer transformações para variáveis Gaussianas. Inversão propriamente dita é realizada em Soares; Diet e Guerreiro (2007), onde a ideia de combinação presente em algoritmos genéticos é utilizada. Regiões locais de simulações estocásticas são combinadas afim de encontrar uma amostra global que maximize a correlação da sísmica sintética com a medida. Como o método utiliza SSD, a continuidade espacial é modelada em qualquer direção desejada pelos variogramas.

2.3 Resumo

Este capítulo revisou o estado da arte em inversão sísmica acústica com modelagem de incerteza. Pontos críticos dos métodos foram considerados e identificados para pesquisa futura. O próximo capítulo irá definir a proposta de pesquisa, apresentar o plano de trabalho e concluir com as perspectivas de contribuição.

Capítulo 3

Modelo Híbrido Bayesiano Geoestatístico

Este trabalho propõe desenvolver um novo modelo para inversão sísmica acústica e elástica com modelagem de incerteza. No capítulo anterior foi apresentado um método que trata o problema utilizando amostragem da distribuição posterior via simulação sequencial direta (Soares; Diet e Guerreiro, 2007). Outro método apresentado utiliza o cálculo do máximo a posteriori (MAP) para determinar o resultado mais provável (Buland e Omre, 2003 a).

A presente proposta tem como objetivo avançar no sentido de aumentar a eficiência do método GSI (Soares; Diet e Guerreiro, 2007) afim de evitar amostrar soluções que não estejam dentro da distribuição posterior. Neste caso o termo distribuição posterior é utilizado para denotar o conjunto de soluções que é consistente com os conhecimentos a priori e com os dados sísmicos. Para melhorar a eficiência do método, mantendo sua capacidade de modelagem da continuidade espacial direcional, serão utilizados resultados de Buland e Omre (2003 a) para guiar a amostragem. Apesar do resultado MAP ser obtido com um método que lineariza o problema e só utiliza distribuições Gaussianas, o GSI mantém o caráter de busca não linear e reproduz a distribuição presente nos dados de poço.

3.1 Resultados Preliminares

Testes realizados com o método para cálculo do MAP obtiveram resultados e tempo de execução comparáveis com métodos implementados na indústria (Figueiredo et al., 2014). Apesar de resultar somente nas médias e variâncias, a parametrização e regularização utilizada na metodologia pode ajudar a restringir a amostragem, de forma a inserir informações da posterior e

tornar a amostragem mais eficiente.

Como a inversão GSI é fundamentada na Simulação Sequencial Direta, os resultados da primeira iteração da GSI são iguais aos resultados de uma SSD utilizando os mesmos parâmetros. Desta forma é possível verificar a melhora na correlação da sísmica sintética com a sísmica original quando se utiliza o resultado do MAP como imagem secundária para SSD. Este teste foi realizado e na primeira iteração do GSI a correlação das sísmicas foi 0, com a SSD utilizando o resultado do MAP como imagem secundária foi obtida uma correlação de 0.61, foram utilizadas populações de 15 amostras para cada método. A Figura 3.1 exibe um comparativo entre as médias de impedância da primeira iteração do GSI e das amostras da SSD com o resultado do MAP. Também são exibidos os dados de impedância acústica de quatro poços.

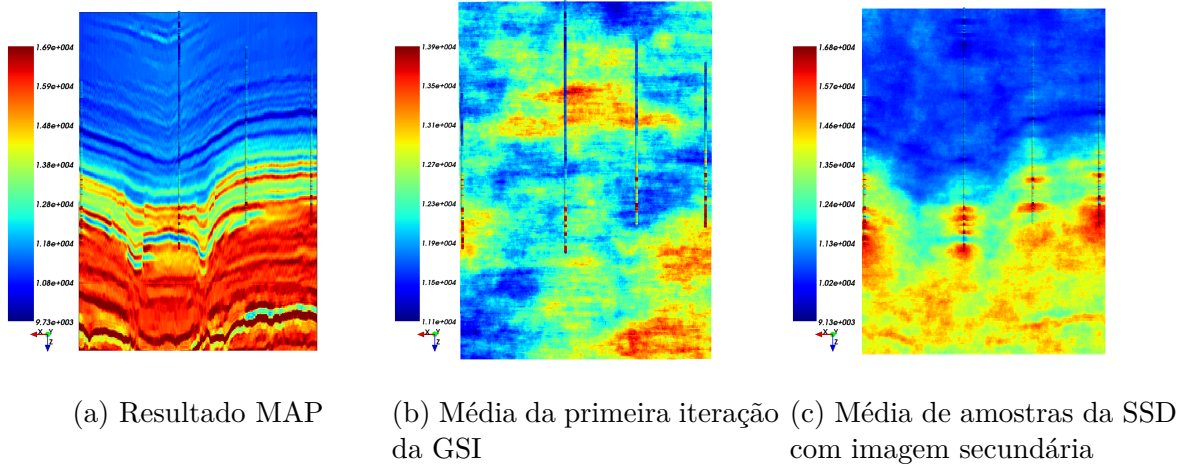


Figura 3.1: Comparação entre imagens de impedância acústica

Na Figura 3.1b é possível notar que a tendência da baixa frequência não é incorporada no resultado, já na Figura 3.1c essa característica fica visível, pois conforme a profundidade (eixo Z) aumenta, os valores de impedância estão mais altos. Da mesma forma que o resultado da inversão MAP na Figura 3.1a. A consistência com os dados dos poços também foi maior, numa análise visual. Os testes foram realizados no software GEOMS2, implementado pelo grupo CERENA da Universidade de Lisboa.

3.2 Plano de trabalho

A Figura 3.2 mostra o cronograma quinzenal das atividades a serem desenvolvidas durante o período do estágio.

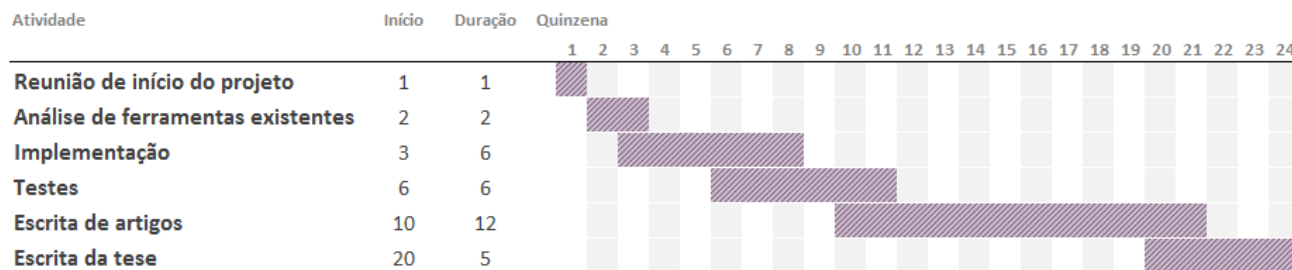


Figura 3.2: Cronograma quinzenal do projeto

Dentre os periódicos que apresentam trabalhos relacionados com esta área, se destacam os seguintes com suas classificações no Qualis para Ciências da Computação:

- IEEE Trans. on Geoscience and Remoting Sensing
- ISSN: 0196-2892 - IEEE - Qualis A1
- Journal of Computational Physics
- ISSN: 0021-9991 - Elsevier - Qualis A1
- Computers & Geosciences
- ISSN: 0098-3004 - Elsevier - Qualis A2
- Computational Geosciences
- ISSN: 1420-0597 - Springer - Qualis B1
- Geophysics
- ISSN: 0016-8033 - Society of Exploration Geophysicists - Qualis A2 Interdisciplinar
- Mathematical Geosciences
- ISSN: 1874-8961 - Springer - Qualis A2 Interdisciplinar

A colaboração com o grupo de pesquisas do Prof. Amílcar Soares, do Instituto Superior Técnico ligado a Universidade de Lisboa, é justificada pois este centro é a principal referência na literatura sobre Simulação Sequencial Direta e inversão Geoestatística (Azevedo et al., 2013; Soares; Diet e Guerreiro, 2007; Soares, 2001). A pesquisa nos dois primeiros anos do projeto de Doutorado foi feita sob orientação do Prof. Mauro Roisenberg, coautor de referências relevantes para inversão Bayesiana (Figueiredo et al., 2013, 2014). Desta forma, espera-se cumprir a proposta de integração sob orientação de especialistas em ambos os métodos. A proposta também

faz parte de uma parceria público-privada com a PETROBRAS, incluindo-se em duas áreas contempladas pelo programa Ciência sem Fronteiras: Ciências Exatas e da Terra; e Petróleo, Gás e Carvão Mineral.

Bibliografia

- Artun, E. e Mohaghegh, S. (2011), Intelligent seismic inversion workflow for high-resolution reservoir characterization, *Computers & Geosciences* **37**(2), 143 – 157.
- Azevedo, L.; Nunes, R.; Soares, A. e Neto, G. (2013), Stochastic seismic AVO inversion, 75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013, EAGE, Londres, Reino Unido.
- Buland, A. e Omre, H. (2003 a), Bayesian linearized avo inversion, *Geophysics* **68**(1), 185–198.
- Buland, A. e Omre, H. (2003 b), Bayesian wavelet estimation from seismic and well data, *Geophysics* **68**(6), 2000–2009.
- Clapp, M. L.; Clapp, R. G. e Biondi, B. L. (2005), Regularized least-squares inversion for 3-D subsalt imaging, SEG Annual Meeting, Society of Exploration Geophysicists, Houston, Texas, pp. 1814–1817.
- Fernández Martínez, J. L.; Fernández Muñiz, M. Z. e Tompkins, M. J. (2012), On the topography of the cost functional in linear and nonlinear inverse problems, *Geophysics* **77**(1), W1–W15.
- Fernández Martínez, J. L.; Mukerji, T.; García Gonzalo, E. e Suman, A. (2012), Reservoir characterization and inversion uncertainty via a family of particle swarm optimizers, *Geophysics* **77**(1), M1–M16.
- Figueiredo, L. P.; Santos, M.; Roisenberg, M. e Neto, G. (2013), Stochastic Bayesian algorithm to a jointly acoustic inversion and wavelet estimation, Society of Exploration Geophysicists, chapter 635, pp. 3273–3277.
- Figueiredo, L. P.; Santos, M.; Roisenberg, M.; Neto, G. e Figueiredo, W. (2014), Bayesian

- framework to wavelet estimation and linearized acoustic inversion, *Geoscience and Remote Sensing Letters*, IEEE **PP**(99), 1–5.
- Hansen, T.; Cordua, K. e Mosegaard, K. (2012), Inverse problems with non-trivial priors: efficient solution through sequential gibbs sampling, *Computational Geosciences* **16**(3), 593–611.
- Isaaks, E. H. e Srivastava, R. M. (1989), *Applied geostatistics*, Oxford University Press, New York.
- Jolliffe, I. T. (2002), *Principal component analysis*, Springer, New York.
- Ma, X. (2002), Simultaneous inversion of prestack seismic data for rock properties using simulated annealing, *GEOPHYSICS* **67**(6), 1877–1885.
- Mallick, S. (1995), Model-based inversion of amplitude-variations-with-offset data using a genetic algorithm, *Geophysics* **60**(4), 939–954.
- Martínez, J. L. F.; Gonzalo, E. G.; Muñoz, Z. F.; Mariethoz, G. e Mukerji, T. (2010), Posterior sampling using particle swarm optimizers and model reduction techniques, *Int. J. Appl. Evol. Comput.* **1**(3), 27–48.
- Sambridge, M. (2013), A parallel tempering algorithm for probabilistic sampling and multimodal optimization, *Geophysical Journal International* .
- Scheidt, C. e Caers, J. (2009), Representing spatial uncertainty using distances and kernels, *Mathematical Geosciences* **41**(4), 397–419.
- Sen, M. K. (2006), *Seismic Inversion*, Society of Petroleum Engineers, Richardson, TX, USA.
- Sen, M. K. e Stoffa, P. L. (1991), Nonlinear one-dimensional seismic waveform inversion using simulated annealing, *Geophysics* **56**(10), 1624–1638.
- Sen, M. K. e Stoffa, P. L. (1995), *Global Optimization Methods in Geophysical Inversion*, Vol. 4 of *Advances in Exploration Geophysics*, Elsevier.
- Smolyak, S. (1963), Quadrature and interpolation formulas for tensor products of certain classes of functions, *Soviet Mathematics, Doklady* **4**, 240–243.

- Soares, A. (2001), Direct sequential simulation and cosimulation, *Mathematical Geology* **33**(8), 911–926.
- Soares, A.; Diet, J. e Guerreiro, L. (2007), Stochastic inversion with a global perturbation method, *Petroleum Geostatistics 2007*, EAGE.
- Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics.
- Tompkins, M. J. (2012), Efficient estimation of nonlinear posterior model covariances using maximally sparse cubature rules, *Geophysics* **77**(5), ID1–ID8.
- Tompkins, M. J.; Fernández Martínez, J. L.; Alumbaugh, D. L. e Mukerji, T. (2011), Scalable uncertainty estimation for nonlinear inverse problems using parameter reduction, constraint mapping, and geometric sampling: Marine controlled-source electromagnetic examples, *Geophysics* **76**(4), F263–F281.
- Tompkins, M. J.; Fernández Martínez, J. L. e Fernández Muñiz, Z. (2013), Comparison of sparse-grid geometric and random sampling methods in nonlinear inverse solution uncertainty estimation, *Geophysical Prospecting* **61**(1), 28–41.
- Zhe, Y. e Hanming, G. (2013), Non-linear prestack seismic inversion with global optimization using an edge-preserving smoothing filter, *Geophysical Prospecting* **61**(4), 747–760.
- Zunino, A.; Lange, K.; Melnikova, Y.; Hansen, T. e Mosegaard, K. (2014), Reservoir modeling combining geostatistics with markov chain monte carlo inversion, E. Pardo-Igúzquiza; C. Guardiola-Albert; J. Heredia; L. Moreno-Merino; J. J. Durán e J. A. Vargas-Guzmán, (Eds.), *Mathematics of Planet Earth, Lecture Notes in Earth System Sciences*, Springer Berlin Heidelberg, pp. 683–687.