

4

Sistema de Identificação de Parâmetros em Dados Sísmicos por Algoritmos Genéticos

Neste capítulo são abordadas as especificações de cada módulo e o Algoritmo Genético que compõem o modelo proposto. Na primeira seção estão descritas as etapas e os dados gerados em cada módulo. Na segunda seção, são definidos os principais componentes do Algoritmo Genético proposto para este trabalho, tais como, a representação dos cromossomos, os operadores genéticos e os processos de decodificação e de avaliação. A terceira e última seção descreve o simulador desenvolvido a partir do modelo proposto neste trabalho.

4.1.

Arquitetura do Modelo Proposto

O método proposto utiliza algoritmos genéticos para identificar propriedades do meio rochoso que melhor caracterizem um conjunto de dados sísmicos observados. A arquitetura do modelo proposto é composta por 4 módulos como pode ser visto na Figura 13.

O primeiro módulo compreende o Algoritmo Genético que define a representação dos cromossomos, os operadores genéticos, a evolução do AG, a decodificação, a avaliação e a seleção dos indivíduos. Na decodificação dos indivíduos é criado um modelo da subsuperfície a partir do qual será gerada a sísmica, sendo que cada indivíduo representa apenas um modelo.

No segundo módulo, é feita a geração de dados sísmicos sintéticos a partir dos indivíduos decodificados e da geometria de aquisição. Para gerar dados sísmicos podem-se utilizar equações aproximadas ou simuladores que calculem os dados sísmicos.

Já no terceiro módulo, a partir do sismograma de referência para o qual se quer obter os parâmetros físicos das rochas, é selecionado no traço sísmico de cada receptor o tempo em que a amplitude da onda capturada foi máxima.

No quarto e último módulo é calculada uma medida de erro entre o sismograma de referência e o sismograma gerado a partir de um indivíduo

decodificado do AG. Essa medida de erro é atribuída ao indivíduo, com a qual este será avaliado na etapa de seleção do AG.

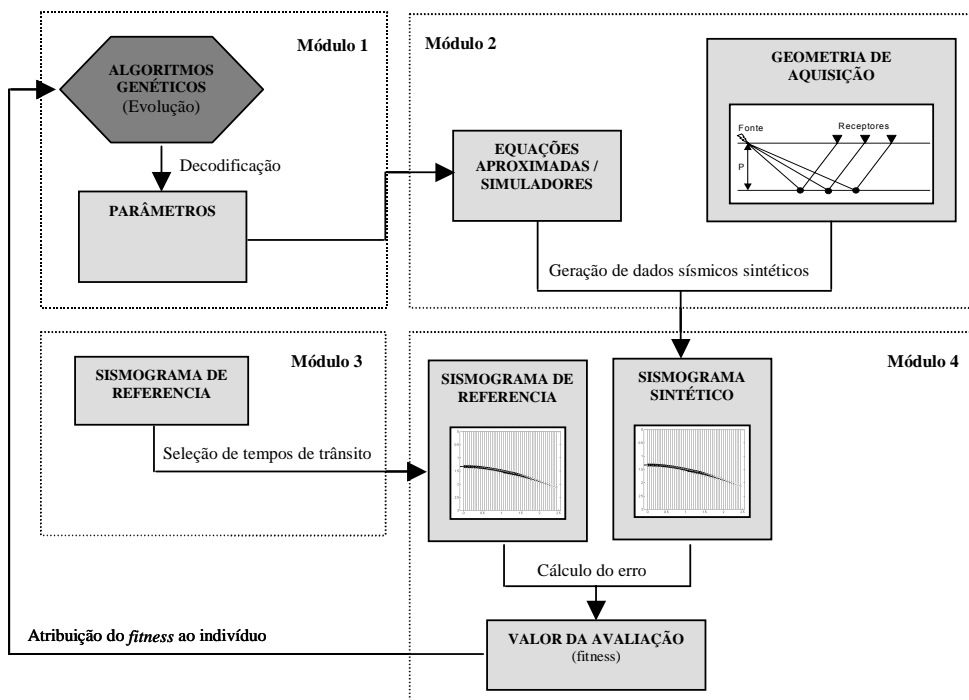


Figura 13 – Descrição do sistema de inversão de parâmetros por Algoritmos Genéticos a partir de dados sísmicos.

O módulo 3 compreende uma tarefa que é executada uma única vez no início das atividades. Já os módulos 2 e 4 são executados para cada indivíduo da população de cada geração. O sistema é dado como finalizado se for atingido um dos critérios de parada – número máximo de experimentos ou valor de erro mínimo atingido.

4.1.1. Módulo 1 – Algoritmo Genético

Este módulo é responsável pela otimização propriamente dita das propriedades do meio, tais como parâmetros físicos e atributos sísmicos. Através do uso de um Algoritmo Genético (AG), a modelagem é evoluída, tendo como resultado final a caracterização das rochas de uma subsuperfície.

A estratégia é gerar vários modelos da subsuperfície e, em seguida, avaliar esses modelos através de uma medida de erro.

O AG é a etapa principal do modelo proposto e está descrito em detalhes na próxima seção.

4.1.2. Módulo 2 – Geração de Dados Sísmicos Sintéticos

Este módulo recebe do Algoritmo Genético o modelo da subsuperfície gerado a partir de um indivíduo decodificado. Este módulo será executado para cada indivíduo da população corrente do AG.

O modelo da subsuperfície pode ser representado por parâmetros físicos (velocidade intervalar, densidade da rocha), geométricos (espessura da camada, número de camadas) ou atributos sísmicos (velocidade rms, tempo para incidência normal). A Figura 14 mostra um modelo da subsuperfície hipotético composto por n camadas horizontais, homogêneas e isotrópicas. As camadas são caracterizadas neste modelo pelos seguintes parâmetros: velocidade intervalar (V_i) e espessura da camada (e). A Tabela 2 mostra os parâmetros de cada camada que serão otimizados pelo método proposto.

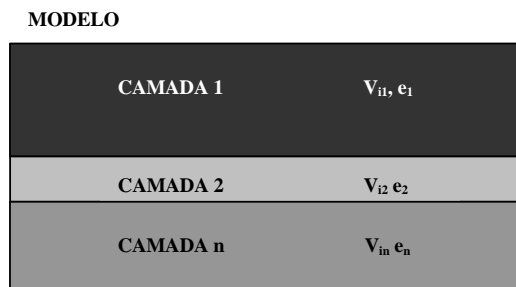


Figura 14 – Exemplo de modelo da subsuperfície com n camadas horizontais homogêneas e isotrópicas.

Índice da camada	Velocidade intervalar	Espessura
1	V_{i1}	e_1
2	V_{i2}	e_2
n	V_{in}	e_n

Tabela 2 – Parâmetros que caracterizam as camadas de um modelo da subsuperfície. A camada de índice 1 tem velocidade intervalar V_{i1} e espessura e_1 .

Além do modelo da subsuperfície, é necessário estabelecer a geometria de aquisição sísmica – procedimento este que é realizado uma única vez logo no início. As principais definições do sistema de aquisição são: a quantidade de receptores, a distância entre o primeiro receptor e a fonte (*offset*), a distância

entre si dos demais receptores, o intervalo de amostragem e o tempo máximo de leitura. O sistema de aquisição definido para gerar os dados sísmicos neste trabalho deve ter a mesma configuração do sistema de aquisição do sismograma de referência.

Para gerar os dados sísmicos sintéticos é preciso selecionar uma equação que calcule os tempos de trânsito a partir de um modelo parametrizado da subsuperfície e da geometria de aquisição.

Simuladores também são empregados para gerar sísmica sintética. Além de combinarem várias equações aproximadas encontradas na literatura, os simuladores permitem que a definição do meio seja mais compatível com a realidade. Normalmente estas ferramentas calculam, além dos tempos de trânsito, as amplitudes das ondas recebidas – dados importantes para uma futura comparação com um sismograma real.

Apesar de todos estes benefícios, todos os simuladores em geral possuem algum tipo de limitação. Um dos problemas é a geração de ruído indesejado, juntamente com os dados sísmicos, causado pelo método utilizado no simulador. Outra questão é a necessidade de um grande poder computacional para gerar os dados devido à complexidade dos cálculos.

Neste trabalho foi utilizada a equação de Dix (1955), descrita no capítulo 2, para calcular os dados sísmicos sintéticos (tempos de trânsito) para os estudos de caso 1 e 2. No estudo de caso 3, foi avaliado o uso dos simuladores *crewes* e *anray* para gerar sismogramas sintéticos.

4.1.3. Módulo 3 – Seleção de Tempos de Trânsito

Neste módulo deve ser selecionado o sismograma de referência, a partir do qual se quer caracterizar o meio rochoso. O sismograma contém as amplitudes das ondas capturadas pelos receptores em um período de tempo.

O capítulo 2 mostra que, ao se gerar dados sísmicos através da equação de Dix (1955), teremos apenas um valor de tempo de trânsito calculado para cada onda capturada, como se a onda chegasse de uma só vez num mesmo instante de tempo, simplificando a realidade. Neste caso, as amplitudes das ondas recebidas não são calculadas, apenas os tempos de trânsito.

Sendo assim, não é possível comparar um conjunto de dados sísmicos que considera somente os tempos de trânsito das ondas recebidas com um sismograma que considera, além dos tempos de trânsito, as amplitudes das

ondas recebidas.

Para que o cálculo da medida do erro entre os dados sísmicos gerados a partir do AG e o sismograma de referência seja possível, é preciso selecionar os eventos neste sismograma. Este processo implica na seleção do tempo de trânsito correspondente à máxima amplitude da onda recebida para cada traço sísmico de cada evento.

A Tabela 3 mostra um sismograma de exemplo obtido a partir de 4 receptores (rec_1 , rec_2 , rec_3 , rec_4) e em 6 instantes de tempo (tt_1 , tt_2 , tt_3 , tt_4 , tt_5 , tt_6). As amplitudes das ondas capturadas em cada receptor rec_j no instante de tempo tt_i formam o conteúdo deste sismograma. Neste exemplo está sendo considerado apenas um evento sísmico, ou seja, o modelo possui apenas uma única camada. As células que estão em cinza escuro contêm as amplitudes máximas de cada onda lida pelos receptores. O tempo de trânsito respectivo à leitura da onda em que a amplitude foi máxima será armazenado em um outro conjunto de dados.

Na Tabela 4 estão os tempos de trânsito selecionados após a extração de características do sismograma de exemplo da Tabela 3. O cálculo do erro no módulo 4 será realizado a partir do conjunto de dados sísmicos mostrados na Tabela 4.

		Receptor			
		rec_1	rec_2	rec_3	rec_4
Tempo de Trânsito	tt_1	0.06	0	0	0
	tt_2	0.08	0.05	0	0
	tt_3	0.07	0.07	0.04	0
	tt_4	0	0.04	0.06	0.04
	tt_5	0	0	0.04	0.05
	tt_6	0	0	0	0.03

Tabela 3 – Exemplo de um sismograma contendo as amplitudes lidas nos receptores em cada instante de tempo. No tempo tt_1 o receptor rec_1 está iniciando a leitura de uma onda e a amplitude lida neste tempo é 0.06. Nenhuma onda está retornando para os demais receptores no instante de tempo tt_1 . As células em cinza escuro contêm as amplitudes máximas de cada onda lida.

	Receptor			
	rec ₁	rec ₂	rec ₃	rec ₄
Camada 1	tt ₂	tt ₃	tt ₄	tt ₅

Tabela 4 – Conjunto de tempos de trânsito obtidos a partir de um sismograma por um processo de extração de características.

Para efetuar a extração de características do sismograma de referência, dois algoritmos foram testados: o Pickevs (SILVA, 2005) e o MaxAmp (SILVA, 2005) ambos desenvolvidos em Matlab®. O MaxAmp lê um sismograma e cria automaticamente outro conjunto de dados, selecionando apenas os tempos de trânsito em que a amplitude da onda capturada foi máxima. Para executar esse algoritmo é necessário informar quantos eventos sísmicos existem no sismograma.

O Pickevs permite que o especialista selecione no gráfico do sismograma os eventos sísmicos e a partir da seleção, é gerado um conjunto de tempos de trânsito em que a amplitude é máxima.

A vantagem do método MaxAmp em relação ao Pickevs é a não dependência de um especialista para selecionar os eventos. Entretanto, apesar do algoritmo Pickevs ser manual e exigir conhecimento técnico para selecionar os eventos, o especialista poderá selecionar apenas os eventos mais significativos no sismograma, ao contrário do MaxAmp, que por fazer a seleção automática poderá gerar mais eventos do que existem realmente. Além disso, caso as semihiperboles se cruzem, o algoritmo MaxAmp poderá selecionar os eventos de forma incorreta.

Neste trabalho, apesar de ambos os algoritmos resultarem em conjuntos de dados bem semelhantes, aplicou-se o Pickevs para ter maior controle na seleção dos eventos.

4.1.4.

Módulo 4 – Cálculo da Medida de Erro

O conjunto de tempos de trânsito gerados no segundo módulo são comparados através de uma medida de erro com os dados sísmicos do terceiro módulo, sismograma de referência. Existem vários métodos que permitem avaliar se a solução gerada no processo de otimização, neste caso o AG, é uma solução boa ou ruim.

Por este trabalho conter três estudos de caso com abordagens diferentes, empregou-se as métricas de acordo com a necessidade de cada problema. Os

estudos de caso 1 e 2 usaram a métrica MAPE para calcular o erro. Já no estudo de caso 3 foi utilizado o RMSE. Abaixo estão descritas as métricas aplicadas em cada estudo de caso.

Estudos de Caso 1 e 2

Nestes estudos de caso, a métrica MAPE (*Mean Absolute Percentage Error*) foi empregada para calcular o erro entre o conjunto de dados sísmicos gerado a partir de um indivíduo decodificado e o sismograma de referência. Esta medida indica o valor médio do erro percentual das previsões sobre todo o conjunto de teste, dado por,

$$MAPE = \frac{\sum_{k=1}^N \left| \frac{a_k - y_k}{a_k} \right|}{N} * 100\% \quad \text{Eq.(5)}$$

onde,

N = número de previsões realizadas.

a_k = saída desejada para a previsão índice 'k'.

y_k = saída prevista para a previsão índice 'k'.

O número de previsões realizadas, neste trabalho, é calculado multiplicando-se o número de receptores pelo número de eventos sísmicos – cada evento corresponde a uma camada de rocha – consideradas no sismograma de referência.

A Tabela 5 e a Tabela 6 exemplificam, respectivamente, um conjunto de dados sísmicos objetivo (saída desejada) e um conjunto de dados sísmicos gerados (saída prevista) a partir de uma solução do AG. O erro é calculado comparando-se as células de mesmos índices dessas tabelas. Para isto, ambos os conjuntos de dados sísmicos precisam ter mesmos tamanhos.

Por exemplo, o tempo de trânsito tt_{11} correspondente ao receptor rec_1 e à camada cmd_1 da Tabela 5 é comparado através da métrica MAPE com o tempo de trânsito tt'_{11} correspondente ao receptor rec_1 e à camada cmd_1 da Tabela 6 (ambas as células em cinza escuro). Os erros parciais de cada células serão somados e divididos pelo número total de previsões realizadas. Por fim, o valor obtido neste cálculo é multiplicado por 100, já que se trata de um erro percentual.

		Receptor					
		rec ₁	rec ₂	rec ₃	rec ₄	rec ₅	rec ₆
Camada de rocha	cmd ₁	tt ₁₁	tt ₁₂	tt ₁₃	tt ₁₄	tt ₁₅	tt ₁₆
	cmd ₂	tt ₂₁	tt ₂₂	tt ₂₃	tt ₂₄	tt ₂₅	tt ₂₆

Tabela 5 – Exemplo de um conjunto de dados sísmicos objetivo. A célula cinza escuro representa o tempo de trânsito tt_{11} , tempo de viagem da fonte ao receptor rec_1 , de uma onda refletida na camada cmd_1 .

		Receptor					
		rec ₁	rec ₂	rec ₃	rec ₄	rec ₅	rec ₆
Camada de rocha	cmd ₁	tt' ₁₁	tt' ₁₂	tt' ₁₃	Tt' ₁₄	tt' ₁₅	tt' ₁₆
	cmd ₂	tt' ₂₁	tt' ₂₂	tt' ₂₃	Tt' ₂₄	tt' ₂₅	tt' ₂₆

Tabela 6 – Exemplo de um conjunto de dados sísmicos gerados. Estes tempos de trânsito foram calculados a partir de um modelo da subsuperfície dado por um indivíduo decodificado do AG.

Após ser calculado o MAPE, este valor é associado ao *fitness* do indivíduo no AG. No processo de seleção do AG, o indivíduo será avaliado pelo seu valor de *fitness*.

Estudo de Caso 3

Neste estudo de caso, a métrica RMSE (*Root Mean Square Error*) foi empregada para calcular o erro entre sismograma gerado a partir de um indivíduo decodificado e o sismograma de referência. Esta métrica indica a raiz do erro quadrático médio, dado por,

$$RMSE = \sqrt{\frac{\sum_{k=1}^N (a_k - y_k)^2}{N}} \quad \text{Eq.(6)}$$

onde,

N = número de previsões realizadas.

a_k = saída desejada para a previsão índice 'k'.

y_k = saída prevista para a previsão índice 'k'.

O número de previsões realizadas, neste estudo de caso, é calculado multiplicando-se o número de linhas pelo número de colunas do sismograma de referência.

A Tabela 7 e a Tabela 8 exemplificam, respectivamente, um sismograma de referência (saída desejada) e um sismograma gerado (saída prevista) a partir de uma solução do AG. O erro é calculado comparando-se as células de mesmos índices dessas tabelas. Para isto, ambos os sismogramas precisam ter mesmas dimensões.

Por exemplo, o valor da amplitude lida a_{11} pelo receptor rec_1 no instante de tempo tt_1 da Tabela 7 é comparado através da métrica RMSE com o valor da amplitude lida a'_{11} pelo receptor rec_1 no instante de tempo tt_1 da Tabela 8 (ambas células em cinza escuro).

		Receptor			
		rec ₁	rec ₂	rec ₃	rec ₄
Tempo de Trânsito	tt ₁	a_{11}	a_{21}	a_{31}	a_{41}
	tt ₂	a_{21}	a_{22}	a_{23}	a_{24}
	tt ₃	a_{31}	a_{32}	a_{33}	a_{34}
	tt ₄	a_{41}	a_{42}	a_{43}	a_{44}
	tt ₅	a_{51}	a_{52}	a_{53}	a_{54}
	tt ₆	a_{61}	a_{62}	a_{63}	a_{64}

Tabela 7 – Sismograma de referência. As linhas da matriz indicam os índices dos instantes de tempo, já as colunas indicam os índices dos receptores. O conteúdo é formado pelos valores de amplitudes das ondas lidas.

		Receptor			
		rec ₁	rec ₂	rec ₃	rec ₄
Tempo de Trânsito	tt ₁	a'_{11}	a'_{21}	a'_{31}	a'_{41}
	tt ₂	a'_{21}	a'_{22}	a'_{23}	a'_{24}
	tt ₃	a'_{31}	a'_{32}	a'_{33}	a'_{34}
	tt ₄	a'_{41}	a'_{42}	a'_{43}	a'_{44}
	tt ₅	a'_{51}	a'_{52}	a'_{53}	a'_{54}
	tt ₆	a'_{61}	a'_{62}	a'_{63}	a'_{64}

Tabela 8 – Sismograma gerado a partir de um modelo da subsuperfície – meio definido por um indivíduo decodificado do AG.

Após ser calculado o RMSE, este valor é associado ao *fitness* do indivíduo no AG. No processo de seleção do AG, o indivíduo será avaliado pelo seu valor de *fitness*.

4.2. Algoritmo Genético Proposto

A parte mais importante do trabalho descrito nesta dissertação é a modelagem do AG. Este é responsável pela otimização dos parâmetros físicos, geométricos ou matemáticos das camadas rochosas que compõem a subsuperfície.

O processo evolucionário realizado pelo AG possui alguns componentes que devem ser desenvolvidos cuidadosamente para o seu correto funcionamento. Os componentes são: representação, operadores genéticos, decodificação e avaliação. Esses componentes foram descritos no capítulo 3 e sua modelagem neste trabalho é detalhada nesta seção.

4.2.1. Estrutura do Cromossomo

A definição da estrutura do cromossomo dos indivíduos é uma atividade essencial no desenvolvimento do Algoritmo Genético. O cromossomo deve conter os parâmetros que serão otimizados pela técnica. Uma vez definida a estrutura de um cromossomo, esta é replicada para os demais indivíduos da população, lembrando que cada indivíduo representará um modelo único da subsuperfície.

Em primeiro lugar, é preciso selecionar a representação do cromossomo. Dentre as diferentes representações, as do tipo binária e real se aplicam a este trabalho. Optou-se pela representação real em todos os estudo de casos, com exceção de parte do cromossomo do terceiro estudo de caso. Um dos motivos que influenciou nesta escolha, foi a redução do tempo computacional da etapa de decodificação, pois seria necessário transformar dados binários em dados reais para depois construir o modelo da subsuperfície.

No estudo de casos foi considerado um número pequeno de camadas, entretanto ao estender o problema atual para muitas camadas, ficaria cada vez mais complexa a representação binária.

Para calcular o número total de parâmetros que o AG irá otimizar, deve-se multiplicar o número de camadas que constam no modelo da subsuperfície pelo

número de propriedades do meio que se quer inverter a partir dos dados sísmicos.

O GACOM (Guimarães, 2004), *framework* utilizado para desenvolver o AG, tem como um recurso inovador a representação do cromossomo por segmentos. Esse conceito permite que sejam definidos para cada segmento, individualmente, os operadores genéticos e a forma de representação.

Neste trabalho, cada parâmetro a ser otimizado no AG será representado por um segmento. Os segmentos possuem tamanhos idênticos, definidos pelo número de camadas que constam no modelo da subsuperfície. Isto quer dizer que cada posição do segmento corresponde a um parâmetro de uma camada do modelo.

Por este trabalho conter três estudos de casos com abordagens diferentes, foi necessário definir uma estrutura de cromossomo para cada estudo. Abaixo estão descritos os cromossomos para cada estudo de caso.

Estudo de Caso 1

O objetivo desse estudo de caso é otimizar a velocidade intervalar (V_i) e a espessura (e) para um número conhecido de camadas.

A Figura 15 mostra a definição da estrutura do cromossomo para este caso. O segmento 1 contém as velocidades das camadas e o segmento 2 contém as espessuras das camadas. Cada posição dos segmentos corresponde a uma camada, por exemplo, para a camada 1 a velocidade intervalar no segmento 1 é V_{i1} e sua espessura no segmento 2 é e_1 .

CROMOSSOMO

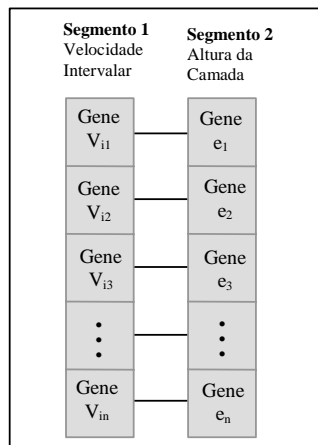


Figura 15 – Estrutura do cromossomo para o primeiro estudo de caso. Este cromossomo representa a velocidade intervalar e a espessura das camadas, parâmetros estes

separados em dois segmentos. Cada posição i dos segmentos corresponde a uma camada do modelo.

Para cada parâmetro representado no cromossomo é preciso definir um intervalo de valores válidos. As soluções possíveis serão geradas dentro destes intervalos.

A Tabela 9 mostra os valores mínimos e máximos para os parâmetros velocidade intervalar e espessura das camadas.

Parâmetro	Valor mínimo	Valor máximo
Velocidade intervalar	1,4 km/s	6,0 km/s
Espessura	0,1 km	3 km

Tabela 9 – Definição dos intervalos de valores mínimo e máximo para cada parâmetro a ser otimizado.

Estudo de Caso 2

O objetivo desse estudo de caso é otimizar a velocidade *root mean square* (V_{rms}) e o tempo para incidência normal (t_0) para um número conhecido de camadas. Estes atributos sísmicos estão descritos no capítulo 2.

A Figura 16 mostra a definição da estrutura do cromossomo para este caso. O segmento 1 contém as velocidades *root mean square* atribuídas para cada camada e o segmento 2 contém os tempos para incidência normal respectivos.

CROMOSSOMO

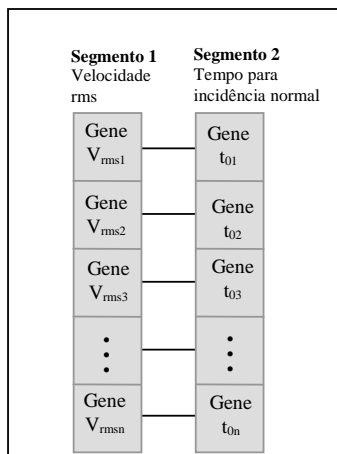


Figura 16 – Estrutura do cromossomo para o segundo estudo de caso. Este cromossomo representa a velocidade rms e o tempo para incidência normal das camadas, parâmetros estes separados em dois segmentos. Cada posição i dos segmentos corresponde a uma camada do modelo.

A Tabela 10 mostra os valores mínimos e máximos para os parâmetros velocidade rms e tempo para incidência normal das camadas.

Parâmetro	Valor mínimo	Valor máximo
Velocidade rms	1,4 km/s	6,0 km/s
Tempo para incidência normal	0,1 s	5 s

Tabela 10 – Definição dos intervalos de valores mínimo e máximo para cada parâmetro a ser otimizado.

Estudo de Caso 3

O objetivo desse estudo de caso é otimizar a velocidade intervalar (V_i), a espessura (e) e a densidade (ρ) para n camadas.

Este caso considera que se desconhece o número de camadas que estão associadas ao sismograma de referência. Isto quer dizer que o número de camadas também será evoluído neste problema.

Para isto, foi utilizado um recurso do GACOM que permite criar uma máscara de ativação. Essa máscara é representada por um segmento do tipo binário. Os demais segmentos do cromossomo possuem representação real. O uso da máscara de ativação neste estudo de caso está detalhado na seção 4.2.3 deste capítulo que explica o processo de decodificação.

A Figura 17 mostra a definição da estrutura do cromossomo para este caso. O segmento 1 contém a máscara de ativação, o segmento 2 contém as velocidades das camadas, o segmento 3 contém as densidades e o segmento 4 contém as espessuras respectivas.

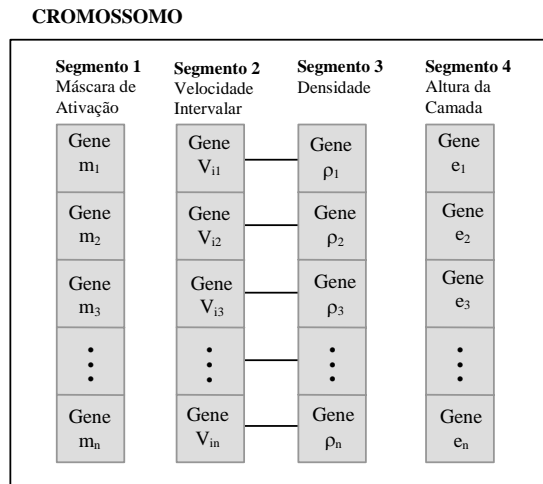


Figura 17 – Estrutura do cromossomo para o terceiro estudo de caso. Este cromossomo representa a velocidade intervalar, a densidade e a espessura das camadas, parâmetros estes separados em três segmentos. O segmento 1 deste cromossomo representa a máscara de ativação. Cada posição i dos segmentos corresponde a uma camada do modelo.

A Tabela 11 mostra os valores mínimos e máximos para os parâmetros velocidade intervalar, densidade e espessura das camadas.

Parâmetro	Valor mínimo	Valor máximo
Velocidade intervalar	1,4 km/s	6,0 km/s
Densidade	0,5 g/cm ³	4 g/cm ³
Espessura	0,1 km	3 km

Tabela 11 – Definição dos intervalos de valores mínimo e máximo para cada parâmetro a ser otimizado.

4.2.2. Operadores Genéticos

Os operadores genéticos utilizados nestes segmentos são operadores reais. No cruzamento dos cromossomos, cada segmento cruza com o mesmo segmento de outro cromossomo. Os operadores utilizados nesse trabalho foram O cruzamento aritmético (*arithmetical crossover*- AX) e a mutação real (*real mutation* - RM).

De acordo com Michalewicz (1996), o operador de cruzamento AX é uma combinação linear de dois vetores genitores. Esse cruzamento pode ser visto em (1), onde p_1 e p_2 são os genitores e f_1 e f_2 os descendentes.

$$\begin{aligned}f_1 &= a * p_1 + (1-a) * p_2 \\f_2 &= a * p_2 + (1-a) * p_1,\end{aligned}\tag{1}$$

onde,

$a = rand \in [0,1]$ (*crossover uniforme*).

Na mutação RM para cada gene do cromossomo é feito o teste de probabilidade, se for verdadeiro, o número real é substituído por um número real aleatório (Michalewicz, 1996). Esse número aleatório é gerado dentro do intervalo de valores pré-definido para cada gene.

4.2.3.

Decodificação

O processo de decodificação do indivíduo consiste basicamente na construção da solução real do problema a partir do cromossomo. Neste trabalho, cada indivíduo decodificado representa um modelo da subsuperfície.

A solução, após ser construída, é levada ao módulo 2 do sistema proposto, no qual serão gerados dados sísmicos sintéticos. Normalmente a avaliação em um AG é feita diretamente com o indivíduo decodificado, entretanto, para a avaliação neste trabalho são considerados os dados sísmicos sintéticos gerados a partir de um indivíduo decodificado.

Os processos de decodificação dos estudos de caso 1 e 2 são iguais, porém diferem do estudo de caso 3. Abaixo está descrita a tarefa de decodificação para cada estudo de caso.

Estudos de Caso 1 e 2

Em ambos os casos 1 e 2 o número de camadas do meio associado ao sismograma de referência é conhecido e será informado no início de todo o processo. Uma vez definido o número de camadas, todos os modelos da subsuperfície, gerados a partir do AG, terão a mesma quantidade de camadas.

É importante lembrar que os segmentos no cromossomo possuem tamanhos idênticos, pois existe um parâmetro em cada segmento para cada camada do modelo.

Considerando um modelo de n camadas em que se quer otimizar m parâmetros para cada camada, o algoritmo de decodificação pode ser definido pelo seguinte pseudo-código:

Para posição $i=1$ até n fazer:

Criar uma nova camada i no modelo

Para o segmento $j=1$ até m fazer:

Ler o valor da posição i no segmento j do cromossomo

Atribuir o valor lido ao parâmetro j da camada i no modelo

Fim – Para

Fim – Para

Estudo de Caso 3

No estudo de caso 3 o número de camadas do meio associado ao sismograma de referência é desconhecido. Este será mais um parâmetro que o AG terá que otimizar. Uma vez que não é conhecido o número de camadas, os modelos da subsuperfície, gerados a partir do AG, terão diferentes números de camadas.

É fixado um número n máximo de camadas que o meio poderá conter. O modelo terá no mínimo uma camada. O tamanho dos segmentos no cromossomo será igual a n .

A máscara de ativação pode ter valores '0' ou '1'. Se a posição i do segmento 1 for '0', então esta camada será descartada, porém se for '1' essa camada será incluída no modelo da subsuperfície que está sendo construído.

Considerando um modelo de n camadas em que se quer otimizar m parâmetros para cada camada, o algoritmo de decodificação pode ser definido pelo seguinte pseudo-código:

Para posição $i=1$ até n fazer:

Ler o valor da posição i no segmento 1 (máscara de ativação)

Se valor lido for igual a '1' fazer:

Para o segmento $j=2$ até m fazer:

Criar uma nova camada i no modelo

Ler o valor da posição i no segmento j do cromossomo

Atribuir o valor lido ao parâmetro j da camada i no modelo

Fim – Para

Fim – Se

Fim – Para

4.3. O Simulador

Para que o modelo proposto pudesse ser avaliado, foi desenvolvida uma ferramenta baseada na descrição do modelo detalhada no capítulo anterior.

As definições da representação do cromossomo, dos operadores genéticos, dos processos de decodificação, avaliação e seleção foram utilizadas no desenvolvimento do simulador.

A parte principal do simulador que contém o Algoritmo Genético foi implementada com o GACOM (*Genetic Algorithms Components*), um *Framework* para modelagem de Algoritmos Genéticos (Guimarães, 2004). O processo de seleção de tempos de trânsito, a geração de dados sísmicos pela equação de Dix e a função de avaliação são atividades desenvolvidas fora do simulador, utilizando a ferramenta Matlab®. O simulador é responsável por controlar a execução de cada uma dessas atividades, enviar os parâmetros necessários aos diversos módulos e tratar os dados de saída ao completar uma atividade.

Para iniciar o simulador, alguns parâmetros precisam ser informados. Estes dados estão divididos em três grupos: parâmetros do Algoritmo Genético, geometria de aquisição sísmica e sismograma de referência.

Os parâmetros de ajuste do AG desenvolvido são: as taxas de mutação e de cruzamento para os operadores genéticos e o intervalo de geração. Ao longo de cada estudo de caso esses parâmetros são ajustados. A combinação desses valores visa aumentar o desempenho do processo de otimização do AG. Também são parâmetros ajustáveis do AG o tamanho da população, o número de gerações e o número de experimentos. Estes são variados no início do estudo de caso e depois são fixados para os testes com as taxas de mutação e cruzamento.

É importante ressaltar a diferença entre os termos rodada e experimento, utilizados neste trabalho. Cada execução do simulador com uma mesma configuração do AG é considerada um experimento. Entretanto, quando um ou mais parâmetros do AG são modificados e o simulador é novamente executado isto é considerada uma nova rodada. Dado que o AG é estocástico, vários experimentos são realizados para que se possa avaliar seu desempenho médio com uma determinada configuração.

Os parâmetros da geometria de aquisição sísmica necessárias para configurar este simulador são: a quantidade de receptores, a distância entre o primeiro receptor e a fonte (*offset*), a distância entre si dos demais receptores, o

intervalo de amostragem e o tempo máximo de leitura. Estes parâmetros são fixos durante a simulação, sendo modificados somente quando outro sismograma de referência é selecionado para uma nova simulação.

O sismograma de referência selecionado deve estar armazenado em um arquivo que possa ser lido pelo Matlab[®].

Nas próximas seções estão detalhados os estudos de caso e os resultados obtidos com este simulador. Para os estudos de caso 1 e 2 foram realizados vários testes para ajustar o algoritmo genético de acordo com o problema. Já no estudo de caso 3, em que é utilizado um simulador para gerar dados sísmicos sintéticos, os resultados são ainda preliminares.