



## Critério da Validação Cruzada Generalizada na Seleção do Parâmetro Ótimo de Regularização em Tomografia Sísmica Interpoços

Naiane Pereira de Oliveira\*, IGEO/UFBA

Amin Bassrei, CPGG/IGEO/UFBA e INCT - Geofísica de Petróleo

Copyright 2008, SBGf - Sociedade Brasileira de Geofísica

*Este texto foi preparado para a apresentação no IV Simpósio Brasileiro de Geofísica, Belém, 14 a 17 de novembro de 2010. Seu conteúdo foi revisado pelo Comitê Técnico do IV SimBGf, mas não necessariamente representa a opinião da SBGf ou de seus associados. É proibida a reprodução total ou parcial deste material para propósitos comerciais sem prévia autorização da SBGf.*

### Abstract

In general the inverse problems in Exploration Geophysics are ill-posed, so that they need some kind of regularization. In this work we deal with acoustic traveltimes tomography, which is non-linear, so that we use Taylor's series in order to linearize the inverse procedure. Due to the ill-posedness we use the regularization by derivative matrices. This approach needs the so called regularization parameter lambda. There are several techniques for the selection of the optimum lambda and in this work we use the Generalized Cross Validation or GCV. After the numerical simulations in one set of synthetic data we conclude that GCV is a feasible and is also a robust technique for the selection of the optimum regularization parameter lambda.

### Introdução

A tomografia sísmica tem sido utilizada em Geofísica de Exploração com o propósito de fornecer imagens em subsuperfície de potenciais reservatórios de hidrocarbonetos. A tomografia sísmica tal que a tomografia médica é classificada como um método de inversão de dados. A tomografia consiste em uma técnica de reconstrução de imagens por meio do mapeamento e soma dos valores de propriedades físicas em determinadas direções. Em outras palavras, é um tipo especial de problema inverso que permite estimar uma função utilizando-se integrais de linha. Neste trabalho, utilizamos a tomografia de transmissão cujos dados de entrada são os tempos de trânsito medidos entre fontes e receptores e os parâmetros de modelo é o vetor de vagarosidades (recíprocos das velocidades) do meio bidimensional. A Inversão tomográfica é um problema mal-posto devido ao fato das condições de existência, unicidade e estabilidade não serem completamente satisfeitas. Desta forma, é requerido um tratamento matemático de modo a garantir soluções mais realistas. Nós utilizamos a regularização por matrizes de derivadas, mais conhecida na literatura como regularização de

Tikhonov. Esta abordagem solicita a seleção de um parâmetro de regularização  $\lambda$  que ofereça a melhor solução do procedimento inverso. Com o intuito de selecionar um valor ótimo desse parâmetro  $\lambda$ , empregamos o critério da Validação Cruzada Generalizada, mais conhecido como GCV, do inglês *Generalized Cross Validation*.

### Metodologia

Desenvolvemos uma metodologia de inversão tomográfica na geometria de aquisição poço a poço, ou interpoços, também chamada de XWP do inglês *Cross Well Profile*. Nessa configuração as fontes e receptores estão dispostos ao longo de poços. A tomografia de tempos de trânsito desconsidera a forma da onda e utiliza apenas o tempo de percurso entre fontes e receptores.

Esses tempos de trânsito formam o vetor dos parâmetros de dados observados  $\mathbf{d}^{obs}$  na inversão. A modelagem direta é obtida com a teoria do traçado de raios que permite modelar os meios geológicos com boa exatidão e eficiência computacional. O meio de interesse é discretizado em células, via de regra com dimensões uniformes. O traçado de raios fornece o comprimento de arco em cada célula. Com essa abordagem, fazemos uma análise para meios heterogêneos, ou seja, a velocidade da onda acústica varia de célula a célula, sendo porém constante dentro de cada célula. Como o meio é heterogêneo o traçado de raios fornecerá raios encurvados, que devem obedecer a Lei de Snell em cada interface entre duas células. Os comprimentos de arcos serão os coeficientes da matriz  $\mathbf{G}$ . Finalmente, os parâmetros do modelo a serem estimados estão contidos no vetor  $\mathbf{m}^{est}$ , e correspondem à vagarosidade estimada, que por seu turno é a saída do procedimento inverso. Como esse problema é não-linear, uma vez que o caminho de cada raio depende das velocidades (ou vagarosidades) do meio, há a necessidade de se linearizar o problema. Substituímos a função não-linear pela sua expansão em série de Taylor, truncando a expressão até o termo de primeira ordem. O processo foi utilizado de forma iterativa, sendo  $k$  a ordem da iteração conforme a expressão:

$$\Delta \mathbf{m}^{est,k} = \mathbf{G}^{k,+} \Delta \mathbf{d}^{obs,k}.$$

Aplicando o método da decomposição em valores singulares o sistema linear é resolvido em cada iteração. Devido ao mal-condicionamento da matriz tomográfica  $G$  utilizamos requisitos adicionais de suavidade da solução, ou a chamada regularização por matrizes de derivadas. A regularização tem um parâmetro de entrada de papel crucial chamado parâmetro ou fator de regularização,  $\lambda$ . Um bom parâmetro  $\lambda$  pode proporcionar um balanceamento ideal entre a minimização do erro nos dados e a suavização dos parâmetros do modelo na solução regularizada. A fim de selecionar um adequado parâmetro  $\lambda$  usamos uma técnica bastante popular chamada de Validação Cruzada Generalizada (GCV do inglês *Generalized Cross Validation*).

### SVD e Regularização

A inversa do operador matricial  $G$  é chamada de pseudo-inversa, ou inversa generalizada, ou ainda de inversa de Moore-Penrose, e denotada por  $G^+$ . Neste trabalho a pseudo-inversa é calculada em cada iteração em  $k$ . Para tanto, recorremos à técnica da decomposição por valores singulares onde a matriz  $G_{MN}$  é decomposta segundo a expressão  $G = U\Sigma V^T$  onde  $U_{MM}$  é a matriz que contém os autovalores ortonormalizados de  $GG^T$ ,  $V_{NN}$  é a matriz que contém os autovalores ortonormalizados de  $G^T G$  e  $\Sigma$  é a matriz diagonal que contém os valores singulares de  $G$  colocados em ordem decrescente. A inversa de Moore-Penrose é definida por uma matriz dada por:  $G^+ = V\Sigma^+ U^T$  onde  $\Sigma^+$  representa a matriz diagonal com os valores recíprocos dos valores singulares não-nulos de  $\Sigma$ . Com a introdução da regularização de Tikhonov termos referentes à suavidade do modelo ( $D_n \mathbf{m}$ ) são inseridos na função objetivo além do erro ( $\mathbf{e}$ ) (Bassrei & Rodi, 1993):

$$\varphi(\mathbf{m}) = \mathbf{e}^T \mathbf{e} + \lambda (D_n \mathbf{m})^T (D_n \mathbf{m}),$$

onde  $\lambda$  é o parâmetro de regularização e  $D_n$  é a matriz derivada de ordem  $n$ . Ao minimizar essa função objetivo obtêm-se:

$$\mathbf{m}^{est} = (G^T G + \lambda D_n^T D_n)^+ G^T \mathbf{d}^{obs}.$$

Para o processo linearizado a solução passa a ser

$$\Delta \mathbf{m}^{est,k} = (G^{T,k} G^k + \lambda D_n^T D_n)^{k,+} G^{T,k} \Delta \mathbf{d}^{obs,k},$$

onde temos para cada iteração a solução estimada é atualizada pela expressão:

$$\mathbf{m}^{est,k+1} = \mathbf{m}^{est,k} + \Delta \mathbf{m}^{est,k}.$$

Para  $\lambda = 0$ , a solução se reduz ao método dos mínimos quadrados. E para  $n=0$ ,  $D_n$  corresponde à matriz identidade o que corresponde a uma regularização de ordem zero, sendo que neste caso a solução se reduz ao método dos mínimos quadrados amortecidos. Para  $D_1$  a regularização é de primeira ordem (primeira derivada) e  $D_2$  temos a regularização de ordem dois (segunda derivada).

### O Critério GCV

A GCV sugere que um bom valor para o parâmetro de regularização deve prever dados do vetor de tempos de trânsito que estejam faltando ou que foram retirados. Mais precisamente, se um elemento do vetor de dados for retirado, então a solução regularizada correspondente deve prever bem essa falta. O critério GCV é baseado em considerações estatísticas, no princípio da validação cruzada que consiste em resolver o problema inverso regularizado omitindo-se um dado do vetor de dados por vez. Em cada realização é empregado um valor diferente e determinado do parâmetro de regularização  $\lambda$ . Se o  $\lambda$  usado for adequado espera-se que o vetor de dados calculados obtidos na modelagem direta para a  $k$ -ésima interação seja bem próximo do vetor de dados observados. Desta forma, o  $\lambda$  ótimo será aquele que minimiza a função (Wahba, 1990):

$$V(\lambda) = \sum_{k=1}^N \{d_k^{obs} - d_k[\mathbf{m}^k]\}^2.$$

Esta é a função de validação cruzada convencional. A mesma pode ser avaliada de um modo mais eficiente, não necessitando resolver explicitamente a problema inverso para cada observação omitida, utiliza-se a seguinte expressão (Wahba, 1990):

$$V(\lambda) = \sum_{k=1}^N \frac{[d_k^{obs} - d_k(\mathbf{m}_\lambda)]^2}{[1 - A_{kk}(\lambda)]^2},$$

onde  $\mathbf{m}_\lambda = G(G^T G + \lambda D_n^T D_n)^+ G^T \mathbf{d}^{obs}$  é a solução do problema inverso para um valor particular de  $\lambda$ . Por seu turno  $A_{kk}$  é o  $i$ -ésimo elemento da matriz diagonal da matriz  $A(\lambda) = G(G^T G + \lambda D_n^T D_n)^+ G^T$ .

A função de validação cruzada não é invariante sob uma transformação ortogonal do sistema linear original. Portanto, o valor de  $\lambda$  que minimiza a função  $V(\lambda)$  para o problema transformado não é o mesmo  $\lambda$  que minimiza a função  $V(\lambda)$  para o problema original, levando resultados diferentes para problemas equivalentes. Por essa razão tal função foi modificada obtendo a função de validação cruzada generalizada ou GCV, dada por (Wahba, 1990):

$$GCV(\lambda) = \frac{\|\mathbf{d}^{obs} - \mathbf{d}(\mathbf{m}_\lambda)\|^2}{\{\text{traço}[1 - A(\lambda)]\}^2}.$$

Na comunidade de geofísicos, o critério GCV já foi aplicado na inversão 1-D de dados eletromagnéticos por Farquharson & Oldenburg (2004) e em sismologia de exploração por Santos (2006) e Santos & Bassrei (2007), no caso uma aplicação em tomografia de difração.

## Resultados

Utilizamos um modelo sintético que representa uma situação geológica típica (Figura 1) para analisar a metodologia empregada de tomografia linearizada. Foram dispostos 30 fontes num poço e 30 receptores no outro poço, de modo a se ter 900 raios ou 900 equações. O meio geológico foi discretizado em 800 blocos quadrados com vagarosidade (recíproco da velocidade) constante em cada bloco, sendo 20 blocos na horizontal e 40 blocos na vertical. Por conseguinte o problema é dito sobredeterminado, isto é, existem mais equações que incógnitas. Conforme se pode ver na Figura 1 existem cinco camadas com velocidades entre 2000 e 3000 *m/s*. São três camadas que representam arenitos com velocidades acústicas de 2800, 2900 e 3000 *m/s*., assim como duas camadas que representam folhelhos com velocidades de 2300 e 2200 *m/s*. Os diferentes valores de velocidade para as rochas estão associados às diferenças nas suas propriedades petrofísicas, tais como a porosidade, densidade e grau de saturação em óleo. Existe ainda uma falha que juntamente com os cinco níveis estratigráficos, constituem uma trapa mista para acumulação de hidrocarboneto (zona de baixa velocidade 2000 *m/s*). O imageamento adequado de estruturas geológicas, tais como falhas, é de interesse para a indústria do petróleo, uma vez que tais estruturas podem atuar como reservatórios de hidrocarbonetos.

Empregamos três ordens de regularização: zero, um e dois e o parâmetro ótimo selecionado deve está associado ao mínimo de cada curva da função GCV. O método usado é o de iteração simples de forma que o GCV é aplicado em cada iteração, por significar cada iteração um problema linear particular, e por sua vez um  $\lambda$  ótimo é escolhido na iteração respectiva. Geramos curvas da função GCV em relação ao parâmetro  $\lambda$  como variável independente para cada exemplo, sendo uma curva para cada ordem de regularização. Na primeira iteração partiu-se de um modelo inicial de velocidade acústica constante igual a 3000 *m/s* equivalendo ao traçado de raios retos.

Para a regularização de ordem zero, a Figura 2 mostra a curva GCV para as 4 iterações realizadas mostrando um  $\lambda$  ótimo associado ao mínimo em cada iteração. As Figuras 3 e 4 mostram a recuperação tomográfica equivalente. O critério de parada foi considerado quando

a variação nos valores de velocidades torna-se mínima conforme se verifica entre as diferentes iterações.

No caso da regularização de ordem um, a Figura 5 mostra a curva GCV para as 4 iterações realizadas mostrando um  $\lambda$  ótimo associado ao mínimo em cada iteração. As Figuras 6 e 7 mostram a recuperação tomográfica equivalente. Tal qual o caso anterior verificou-se que 4 iterações são suficientes.

Finalmente, para a regularização de ordem dois, a Figura 8 mostra a curva GCV para as 4 iterações realizadas mostrando um  $\lambda$  ótimo associado ao mínimo em cada iteração. As Figuras 9 e 10 mostram a recuperação tomográfica equivalente.

Note-se que para as três ordens a imagem recuperada na quarta iteração mostra boa coerência com o modelo verdadeiro.

## Conclusões

Neste trabalho, analisou-se a aplicabilidade do critério GCV para a escolha do fator ótimo de regularização aplicada em tomografia linearizada de tempos de trânsito com um modelo sintético. Os modelos recuperados apresentam boa concordância com os respectivos modelos verdadeiros, permitindo validar a técnica GCV. As simulações permitem concluir que apesar do critério GCV apresentar baixo custo computacional, para se adquirir bons resultados é requerida a qualidade dos dados de entrada, além da boa cobertura, isto é, uma boa iluminação do meio pela passagem dos raios pelo mesmo.

## Agradecimentos

Naiane Pereira de Oliveira agradece ao PIBIC/UFBA pelas bolsas de IC do CNPq e da FAPESB. Também agradece ao ANP/PRH-08 pela atual bolsa de graduação. Amin Bassrei agradece ao CNPq pela bolsa de produtividade de pesquisa. Os autores agradecem à FINEP pelo apoio à Rede CT-PETRO de Geofísica de Exploração – Rede 01 – Fase 3, e ao LAGEP pela utilização dos recursos computacionais.

## Referências

- Bassrei, A. & Rodi, W.L., 1993. Regularization and inversion of linear geophysical data. In: Proceedings of the Third International Congress of the Brazilian Geophysical Society, Rio de Janeiro, Brazil, vol. 1, pp. 111-116.
- Farquharson, C.G. & Oldenburg, D. W., 2004. A comparison of automatic techniques for estimating the regularization parameter in non-linear problems. *Geophysical Journal International*.

Santos, E. T. F., 2006. Inversão tomográfica sísmica anisotrópica com regularização ótima. Tese de Doutorado, Universidade Federal da Bahia, Salvador, Brasil.

Santos, E. T. F. & Bassrei, A., 2007. Application of GCV in geophysical diffraction tomography. 69th European Association of Geoscientists and Engineers Conference & Exhibition, Londres, Reino Unido, 11 a 14 de julho de 2007, CD-ROM, 5 pp.

Wahba, G., 1990. Spline Models for Observational Data, SIAM, Filadelfia.

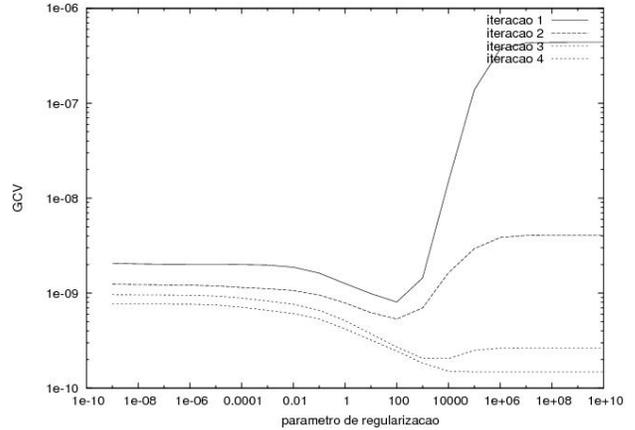


Figura 2 – Curva GCV em função do parâmetro de regularização, para regularização de ordem 0, para diferentes iterações de traçado de raio .

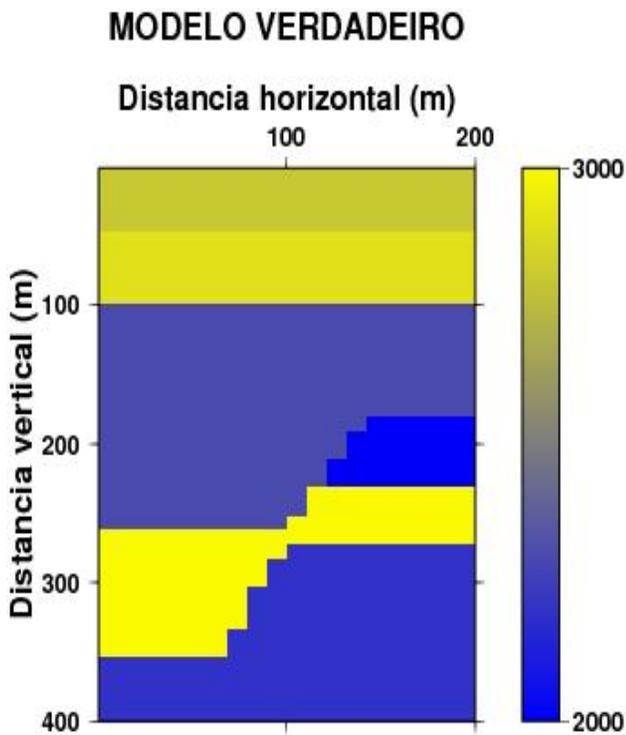


Figura 1 – Modelo geológico verdadeiro com velocidades em m/s.

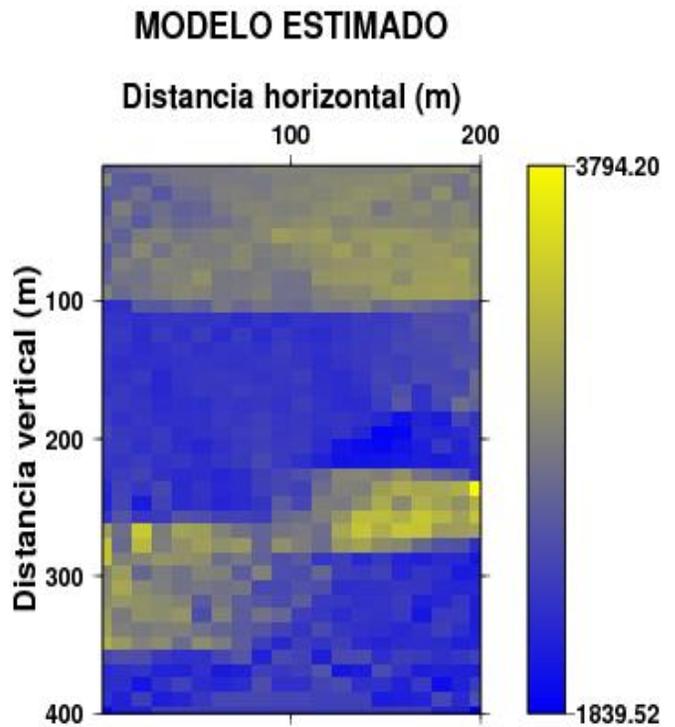


Figura 3 – Modelo estimado na primeira iteração para regularização de ordem 0. Velocidades em m/s.

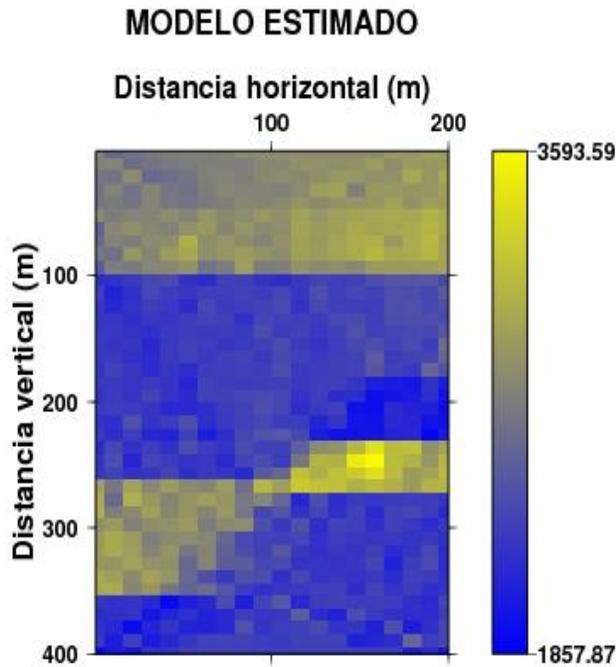


Figura 4 – Modelo estimado na quarta iteração para regularização de ordem 0. Velocidades em m/s.

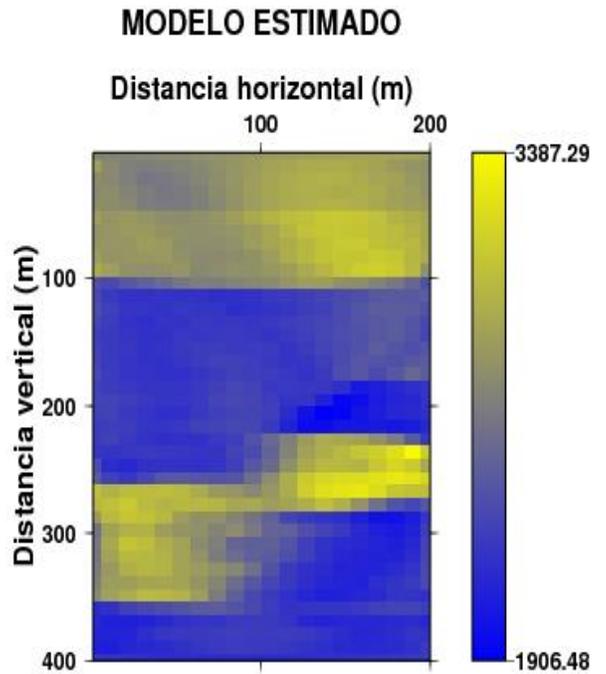


Figura 6 – Modelo estimado na primeira iteração para regularização de ordem 1. Velocidades em Ms/sm/s.

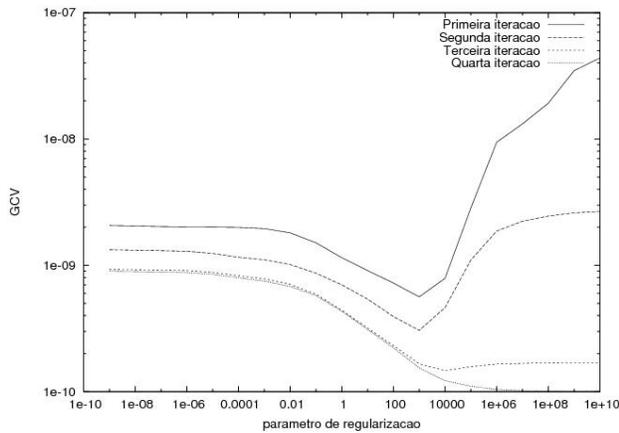


Figura 5 – Curva GCV em função do parâmetro de regularização, para regularização de ordem 1, para diferentes iterações de traçado de raio .

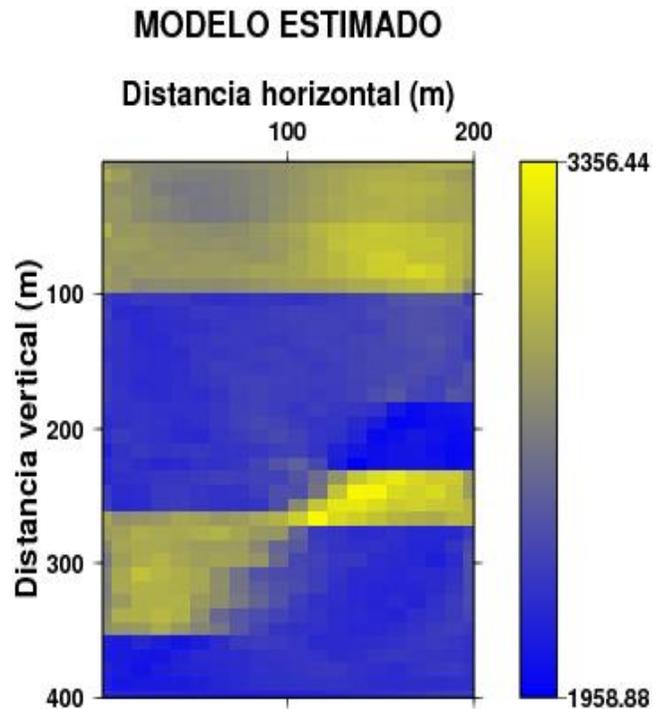


Figura 7 – Modelo Estimado na quarta iteração para regularização de ordem 1. Velocidades em m/s.

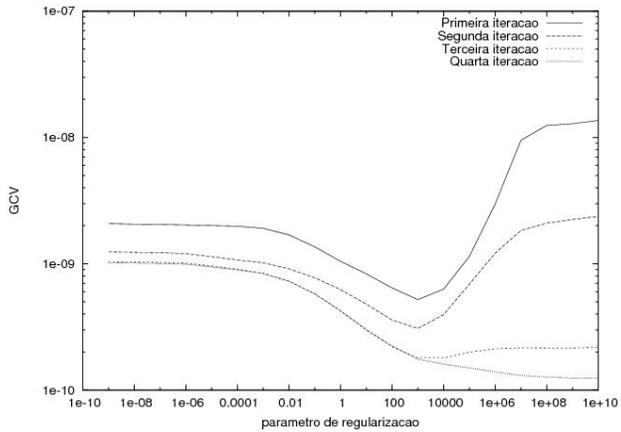


Figura 8 – Curva GCV em função do parâmetro de regularização, para regularização de ordem 2, para diferentes iterações de traçado de raio .

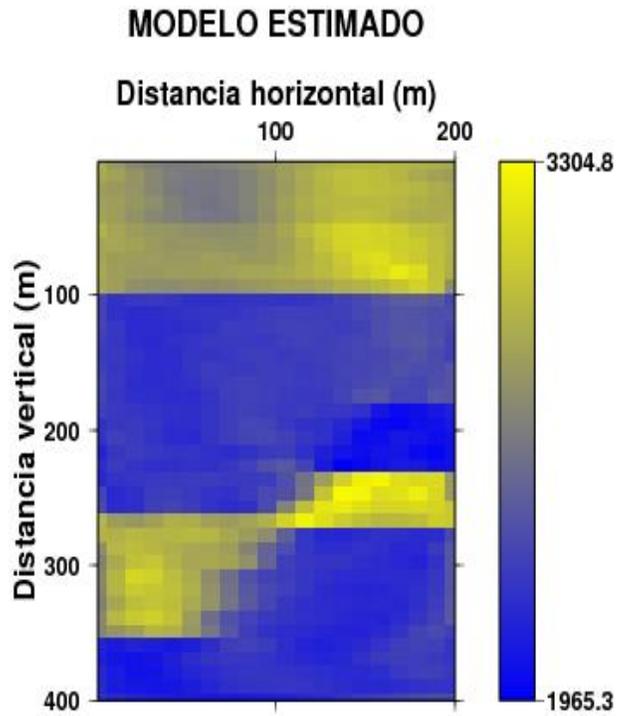


Figura 10 – Modelo Estimado na quarta iteração para regularização de ordem 2. Velocidades em m/s.

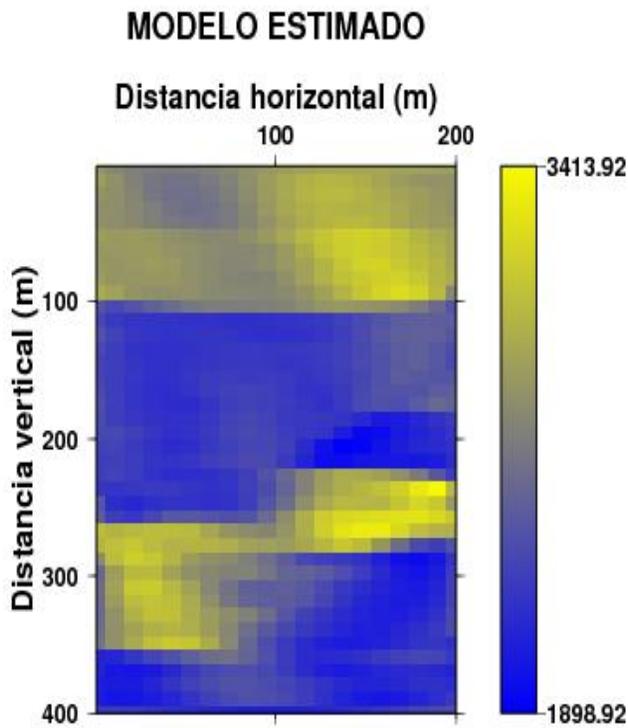


Figura 9 – Modelo estimado na primeira iteração para regularização de ordem 2. Velocidades em m/s.