



Crítério da Validação Cruzada Generalizada na Seleção do Parâmetro Ótimo de Regularização na Inversão Paralelizada de Dados Gravimétricos 3D de Bacias Sedimentares com Densidade Variável

Oscar F. Mojica, IGEO/UFBA; Amin Bassrei, CPGG/IGEO/UFBA & INCT-GP

Copyright 2014, SBGf - Sociedade Brasileira de Geofísica.

Este texto foi preparado para a apresentação no VI Simpósio Brasileiro de Geofísica, Porto Alegre, 14 a 16 de outubro de 2014. Seu conteúdo foi revisado pelo Comitê Técnico do VI 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.

Resumo

This paper addresses the problem of gravity inversion to determinate the 3D basement relief of sedimentary basins. It is a nonlinear optimization problem where the gravity anomalies attributable to basement interfaces above which the density contrast varies continuously with the depth are analyzed. We use the Levenberg-Marquardt (LM) algorithm which needs an key input parameter called the regularization parameter λ to obtain the linearized solution to the corresponding nonlinear problem at each iteration. The generalized cross-validation (GCV) critério for the choice of the regularization parameter was adopted in this work. We present simulation results with a synthetic model of a complex sedimentary basin taking advantage of a parallel LM algorithm implemented using MPI (Message Passing Interface). The applicability and efficacy of the GCV technique for the selection of the optimum regularization parameter is demonstrated.

Introdução

Uma das aplicações importantes do método gravimétrico é na estimativa da profundidade do contato sedimento - embasamento numa dada bacia sedimentar, uma vez que o entendimento da estrutura do embasamento da bacia desempenha um papel significativo na compreensão do sistema petrolífero. Em geral os métodos existentes para a estimação do relevo do embasamento de uma bacia sedimentar podem ser agrupados em duas categorias para as quais casos bidimensionais e tridimensionais podem ser reconhecidos. Um primeiro grupo considera que o contraste de densidade entre sedimentos e embasamento é constante. O segundo grupo assume uma variação com a profundidade do contraste de densidade devida à compactação dos sedimentos. Dentro deste último Chakravarthi & Sundararajan (2007) desenvolveram um esquema de inversão baseado no algoritmo de Marquardt para estimar simultaneamente a anomalia gravimétrica regional e o relevo 3D do embasamento de bacias sedimentares nas quais o contraste de densidade decai com a profundidade segundo uma lei parabólica conhecida. Neste trabalho é usada uma adaptação desta abordagem introduzindo o critério da validação cruzada generalizada GCV para a seleção do parâmetro ótimo de regularização λ que ofereça a melhor solução do

procedimento inverso a cada iteração. A fim de minimizar o tempo de computo requerido na solução deste problema de otimização não linear uma implementação em paralelo do esquema de inversão, através da interface MPI é considerada. Os testes foram realizados no cluster Aguiá, do CPGG-UFBA, utilizando até 16 nós idênticos. Cada nó consiste em um processador Intel Xeon E5260 2.40GHz, com cache de 12Mb.

Modelagem

Admitindo que o contraste de densidade entre o embasamento e o pacote sedimentar da bacia varia parabólicamente com a profundidade z (Chakravarthi et al. 2002)

$$\Delta\rho(z) = \frac{\Delta\rho_0^3}{(\Delta\rho_0 - \alpha z)^2}, \quad (1)$$

sendo $\Delta\rho_0$ o contraste de densidade na superfície da Terra em g/cm^3 e α uma constante que controla o gradiente do contraste de densidade com a profundidade, expressa em $g.cm^{-3}/km$.

A bacia é representada por uma série de prismas retangulares de dimensões horizontais conhecidas dx e dy com seus topos coincidindo com a superfície da Terra e as suas bases com a interface do embasamento. O campo gravimétrico observado é reduzido ao plano horizontal e é assumido que as medidas estão disponíveis nos nós de uma malha regular com coordenadas x e y correspondendo com as coordenadas horizontais dos centros dos prismas (Figura 1)

Sendo $\mathbf{g} = (g_1, \dots, g_M)$ um vetor M -dimensional contendo as observações da componente vertical da anomalia gravimétrica presumivelmente produzidas pelo relevo do embasamento da bacia sedimentar, os parâmetros a serem estimados, vetor \mathbf{z} contendo as profundidades do embasamento em N pontos discretos da malha ($N = M$) estão relacionados aos dados gravimétricos através da relação não linear dada pela equação:

$$g_i(x_i, y_i, z_i) = \sum_{j=1}^N f_i(z_j, \Delta\rho_0, \alpha), \quad i = 1, \dots, M. \quad (2)$$

A componente vertical do campo gravimétrico $f_i = g_{prisma}$ no i -ésimo ponto (x_i, y_i, z_i) produzida por um único prisma é expressada analiticamente como:

$$g_{prisma} = \gamma \int_{z_1}^{z_2} \int_{y_{o_j-dy/2}}^{y_{o_j+dy/2}} \int_{x_{o_j-dx/2}}^{x_{o_j+dx/2}} \frac{\Delta\rho_0^3}{(\Delta\rho_0 - \alpha z')^2}$$

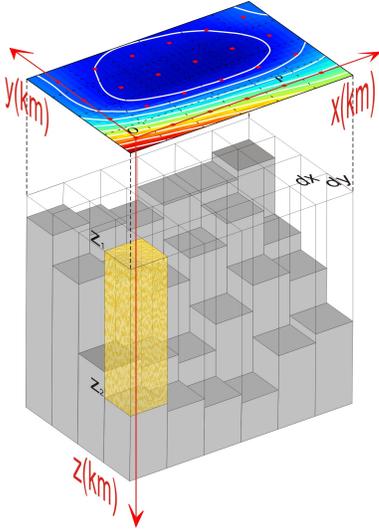


Figura 1: Representação esquemática do embasamento 3D de uma bacia sedimentar simples. A bacia é representada por um conjunto de prismas verticais 3D cujas bases são os parâmetros a ser estimados. A anomalia gravimétrica atribuída à bacia é amostrada em cada nó de uma malha regular (pontos vermelhos).

$$\times \frac{z_i - z'_j}{[(x_i - x'_j)^2 + (y_i - y'_j)^2 + (z_i - z'_j)^2]^{3/2}} dx' dy' dz', \quad (3)$$

sendo γ a constante gravitacional de Newton e x_{oj} e y_{oj} as coordenadas do centro do j -ésimo prisma em relação a x e y . Chavakarthy et al. (2002) apresentaram uma expressão fechada para a integral da equação (3) usada neste trabalho. Pelo princípio de superposição, a anomalia de gravidade da bacia em qualquer ponto da malha pode ser aproximada pelo somatório dos efeitos de todos os prismas.

Inversão linearizada via Levenberg-Marquardt (LM)

O algoritmo utilizado no método de inversão linearizada inicia-se com a modelagem do campo gravitacional, que é o dado observado \mathbf{d}_{obs} . Partindo da premissa que os valores de $\Delta\rho_0$, α e a base do embasamento são conhecidos. A inversão começa com um chute inicial dos valores da profundidade, e assim, o campo é calculado. Os passos envolvidos no método de inversão linearizada usando o algoritmo de Levenberg-Marquardt são os seguintes:

1. Inicialização dos parâmetros usados na lei parabólica ($\Delta\rho_0$, α) assumidos conhecidos e o valores do critério de parada ε e tol .
2. Definição da profundidades iniciais do embasamento \mathbf{m}_0 e calculo do campo \mathbf{d}_{calc} . Estimativa do erro rms entre os parâmetros dos dados e_d e o resíduo entre os parâmetros dos modelos $\delta\mathbf{g}$.
3. Calculo da matriz jacobiana J usando diferenças finitas. Calculo da matriz $J^T J$.
4. Implementação da inversão propriamente dita . Resolvendo $\delta\mathbf{m} = (J^T J + \lambda I)^+ J^T \delta\mathbf{g}$ usando SVD na obtenção da pseudo-inversa e o critério GCV para a seleção do λ .

5. Atualização do modelo $\mathbf{m}_k = \mathbf{m}_0 + \delta\mathbf{m}$ e determinação do erro rms (e_d) usando \mathbf{m}_k . Se e_d ou sua variação relativa são menores que os valores predefinidos ε e tol o algoritmo finaliza e \mathbf{m}_k é o modelo estimado, se não, faz \mathbf{m}_0 receber o novo modelo \mathbf{m}_k e reinicializa, até que algum dos critérios de parada seja alcançado.

Critério GCV

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 λ . Se o λ usado for adequado espera-se que o vetor de dados calculados obtidos na modelagem direta para a k -ésima iteração seja bem próximo do vetor de dados observados. Desta forma, o λ ótimo será aquele que minimiza a função (Wahba, 1990):

$$CV(\lambda) = \sum_{k=1}^N \{\mathbf{d}_k^{obs} - \mathbf{d}_k[\mathbf{m}^k]\}^2, \quad (4)$$

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

$$CV(\lambda) = \frac{\sum_{k=1}^N \{\mathbf{d}_k^{obs} - \mathbf{d}_k(\mathbf{m}_\lambda)\}^2}{[1 - A_{kk}(\lambda)]^2}, \quad (5)$$

onde $\mathbf{m}_\lambda = (J^T J + \lambda D_n^T D_n)^+ J^T \mathbf{d}_{obs}$ é a solução do problema inverso para um valor particular de λ e A_{kk} é o i -ésimo elemento da matriz diagonal $A(\lambda) = J(J^T J + \lambda D_n^T D_n)^+ J^T$. A função de validação cruzada não é invariante sob uma transformação ortogonal do sistema linear original. Portanto, o valor de λ que minimiza a função $CV(\lambda)$ para o problema transformado não é o mesmo λ que minimiza a função $CV(\lambda)$ 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):

$$GCV(\lambda) = \frac{\|\mathbf{d}^{obs} - \mathbf{d}(\mathbf{m}_\lambda)\|^2}{\{Tr[1 - A(\lambda)]\}^2}, \quad (6)$$

O critério GCV tem sido aplicado recentemente na solução de alguns problemas inversos não lineares em geofísica. Farquharson & Oldenburg (2004) o aplicaram na inversão 1-D de dados eletromagnéticos, Santos & Bassrei (2007) e Oliveira & Bassrei (2010) em sismologia de exploração, no caso com aplicações em tomografia de difração e tomografia sísmica interpoços.

Paralelização do algoritmo LM

Dois níveis de paralelização foram introduzidos no algoritmo LM usado em nossa aplicação em ordem de reduzir o tempo de computo. Um programa em paralelo consiste em processadores autônomos executando seu próprio código, sendo os códigos executados por cada

processador não necessariamente iguais. Um destes processadores monitora a execução global do programa e é designado como mestre. Os demais processadores são designados como escravos.

1. Paralelização do cálculo da matriz jacobiana. O cálculo da matriz jacobiana tem um grande número de computações independentes que podem ser executadas em paralelo, em nosso caso usamos uma distribuição da matriz jacobiana em blocos de filas onde cada processador é encarregado do cálculo de um bloco. A construção da matriz jacobiana é feita no processador mestre adicionando as componentes correspondentes da matriz (blocos) computadas em cada processador escravo.
2. Paralelização da escolha do parâmetro de regularização ótimo, utilizando GCV. O critério GCV é computacionalmente caro devido à inversão de uma matriz para cada λ via SVD como parte da rotina de escolha do parâmetro ótimo. Em nossa implementação um intervalo para este parâmetro é escolhido e cada processador executa uma inversão usando um determinado valor de λ . Com P processadores disponíveis P inversões usando os primeiros P valores de λ são efetuadas em paralelo, com o primeiro processador usando o primeiro valor de λ , o segundo usando o segundo valor e assim por diante. Cada processador calcula a função $GCV(\lambda)$ para o seu dado λ e atualiza o modelo. Operações de redução são aplicadas sobre o conjunto de processadores e o resultado é disponibilizado no processador mestre, que determina o parâmetro ótimo e o novo modelo.

Diferentes números de nós foram considerados na execução do algoritmo. A Figura 2 mostra o ganho computacional obtido por o uso de vários nós. É possível perceber uma queda progressiva no tempo total de execução quando o número de nós é aumentado.

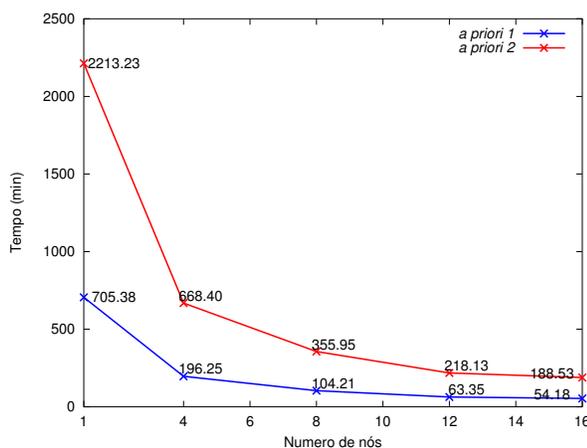


Figura 2: Tempo de execução do algoritmo com número variável de nós

Resultados

Utilizamos um modelo sintético de uma bacia sedimentar de complexo arcabouço estrutural (Figura 3) para analisar

a metodologia empregada na seleção do parâmetro de regularização. A bacia é constituída por duas sub-bacias com direções norte-sul e noroeste-sudoeste interconectadas. A sub-bacia norte apresenta ma geometria romboédrica alongada em direção norte e é limitada por rampas íngremes correspondentes a falhamentos, enquanto que a sub-bacia sul é uma típica bacia rifte conhecida como meio-graben com sua borda leste definida por uma falha de direção noroeste-sudeste.

A bacia simulada é coberta por uma malha retangular de 28×50 prismas nas direções leste e norte. As dimensões dos prismas escolhidas foram de 1.5 km em ambas direções com os seus topos coincidindo com o plano $z=0$ km. A anomalia de gravidade do modelo foi calculada em 1400 estações presumindo uma variação parabólica de densidade definida por os parâmetros $\Delta\rho_0 = -0.67 \text{ g/cm}^3$ e $\alpha = 0.026 \text{ g.cm}^{-3}/\text{km}$. Adotamos um modelo a priori para o qual a profundidade inicial é definida por uma fórmula teórica (Chavakarhi, 2007) que assume que o valor de gravidade medido em cada nó da malha é dado por uma placa horizontal 2D infinita com contraste de densidade variando parabólicamente (primeiro a priori). Também consideramos um modelo a priori com profundidade inicial constante igual a 2.5 km (segundo a priori). Todas as simulações foram realizadas usando dados contaminados com ruído. Basicamente foi adicionado ruído gaussiano pseudo-aleatório com média nula e desvio padrão de 0,134 mGal (Figura 4).

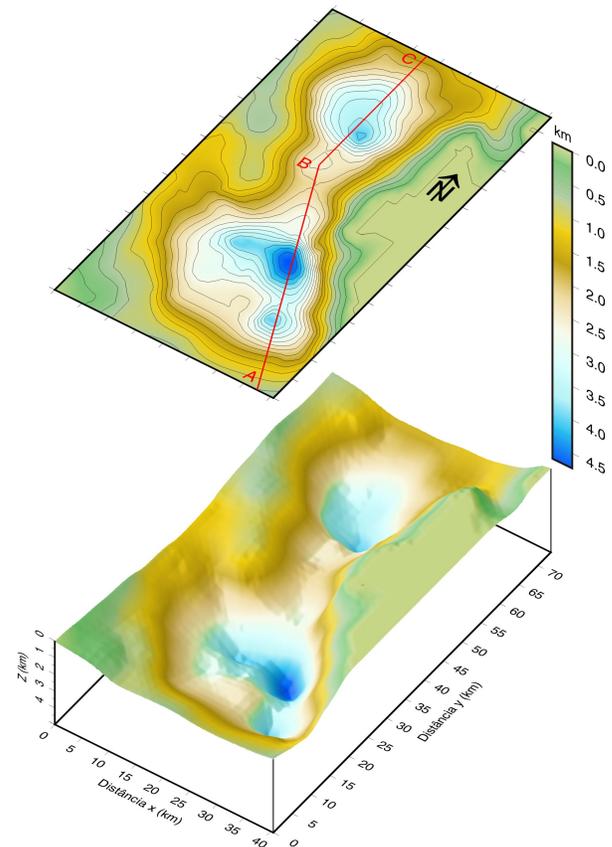


Figura 3: Mapa de contorno e vista em perspectiva do relevo verdadeiro do embasamento cuja profundidade máxima atinge 4,72 km.

A Figura 5 mostra o mapa de contorno e a vista em perspectiva das estimativas das profundidades do relevo do embasamento da bacia simulada na Figura 3 quando os dados foram contaminados com ruído pseudo-aleatório e o primeiro *a priori* foi usado. Esta estimativa foi obtida após só cinco iterações usando o critério GCV.

A Figura 6 apresenta anomalia gravimétrica produzida por o relevo do embasamento estimado mostrado na Figura 5 junto à anomalia gravimétrica observada corrompida por ruído. A Figura 7 mostra os relevos verdadeiro e estimado ao longo do perfil A-B-C usando o primeiro *a priori*. A localização do perfil pode ser observada nas Figuras 3 e 5. A Figura 8 mostra as curvas GCV para a inversão realizada usando o primeiro *a priori*. Pode ser observado que o λ ótimo em cada iteração esta associado ao mínimo da curva. Igualmente a Figura 9 mostra o mapa de contorno e a vista em perspectiva das profundidades estimadas pela inversão para o relevo do embasamento simulado quando os dados foram contaminados com ruído pseudo-aleatório e foi usado o segundo *a priori*. Esta estimativa foi obtida após dezessete iterações usando o critério GCV. Neste caso os resultados obtidos foram bastante semelhantes aos alcançados usando o primeiro *a priori*, mostrando o bom desempenho do critério GCV na escolha do parâmetro de regularização ótimo.

A Figura 10 apresenta anomalia gravimétrica produzida por o relevo do embasamento estimado mostrado na Figura 9 junto à anomalia gravimétrica observada corrompida por ruído. A Figura 11 mostra os relevos verdadeiro e estimado usando o segundo *a priori*, ao longo do perfil A-B-C. A localização do perfil pode ser observada nas Figuras 3 e 9.

A Figura 12 mostra as curvas GCV para as dezessete iterações realizadas usando o segundo *a priori*, o λ ótimo é associado ao mínimo em cada iteração.

Note-se que para ambas simulações se obtiveram resultados consistentes na estrutura estimada da bacia, evidenciando-se o bom desempenho do critério GCV na escolha do parâmetro de otimização.

Conclusões

Neste trabalho, é apresentado o critério GCV para a escolha do fator ótimo de regularização na inversão gravimétrica 3D de bacias sedimentares com densidade variável. O critério foi aplicado em um processo iterativo para o qual a cada iteração é estimado um valor para o parâmetro de regularização. Valores adequados para o parâmetro de regularização foram determinados com sucesso para um modelo sintético de características estruturais relevantes. Os modelos recuperados usando diferentes *a priori* apresentaram boa concordância com o respectivo modelo verdadeiro, permitindo validar a técnica GCV. Os dois níveis de paralelismo implementados em nosso algoritmo - paralelização do cálculo da matriz Jacobiana e paralelização da busca do parâmetro de regularização ótimo, reduziram significativamente o alto custo computacional inerente à dimensão do problema. Os resultados obtidos na implementação mostram a eficácia e eficiência do algoritmo proposto e seu uso pode ser estendido à inversão de dados reais.

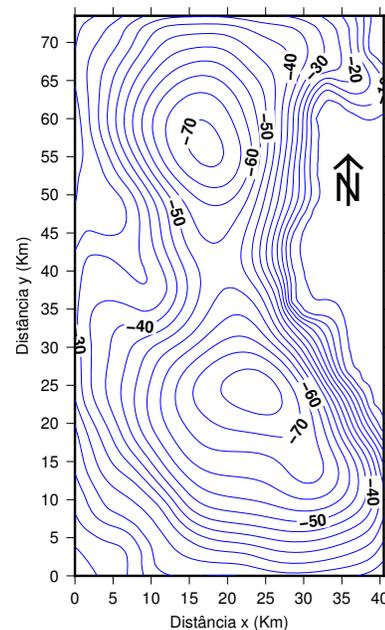


Figura 4: Anomalia gravimétrica contaminada com ruído devida à bacia sedimentar simulada na Figura 3.

Referências

- Chakravarthi, V.; Raghuram, H. M. e Singh, S. B. (2002) Forward gravity modeling of density interfaces above which the density contrast varies continuously with depth. *Computers & Geosciences*, v. 28, pp. 53-57.
- Chakravarthi, V. e Sundararajan, N. (2007) 3D gravity inversion of basement relief - A depth-dependent density approach. *Geophysics*, v. 72, n. 2, pp. 123-132.
- Farquharson, C.G. e 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. e 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.
- Oliveira, N. P. e Bassrei, A. (2010) Critério da Validação Cruzada Generalizada na Seleção do Parâmetro Ótimo de Regularização em Tomografia Sísmica Interpoços. VI Simpósio Brasileiro de Geofísica.
- Wahba, G (1990) *Spline Models for Observational Data*, SIAM, Filadélfia.
- Agradecimentos**
- Os autores agradecem à FINEP pelo apoio à Rede Cooperativa de Pesquisa em Geofísica de Exploração (Rede 01-Fases 3 e 5) e ao CNPq e PETROBRAS pelo apoio ao Instituto Nacional de Ciência e Tecnologia de Geofísica de Petróleo (INCT-GP). O. F. Mojica agradece a CAPES pela bolsa de doutorado. A. Bassrei agradece ao CNPq pelos projetos 307.427/2010-2 e 308.690/2013-3 (bolsa de produtividade de pesquisa).

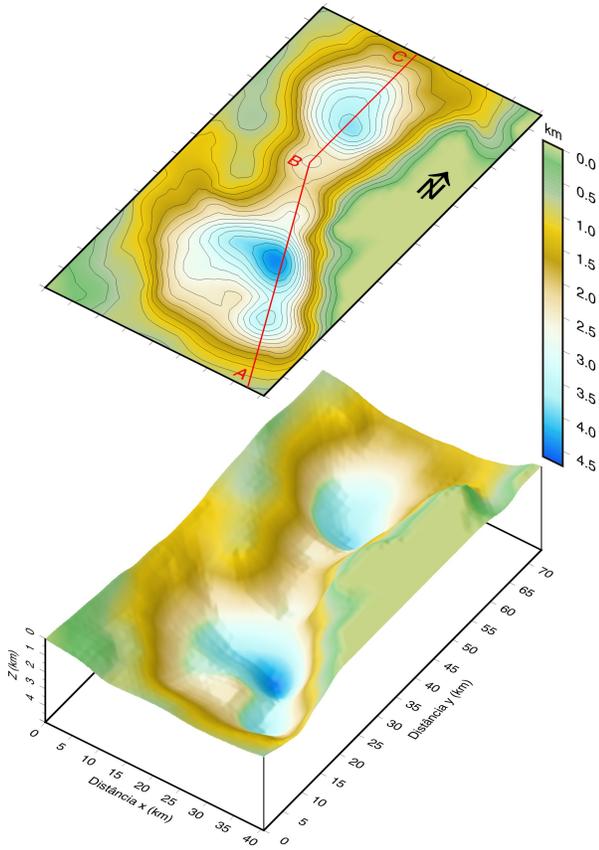


Figura 5: Mapa de contorno e vista em perspectiva do relevo estimado do embasamento usando o primeiro *a priori*.

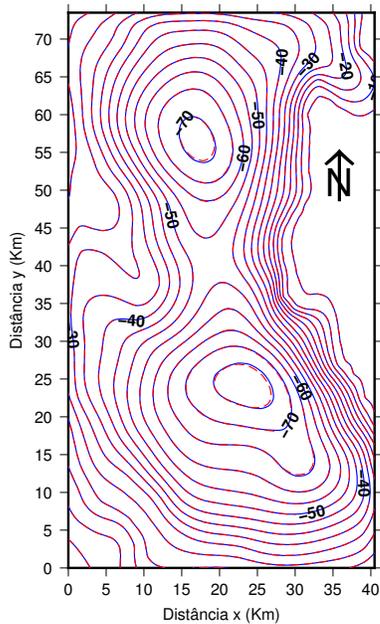


Figura 6: Anomalia gravimétrica contaminada com ruído (linhas contínuas azuis) devida à bacia sedimentar simulada na Figura 3 e a anomalia calculada (linhas tracejadas vermelhas) pela solução da Figura 5.

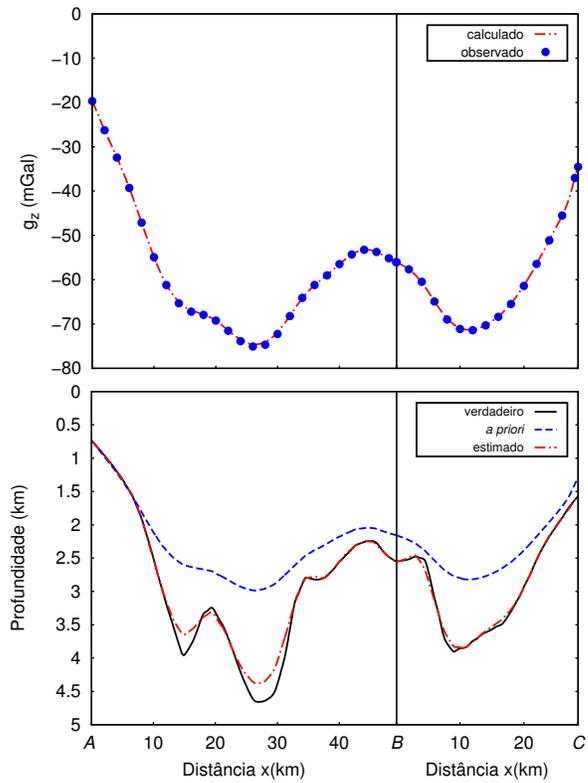


Figura 7: Estruturas verdadeira, inicial e estimada ao longo do perfil A-B-C mostrado nas Figuras 3 e 5

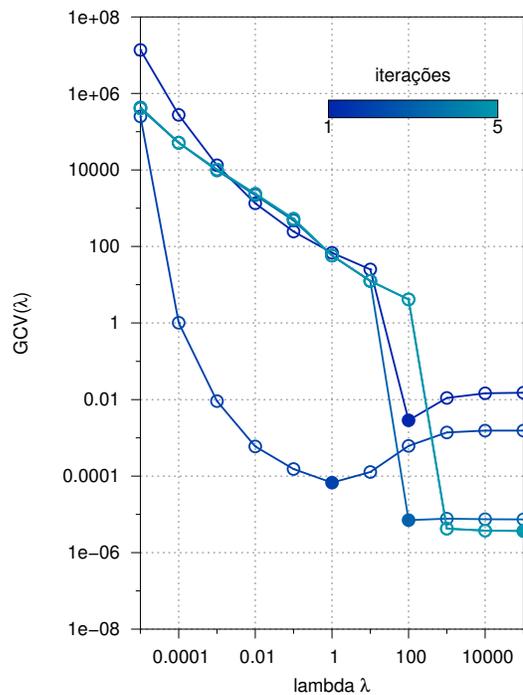


Figura 8: Curva GCV em função do parâmetro de regularização, para regularização de ordem 0, para as diferentes iterações da inversão usando o primeiro *a priori*. O λ ótimo é representado pelo círculo preenchido.

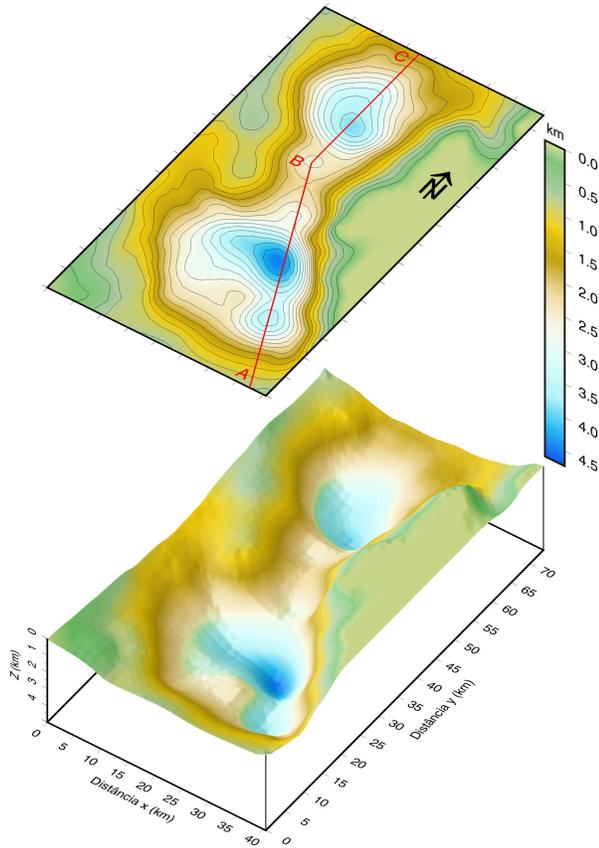


Figura 9: Mapa de contorno e vista em perspectiva do relevo estimado do embasamento usando o segundo *a priori*.

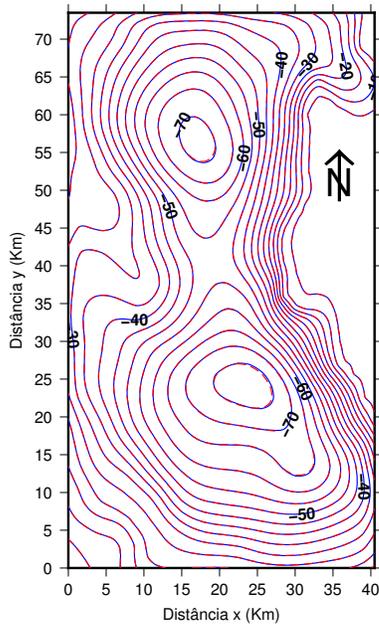


Figura 10: Anomalia gravimétrica contaminada com ruído (linhas contínuas azuis) devida à bacia sedimentar simulada na Figura 3 e a anomalia calculada (linhas tracejadas vermelhas) pela solução da Figura 9.

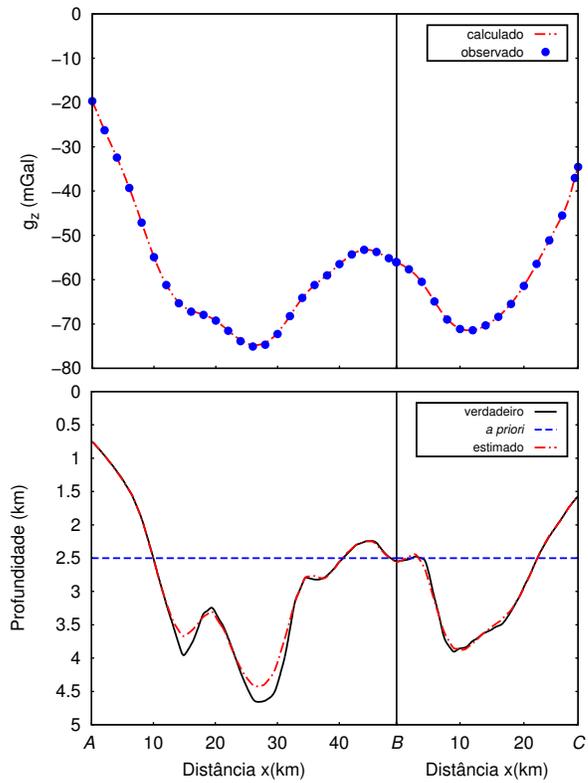


Figura 11: Estruturas verdadeira, inicial e estimada ao longo do perfil A-B-C mostrado nas Figuras 3 e 9

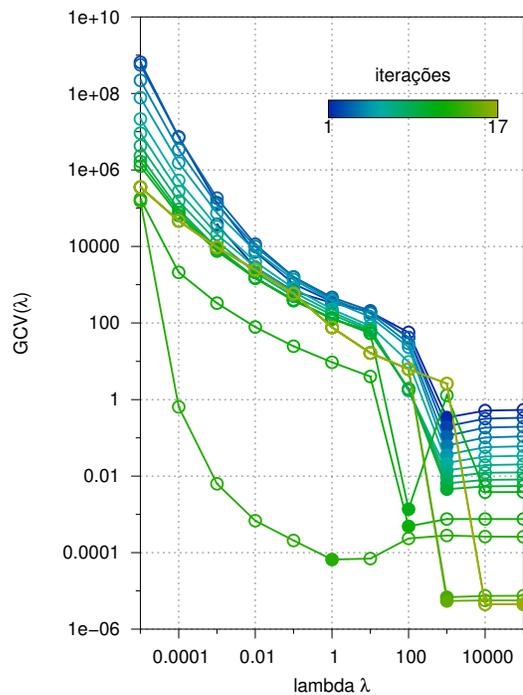


Figura 12: Curva GCV em função do parâmetro de regularização, para regularização de ordem 0, para as diferentes iterações da inversão usando o segundo *a priori*. O λ ótimo é representado pelo círculo preenchido.