



HETEROGENEOUS STOCHASTIC MULTISCALE

Ianyqui Falcão Costa

Liliane de Allan Fonseca

Ézio da Rocha Araújo

ianyqui.costa@gmail.com

liliane.fonseca@ufpe.br

ezio@ufpe.br

Universidade Federal de Pernambuco, UFPE.

Av. Prof. Moraes Rego, s/n, Cidade Universitária, Recife, PE.

Abstract. *This work brings together theoretical formulation and computational strategies introduced elsewhere (Koutsourelakis (2007), Weinan e Engquist (2003), Nonenmacher (2011)) for upscaling in random heterogeneous mediums. By heterogeneous it means both multi-physics models at different scales and random parameters at distinct locations in the physical domain, probably correlated, e.g. Stokes-Brinkman and fractures at microscale and heterogeneous Darcy at macroscale when the target problem is hydrocarbon reservoir simulation (hyperbolic problem) as in (Pal, 2011), or a matrix with aleatoric inclusions at microscale and homogeneous matrix at macroscale when the target problem is solid and structural mechanics. The approach uses a mixing of information theory and statistical mechanics in order to extract relevant probabilistic information from microscale and adaptively control the stochastic distance between two scale responses. A goal-oriented upscaling procedure is defined to guarantee equivalence between micro and macroscale target to specific output, and its generalization to a mathematical sound multi-goal-oriented response where users controls distinct accuracies of specific target responses. Preliminary applications in one-dimension domain for elliptic equations are presented. Applications use a number of realizations from microscale parameters distributions producing a reduced number of equivalent macroscale realizations, as required by user desired accuracy. The full formulation requires solving a difficult optimization problem for an extended Lagrangian formulation that is solved with a new Parallel Deterministic Annealing. Mathematical formulation is such that parallelization is done over realizations so that the overall computation cost of the stochastic upscaling is on the same order of a deterministic one.*

Keywords: *Reservoir, Stochastic, Heterogeneity, Multiphysics, Upscaling, Goal-oriented*

INTRODUÇÃO

Este texto descreve os aspectos matemáticos e computacionais de uma formulação para transposição de escalas regidas por distintas físicas estocasticamente caracterizadas, em geral de uma microescala cujas equações estejam discretizadas por uma malha densa, ou *micromalha*, que capture adequadamente as heterogeneidades incertas dos parâmetros, para macroescalas (*upscaling*), representadas por malhas esparsas, ou *macromalhas*, apropriadas à execução de centenas ou milhares de simulações de amostras (realizações) probabilísticas desses parâmetros incertos. Assume-se, por consequência, que o sistema de equações resultante da discretização na microescala seja proibitivamente grande para a capacidade dos computadores disponíveis.

Embora a formulação seja bastante geral (no sentido, por exemplo, em que se aplica à solução de sistemas de equações diferenciais elípticas, parabólicas e hiperbólicas), o maior interesse está na simulação de reservatórios de hidrocarbonetos. Imagina-se que o reservatório esteja descrito por um número finito de *realizações geoestatísticas* de distintas distribuições (variogramas) arbitradas para a descrição das litofáceis de um determinado *cenário geológico*. Também, e muito importante, são indistintas as técnicas de discretização no espaço e no tempo (elementos ou volumes finitos, diferenças finitas, etc.) como também as características da solução e dos algoritmos utilizados para encontrá-las (soluções com ou sem linearidades, algoritmos iterativos implícitos, explícitos, etc.).

A formulação imita Koutsourelakis (2007), adicionando-lhe novos elementos, como por exemplo uma mudança de física; utiliza uma generalização dos princípios apresentados em Weinan e Engquist (2003), adicionando frutíferas ideias de malhas orientadas aos objetivos desenvolvidos na tecnologia do método dos elementos finitos (*goal-oriented FEM*) e estendidas a elementos finitos para múltiplas escalas em Nonnenmacher (2011). A transposição da micro para a macroescala traz inevitáveis perdas de resolução ou de acurácia das respostas. Entretanto, essas perdas podem ser substancialmente reduzidas se forem efetuadas transições para distintas macroescalas com distintos objetivos e acurácias, daí a noção aqui apresentada de *upscaling orientado aos objetivos* da análise.

Na técnica ora apresentada, são utilizadas pelo menos duas medidas para a mensuração da qualidade da transposição de escalas e de físicas. Com uma primeira medida, de forma similar às decomposições espectrais de covariâncias (a transformação de Kosambi–Karhunen–Loève, KKL, por exemplo), avalia-se a retenção na macroescala das informações probabilísticas relevantes ao problema, mas não apenas da covariância. Essa importância das informações retidas, sendo dependente do problema – dos termos de fonte, do objetivo do problema, etc. – exige uma medida ou um conjunto delas que julgue a acurácia da resposta desejada frente à resposta na microescala. O uso de mais de uma medida contrasta com as técnicas usuais de *upscaling* e, equivalentemente, das técnicas de redução de base.

O problema final de extremização condicional é resolvido com uma formulação Lagrangeana Aumentada, e com a elaboração de um algoritmo paralelo, o PDA – *Parallel Deterministic Annealing*, uma versão paralela e estendida do DA- *Deterministic Annealing*, descritos em Koutsourelakis (2007), e originado em Rose (1991, 1994, 1998). O PDA é naturalmente paralelizável sobre realizações geoestatísticas. Essas variantes do probabilístico *Simulating Annealing* são as naturais para essa classe de problemas haja vista a ocorrência espontânea de possíveis distribuições exponenciais na sua solução. Claramente, outros algoritmos podem ser eficientemente utilizados.

Embora a formulação seja bastante geral, neste momento, são apenas apresentadas algumas primeiras aplicações em problemas unidimensionais. Nesses problemas, procura-se ilustrar os principais aspectos da formulação e dos algoritmos, facilmente generalizados para outras dimensões. A principal preocupação nesse texto é com a qualidade da formulação, dos métodos e algoritmos desenvolvidos.

1 METODOLOGIA

1.1 Generalidades

Seja o reservatório descrito por um conjunto de N_R realizações geoestatísticas coletivamente agregadas no ensemble $\mathbf{X}$, identificando malhas com um número N de células. Confundiremos $\mathbf{X}$ com o campo randômico que ele representa. Reconhece-se que o campo $\mathbf{X}$ pode conter mais informações do que necessário para a solução de um determinado problema de simulação de reservatório. O objetivo é descrever esse campo por vetores randômicos $\mathbf{Y}$ com $n < N$ e, possivelmente, $n_R < N_R$, realizações da macroescala, de tal forma que a resposta ao problema específico possa ser encontrada com menor esforço computacional, dentro de aproximações pré-estabelecidas. Imagina-se que $\mathbf{X}$ possa ser tão detalhada quanto uma *malha geocelular*, e que contenha informações na *microescala* que possam ser comprimidas em uma *macroescala* desejada para o *objetivo específico*.

Se as propriedades descritas por $\mathbf{X}$ podem ser compactadas em $\mathbf{Y}$ para a solução de um problema específico, o que se precisa determinar são quais as informações em $\mathbf{X}$ que devem ser *estatisticamente* preservadas, de tal forma que a distorção na estatística da resposta desejada possa ser controlada. Pelo menos duas medidas precisam, dessa forma, ser estabelecidas. Uma da fidelidade da representação de $\mathbf{X}$ por $\mathbf{Y}$, e outra da acurácia estatística da resposta desejada. Notoriamente, as físicas na microescala e na macroescala não precisam, ou não devam, em muitos casos, serem as mesmas.

Neste texto, sem perda de generalidade, não será abordada *explicitamente* a escala temporal. Isso será feito em um futuro texto por ocasião da aplicação da metodologia ao ajuste de histórico como componente do controle estocástico (*closed loop*, malha fechada) de gerenciamento e desenvolvimento.

As técnicas espectrais tradicionais, como as expansões KKL, a análise dos componentes principais (PCA) e o escalonamento multidimensional (MDS), e suas variantes que não são lineares, procuram preservar a estrutura da covariância com o uso de uma hierarquia de aproximações dada por uma sequência de autopares. Isso não maximiza a compactação das informações, uma vez que tanto preservam modos desnecessários como omitem modos associados a autovalores que podem ser importantes para a resposta específica, como também não levam em conta as características das excitações externas.

A ideia por trás da metodologia descrita é que não se podem escolher elementos representativos de volume (REV) nem fazer uso de periodicidade de informações na microescala, dado ao caráter essencialmente estocástico em todas as escalas das heterogeneidades e que, provavelmente, não se pode garantir alguma separação de escalas. Desta forma, a mudança de escalas deve ser feita com base em medidas de aproximação, possivelmente, não tão aproximadas quanto se queira.

Em uma metodologia ideal, a escolha de n_r e de n deveria ser adaptativa. A qualidade das duas medidas deveria ser aferida *a posteriori*, e ajustada a exigências de aproximações. Neste texto, a determinação de n_r é de fato feita *a posteriori*, enquanto que a escolha da malha geométrica, e por consequência de n , é feita *a priori*, com base em estimativas de recursos computacionais. Um conjunto adjacente de células é substituído por uma célula equivalente (macro célula, macro elemento), a ser estocasticamente homogeneizada. Uma que garanta que a descrição estatística da resposta possua a acurácia desejada. Costuma-se denominar tais procedimentos de local para local ou, em um esforço de linguagem, local-local (Weinan e Engquist, 2003). É evidente que um procedimento global-local pode ser igualmente empregado à custa de maior esforço computacional na geração da macromalha.

A mudança de escala é feita por assimilação de *dados prévios dos modelos, dados observados*, ou ambos, tornando a técnica também adequada a controles estocásticos, como por exemplo, no controle de malhas fechadas com previsão e assimilação de dados de produção, em qualquer escala de tempo. Um dos problemas principais do controle estocástico é o esforço computacional nas duas fases, previsão e assimilação. Neste caso, as técnicas de parametrização, como por exemplo, as decomposições espectrais, podem ser convenientemente substituídas por macromalhas, estocasticamente equivalentes. Tanto métodos Bayesianos, como MCMC e RML, como estimadores clássicos, como os de máxima verossimilhança, regularizados (máximo *a posteriori*, por exemplo), são beneficiados com a economia computacional da mudança de escala, do uso de uma macromalha. Isso sintetiza a importância do presente estudo.

1.2 Indicador de informação mútua entre os campos randômicos

A metodologia seguida é aproveitada em seus aspectos mais importantes de Koutsourelakis (2007), e generalizada para múltiplas físicas. Está baseada em modelos da mecânica estatística (Hill, 1956), e da teoria da informação originadas principalmente em Shannon (1948), com algoritmos específicos desenvolvidos em Rose (1991, 1994, 1998).

A substituição do campo $\mathbf{X}$ por $\mathbf{Y}$, de menor dimensão, implica na possibilidade de atribuir o mesmo valor a grandes regiões do domínio. O conceito de *indicador de informação mútua* é introduzido Shannon (1948) como uma medida da informação média que o conhecimento de $\mathbf{Y}$ pode proporcionar de $\mathbf{X}$, e vice-versa.

A hipótese de que a grande regiões do domínio heterogêneo de interesse possam ser atribuídos os mesmos valores possibilita a descrição do campo randômico $\mathbf{X}$ pelo campo de menor dimensão $\mathbf{Y}$. Uma medida da capacidade de $\mathbf{Y}$ aproximar $\mathbf{X}$ pode ser dada pelo *indicador de informação mútua* (Shannon, 1948), dado por,

$$I(\mathbf{X}, \mathbf{Y}) = \iint p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)} dx dy, \quad (1)$$

onde $p_{XY}(x, y)$ é densidade conjunta de $(\mathbf{X}, \mathbf{Y})$, e $p_X(x)$ e $p_Y(y)$ são as respectivas distribuições marginais. O seu valor mínimo é zero quando os dois campos são independentes, ou seja, $p(\mathbf{X}, \mathbf{Y}) = p_X(\mathbf{X})p_Y(\mathbf{Y})$. O seu valor máximo é conhecido como a *entropia* de $\mathbf{X}$, que se dá quando os dois campos são idênticos $\mathbf{X} \equiv \mathbf{Y}$. Definindo a *entropia conjunta* como o próprio indicador, e desenvolvendo,

$$H(\mathbf{X}, \mathbf{Y}) = -\int p_X(x) \log p_X(x) dx + \iint p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{p_Y(y)} dx dy, \quad (2)$$

usando o fato de que $\iint p_{XY}(x, y) \log p_X(x) dx dy = \int p_X(x) \log p_X(x) dx$. O primeiro termo no segundo membro da eq. (2) é a entropia (Planck (1948)) de $\mathbf{X}$,

$$H(\mathbf{X}) = -\int p_X(x) \log p_X(x) dx, \quad (3)$$

enquanto que o segundo é a *entropia condicional* ou *equivoco* (tradução livre para *equivocation*),

$$H_Y(\mathbf{X}) = -\iint p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{p_Y(y)} dx dy, \quad (4)$$

que se anula. Tem-se então que $I(\mathbf{X}, \mathbf{Y}) = H(\mathbf{X}) - H_Y(\mathbf{X})$ e $I(\mathbf{X}) = H(\mathbf{X})$. Em teoria da informação, $I(\mathbf{X}, \mathbf{Y})$ é também conhecido como a taxa R de transmissão de informações de um canal contínuo (Shannon, 1948).

A conveniência do uso de medidas logarítmicas, geralmente de base 2 em teoria da informação, é justificada por praticidade, propriedade intuitiva e conveniência matemática. Entretanto, a entropia e o indicador de informação mútua, embora possam medir a incerteza, não são medidas probabilísticas. Aparentemente, uma medida probabilística pode ser obtida dividindo o indicador pela entropia de $\mathbf{X}$. Se a entropia de $\mathbf{Y}$ é menor do que a de $\mathbf{X}$, a compressão pode ser feita a taxas maiores. Outras medidas não probabilísticas apropriadas de proximidade dos dois campos estão sendo desenvolvidas com base no conceito de *entropia relativa* ou *divergência KL*.

1.3 Fidelidade da macroescala à microescala

Parece evidente que, afora algumas patologias atípicas, a qualidade da representação na macroescala deve ser avaliada com medidas que comparem respostas específicas, e não todas as possíveis respostas. Como, por exemplo, a capacidade de a macroescala aproximar o ajuste ao histórico de determinada grandeza produzida pela microescala, é de grande interesse. Difícil imaginar que uma única macroescala possa, em geral, substituir satisfatoriamente a microescala na solução de todos os problemas de interesse. Então, faz sentido imaginar que, de maneira ótima, podem-se ter modelos específicos na macroescala para reproduzir respostas específicas.

A medida de fidelidade é conhecida como *medida de distorção* da resposta fornecida pelos dois campos. Essa medida pode ser escalar, vetorial ou matricial. Sem perda de generalidade, será assumida que a medida é escalar, vez que as outras medidas deverão sempre ser transformadas em uma ou em um conjunto de medidas escalares. Para exemplificar, seja a distância dada por,

$$d(\mathbf{X}, \mathbf{Y}) = (r(\mathbf{X}) - r(\mathbf{Y}))^2, \quad (5)$$

onde $r(\mathbf{X})$ é a resposta devida à microescala e $r(\mathbf{Y})$ à macroescala, geralmente funções implicitamente definidas por um simulador numérico. Para cada dada distância definida como na eq. (5), haverá necessidade do seu gradiente em relação a $\mathbf{Y}$.

Não há estatísticas na eq. (5), ou seja, as funções randômicas $r(\mathbf{X}): \mathfrak{R}^N \rightarrow \mathfrak{R}$, e $r(\mathbf{Y}): \mathfrak{R}^n \rightarrow \mathfrak{R}$ devem ser compreendidas como um ensemble, ou conjunto, de amostras, ou realizações, $r(X_i), i=1, \dots, N_R$, e $r(Y_i), i=1, \dots, n_R$. Neste sentido, $d(\mathbf{X}, \mathbf{Y}): \mathfrak{R}^N \times \mathfrak{R}^n \rightarrow \mathfrak{R}^+$ é uma coleção de $N_R \times n_R$ amostras em uma dimensão. Além disso, será admitido, sem perda de generalidade, que a microescala está representada apenas por uma variável randômica Y . Isso significa que a macromalha possui um único elemento, célula ou bloco, com uma única propriedade que, essencialmente, é sua melhor representação estocástica. A distância, neste caso, será uma aplicação $d(\mathbf{X}, Y): \mathfrak{R}^N \times \mathfrak{R} \rightarrow \mathfrak{R}^+$.

É importante observar, neste estágio, que não se faz referências às físicas que prevalecem no domínio geométrico. É possível, por exemplo, avaliar $r(\mathbf{X})$ numa complexa física dos carbonatos, utilizando *Darcy-Stokes* ou *Stokes-Brinkman* com fraturas *Darcyanas*, ou mesmo em modelos composicionais, e $r(\mathbf{Y})$ na física do modelo *Black-Oil*.

Além disso, uma parte ou todo o domínio na microescala pode ter sido transformado para uma representação espectral, como por exemplo, a representação de KKL e suas variantes.

O objetivo é duplo. Primeiro, é desejável um esquema de compressão de $\mathbf{X}$ para Y que leve a uma distorção mínima da resposta desejada e, em seguida, determinar o seu valor. Determinar a transformação ótima envolve avaliar a distribuição *a posteriori* de Y , $p_Y(y)$.

Importante, do ponto de vista prático, é que uma vez especificada a distorção desejada, o esquema permitirá determinar várias transformações subótimas e suas respectivas distorções.

Para o esquema estocástico estar completo, é necessário avaliar a distorção de forma probabilística. Uma possibilidade é a distorção D ser a esperança matemática da distância, mas, funções mais significativas para certas aplicações, como percentis e medidas de risco sobre as caudas da densidade, podem ser escolhidas,

$$D = E[d(\mathbf{X}, Y)]. \quad (6)$$

Por hipótese, o campo que descreve a microescala é dado pelo ensemble de N_R realizações geoestatísticas. Assim, a avaliação da distorção, a sua esperança, ou qualquer outra medida estocástica conveniente, deve ser feita usando a técnica de Monte Carlo, ou alguma variante simplificadora. Uma vez que $\mathbf{X}$ já está amostrado é possível escrever, usando a relação entre a densidade conjunta e a densidade condicional,

$$D = E[d(\mathbf{X}, Y)] = \sum_{\mathbf{X}, Y} p_{\mathbf{X}Y}(\mathbf{X}, Y) d(\mathbf{X}, Y) = \sum_{\mathbf{X}, Y} p_{\mathbf{X}}(\mathbf{X}) p_{\mathbf{X}Y}(Y / \mathbf{X}) d(\mathbf{X}, Y), \quad (7)$$

onde $p_{\mathbf{X}Y}(Y / \mathbf{X})$ é a densidade condicional de Y . O seu mínimo com respeito à $p_{\mathbf{X}Y}(Y / \mathbf{X})$, será uma função *delta de Dirac* que atribui a cada x de $\mathbf{X}$ o Y que minimiza $d(\mathbf{X}, Y)$.

Esse é um problema de otimização que exige mais esforço computacional do que desejado. Em lugar disso, procura-se resolver o *Lagrangeano aumentado* com a restrição de que o indicador $I(\mathbf{X}, Y)$ seja menor ou igual a um valor pré-especificado $I(\mathbf{X}, Y) \leq R$.

Com esse objetivo, podemos desenvolver a eq. (1),

$$I(\mathbf{X}, Y) = \sum_{\mathbf{X}, Y} p_{\mathbf{X}}(\mathbf{X}) p_{\mathbf{X}Y}(Y / \mathbf{X}) \log \frac{p_{\mathbf{X}}(\mathbf{X}) p_{\mathbf{X}Y}(Y / \mathbf{X})}{p_{\mathbf{X}}(\mathbf{X}) \sum_{\mathbf{X}} p_{\mathbf{X}}(\mathbf{X}) p_{\mathbf{X}Y}(Y / \mathbf{X})}, \quad (8)$$

obtendo, após cancelamento de $p_{\mathbf{X}}(\mathbf{X})$,

$$I(\mathbf{X}, Y) = \sum_{\mathbf{X}, Y} p_{\mathbf{X}}(\mathbf{X}) p_{\mathbf{X}Y}(Y / \mathbf{X}) \log \frac{p_{\mathbf{X}Y}(Y / \mathbf{X})}{q_Y(Y)}, \quad (9)$$

onde,

$$q_Y(Y) = \sum_{\mathbf{X}} p_{\mathbf{X}}(\mathbf{X}) p_{\mathbf{X}Y}(Y / \mathbf{X}), \quad (10)$$

é a marginal de Y .

A função Lagrangeana aumentada a ser minimizada será então,

$$F = D + T(I(\mathbf{X}, Y) - R), \quad (11)$$

onde $T \geq 0$ é o multiplicador de Lagrange. O mínimo se dará para derivada de F , com respeito a $p_{\mathbf{X}Y}(Y / \mathbf{X})$, igual a zero, e $T(I(\mathbf{X}, Y) - R) = 0$, mais conhecidas como condições de Kuhn-Tucker. Desta forma, mantendo T maior que zero e aumentando gradualmente R , serão obtidas realizações de Y mais informativas sobre as realizações de $\mathbf{X}$. Esse problema de minimização é equivalente à minimização da função Lagrangeana complementar,

$$F^c = I(\mathbf{X}, Y) + \beta(D - D_0), \quad (12)$$

para uma distorção D_0 especificada. As duas funções são simultaneamente minimizadas para $T = \beta^{-1}$. Ao multiplicador de Lagrange T pode-se atribuir o mesmo significado da temperatura no método conhecido como *Simulating Annealing*, com origem na Mecânica Estatística. O algoritmo descrito adiante toma partido dessa similaridade.

A distribuição condicional ótima é a de *Gibbs*, obtida anulando o gradiente de F ,

$$p_{\mathbf{X}Y}^*(Y / \mathbf{X}) = \frac{q_Y(Y) \exp(-\frac{d(\mathbf{X}, Y)}{T})}{Z(\mathbf{X})}, \quad (13)$$

onde $Z(\mathbf{X}) = \sum_{\mathbf{X}} q_Y(Y) \exp(-\frac{d(\mathbf{X}, Y)}{T})$ é uma constante normalizadora. Ela converge para o delta de Dirac quando T vai à zero. O mínimo correspondente de F será,

$$F^* = -T \sum_{\mathbf{X}} p(\mathbf{X}) \log Z(\mathbf{X}) + TR = -T \sum_{\mathbf{X}} p(\mathbf{X}) \log \sum_{\mathbf{X}} q_Y(Y) \exp(-\frac{d(\mathbf{X}, Y)}{T}) + TR. \quad (14)$$

Pode-se eliminar o produto TR da expressão acima sem nenhuma perda, uma vez que apenas interessam os valores das variáveis que a minimizam, quais sejam, as realizações Y e suas probabilidades de ocorrência dadas pela marginal $q_Y(Y)$. As realizações $\mathbf{X}$, por sua vez, supõem-se que sejam fornecidas associadas às suas probabilidades. As realizações geoestatísticas, em geral, não vêm ranqueadas. Neste caso, é usual assumir a mesma probabilidade $1/N_R$ para cada uma.

O problema posto consiste em minimizar F^* em relação à variável randômica Y submetida à restrição de que $\sum_i q_i = 1$, ou seja, minimizar o Lagrangeano aumentado $F^{**}(y_i, q_i, \lambda)$ dado por,

$$F^{**}(y_i, q_i, \lambda) = F^*(y_i, q_i) - \lambda(\sum_i q_i - 1), \quad (15)$$

onde $y_i \in Y$ são as realizações na macroescala, λ o multiplicador de Lagrange, e q_i a marginal de y_i . Tomando o gradiente em q_i ,

$$\frac{\partial F^{**}}{\partial q_i} = T \sum \frac{1}{Z(\mathbf{X})} \exp\left(\frac{d(\mathbf{X}, y_i)}{T}\right) = 0, \quad (16)$$

e, levando em conta a eq. (3) e a eq. (10), resulta em que

$$\sum_{\mathbf{X}} p_{\mathbf{X}}(\mathbf{X}) \frac{p_{\mathbf{XY}}(Y / \mathbf{X})}{q_Y(Y)} = \frac{\lambda}{T} = 1. \quad (17)$$

Tomando agora o gradiente em y_i ,

$$\frac{\partial F^{**}}{\partial y_i} = \sum p_{\mathbf{X}}(\mathbf{X}) p_{\mathbf{XY}}^*(y_i / \mathbf{X}) \frac{\partial d(\mathbf{X}, y_i)}{\partial y_i} = 0, \quad (18)$$

pelo qual se obtém

$$r(y_i) = \frac{\sum_{\mathbf{X}} p_{\mathbf{X}}(\mathbf{X}) p_{\mathbf{XY}}(y_i / \mathbf{X}) r(\mathbf{X})}{q_Y(y_i)}.. \quad (19)$$

Está-se agora em condições de formular algoritmos para determinar Y e sua marginal associada $q_Y(Y)$.

1.4 O Algoritmo PDA

O algoritmo abaixo é uma versão generalizada do denominado *Deterministic Simulated Annealing*, ou DA, elaborado e discutido em Koutsourelakis (2007) e Rose (1991, 1994, 1998). Ele encontra Y e sua marginal associada $q_Y(Y)$. É considerado uma versão determinística do clássico *Stochastic Simulated Annealing*.

Os detalhes dados são maiores do que os normalmente fornecidos na literatura, permitindo sua imediata programação. Além disso, as computações estão ordenadas de tal

forma que os cálculos para cada realização da microescala sejam feitos em paralelo. Assim, incentiva-se o uso da malha geocelular, supostamente a descrição em menor escala, para realizar a mudança de escala estocástica, com computações paralelas.

Algoritmo PDA: Parallel Deterministic Annealing

- (1) Inicie com $K=1$, um valor suficientemente alto $T=T_{\max}$, um fator $\alpha < 1$, y_1 arbitrário, $q_1=1$, ε_1 , ε_2 tolerâncias para convergências, e T_f um valor final próximo de zero,
- (2) Para $k=1, 2, \dots, K$, em paralelo para cada realização da microescala.
 - a. Calcule as N_R distâncias a y_k , $d(X_j, y_k) = (r(X_j) - r(y_k))^2$, eq. (5),
 - b. Calcule os N_R denominadores, $Z(X_j) = \sum_{i=1}^K q_i \exp(-\frac{d(X_j, y_i)}{T})$,
 - c. Calcule as N_R densidades condicionais, $p(y_k / X_j) = \frac{q_k \exp(-\frac{d(X_j, y_k)}{T})}{Z(X_j)}$, eq. (13),
 - d. Estime (Monte Carlo) a marginal de y_k , $q_k = \sum_{j=1}^{N_R} p(X_j) p(y_k / X_j)$, ver eq. (10),
 - e. Estime (Monte Carlo), $G_k = \sum_{j=1}^{N_R} p(X_j) p(y_k / X_j) r(X_j)$, ver eq. (19),
 - f. Resolva o problema unidimensional $y_k = r^{-1}(\frac{G_k}{q_k})$, ver eq. (19).
- (3) Verifique com ε_1 , a convergência para y_k e q_k . Se não satisfeita, repita a etapa (2). De outra forma, continue para a etapa (4),
- (4) Estime, usando Monte Carlo, a distorção média e o indicador, (em paralelo),
 - a. A distorção média, $D = \sum_{k=1}^K \sum_{j=1}^{N_R} p(X_j) p(y_k / X_j) d(X_j, y_k)$,
 - b. O indicador, $I = \sum_{k=1}^K \sum_{j=1}^{N_R} p(X_j) p(y_k / X_j) \log \frac{p(y_k / X_j)}{q_k}$,
- (5) Reduza a temperatura T , $T \leftarrow \alpha T \geq T_f$.
- (6) Dobre o número de realizações y_k , ou seja, faça $K \leftarrow 2K$ e inicialize $y_{K+k} = y_k$, $q_{K+k} = q_k$, $k=1, \dots, K$.
- (7) Elimine as redundâncias com ε_2 em Y , determine $n_R \leq K$, e consolide $q_k, k=1, \dots, n_R$.

Existem versões sequenciais mais atuais deste algoritmo, que serão estudadas posteriormente. Usualmente o resfriamento deve ser lento, como, por exemplo, reduzindo a temperatura por seus milésimos, ou seja, $\alpha = 1/1.001$. Os modernos métodos de Homotopia podem ser utilizados de forma equivalente.

O que distingue esse algoritmo do *Simulated Annealing* é a introdução de um algoritmo determinístico (*steepest descent*) a cada nível de temperatura. Para algumas distâncias d e respostas r a eq. (18) não possui um único ponto fixo, portanto, não oferece resposta única. Nesses casos, deve-se escolher a de menor distorção média D .

2 APLICAÇÕES

Exemplo 1. Com o intuito de avaliar a qualidade da formulação e a acurácia possível de seus resultados, considere (Koutsourelakis, 2007) a equação diferencial de segunda ordem com uma variável a coeficientes estocásticos com $\omega \in \Omega$, Ω o espaço amostral,

$$\frac{d}{dx} \left(a(x, \omega) \frac{dp}{dx} \right) = 0, \quad x \in [0, 1], \quad (20)$$

com condições de contorno mistas, $p(0) = 0$ e $v(1) = a(1) \frac{dp}{dx} \Big|_{x=1} = 1$, correspondendo a um fluxo monofásico em regime permanente em um meio poroso com pressão e velocidade prescritas; ou uma barra elástica com forças e deslocamentos prescritos. A variabilidade do parâmetro $a(x, \omega)$ é dada por

$$a(x, \omega) = 1 + 0.5 \cos \left(2\pi \frac{x}{x_0} + \phi(\omega) \right), \quad (21)$$

onde o ângulo de fase $\phi(\omega)$ é uniformemente distribuído no intervalo $[0, 2\pi]$ e a constante x_0 caracteriza o comprimento de correlação (escala) da heterogeneidade espacial. Para cada realização, há um parâmetro equivalente, ou efetivo, dado pela média harmônica dos seus valores no domínio. Para pequenos valores do comprimento de correlação a variância é pequena e a distribuição do parâmetro efetivo pode ser dada pela sua média harmônica apenas. O problema consiste em encontrar a distribuição equivalente para a determinação de $p(1)$, quando o comprimento de correlação for grande, tipicamente maior do que o domínio. O problema é resolvido com as discretizações, para o espaço geométrico e para o espaço amostral, coincidentes. Sendo a distribuição de massas dada pelo ensemble de realizações $\mathbf{X}$ sobre uma dada discretização, o problema consiste em encontrar a variável aleatória Y do parâmetro equivalente.

A micromalha possui 1000 elementos. Para $x_0 = 1/100$ o campo estocástico é descrito por 100 realizações $\mathbf{X}$ da microescala, e calculadas soluções $p(1)$ com as quais são estimadas as Y realizações da macroescala orientadas para a estimativa de $p(1)$. O problema é essencialmente determinístico e o resultado é uma única realização com o parâmetro dado pela média harmônica. A esperança da distorção é nula.

Tomando $x_0 = 10$, maior do que a extensão do domínio geométrico, e simulando 1000 realizações da microescala, foram calculadas novamente as macro-realizações, com o mesmo objetivo de determinar $p(1)$. A Fig. 1 mostra o gráfico da informação mútua pela distorção média. Para evitar que o algoritmo permaneça aprisionado em um mínimo local, foram introduzidos ruídos Gaussianos nas realizações da macroescala. A temperatura original utilizada foi $T = 100$ com um resfriamento lento dado por $\alpha = 1/1.001$. A Fig. 2 mostra a densidade das propriedades efetivas calculadas para a macroescala. As distribuições nas duas escalas são mostradas na Fig. 3. Fez-se exigência de pouca precisão nas análises, melhores resultados são obtidos com maiores precisões e com maiores N_R .

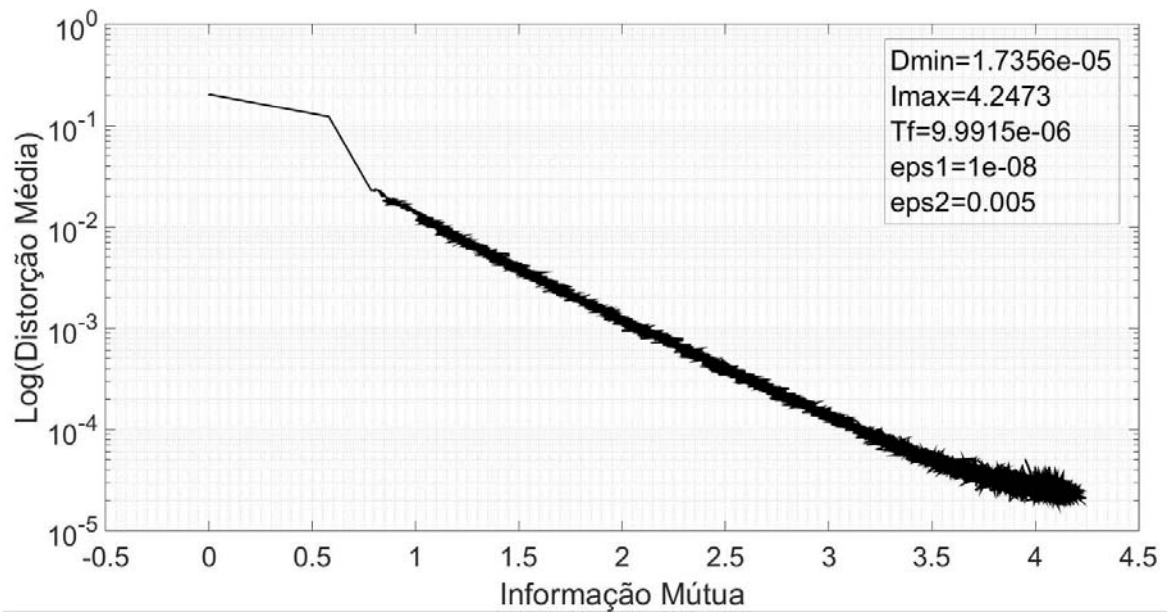


Figura 1. Curva da Distorção Média com a Informação Mútua para o Exemplo 1.

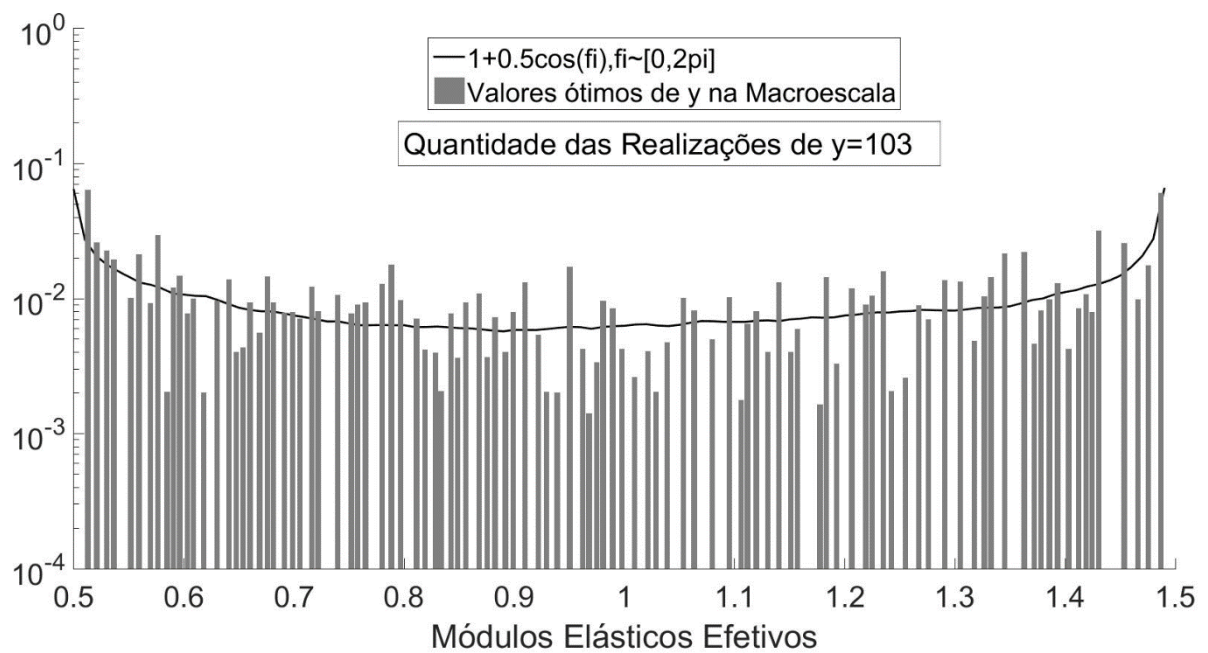


Figura 2. Densidades das realizações nas duas escalas, para o Exemplo 1.

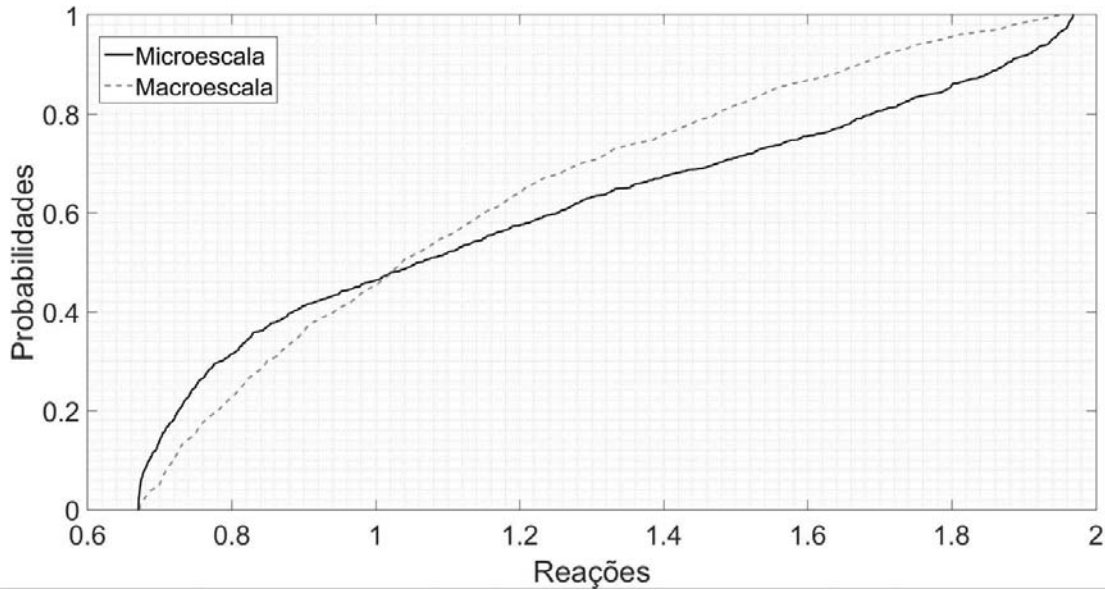


Figura 3. Distribuição das realizações nas duas escalas, para o Exemplo 1.

Exemplo 2. Este exemplo realiza um *upscaling* de uma micromalha de 1000 elementos finitos para uma macromalha de 10 elementos finitos de igual comprimento (baseado em Koutsourelakis, 2007). A equação diferencial é da mesma forma da Eq. (20), também no intervalo unitário, com condições de contorno de Dirichlet, $p(0)=0$ e $p(1)$ no intervalo $[0,3]$, discretizado, para efeito de cálculo, em 10 subintervalos uniformes. O parâmetro estocástico, no entanto, não é linear, sendo função da derivada de p , na forma $a(p', x, \omega)$, nos seguintes moldes:

$$a(p', x, \omega) = p' \quad \text{para } v(x, \omega) \leq 0,3 + 0,3\Phi(Z_v(\omega)), \quad \omega \in [0,1], \quad (22)$$

$$a(p', x, \omega) = \frac{H(\omega)}{1 + H(\omega)}, \quad H(\omega) = \Phi(Z_H(\omega)), \quad \text{para } v(x, \omega) > 0,3 + 0,3\Phi(Z_v(\omega)), \quad \omega \in [0,1], \quad (23)$$

onde $Z_v(\omega)$ e $Z_H(\omega)$ são campos Gaussianos independentes com a mesma autocorrelação espacial dada por $\rho(\Delta x) = \exp\left(-\frac{|\Delta x|}{x_0}\right)$ e $\Phi(\omega)$ é a distribuição acumulada normal, padrão, e, neste exemplo, $x_0 = 1/100$. Foi adotada uma micromalha de 1000 elementos.

O objetivo do *upscaling* é a determinação de $p(1) \in [0,3]$. Entretanto, deseja-se que a física na macroescala seja linearizada, ou seja, $a(p', x, \omega) = a(x, \omega)$ na Eq. (23), tornando os algoritmos de solução ainda mais eficientes. Na micromalha, resolvem-se sucessivos sistemas 1000×1000 com o método de Newton-Raphson, enquanto a resposta da macromalha reduz-se à solução de sistemas lineares 10×10 . As Figs. 4a a 4e mostram as densidades (massa) nas duas malhas para cada extremo nos subintervalos de *upscaling*. Percebe-se que quanto maior os valores médios são, tão melhores são os resultados, ao mesmo tempo em que há uma redução das incertezas. Essa redução resulta de a macromalha ser específica para a medida utilizada,

$$d(\mathbf{X}, Y) = \sum_{j=1}^{10} (r_j(\mathbf{X}) - r(Y))^2, \quad (24)$$

porém, com uma única distribuição Y para todos os 10 macroelementos. Claramente, melhores valores são obtidos se os parâmetros forem escalados macroelemento por macroelemento, como se espera em aplicações em reservatórios descritos geoestatisticamente.

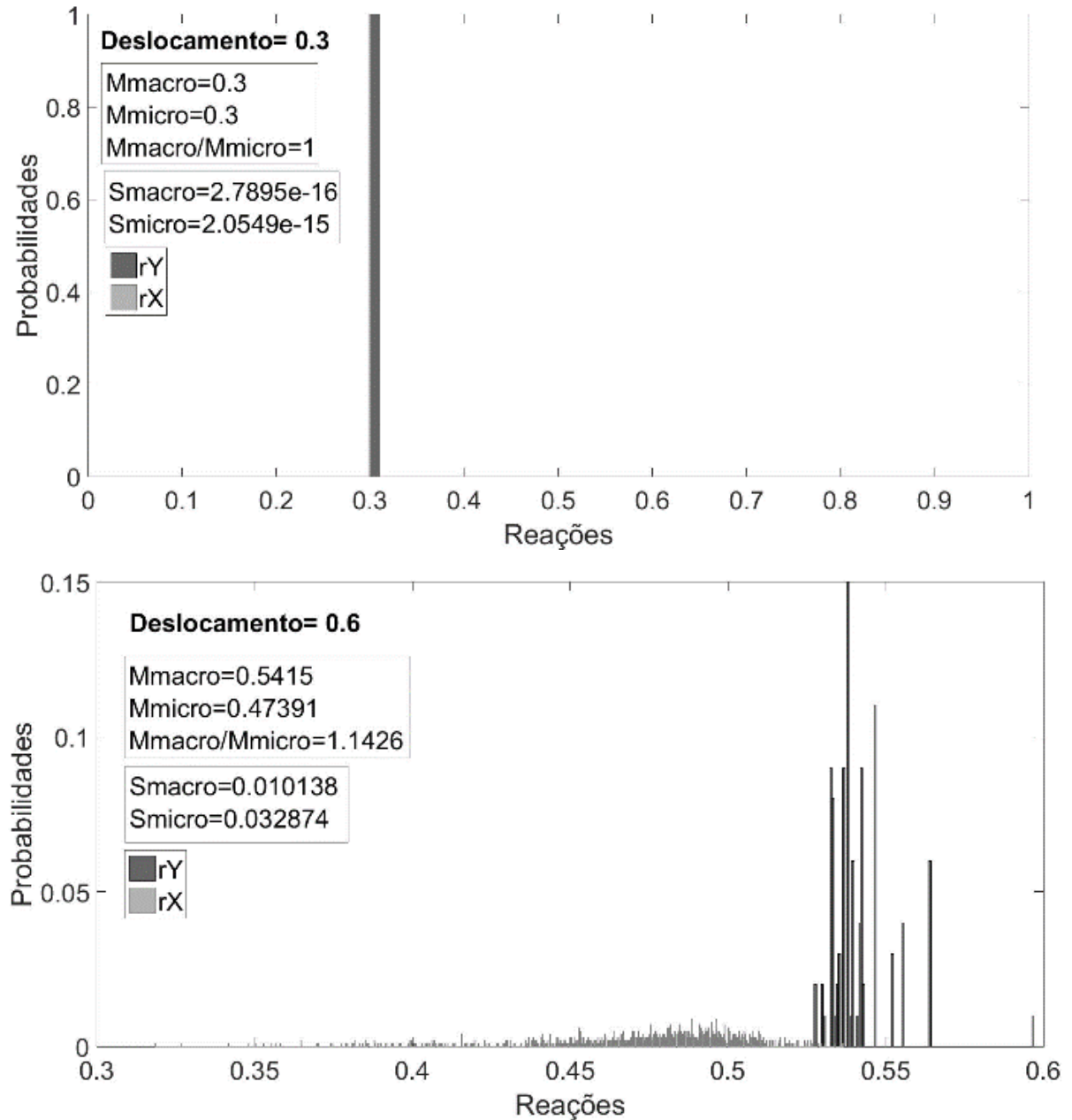


Figura 4a – Distribuições nas duas escalas para $p(1) = 0,3$ e $p(1) = 0,6$; indicando valores médios e desvios padrão para $v(0)$.

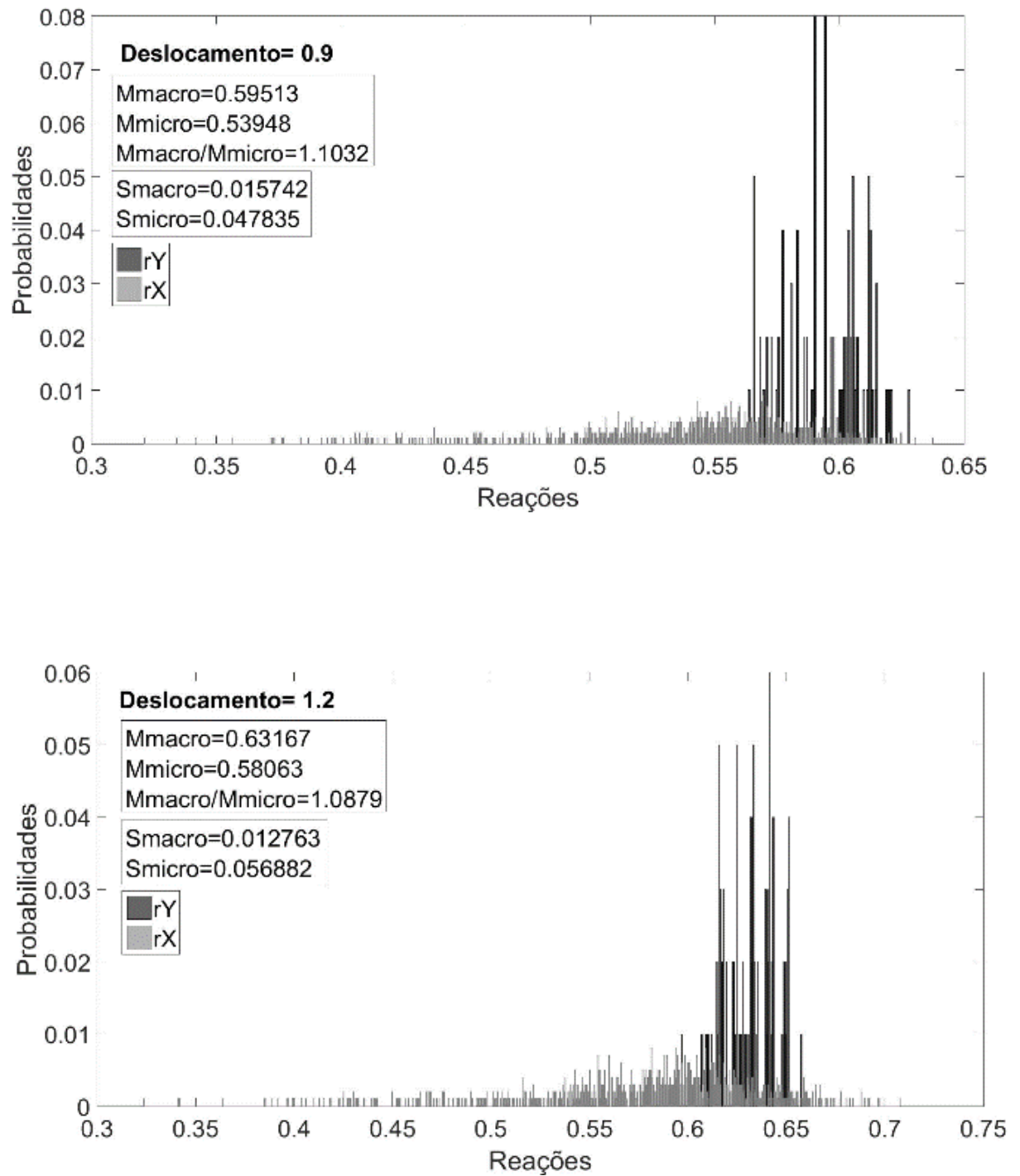


Figura 4b – Distribuições nas duas escalas para $p(1) = 0,9$ e $p(1) = 1,2$; indicando valores médios e desvios padrão para $v(0)$.

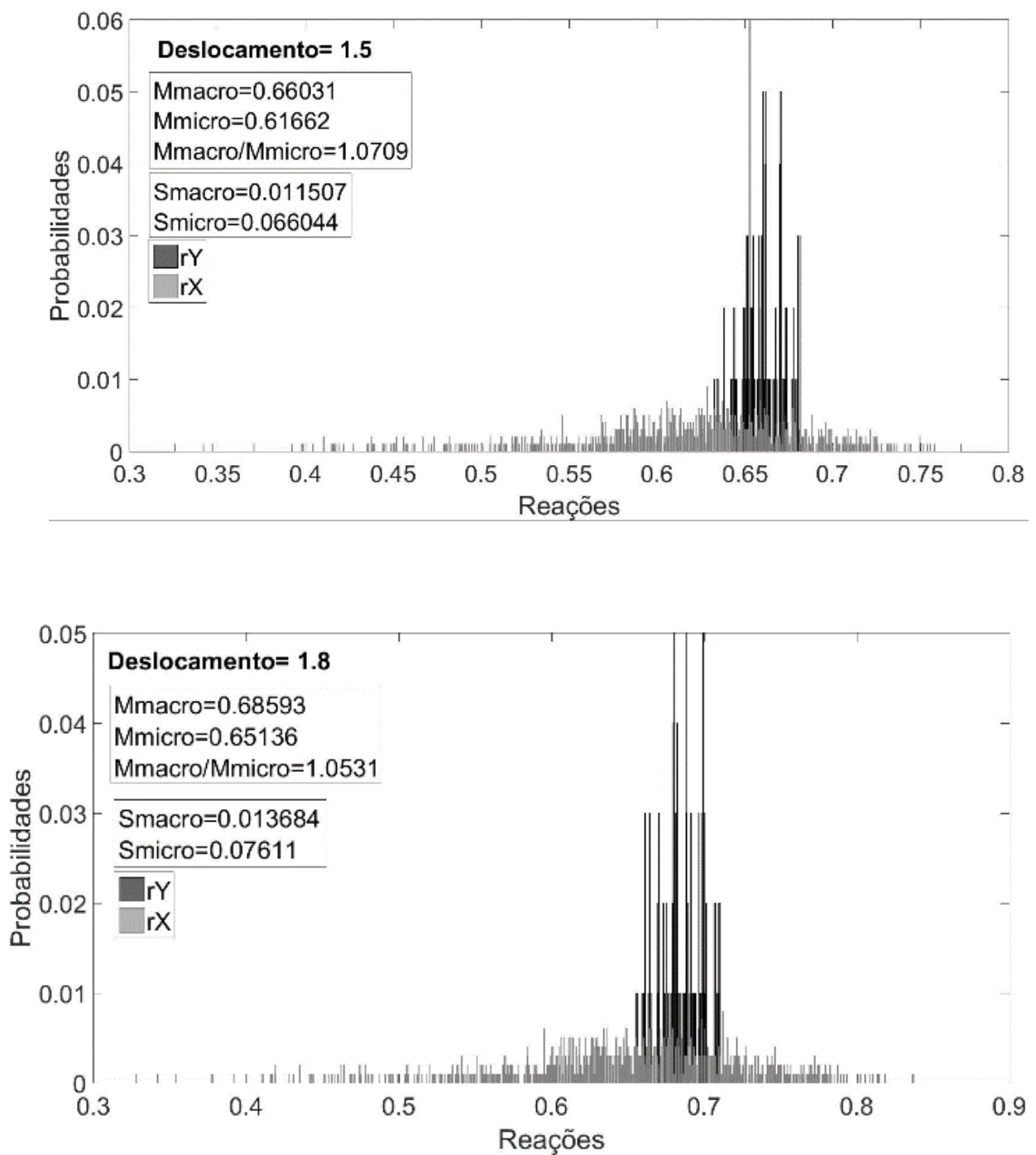


Figura 4c – Distribuições nas duas escalas para $p(1) = 1,5$ e $p(1) = 1,8$; indicando valores médios e desvios padrão para $v(0)$.

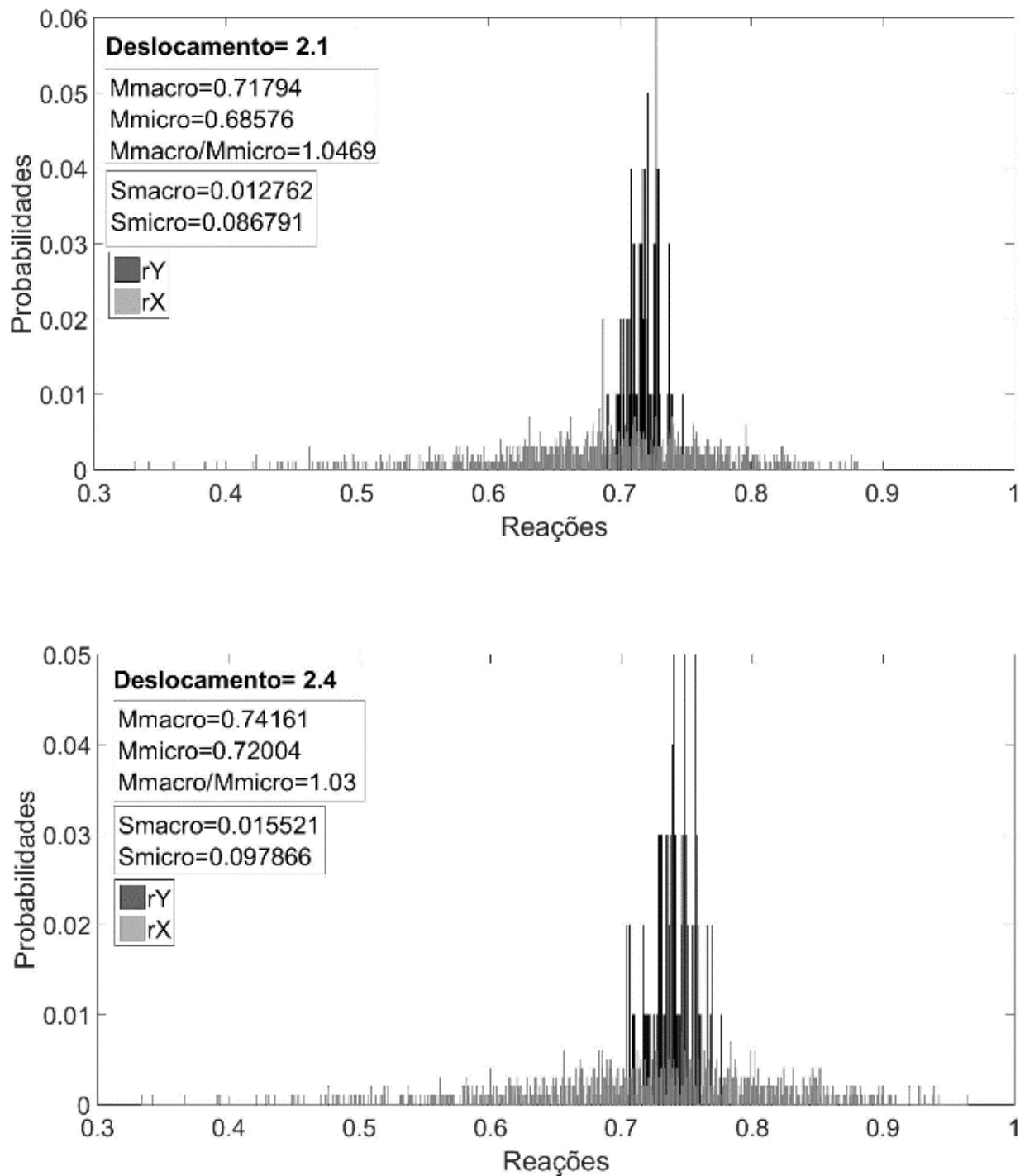


Figura 4d – Distribuições nas duas escalas para $p(1) = 2,1$ e $p(1) = 2,4$; indicando valores médios e desvios padrão para $v(0)$.

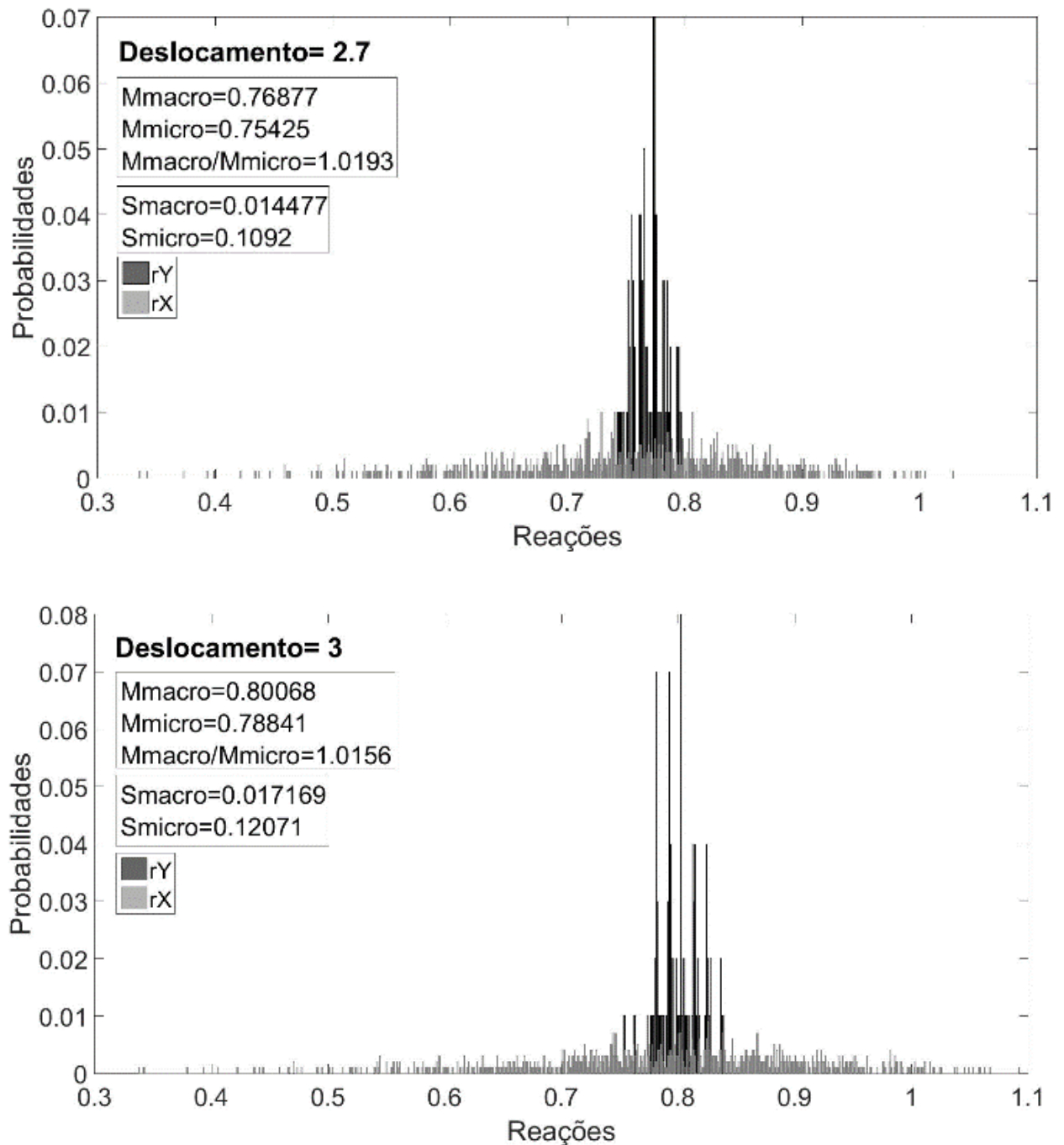


Figura 4e – Distribuição nas duas escalas para $p(1) = 2,7$, $p(1) = 3,0$; indicando valores médios e desvios padrão para $v(0)$.

3 CONCLUSÕES

Foi apresentada uma formulação, de fato uma pequena extensão da apresentada em Koutsourelakis (2007), para efetuar mudanças de escalas estocásticas de reservatórios

descritos geoestatisticamente. A formulação e a metodologia permitem mudanças de escalas com técnicas local-local, e local-global, como definidos em Weinan e Engquist (2003), com simultânea mudança de física, com acurácia controlada, e incorporando o conceito de *goal-oriented upscaling* contido em Nonnenmacher (2011). Foram apresentados exemplos preliminares muito animadores. Os autores estão desenvolvendo aplicações mais realistas em reservatórios bidimensionais para aplicações finais em otimização e controle estocástico de reservatórios, os quais serão relatados futuramente. Está sendo desenvolvida uma formulação diferente para o problema de otimização que permite utilizar outros algoritmos estocástico-determinísticos, com grande potencial de redução do número de avaliação das medidas, de fuga do paradigma da densidade de Gibbs e de sua estacionariedade temporal.

AGRADECIMENTOS

Os autores agradecem à UFPE e ao apoio financeiro parcial da PETROBRAS através da rede SIGER, através do projeto UFPE/PETROBRAS/SIGER2.

REFERENCIAS

- Hill, T. L. 1956. Statistical Mechanics, Principles and Selected Applications, Dover Publications, Inc., New York.
- Koutsourelakis, P. S., 2007. Stochastic upscaling in solid mechanics: An exercise in machine Learning. *Journal of Computational Physics*, 226, 301-325.
- Krogstad, S., Lie, K.-A., Nielsen, H. M., Natvig, J. R., Skaflestad, B., Aarnes, J. E., 2009. A Multiscale Mixed Finite-Element Solver for Three-Phase Black-Oil Flow. *SPE 118993*.
- Nonnenmacher, A., 2011. Adaptive Finite Element Methods for Multiscale Partial Differential Equations. PhD. Thesis, EPFL.
- Pal, M., 2011. A Unified Approach to Simulation and Upscaling of Single-Phase Flow through Vuggy Carbonates, *Int. J. Numer. Meth. Fluids*, on-line publication.
- Planck, M. K. E. L., Wissenschaftliche Selbstbiographie, Barth, Leipzig, 1948.
- Rose, K., 1991. Deterministic annealing clustering and optimization. PhD. Thesis, Caltech, Pasadena, CA.
- Rose, K., 1994. A Mapping approach to rate-distortion computation and analysis, *IEEE Transactions in Information Theory*, vol. 40, pp.1939-1952.
- Rose, K., 1998. Deterministic annealing for clustering, compression, classification regression and related optimization problems. *Proceedings of the IEEE*, vol. 86 (11), pp. 2210-2239.
- Shannon, C. E., 1948. A Mathematical Theory of Communication. Reimpresso com correções a partir de *The Bell System Technical Journal*, Vol. 27, pp. 379-423, pp. 623-656, July, October.
- Weinan, E., Engquist, B., 2003. The heterogeneous multi-scale methods. *Communication in Mathematical Sciences*, Vol.1 (1), pp. 87-132.