

Estimação de Funções de Pedotransferência: Redes Neurais Recorrentes de Pesos Variáveis no Espaço para Predição de Propriedades do Solo

Daniel Takata Gomes, Emanuel Pimentel Barbosa e Luis Carlos Timm

Resumo—O objetivo do artigo é propor um novo modelo de regressão que relacione variáveis do solo de medição difícil ou complexa com outras variáveis de medição mais fácil e menos dispendiosa, visando a predição da primeira com base em dados das demais. As medições são tomadas ao longo de linhas do solo chamadas transeções espaciais ou transects. O estudo dessas relações (funções de pedotransferência), entretanto, apresenta a complexidade da presença simultânea de 3 elementos: dependência espacial dos dados, não-homogeneidade do solo e não-linearidade da relação.

Os principais modelos usualmente considerados na literatura para tais relações (principalmente espaço de estado linear e redes neurais feedforward) têm a limitação de expressarem apenas duas das três características do problema. De modo a superar essa limitação, é proposto aqui um modelo de regressão para a função de pedotransferência baseado em rede neural recorrente (a realimentação ajuda a expressar melhor a dependência espacial), mas, ao contrário de uma rede padrão de pesos fixos, com pesos variando suavemente ao longo do espaço, de modo a incorporar a não-homogeneidade do solo. O algoritmo desenvolvido para estimação do modelo e predição é baseado em extensão não-linear de 2ª ordem do filtro de Kalman, o que corrige substancialmente as deficiências do FK estendido de 1ª ordem. As vantagens comparativas do modelo proposto em relação aos outros modelos propostos na literatura é mostrada, considerando-se diferentes medidas de performance de predição para os extremos do transect.

Index Terms—regressão não-linear, redes neurais recorrentes, funções de pedotransferência, propriedades do solo, transeção espacial, filtro de Kalman.

I. INTRODUÇÃO

O estudo das relações entre os atributos (variáveis) do solo é assunto de grande importância na área agrônômica, de modo a buscar um manejo mais racional e adequado dos recursos naturais do meio ambiente e uma melhoria da produtividade agrícola.

Neste artigo, o principal objetivo é propor um novo modelo de regressão que relacione uma variável do solo de medição complexa (como Nitrogênio Total) com outras variáveis de medição simples (como Carbono Orgânico) no intuito de prever a primeira com base na medição das demais. Essas medições são tomadas ao longo de linhas do solo ditas transects. Tal relação é denominada função de pedotransferência (Bouma, J., 1989), a qual apresenta uma certa dificuldade em sua modelagem devido a presença simultânea de três fatores principais: a dependência espacial dos dados, a não homogeneidade do solo e a não

linearidade da relação em si.

O artigo compreende 5 partes principais, incluindo esta breve introdução, cada uma delas organizada em uma seção distinta, a saber. Primeiro, uma análise preliminar do problema é apresentada na seção II, considerando os principais tipos de modelos propostos na literatura para funções de pedotransferência (Pachepsky et al., 1996; Wosten, J., 1997; Wosten et al., 2001; McBratney et al., 2002; Nielsen & Wendroth, 2003; Timm, L., Barbosa, E. P. et al. 2003). Entre esses, alguns modelos de regressão, espaço de estado linear e redes neurais feedforward são comparados, valendo-se de dados experimentais, com vistas a constatação de suas potencialidades e limitações (eles expressam apenas 2 das 3 características mencionadas presentes no problema).

Com o intuito de superar essas limitações, é apresentado na seção III um novo modelo de regressão para função de pedotransferência que incorpora todas as 3 características desejáveis. Este modelo é baseado em rede neural recorrente com pesos variando suavemente ao longo do espaço segundo um passeio aleatório com pequena variância, visando expressar a não-homogeneidade do solo. O correspondente algoritmo de estimação proposto é baseado em uma extensão não-linear de 2ª ordem do filtro de Kalman, o que corrige substancialmente as limitações do filtro de Kalman estendido de 1ª ordem. Uma comparação da performance preditiva do modelo proposto em relação aos outros modelos sugeridos na literatura é apresentada na seção IV, onde diferentes medidas de performance são consideradas para todos os modelos nos extremos do transect (séries espaciais do Nitrogênio e do Carbono).

As principais conclusões e discussão final são apresentadas na seção V. Após os agradecimentos (seção VI), um apêndice é apresentado com os detalhes do algoritmo de estimação (seção VII), seguido pelas referências bibliográficas.

II. MOTIVAÇÃO AO PROBLEMA: DADOS EXPERIMENTAIS E ANÁLISE PRELIMINAR

Os dados das variáveis de solo (Nitrogênio Total e Carbono Orgânico) foram coletados na estação experimental da EMBRAPA, em Jaguariúna, S.P. As medidas foram tomadas ao longo de uma linha do solo ou transect de 194 m, com 97 pontos amostrais igualmente espaçados, resultando em duas séries espaciais, mostradas nas figs. 1 e 2. A dependência espacial (linear) do Nitrogênio é mostrada através de sua auto-correlação (simples e parcial, fig. 3 e 4), o que sugere auto-dependência de primeira ordem.

D. T. Gomes e E. P. Barbosa: Dep. Estatística, Imecc/Unicamp, CP 6065, Campinas, SP. e-mail: takata@ime.unicamp.br; emanuel@ime.unicamp.br

L. C. Timm: Dep. Eng. Rural, FAEM/UFPel, CP 354, Pelotas, RS. e-mail: lctimm@ufpel.edu.br

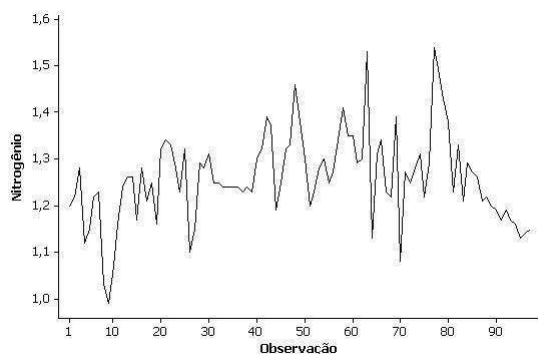


Figura 1. Série espacial do Nitrogênio

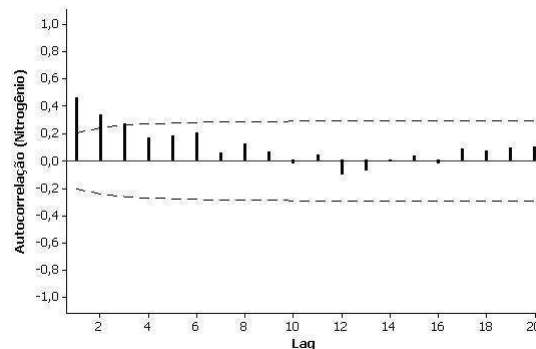


Figura 3. Autocorrelação simples do Nitrogênio

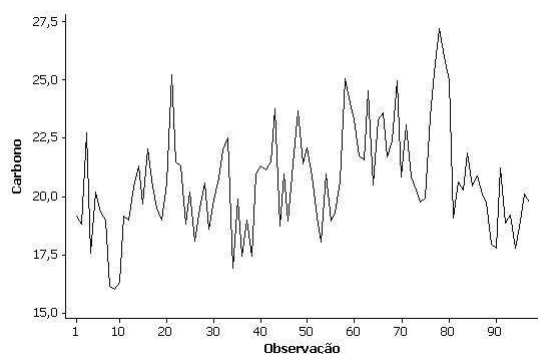


Figura 2. Série espacial do Carbono

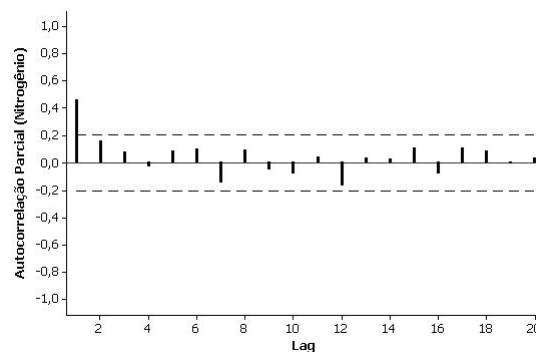


Figura 4. Autocorrelação parcial do Nitrogênio

Em uma análise preliminar, três diferentes tipos de modelos foram considerados. Primeiro, regressão linear (escalar) em duas versões: regressão com erros auto-regressivos de primeira ordem (estimado por mínimos quadrados generalizados como em Greene, W., 2003), onde o regressor é o Carbono, e regressão com coeficientes variando no espaço (estimado pelo filtro de Kalman em forma Bayesiana, como em West & Harrison, 1997, ou Pole, West & Harrison, 1994), onde os regressores são o Carbono (coeficientes estimados mostrados na fig. 6) e Nitrogênio defasado. Desta figura, parece claro que o coeficiente de regressão da relação entre Nitrogênio e Carbono varia fortemente com o espaço, o que expressa a não-homogeneidade do solo.

Segundo, modelos auto-regressivos vetoriais em forma VAR padrão, onde cada variável é usada como resposta contra todas as versões defasadas de todas as variáveis (estimado por métodos de mínimos quadrados, como em Greene, W., 2003) e em forma VAR estrutural, ou seja, espaço de estado linear, formado por uma equação de observação linear e um sistema de evolução Markoviano (estimado pelo filtro de Kalman em conjunto com o algoritmo EM, como em Shumway & Stoffer, 2000). É também considerada uma versão corrigida do VAR padrão, onde a variável Carbono é defasada (antes do ajuste) uma unidade à frente, visando ter um modelo final (após a defasagem uma unidade atrás) com o Carbono sem defasagem, que é um melhor regressor que o Carbono defasado.

Terceiro, modelos de regressão não-linear (não-paramétrica), não apenas modelos aditivos - GAM (estimado pelo algoritmo backfitting, como em Hastie et al., 2001), mas também redes neurais feedforward e recorrentes (estimadas pelos algoritmos Backpropagation padrão e Backpropagation Through Time - BPTT, respectivamente, como em Haykin, 1999, e Williams & Zipser, 1995).

A não-linearidade da relação Carbono \times Nitrogênio é mostrada na fig. 5 através do uso de um suavizador loess (Cleveland, W. S., 1979; Hastie et al., 2001). A normalidade aproximada dos dados do Nitrogênio pode ser vista pelo gráfico de probabilidade da normal (fig. 7).

Os modelos foram ajustados em duas versões. Na primeira, as últimas 10 observações da série do Nitrogênio foram omitidas, visando sua predição. Na segunda, as primeiras 10 observações do Nitrogênio foram omitidas, visando sua predição. Em ambas as versões, o conjunto restante de observações (87 pontos para cada série) foi dividido em duas partes: um conjunto de treinamento ou ajuste com 77 pontos cada e um conjunto de validação com 10 pontos (onde a regra de parada do algoritmo backpropagation, chamada “early stopping”, é implementada visando evitar sobre-ajuste).

As medidas de performance preditiva consideradas para comparação são o Erro Médio Quadrático (MSE), o Erro Médio Absoluto (MAE) e o Erro Médio Absoluto Percentual (MAPE). Essas medidas de performance foram calculadas para todos os modelos considerados nos dois

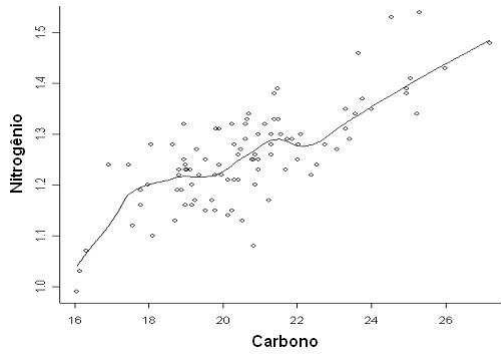


Figura 5. Dispersão Nitrogênio × Carbono

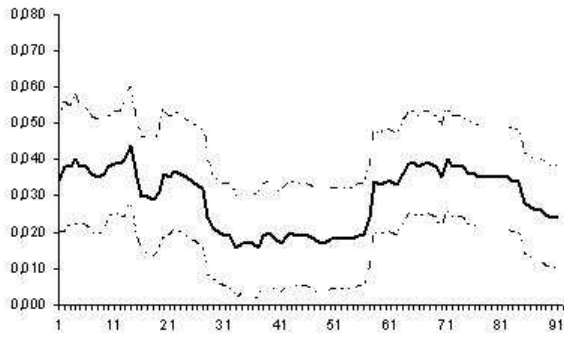


Figura 6. Evolução dos coeficientes do Carbono

experimentos ou versões, e os resultados são apresentados nas tabelas 1 e 2. Na tabela I, é claro que a melhor performance de predição para os últimos pontos do transect é obtida com o modelo de espaço de estado linear (VAR estrutural), independentemente da medida considerada. Para os primeiros 10 pontos, a melhor predição é obtida pela rede neural recorrente (de tipo Elman, apresentada na fig. 8), como mostrado na tabela II. Esse tipo de arquitetura ou topologia mostrou melhor performance que redes com realimentação a partir da última camada (tipo Jordan). A conclusão geral dessa análise preliminar é que os melhores preditores são as redes recorrentes (ao invés de redes feedforward) e/ou o modelo de espaço de estado (VAR estrutural), ambos sugeridos na literatura para esses tipos de dados. No entanto, ambos os modelos não são capazes de superar todos os outros nos dois experimentos, ou seja, nos dois extremos das séries. Isso acontece porque nenhum dos modelos é capaz de lidar ao mesmo tempo com a não-linearidade da relação e parâmetros variando no espaço (heterogeneidade do solo). Embora ambos os modelos incorporem a dependência espacial dos dados, as redes não lidam com a heterogeneidade do solo e o VAR estrutural não contempla a não-linearidade da relação. Por causa dessas limitações, novos modelos são propostos na próxima seção.

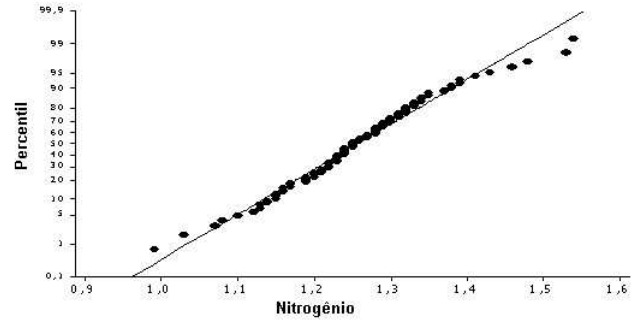


Figura 7. Gráfico de normalidade do Nitrogênio

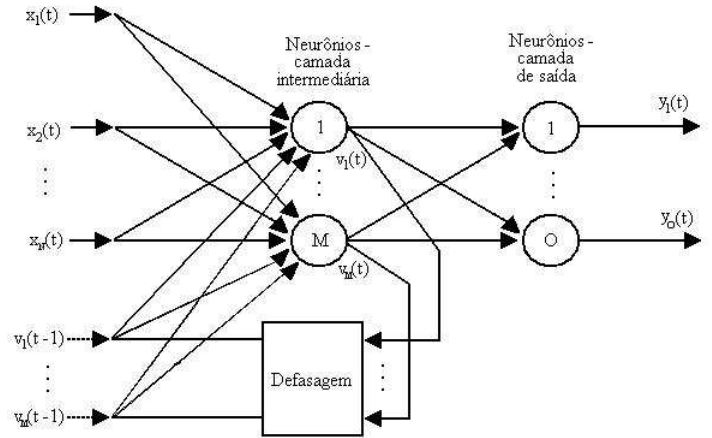


Figura 8. Rede recorrente tipo Elman

III. MODELO PROPOSTO: REDE NEURAL RECORRENTE COM PESOS VARIÁVEIS NO ESPAÇO

O objetivo do modelo proposto é unir a capacidade não-linear das redes neurais com a possibilidade de captação da heterogeneidade do solo pelo modelo de espaço de estado. A combinação dos modelos é feita de modo a representar a estrutura da rede em forma de espaço de estado, com os pesos da rede representando as variáveis de estado a serem estimadas. O modelo é composto pela equação de observação não-linear,

$$y(i) = f(\mathbf{w}(i), \mathbf{x}(i)) + v(i), \quad v(i) \sim N(0; V)$$

e pela equação de evolução (ou equação do sistema),

$$\mathbf{w}(i) = \mathbf{w}(i-1) + \mathbf{u}(i), \quad \mathbf{u}(i) \sim N(0; \Sigma)$$

onde $y(i)$ é o valor observado ou medido do Nitrogênio Total no i -ésimo ponto do transect, $f(\cdot)$ é a função de mapeamento da rede, $\mathbf{x}(i)$ é o vetor dado pelas variáveis exógenas (Carbono Orgânico e Nitrogênio Total defasado) mais as realimentações da rede (ver fig. 8) e $\mathbf{w}(i)$ é o vetor dos pesos da rede. A função de mapeamento f é caracterizada por

Modelos de predição	MSE	MAE	MAPE
Regressão linear escalar com erros AR(1)	0,00389	0,04953	0,04279
Regressão linear escalar c/ coeficientes variáveis	0,00288	0,04610	0,03960
VAR padrão	0,00713	0,07993	0,06390
VAR corrigido	0,00350	0,04791	0,03905
VAR estrutural	0,00096	0,02691	0,02302
GAM/lowess	0,00361	0,05014	0,04084
Rede feedforward	0,00313	0,04308	0,03727
Rede recorrente (Elman)	0,00279	0,04154	0,03599

TABLE I

MEDIDAS DE PERFORMANCE PREDITIVA - 10 ÚLTIMOS PONTOS
(ANÁLISE PRELIMINAR)

Modelos de predição	MSE	MAE	MAPE
Regressão linear escalar com erros AR(1)	0,00475	0,05408	0,04601
Regressão linear escalar c/ coeficientes variáveis	0,00407	0,05280	0,04589
VAR padrão	0,00713	0,07993	0,06390
VAR corrigido	0,00423	0,05082	0,04358
VAR estrutural	0,00314	0,04799	0,04192
GAM/splines	0,00793	0,06650	0,05583
Rede feedforward	0,00344	0,04536	0,038987
Rede recorrente (Elman)	0,00213	0,03323	0,028270

TABLE II

MEDIDAS DE PERFORMANCE PREDITIVA - 10 PRIMEIROS PONTOS
(ANÁLISE PRELIMINAR)

$$f(\mathbf{w}(i), \mathbf{x}(i)) = \sum_{j=0}^k W_j(i) g \left(\sum_{l=0}^n w_{jl}(i) x_l(i) \right)$$

onde $g(\cdot)$ é uma função de ativação sigmoideal ou função de ligação do tipo $g(x) = (1 + e^{-x})^{-1}$, n é o número de entradas, k é o número de neurônios na camada intermediária, $w_{jl}(i)$ e $W_j(i)$ são os pesos da rede no ponto i que formam o vetor $\mathbf{w}(i)$, e $x_l(i)$ são as entradas da rede que formam o vetor $\mathbf{x}(i)$. Ambas as sequências aleatórias $v(i)$ e $\mathbf{u}(i)$ são por suposição independente e normalmente distribuídas com variâncias constantes.

Para a estimação dos pesos, é proposto um algoritmo baseado em uma extensão de 2ª ordem do filtro de Kalman, envolvendo o Jacobiano e a Hessiana da função f . O filtro é considerado em forma Bayesiana, com distribuição a priori normal multivariada-gama inversa para o par (\mathbf{w}, V) , com as variâncias de $\mathbf{u}(i)$ especificadas através de fatores de desconto (West & Harrison, 1997; Pole, West & Harrison, 1994). Com esse esquema de evolução suave e controlada para os pesos, fica introduzida a restrição de que as observações mais recentes (tempo) ou mais próximas (espaço) terão maior influência na construção da função

de predição, o que dá versatilidade no trato de séries não-estacionárias ou não-homogêneas. As equações do algoritmo estão apresentadas em detalhes no apêndice no fim do artigo.

IV. COMPARAÇÃO DE MODELOS

Os dois modelos que mostraram melhor performances preditivas na análise preliminar da seção 2, rede neural recorrente padrão (tipo Elman) e espaço de estado linear (VAR estrutural), são comparados aqui com nosso modelo proposto baseado em rede neural recorrente com pesos variando ao longo do espaço, e os resultados são apresentados nas tabelas III e IV. Como a rede recorrente padrão é um caso particular do nosso modelo quando os pesos são fixos (as variâncias de Σ são nulas, ou equivalentemente, o fator de desconto é 1), o modelo padrão é implementado aqui em duas versões: usando o Toolbox de Redes Neurais do MATLAB e usando nosso algoritmo proposto baseado em extensão de 2ª ordem do filtro de Kalman. Dessas duas tabelas, é óbvio que nosso modelo apresenta melhor performance preditiva que os outros dois modelos concorrentes, independentemente da particular medida de performance considerada.

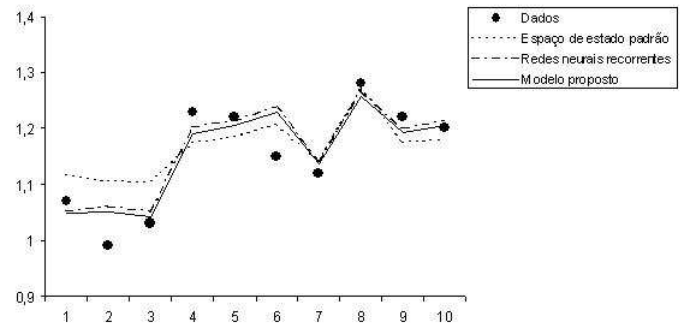


Figura 9. Comparação de modelos (10 primeiros pontos)

V. CONCLUSÕES

O objetivo principal do artigo foi alcançado, ou seja, apresentar um novo e mais adequado modelo de regressão capaz de expressar simultaneamente os três fatores de complexidade do problema (dependência espacial dos dados, não-homogeneidade do solo e não-linearidade da relação), bem como um correspondente algoritmo de estimação, superando as limitações dos dois principais modelos usados para tal tarefa (espaço de estado linear e rede de pesos fixos), conforme mostrado nas tabelas III e IV. O novo modelo introduzido aqui generaliza algumas redes neurais recorrentes padrão, como é o caso da rede Elman, o que aumenta seu potencial de aplicações para predição de séries espaciais ou temporais, não apenas em ciência do solo mas também em outras áreas. Também, como as redes recorrentes com pesos constantes são um caso particular, sua estimação com o filtro de Kalman estendido de 2ª ordem da maneira considerada aqui é um algoritmo de estimação mais simples e eficiente (mais rápido) do que o algoritmo Backpropagation Through Time - BPTT considerado na

Modelos de predição	MSE	MAE	MAPE
Rede recorrente (Elman/gradiente)	0,00279	0,04154	0,03599
Espaço de estado linear (VAR estrutural)	0,00096	0,02691	0,02302
Rede Elman com pesos variáveis (EKF 2ª ordem)	0,00081	0,02334	0,02005
Rede Elman (EKF 2ª ordem, $\Sigma = 0$)	0,00257	0,04032	0,03311

TABLE III

MEDIDAS DE PERFORMANCE PREDITIVA - 10 ÚLTIMOS PONTOS
(COMPARAÇÃO FINAL)

Modelos de predição	MSE	MAE	MAPE
Rede recorrente (Elman/gradiente)	0,00213	0,03323	0,02827
Espaço de estado linear (VAR estrutural)	0,00314	0,04799	0,04192
Rede Elman com pesos variáveis (EKF 2ª ordem)	0,00141	0,03002	0,02599
Rede Elman (EKF 2ª ordem, $\Sigma = 0$)	0,00157	0,03015	0,02614

TABLE IV

MEDIDAS DE PERFORMANCE PREDITIVA - 10 PRIMEIROS PONTOS
(COMPARAÇÃO FINAL)

literatura de engenharia (Williams & Zipser, 1995; Haykin, 1999). Vale também destacar que o algoritmo de 2ª ordem aqui introduzido corrige substancialmente as deficiências do tradicional filtro de Kalman estendido de 1ª ordem.

VI. AGRADECIMENTOS

O primeiro e o terceiro autores agradecem a FAPESP pelo apoio financeiro sob a forma de bolsa mestrado, em andamento, e bolsa pós-doutorado, respectivamente. Os autores são também gratos aos engenheiros agrônomos Manoel D. Souza e José F. Dynia, da EMBRAPA, CNPMA, Jaguariúna, S.P., pelo fornecimento dos dados experimentais aqui utilizados.

VII. APÊNDICE: ALGORITMO FILTRO DE KALMAN ESTENDIDO DE SEGUNDA ORDEM PARA O MODELO PROPOSTO DA SEÇÃO III

Expansão em série de Taylor de segunda ordem da função $f(\cdot)$:

$$y(i) \approx f(\mathbf{a}(i)) + J(\mathbf{a}(i))(\mathbf{w}(i) - \mathbf{a}(i)) + \frac{1}{2}(\mathbf{w}(i) - \mathbf{a}(i))' H(\mathbf{a}(i))(\mathbf{w}(i) - \mathbf{a}(i))$$

onde

$$J(\mathbf{a}(i)) = \left\{ \frac{\partial f(\mathbf{w}(i))}{\partial \mathbf{w}(i)} \right\}_{\mathbf{w}(i)=\mathbf{a}(i)}$$

$$H(\mathbf{a}(i)) = \left\{ \frac{\partial^2 f(\mathbf{w}(i))}{\partial \mathbf{w}(i) \partial \mathbf{w}(i)'} \right\}_{\mathbf{w}(i)=\mathbf{a}(i)}$$

Estimação dos pesos ($\mathbf{w}(i)$):

Inicialização:

- $\mathbf{m}(0) \in [-1, +1]$, como o usual para redes neurais.
- $C(0) = \sigma^2 I$, com σ^2 grande (digamos 10^3).
- $s(0)$ grande (digamos 10^3).
- $n(0) = 0$.
- δ (fator de desconto) $\in [0.95, 1]$.

Para $i = 1, 2, \dots, N$

$$\mathbf{a}(i) = \mathbf{m}(i-1)$$

$$R(i) = \frac{C(i-1)}{\delta}$$

$$\mathbf{m}(i) = \mathbf{a}(i) + A(i)e(i)$$

$$C(i) = (R(i) - A(i)Q(i)A'(i)) \frac{s(i)}{s(i-1)}$$

$$s(i) = s(i-1) \frac{\left(n(i-1) + \frac{e^2(i)}{Q(i)} \right)}{n(i)}$$

$$n(i) = n(i-1) + 1$$

onde

$$A(i) = R(i)J(\mathbf{a}(i))Q^{-1}(i)$$

$$Q(i) = J'(\mathbf{a}(i))R(i)J(\mathbf{a}(i)) + \frac{1}{2} \text{Tr} \{ H(\mathbf{a}(i))R(i) \}^2 + s(i-1)$$

$$f_1(i) = f(\mathbf{a}(i)) + \frac{1}{2} \text{Tr} \{ J(\mathbf{a}(i))R(i) \}$$

$$e(i) = y(i) - f_1(i)$$

Para fórmula dos momentos da forma quadrática resultante da expansão de $f(\cdot)$, que aparecem nas duas últimas equações acima, ver por exemplo McCulloch, C. e Searle, S. (2001). Para outros detalhes sobre o filtro de Kalman em forma Bayesiana e o fator de desconto, ver West, M. e Harrison, J. (1997) ou Pole, West & Harrison (1995); para uma discussão geral dessa categoria de algoritmo (EKF de ordem superior), ver por exemplo Ristic, B., Arulampalam, S. e Gordon, N. (2004).

REFERÊNCIAS

- [1] Bouma, J., *Using soil data for quantitative land evaluation*, Advances in Soil Science, 9, p. 177-213, 1989.
- [2] Cleveland, W. S., *Robust locally weighted regression and smoothing scatterplots*, Journal of the American Statistical Association, 74, p. 829-836, 1979.
- [3] Gomes, D. T., Barbosa, E. P., Timm, L. C., *Pedotransfer Function Estimation via Recurrent Neural Nets with Space-Varying Weights*, RP 50/04, Imecc/Unicamp, 2004.
- [4] Greene, W. H., *Econometrics Analysis*, 5th. edition, Prentice Hall, New Jersey, 2003.
- [5] Hastie, T. J., Tibshirani, R. J., Friedman, J., *The Elements of Statistical Learning*, Springer Verlag, New York, 2001.
- [6] Haykin, S., *Neural Networks - A Comprehensive Foundation*, 2nd edition. Prentice Hall, New Jersey, 1999.
- [7] McBratney, A. B., Minasny, B., Cattle, S. R., Vervoort, R. W., *From pedotransfer functions to soil inference systems*, Geoderma, 109, p. 41-73, 2002.
- [8] McCulloch, C., Searle, S. R., *Generalized, Linear, and Mixed Models*, Wiley, 2001.
- [9] Nielsen, D. R., Wendroth, O., *Spatial and temporal statistics: sampling field soils and their vegetation*, Catena Verlag, Reiskirchen, 2003.
- [10] Pachepsky, Y. A., Timlin, D. J., Varallyay, G., *Artificial Neural Networks to estimate soil water retention from easily measurable data*, Soil Sci. Soc. Am. J., 60, p. 727-733, 1996.

- [11] Pole, A., West, M., Harrison, P. J., *Applied Bayesian Forecasting and Time Series Analysis*, Chapman & Hall, New York, 1994.
- [12] Ristic, B., Arulampalam, S., Gordon, N., *Beyond the Kalman Filter*, Artech House, London., 2004.
- [13] Shumway, R. H., Stoffer, D. S., *Time Series Analysis and Its Applications*, Springer Verlag, New York, 2000.
- [14] Timm, L. C., Barbosa, E. P., Souza, M. D., Dynia, J. F., Reichardt, K., *State-Space Analysis of Soil Data: An Approach Based on Space-Varying Regression Models*, Scientia Agricola, 60, p. 371-376, 2003.
- [15] West, M., Harrison, P. J., *Bayesian Forecasting and Dynamic Models*, Springer, London, 1997.
- [16] Williams, R. J., Zipser, D., *Gradient-Based Learning Algorithms for Recurrent Networks and Their Computational Complexity*, In: Backpropagation: theory, architectures and applications. Ed: Chauvyn, Y., Rumelhart, D. E., Erlbaum, Hillsdale, New Jersey (1995).
- [17] Wosten, J. H. M., *Pedotransfer functions to evaluate soil quality*. In: Soil quality for crop production and ecosystem health. (edited by E. G. Gregorich and M. R. Carter), Elsevier, pp 221-245, 1997.
- [18] Wosten, J., Pachepsky, Y., Rawls, W., *Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics*, Journal of Hydrology, vol. 251, pp 123-150, 2001.