Recurrent Neural Net Regression Models with space-varying coefficients for pedotransfer function estimation and prediction of soil properties

Daniel Takata Gomes and Emanuel Pimentel Barbosa
Dep. Estatística, Imecc/Unicamp, CP 6065, Campinas, SP.
e-mail: takata@ime.unicamp.br; emanuel@ime.unicamp.br
Luis Carlos Timm - Dep. Eng. Rural, FAEM/UFPel
CP 354, Pelotas, RS. e-mail: lctimm@ufpel.edu.br

Abstract

The paper aim is to propose a new regression model for relating soil variables of difficult or complex measurement with other variables easier to measure, in order to predict the first one based on data about the last ones. The measurements are taken along soil lines called transects. The study of these relations (pedotransfer functions) presents the complexity of simultaneous presence of 3 elements: data spatial dependence, soil non-homogeneity and non-linearity of the relationship.

The main models usually considered in the literature for such relations (namely, linear state-space and feedforward neural nets) have the limitation of expressing only two of these 3 characteristics of the problem. In order to overcome such limitations, it is proposed here a regression model for pedotransfer mapping based on recurrent neural nets (the feedback helps to better express the spatial dependence), but with weights varying smoothly along the space in order to incorporate the soil non-homogeneity. The algorithm developed for model estimation and prediction is based on a second order non-linear extension of the Kalman filter in Bayesian form. The comparative advantages of the proposed model in relation to the other ones are shown, considering different prediction performance measures for the transect extremes.

Keywords: non-linear regression, recurrent neural nets, pedotransfer functions, soil properties, transect, Kalman filter extension.

1 Introduction

The study of the soil properties is a subject of great interest in the agronomic area. In this paper the main objective is to propose a new regression model to relate a soil variable of complex measurement (such as the total Nitrogen) with other variables of simpler measure (such as the organic Carbon) in order to predict the first one based on data about the second. These measurements are taken along soil lines called transects. Such relationships, called pedotransfer functions (Bouma, J., 1989) present a non-trivial modelling due to the simultaneous presence of 3 main elements: data spatial dependence, soil non-homogeneity and non-linearity of the relationship.

The paper comprises 6 main parts, including this brief introduction, each one organized in a distinct section, as follows. First, a preliminary analysis of the problem is presented in section 2, considering the main types of models proposed in the literature for pedotransfer functions (Pachepsky et al., 1996; Wösten, J., 1997; Wösten et al., 2001; McBratney et al., 2002; Nielsen & Wendroth, 2003; Timm, L., Barbosa, E. P. et al. 2003); among these, some regression models, linear state-space and feedforward neural nets are tested with experimental data in order to check their potentialities and limitations (they express only 2 of the 3 mentioned characteristics present in the problem).

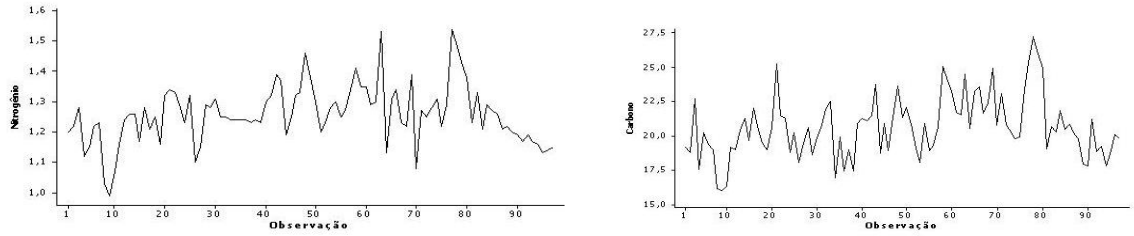
In order to overcome theses limitations, it is presented in section 3 a new regression model for pedotransfer mapping that incorporates all the 3 desired characteristics. This model is based on recurrent neural nets with weights varying smoothly along the space (transect) in order to express the soil non-homogeneity, and the corresponding estimation algorithm proposed is based on a second-order non-linear extension of the Kalman filter. A comparison of the predictive performance of the proposed model in relation to the other ones is presented in section 4, where different performance measures are considered for all models at the transect (spatial series of Nitrogen and Carbon) extremes.

The main conclusions and final discussion are presented in section 5. After the acknowledgments, the bibliographical references are presented, followed by an appendix with the estimation algorithm details (section 6).

2 Experimental Data and Preliminary Analysis

The data about the soil variables (total Nitrogen and organic Carbon) were collected at the Experimental Unit of EMBRAPA - the Brazilian Institute for Agricultural Research, in Jaguariúna, S.P. The measurements were taken along a soil line or transect of 194 m, with 97 sampling points equally spaced, resulting in two spatial series, shown at fig. 1(a) and fig. 1(b). The Nitrogen spatial (linear) dependence is shown through its auto-correlation (simple and partial, fig. 2(a) and 2(b)), which suggest a first order type of auto-dependence.

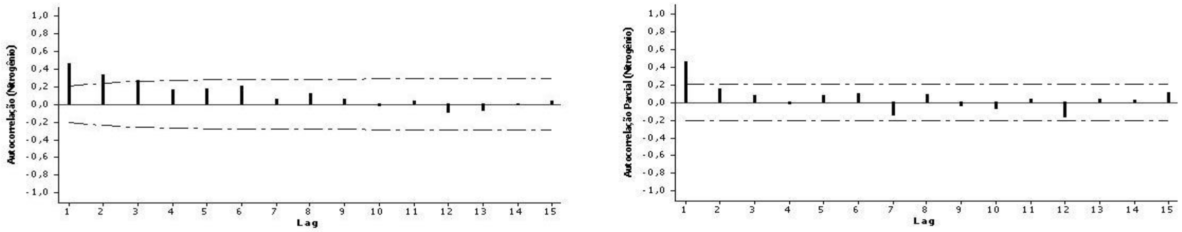
In a preliminary analysis, three different types of models were considered. First, (scalar) linear regression in two different versions: regression with first order auto-regressive errors (estimated by generalized least



(a) Nitrogen

(b) Carbon

Figure 1: Spatial series



(a) N simple auto-correlations

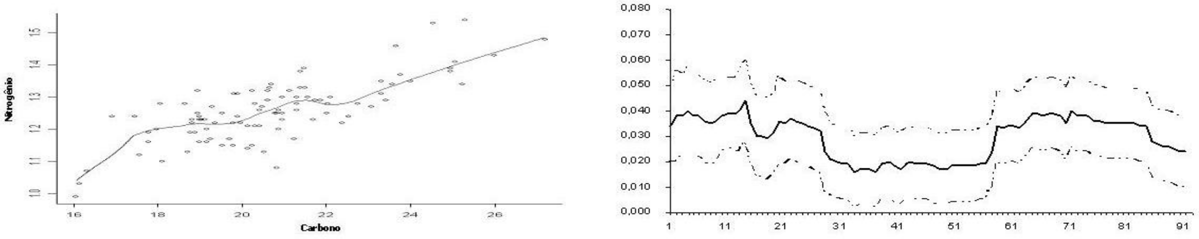
(b) N partial auto-correlations

Figure 2: Nitrogen auto-correlations

squares, as in Greene, W., 2003), where the regressor is the Carbon, and regression with coefficients varying in space (estimated by the Kalman filter in Bayesian form, as in West and Harrison, 1997), where the regressors are the Carbon (estimated coefficients shown at fig. 3(b)) and the lagged Nitrogen. From this figure, it is clear that the regression coefficient of the Nitrogen versus Carbon relation changes markedly with space, which express the soil non-homogeneity.

Second, vector auto-regressive models in standard VAR form, where each variable is regressed against the lagged versions of all variables (estimated by least squares methods as in Greene, W., 2003) and in structural VAR form, that is, linear state-space, formed by a linear observation equation and a Markovian system evolution (estimated by the Kalman filter coupled with the EM algorithm, as in Shumway & Stoffer, 2000). It is also considered a corrected version of the standard VAR where the Carbon variable is lagged (previously) 1 unit forward in order to have a final model (after the backward lagging) with the Carbon without lagging, which is a better regressor than the lagged Carbon.

Third, non-linear regression models (non-parametric), not only additive models - GAM (estimated by the backfitting algorithm, as in Hastie et al., 2001), but also feedforward and recurrent neural nets (estimated



(a) N × C scatterplot

(b) C coefficients evolution

Figure 3: Nitrogen × Carbon dependence

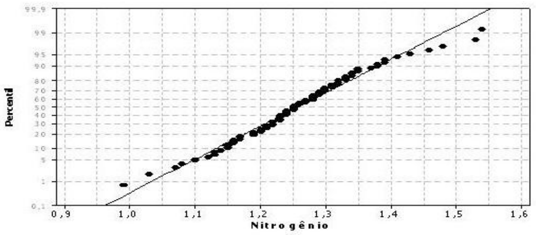


Figure 4: Nitrogen normal probability plot

by the backpropagation and BPTT algorithms, respectively, as in Haykin, S., 1999).

The non-linearity of the Carbon × Nitrogen relationship is show at fig. 3(a) through the use of the loess smoother (Hastie et al., 2001). The approximate normality of the Nitrogen data can be seen by the normal probability plot (fig. 4).

The models were fitted in two versions. In the first one, the last 10 observations of the Nitrogen series were omitted in order to make their prediction. In the second, the first 10 observations were omitted in order to predict them. In both versions, the remaining observation set (87 points for each series) was divided into two parts: a training or fitting set with 77 points each and a validation set with 10 points (where the stopping rule of the backpropagation algorithm, called "early stopping", is implemented in order to avoid overfitting).

The prediction performance measures considered for comparison purposes are the Mean Square Error - MSE, the Mean Absolute Error - MAE and the Mean Absolute Percentual Error - MAPE. These performance measures were calculated for all models considered in the two experiments or versions, and the results are presented at tab. 1 and tab. 2. From tab. 1 it is clear that the best prediction performance for the last transect points is obtained with the linear state-space model, independently of the measure considered. For the first 10 points, the better prediction is obtained through the recurrent neural net (of Elman type architecture, shown at fig. 5), as shown in the tab. 2. This kind of architecture or topology of network with feedback from the intermediate stage has shown better performance than the nets with feedback from the last stage (Jordan type). The general conclusion from this preliminary analysis is that the best predictors are the recurrent nets (instead of the feedforward nets) and/or the state-space model, both suggested in the literature for this kind of data. However, both models are not able to outperform all the others in the two experiments, that is, at the two extremes of the series. This happens because neither models are able to deal at the same time with non-linearity in the relationship and parameters changing in space (soil heterogeneity). Because of these limitations, new models are proposed at the next section.

Prediction models		MSE	MAE	MAPE
Scalar linear regression	With AR(1) errors	0.00389	0.04953	0.04279
	With varying coefficients	0.00288	0.04610	0.03960
Vector auto-regression VAR model	Standard VAR model	0.00713	0.07993	0.06390
	Corrected VAR	0.00350	0.04791	0.03905
	Structural VAR	0.00096	0.02691	0.02302
Nonlinear regression (non-parametric)	GAM/lowess	0.00361	0.05014	0.04084
	Feedforward neural nets	0.00313	0.04308	0.03727
	Recurrent neural nets (Elman)	0.00279	0.04154	0.03599

Table 1: Predictive performance measures (last 10 points)

Prediction models		MSE	MAE	MAPE
Scalar linear regression	With AR(1) errors	0.00475	0.05408	0.04601
	With varying coefficients	0.00407	0.05280	0.04589
Vector auto-regression VAR model	Standard VAR model	0.00713	0.07993	0.06390
	Corrected VAR	0.00423	0.05082	0.04358
	Structural VAR	0.00314	0.04799	0.04192
Nonlinear regression (non-parametric)	GAM/splines	0.00793	0.06650	0.05583
	Feedforward neural nets	0.00344	0.04536	0.03898
	Recurrent neural nets (Elman)	0.00213	0.03323	0.02827

Table 2: Predictive performance measures (first 10 points)

3 Proposed Regression Model: Recurrent Neural Net with space varying coefficients

The aim of the proposed model is to join the non-linear modelling ability of the neural nets with the capacity of dealing with the soil heterogeneity of the state-space model. The models combination is made in order to represent the network structure in state-space form, with the net weights being represented by the state variables to be estimated. The model is given by the non-linear observation equation,

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

and the system or evolution equation,

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

where $y(i)$ is the observed or measured value of total Nitrogen at the i^{th} point in the transect, $f(.)$ is the network mapping function, $\mathbf{x}(i)$ is the vector given by the exogenous variables (organic Carbonic and lagged total Nitrogen) plus the net feedbacks (see fig. 5) and $\mathbf{w}(i)$ is the vector of net weights. The mapping function f is given by

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

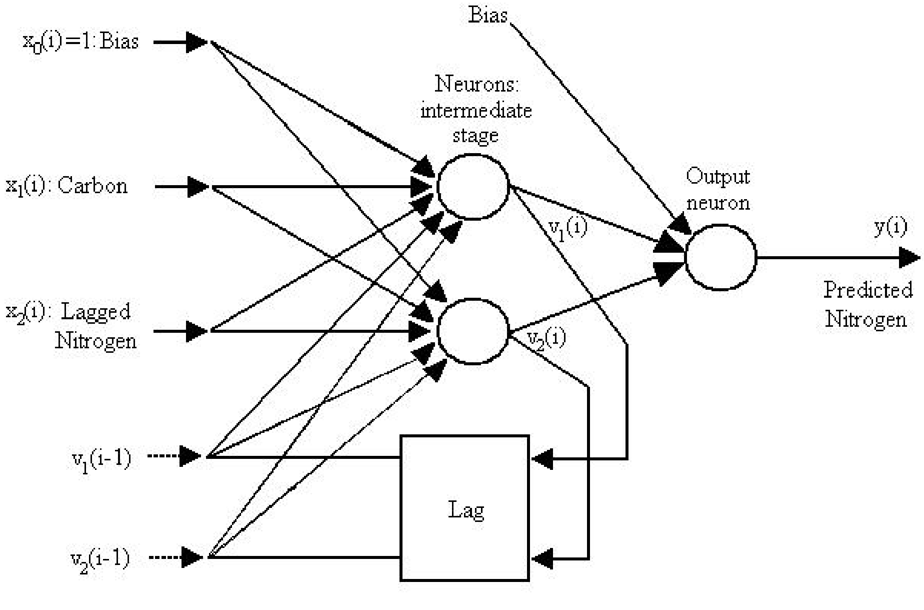


Figure 5: Recurrent neural network (Elman) for the Nitrogen spatial series

where $g(.)$ is a sigmoidal activation or link function of the type $g(x) = (1 + e^{-x})^{-1}$, n is the number of inputs, k is the number of neurons in the intermediate layer, $w_{jl}(i)$ and $W_j(i)$ are the net weights at the point i which form the vector $\mathbf{w}(i)$, and $x_l(i)$ are the net inputs which form the vector $\mathbf{x}(i)$. Both random sequences $v(i)$ and $\mathbf{u}(i)$ are supposed to be independent normal variables with constant variances.

For the weight estimation, it is proposed an algorithm based on a second order extension of the Kalman filter, involving the Jacobian and the Hessian of the f function. The filter is considered in Bayesian form, with multivariate normal-inverse gamma prior distribution for the pair $(\mathbf{w}, V)$, with the variances of $\mathbf{u}(i)$ specified through discount factors (West & Harrison, 1997). The full algorithm equations are presented in detail in the appendix at the end of the paper.

4 Prediction Models Comparison

Prediction models		MSE	MAE	MAPE
Standard models with latent structure	Recurrent neural net (Elman/gradient)	0.00279	0.04154	0.03599
	Linear state-space (structural VAR)	0.00096	0.02691	0.02302
Proposed model and its particular case	Elman network with varying weights (2nd. order EKF)	0.00081	0.02334	0.02005
	Elman network (2nd. order EKF, $\Sigma = 0$)	0.00257	0.04032	0.03311

Table 3: Predictive performance measures (last 10 points)

Prediction models		MSE	MAE	MAPE
Standard models with latent structure	Recurrent neural net (Elman/gradient)	0.00213	0.03323	0.02827
	Linear state-space (structural VAR)	0.00314	0.04799	0.04192
Proposed model and its particular case	Elman network with varying weights (2nd. order EKF)	0.00141	0.03002	0.02599
	Elman network (2nd. order EKF, $\Sigma = 0$)	0.00157	0.03015	0.02614

Table 4: Predictive performance measures (first 10 points)

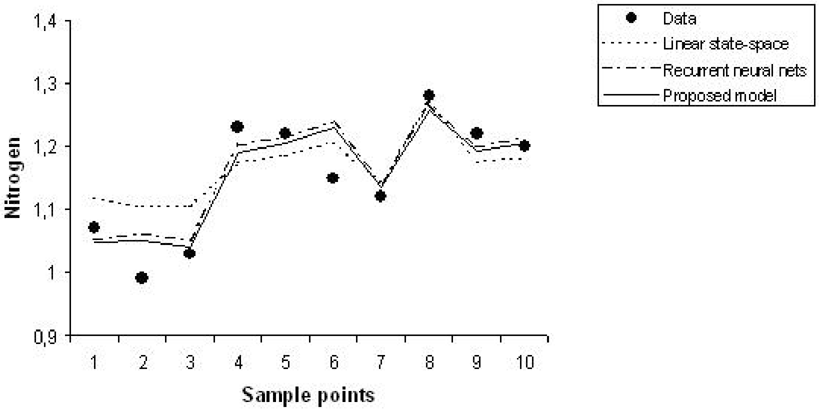


Figure 6: Prediction comparison (first 10 points)

The two models that have shown better predictive performances from the preliminary analysis of section 2, the standard recurrent neural net (of Elman type) and the linear state-space (structural VAR model), are compared here with our proposed model based on recurrent nets with space-varying weights, and the results are presented in tables 3 and 4. Since the standard recurrent net is a particular case of our model when the eights are fixed (the variances in Σ are null, or equivalently, the discount factor is 1), this standard model is implemented here in two versions: using the MATLAB Neural Net Toolbox and using our proposed algorithm based on the second order extended Kalman filter. From these two tables, it is clear that our model presents better predictive performance than the other two competing models, independently of the particular measure of performance considered.

5 Conclusions and Final Discussion

The main aim of this paper was achieved, that is, to present a new and more adequate regression model capable of expressing simultaneously the three complexity factors present in the pedotransfer mapping problem (data spatial dependence, soil non-homogeneity and non-linearity of the relationship) as well a corresponding estimation algorithm, overcoming therefore the drawbacks of the two main models used for such task. The new model introduced here generalizes some standard recurrent neural nets, which increase their potential of applicabilities for prediction of spatial or time series, not only in soil sciences but also in other areas. Also, since the constant weights recurrent net is a particular case, its estimation with the 2nd. order Kalman filter in the way considered here is a simpler and more efficient learning algorithm than the usual back-propagation through time - BPTT method considered in the engineering literature (Haykin, S., 1999).

Acknowledgments The authors are grateful to Dr. Manoel D. Souza and Dr. José F. Dynia from EMBRAPA/CNPMA, Jaguariúna, S.P., for providing the experimental data that we have used in this paper. Also, the first and the third authors would like to thanks FAPESP for financial support under the form of master grant, under way, and pos-doc grant, respectively.

References

[1] Bouma, J. (1989). *Using soil data for quantitative land evaluation.* Advances in Soil Science, 9, p. 177-213.

[2] Greene, W. H. (2003). *Econometrics Analysis.* 5th. edition. Prentice Hall, New Jersey.

[3] Hastie, T. J., Tibshirani, R. J., Friedman, J. (2001). *The Elements of Statistical Learning.* Springer Verlag, New York.

[4] Haykin, S. (1999). *Neural Networks - A Comprehensive Foundation.* 2nd edition. Prentice Hall, New Jersey.

[5] McBratney, A. B., Minasny, B., Cattle, S. R., Vervoort, R. W. (2002). *From pedotransfer functions to soil inference systems.* Geoderma, 109, p. 41-73.

[6] McCulloch, C., Searle, S. R. (2001). *Generalized, Linear, and Mixed Models.* Wiley.

[7] Nielsen, D. R., Wendroth, O. (2003). *Spatial and temporal statistics: sampling field soils and their vegetation.* Catena Verlag, Reiskirchen.

[8] Pachepsky, Y. A., Timlin, D. J., Varallyay, G. (1996). *Artificial Neural Networks to estimate soil water retention from easily measurable data.* Soil Sci. Soc. Am. J., 60, p. 727-733.

[9] Ristic, B., Arulampalam, S., Gordon, N. (2004). *Beyond the Kalman Filter.* Artech House, London.

[10] Shumway, R. H., Stoffer, D. S. (2000). *Time Series Analysis and Its Applications.* Springer Verlag, New York.

[11] Timm, L. C., Barbosa, E. P., Souza, M. D., Dynia, J. F., Reichardt, K. (2003). *State-Space Analysis of Soil Data: An Approach Based on Space-Varying Regression Models.* Scientia Agricola, 60. p. 371-376.

[12] West, M., Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Model.* Springer, London.

[13] Wösten, J. H. M. (1997). *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, p. 221-245.

[14] Wösten, J. H. M., Pachepsky, Y. A., Rawls, W. J. (2001). *Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics.* Journal of Hydrology, 251, p. 123-150.

6 Appendix: A Second Order Extended Kalman Filter Algorithm for the Proposed Model of Section 3

Second order Taylor series expansion of the $f(.)$ mapping:

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

where

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

Weights $(\mathbf{w}\,(i))$ estimation:

Initialization:

- $\mathbf{w}\,(0) \in [-1, +1]$, as usual for neural nets weights.
- $C\,(0) = \sigma^2 I$, with σ^2 large (say 10^3).
- $s\,(0)$ large (say 10^3).
- $n\,(0) = 0$.
- δ (discount factor) $\in [0.95, 1]$.

For $i = 1, 2, \ldots, N$

$$\begin{aligned} \mathbf{a}(i) &= \mathbf{w}(i-1) \\ R(i) &= \frac{C(i-1)}{\delta} \\ \mathbf{w}(i) &= \mathbf{a}(i) + A(i)\,\overbrace{(y(i) - f_1(i))}^{e(i)} \\ C(i) &= (R(i) - A(i)\,Q(i)\,A'(i))\,\frac{s(i)}{s(i-1)} \\ s(i) &= s(i-1)\,\frac{\left(n(i-1) + \frac{e^2(i)}{Q(i)}\right)}{n(i)} \\ n(i) &= n(i-1) + 1 \end{aligned}$$

where

$$\begin{aligned}
A(i) &= R(i)J(\mathbf{a}(i))Q^{-1}(i) \\
Q(i) &= J'(\mathbf{a}(i))R(i)J(\mathbf{a}(i)) + \frac{1}{2}\,\mathrm{Tr}\left\{H(\mathbf{a}(i))R(i)\right\}^2 + s(i-1) \\
f_1(i) &= f(\mathbf{a}(i)) + \frac{1}{2}\,\mathrm{Tr}\left\{J(\mathbf{a}(i))R(i)\right\}
\end{aligned}$$

For the moments formulae of quadratic form resulting from the expansion of $f(.)$, which appear in the last two equations above, see for instance McCulloch, C. and Searle, S. (2001). For other details about the Kalman filter in Bayesian form and the discount factor, see West, M. and Harrison, J. (1997); for a general discussion of this category of algorithm (higher-order EKF), see for instance Ristic, B., Arulampalam, S. and Gordon, N. (2004).