ISSN 2238-1007

UNCERTAINTIES 2012

Proceedings of the 1st International Symposium on Uncertainty Quantification and Stochastic Modeling
February 26th to March 2nd, 2012, Maresias, São Sebastião, SP, Brazil

Data assimilation by neural network emulating representer method applied to the wave equation

Helaine C. M. Furtado[1], **Haroldo F. de Campos Velho**[2], **Elbert E. N. Macau**[2]

[1] Pos Applied Computing, helaine.furtado@lac.inpe.br
[2] Computing and Applied Mathematics Laboratory (LAC),haroldo@lac.inpe.br and elbert@lac.inpe.br
National Intitute For Space Research - INPE
Avenida do Astronautas, 1758, São José dos Campos, SP, Brazil

Abstract. *Description of a physical phenomenon through differential equations has errors involved, since the mathematical model is always an approximation of reality. For an operational prediction system, one strategy to deal with uncertainties from the modeling and observation errors is to add some information from the real dynamics into mathematical model. This aditional information consists of observations on the phenomenon. However, the observational data insertion should be done carefully, for avoiding a worse performance of the prediction. Technical data assimilation are tools to combine data from physical-mathematics model with observational data to obtain a better forecast. Two data assimilation methods are compared here: the Kalman Filter method, and artificial neural network. Artificial neural networks appear as a novel method in the context for data assimilation. The performance of the methods is evaluated under application to wave propagation model (Bennet,2002).*

Keywords. *data assimilation, neural network, variational method, representer method, wave equation.*

1 INTRODUCTION

In recent years, data assimilation methods have been developed and used in many research areas, such as: numerical weather prediction (NWP) (Daley, 1993; Kalnay, 2003); ocean circulation forecasting (Bennett, 2002); air monitoring (Zannetti, 1990); and space weather models (Garner, 1999; Schunk et.all, 2004). However, the first application of data assimilation techniques was in meteorology, and today it is a key component of numerical weather and climate forecasts. Satellites and *in situ* measurements are routinely providing new atmospheric and oceanographic data and bringing the daily practice of physical oceanography closer to that of dynamic meteorology (Belyaev, 2000). Essentially, NWP consists of the integration of Navier-Stokes equation using numerical procedures. Therefore, after some time-steps, there is a disagreement due to small uncertainties in the a specification of the initial conditions and model errors. In other words, the sensitive dependence on the initial conditions causes the forecasting error to grow exponentially with time (Grebogi, 1987). The problem of estimating the initial conditions is called *data assimilation* that improves the estimate of the ocean and atmosphere physical state by combining the data from measurements and from dynamic models in an optimal way.

Many methods have been developed for data assimilation (Kalnay, 2003; Härter, 2008). They have different strategies to combine the forecasting (*background*) and observations. From a mathematical point view, the assimilation process can be represented by:

$$\mathbf{x}^a = \begin{cases} \mathbf{x}^f + \mathbf{K}(\mathbf{y} - \mathbf{H}[\mathbf{x}^f]), & \text{estimation theory} \\ \nabla J(\mathbf{x}), & \text{variational method} \\ \mathbf{F}_w(\mathbf{x}^f, \mathbf{y}^o), & \text{artificial intelligence} \\ \mathbf{x}^f + \sum_{m=1}^{M} \beta_m r_m(x,t), & \text{representer method} \end{cases} \tag{1}$$

For methods based on the estimation theory, it can be cited the Optimal interpolation (Daley, 1993), Kalman Filter (KF) (Kalnay, 2003), and Particle Filter (PF) (Nakano, 2007). In the analysis $\mathbf{x}^a$ for KF, $\mathbf{x}^f$ is the forecasting (from mathematical model, also kown as *background* field); $\mathbf{K}$ is the weight matrix (Kalman gain), generally computed from the covariance matrix of the predictions errors from forecasting and observation; $\mathbf{y}$ denotes the observation; $\mathbf{H}$ represents the observation system. For the variational methods, the analysis is obtained by the minimizing a functional. This method will be described in the section 3.1. For the method based on artificial intelligence analysis is determined by mapping the input data with the desired output data that described in the section 3.2. The representer method is present in the section 3.1.

2 WAVE EQUATION: A TESTING MODEL

Data assimilation is a necessary process for operational prediction systems. The insertion of observational data can degenerate the prediction. Therefore, techniques allowing a soft introduction of the observations in a mathematical prediction model are called data assimilation. Two schemes will be compare for data assimilation, and the testing model is the linear wave equation, following the tests described by Bennet (2002).

2.1 Linear wave equation

The governing equation is given by:

$$\frac{\partial \eta_F}{\partial t} + c\frac{\partial \eta_F}{\partial x} = F(x,t) \tag{2}$$

where η is the displacement, c is the constant phase speed, F is the external forcing, t is time, and x is space. The subscript F indicates the forward solution (or background). The equation 2 can be thought of as a convection equation, where η is concentration, and c is the convection velocity.

By adding observations data into the model 2 at isolated points in time and space, where observations are imperfect measures of the variable $\eta(x,t)$, the problem becomes overdetermined, i. e., there are no smooth solutions that satisfy the model and observations simultaneously (Bennett, 2002). Therefore, the problem is to determine the solution through a building held by weighted least squares adjustment between observations and model.

2.2 Data

It assumes a finite number of observations collected within a spatial domain ($0 \leq x \leq L$), and temporal domain ($0 \leq t \leq T$). These data are imperfect point measurements of the independent variable $\eta(x,t)$ collected at M points in space and time (x_m,t_m):

$$d_m = \eta(x_m,t_m) + \varepsilon_m \quad 1 \leq m \leq M \tag{3}$$

where $\eta(x,t)$ is the unknown, the true displacement field, and ε_m is the measurement error. Considering that the governing equation, the forcing, the initial condition, and the observations contain errors, a perfect consistency with model and observations is not expect, therefore,

$$\eta_F(x,t) \neq d_m \quad 1 \leq m \leq M \ , \tag{4}$$

or for the discrete case:

$$d_m = \eta_{p_m}^{k_m} + \varepsilon_m \quad 1 \leq m \leq M \ . \tag{5}$$

Figure 1 illustrates the domain of integration of the model, with grid points and the observations, where p is counter space, and k is counter temporal.

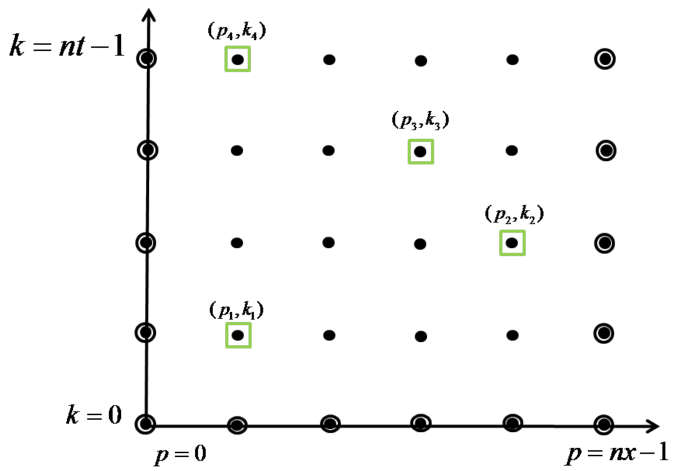


Figure 1: Outline of the model grid points (represented by black circles), the observations represented by the green square, which are distributed at the grid point.

2.3 The residuals

The true displacement $\eta(x,t)$ must satisfy the equations below

$$\frac{\partial \eta_F}{\partial t} + c\frac{\partial \eta_F}{\partial x} = F(x,t) + f(x,t) \quad 0 \leq x \leq L \tag{6}$$

$$\eta(x,0) = I(x) + i(x) \quad 0 \leq t \leq T \tag{7}$$

where $f(x,t)$ represents the error in forcing, $i(x)$ the error in the initial condition. For the periodic boundary condition we have $\eta(0,t) = \eta(L,t)$ com $0 \leq t \leq T$.

Therefore, the equations (3), (6), and (7) correspond to residual observation, modeling, and initial condition, respectively.

ISSN 2238-1007

Proceedings of the 1st International Symposium on Uncertainty Quantification and Stochastic Modeling
February 26th to March 2nd, 2012, Maresias, São Sebastião, SP, Brazil

3 DATA ASSIMILATION METHODS

Two schemes for data assimilation are described: the representer method (based on variational calculus), and assimilation by artificial neural networks (using multi-layer percpetron: MLP). The model for testing and comparison is the evolution wave equation.

3.1 Variational method: the representer approach

Let a forcing $F(x,t)$, an initial condition $I(x)$, and a periodic boundary condition. There is a unique forward solution, $\eta_F(x,t)$, to the linear wave equation. Those arguments also imply that for any choice of $F(x,t)+f(x,t)$, and $I(x)+i(x)$, there is unique true solution $\eta(x,t)$ (Bennett, 2002).

However, the error fields ($f(x,t)$, $i(x)$, and ε_m) are unknown. At this point, we define an optimal solution $\hat{\eta}(x,t)$ to be the solution corresponding to the smallest $f(x,t)$, $i(x)$, and ε_m in a weighted, least-squares sense. The objective is to seek the minimum of the quadratic penalty function as follows:

$$J = J[\eta(x,t)] = W_f \int_0^T \int_0^L \{f(x,t)\}^2 dt dx + W_i \int_0^L \{i(x)\}^2 dx + w \sum_{m=1}^{M} \{\varepsilon_m\}^2 \tag{8}$$

where W_f, W_i, and w are positive and constant. These weights are the inverse covariance operators of error model, initial condition, and observations, respectively (Chua, 2001).

Rewriting the equation 8 with $f(x,t)$, $i(x)$, and ε_m replaced by their definitions makes dependece on $\eta(x,t)$, given by:

$$J = J[\eta(x,t)] \equiv W_f \int_0^T \int_0^L \left\{ \frac{\partial \eta}{\partial t} + c\frac{\partial \eta}{\partial x} - F(x,t) \right\}^2 dt dx + W_i \int_0^L \{\eta(x,0) - I(x)\}^2 dx + w \sum_{m=1}^{M} \{\eta(x_m,t_m) - d_m\}^2 \tag{9}$$

The objective function (9) is one single scalar number for each field $\eta(x,t)$, while $\eta(x,t)$ is a field of values for $0 \le x \le L$ e $0 \le t \le T$. The minimization of functional (9) is done through the calculus of variations. It was defined $\hat{\eta}(x,t)$ as the local extremum, that is, it corresponds to the smallest value of the cost functional, and thus to the smallest $f(x,t)$, $i(x)$, and ε_m. As the penalty functional is quadratic, the local extreme is a global extreme, and functional is not negative. Therefore, your extreme is a global minimum.

Now, considering the Taylor series expansion of J, around the point $\hat{\eta}$:

$$J[\hat{\eta} + \delta\eta] = J[\hat{\eta}] + \frac{\partial}{\partial \hat{\eta}} J(\hat{\eta})\delta\eta + \frac{1}{2!}\frac{\partial^2}{\partial^2 \hat{\eta}} J(\hat{\eta})(\delta\eta)^2 + \ldots \tag{10}$$

the above functional can be written as:

$$J[\hat{\eta} + \delta\eta] = J[\hat{\eta}] + \bigtriangledown J(\eta)\delta\eta + O(\delta\eta)^2 \; . \tag{11}$$

and considering a small quantity $\delta\eta(x,t)$, the following approximation works:

$$J[\hat{\eta} + \delta\eta] - J[\hat{\eta}] \sim \bigtriangledown J(\eta)\delta\eta \tag{12}$$

The first variation for the functional J is written by:

$$\delta J \equiv J[\hat{\eta} + \delta\eta] - J[\hat{\eta}] \; . \tag{13}$$

The explicit form for functional (9) is written by $J[\hat{\eta}]$ as:

$$J[\hat{\eta}] = W_f \int_0^T \int_0^L \left\{ \frac{\partial \hat{\eta}}{\partial t} + c\frac{\partial \hat{\eta}}{\partial x} - F(x,t) \right\}^2 dt dx + W_i \int_0^L \{\hat{\eta}(x,0) - I(x)\}^2 dx + w \sum_{m=1}^{M} \{\hat{\eta}(x_m,t_m) - d_m\}^2 \tag{14}$$

and the expression for $J[\hat{\eta} + \delta\eta]$ is given by:

$$J[\hat{\eta} + \delta\eta] = W_f \int_0^T \int_0^L \left\{ \frac{\partial \hat{\eta}}{\partial t} + \frac{\partial \delta\eta}{\partial t} + c\frac{\partial \hat{\eta}}{\partial x} + c\frac{\partial \delta\eta}{\partial x} - F \right\}^2 dt dx + W_i \int_0^L \{\hat{\eta}(x,0) + \delta\eta(x,0) - I(x)\}^2 dx + w \sum_{m=1}^{M} \{\hat{\eta}(x_m,t_m) + \delta\eta(x_m,t_m) - d_m\}^2 \; . \tag{15}$$

Subtracting the equations (14) from (15), and neglecting the variations of the second order, yields:

$$\delta J = 2W_f \int_0^T dt \int_0^L dx \left\{ \frac{\partial \hat{\eta}}{\partial t} c\frac{\partial \delta\eta}{\partial x} - F(x,t) \right\} \left\{ \frac{\partial \delta\eta}{\partial t} + c\frac{\partial \delta\eta}{\partial x} \right\} + 2W_i \int_0^L dx \{\hat{\eta}(x,0) - I(x)\}\, \delta\eta(x,0)\} + 2w \sum_{m=1}^{M} \{\hat{\eta}(x_m,t_m) - d_m\} \delta(x_m,t_m) + O(\delta\eta)^2 \; . \tag{16}$$

Once computed the variation of the functional J, the next step is to determine the Euler-Lagrange equation associated.

ISSN 2238-1007

3.1.1 Euler-Lagrange equations

Defining the weighted residual:

$$\lambda(x,t) \equiv W_f \left\{ \frac{\partial \hat{\eta}}{\partial t} + c\frac{\partial \hat{\eta}}{\partial x} - F(x,t) \right\} \tag{17}$$

and using the above definition of the weighted residual, integrating by parts, neglecting the second order terms, equation (16) becomes:

$$\frac{\delta J}{2} = \int_0^T dt \int_0^L dx\lambda \left\{ \frac{\partial \delta\eta}{\partial t} \right\} + \int_0^T dt \int_0^L dx\lambda \left\{ c\frac{\partial \delta\eta}{\partial x} \right\} + W_i \int_0^L dx\{\hat{\eta}(x,0) - I(x)\}\delta\eta(x,0) + \\ w\sum_{m=1}^{M} \{\eta(x_m,t_m) - d_m\}\delta\eta(x_m,t_m) + O(\delta\eta)^2 \ . \tag{18}$$

When $\hat{\eta}(x,t)$ is the local extremum of J: $\delta J = O(\delta\eta)^2$. Thus, the sum of the remaining terms must be zero. For eliminating the terms $\partial\delta\eta/\partial t$ and $\partial\delta\eta/\partial x$ from (18), the integration by parts is used to shift the derivative on the variation to the weighted residual. Considering the time derivative term:

$$\int_0^L dx \int_0^T \lambda \left\{ c\frac{\partial \delta\eta}{\partial t} \right\} dt = \int_0^L dx \left\{ [\lambda(x,T)\delta\eta(x,T) + (-\lambda(x,0)\delta\eta(x,0))] + \int_0^T -\frac{\partial\lambda}{\partial t}\delta\eta dt \right\} = \\ \int_0^T dt \int_0^L \left\{ -\frac{\partial\lambda}{\partial t}\delta\eta \right\} dx + \int_0^L \{\lambda(x,T)\delta\eta(x,T)\}dx + \int_0^L \{-\lambda(x,0)\delta\eta(x,0)\}dx \ . \tag{19}$$

Similarly, for the spatial derivative:

$$\int_0^T dt \int_0^L dx\lambda \left\{ c\frac{\partial \delta\eta}{\partial x} \right\} = \int_0^T dt \int_0^L dx \left\{ -c\frac{\partial\lambda}{\partial x}\delta\eta \right\} + \int_0^T dt\{c\lambda(L,t)\delta\eta(L,t)\} + \int_0^T dt\{-c\lambda(0,t)\delta\eta(0,t)\} \ . \tag{20}$$

The boundary terms are equal and opposite, due to the periodic boundary condition. Equation (20) can be written as:

$$\int_0^T dt \int_0^L dx\lambda \left\{ c\frac{\partial \delta\eta}{\partial x} \right\} = \int_0^T dt \int_0^L dx \left\{ -c\frac{\partial\lambda}{\partial x}\delta\eta \right\} \ . \tag{21}$$

Finally, the filtering property of the Dirac delta is employed to eliminate $\delta(x_m,t_m)$:

$$w\sum_{m=1}^{M} \{\eta(x_m,t_m) - d_m\}\delta\eta(x_m,t_m) = \int_0^T dt \int_0^L dx \sum_{m=1}^{M} \{\eta(x_m,t_m) - d_m\}\delta\eta(x,t)\delta(x-x_m)\delta(t-t_m)\} \tag{22}$$

where the last two term in the equation above denote Dirac delta functions. Substituting the equations (19), (21), and (22) into (18), yields:

$$0 = \int_0^T dt \int_0^L dx \left\{ -\frac{\partial\lambda}{\partial t}\delta\eta \right\} + \int_0^L dx\{\lambda(x,T)\delta\eta(x,T)\} + \int_0^L dx\{-\lambda(x,0)\delta\eta(x,0)\} + \int_0^T dt \int_0^L dx \left\{ -c\frac{\partial\lambda}{\partial x}\delta\eta \right\} \\ + W_i \int_0^L dx\{\hat{\eta}(x,0) - I(x)\}\delta\eta(x,0) + \int_0^T dt \int_0^L dxw \sum_{m=1}^{M} \{\eta(x_m,t_m) - d_m\}\delta\eta(x,t)\delta(x-x_m)\delta(t-t_m) \ . \tag{23}$$

Rearranging terms in equation (23):

$$0 = \int_0^T dt \int_0^L dx \left\{ -\frac{\partial\lambda}{\partial t} - c\frac{\partial\lambda}{\partial x} + w\sum_{m=1}^{M} \{\eta(x_m,t_m) - d_m\}\delta(x-x_m)\delta(t-t_m) \right\} \delta\eta \\ + \int_0^L dx\{\lambda(x,t)\}\delta\eta(x,T) + \int_0^L dx\{-\lambda(x,0) + W_i(\hat{\eta}(x,0) - I(x))\}\delta\eta(x,0) \ . \tag{24}$$

Equalling to zero each term from the equation above:

$$-\frac{\partial\lambda}{\partial t} - c\frac{\partial\lambda}{\partial x} + w\sum_{m=1}^{M} \{\eta(x_m,t_m) - d_m\}\delta(x-x_m)\delta(t-t_m)\} = 0 \tag{25}$$

ISSN 2238-1007

Proceedings of the 1st International Symposium on Uncertainty Quantification and Stochastic Modeling
February 26th to March 2nd, 2012, Maresias, São Sebastião, SP, Brazil

with $0 \leq x \leq L$ and $0 \leq t \leq T$, and

$$\lambda(x,T) = 0 \tag{26}$$

$$-\lambda(x,0) + W_i\{\hat{\eta}(x,0) - I(x)\} = 0 \tag{27}$$

In general, equations (17) and (25), along with the periodic boundary conditions, constitute the Euler-Lagrange equations for local extrema of the penalty functional. After rearranging, the Euler-Lagrange equations for the local extremum $\hat{\eta}$ of the penalty functional $J[\eta]$ can be written as a "backwards" problem:

$$-\frac{\partial \lambda}{\partial t} - c\frac{\partial \lambda}{\partial x} = -w\sum_{m=1}^{M}\{\hat{\eta}(x_m,t_m) - d_m\}\delta(x-x_m)\delta(t-t_m)\} \tag{28a}$$

$$\lambda(x,T) = 0 \tag{28b}$$

$$\lambda(0,t) = \lambda(L,t) \tag{28c}$$

The "forwards" problem given by:

$$\frac{\partial \hat{\eta}}{\partial t} + c\frac{\partial \hat{\eta}}{\partial x} = F(x,t) + \frac{\lambda(x,t)}{W_f} \tag{29a}$$

$$\hat{\eta}(x,0) = I(x) + \frac{\lambda(x,0)}{W_f} \tag{29b}$$

$$\hat{\eta}(0,t) = \hat{\eta}(L,t) \tag{29c}$$

from equations (29), the best estimates for f and i are given by:

$$\hat{f}(x,t) \equiv W_f^{-1}\lambda(x,t) \text{ e } \hat{i}(x,0) \equiv W_i^{-1}\lambda(x,0)\ . \tag{30}$$

With the adjoint equations (28) and (29) the representer method is introduced.

3.1.2 Representer method

For catching a sequential solution of the "backward" or "adjoint" problem and the "forward" problem, it is importante to uncouple the Euler-Lagrange equations. The so-called "adjoint-representer" and "representer" functions facilititate this uncoupling. There are M "representer" functions, one for each scalar datum, and these are denoted $r_m(x,t)$, $1 \leq m \leq M$. The mth representer function has an adjoint-representer satisfying the so-called adjoint equation, except forced only with a single, bare impulse located at the space-time position of mth scalar datum:

$$-\frac{\partial \alpha_m}{\partial t} - c\frac{\partial \alpha_m}{\partial x} = \delta(x-x_m)\delta(t-t_m) \text{ com } 0 \leq x \leq L,\ 0 \leq t \leq T \tag{31a}$$

$$\alpha_m(x,T) = 0\ , \tag{31b}$$

$$\alpha_m(0,t) = \alpha_m(L,t)\ . \tag{31c}$$

The name of variable is changed from $\lambda_m(x,T)$ to $\alpha_m(x,T)$ to indicate the different forcing, and also to indicate the dependence of the adjoint-representer on the space-time location of the $m$ht scalar datum. Since there is impulse, equation (31a) can be integrated backward from the final condition (31b) yielding the adjoint-representer, $\alpha_m(x,t)$.

The representer satisfies the forward equation (29a) and the initial condition, except replacing the adjoint field on the right hand side of equation (31a) with the adjoint-representer field, and with no prior estimate of the forcing or the initial condition:

$$\frac{\partial r_m}{\partial t} + c\frac{\partial r_m}{\partial x} = \frac{\alpha_m(x,t)}{W_f},\ \ 0 \leq x \leq L,\ 0 \leq t \leq T \tag{32a}$$

$$r_m(x,0) = \frac{\alpha_m(x,0)}{W_i},\ \ 0 \leq x \leq L\ , \tag{32b}$$

$$r_m(x,0) = r_m(L,T),\ \ 0 \leq t \leq T\ . \tag{32c}$$

The optimal solution is assumed to be the sum of the prior estimate and a linear combination of representers:

$$\hat{\eta}(x,t) = \eta_F(x,t) + \sum_{m=1}^{M}\beta_m r_m(x,t) \tag{33}$$

where β_m are unknown constants. The response of the optimal solution to the forcing $F(x,t)$, with the initial condition $I(x)$, is made by the forward solution $\eta_F(x,t)$, while the terms in the summation represent a sequence of "corrections", one for each scalar datum.

Table 1 presents the algorithm for the method of assimilation for the 1D wave model.

ISSN 2238-1007

Table 1: Algorithm representer

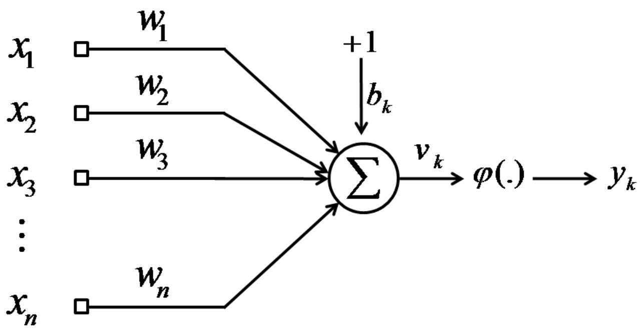
Algorithm representer to one dimension
1. Compute $\eta_F(x,t)$ by numerical integration of the equation 2
2. Compute the innovation **h**, according to:
$h = \sum_{m=1}^{M}(d_m - \eta_F(x_m,t_m))$
where: $\mathbf{d}_m$ represents the observation vector.
3. Compute the adjoint-representer α_m according to equation 31a.
4. Compute the representer $r_m(x,t)$ to $1 \leq m \leq M$
according to equation 32a.
5. Compute the covariance matrix
$r_m(x_j,t_j)$, with $m = 1,2,\ldots,M$ and $j = 1,2,\ldots,M \Rightarrow \mathbf{R}_{M\times M}$
$\mathbf{P} = (\mathbf{R} + \mathbf{w}^{-1}\mathbf{I})$
6. Determine the coefficients of the expansion of the increment analysis ρ.
$\rho = \sum_{m=1}^{M}\beta_m r_m$, $\mathbf{b} = [\beta_1\ \beta_2\ \ldots\ \beta_m]^T$
solving the linear system:
Pb = h, $\mathbf{h} = [h_1\ h_2\ \ldots\ h_m]^T$
7.Compute the analysis given by the following equation:
$\hat{\eta}(x,t) = \eta_F(x,t) + \sum_{m=1}^{M}\beta_m r_m(x,t)$

3.2 Artificial neural network

Artificial Neural Networks (ANN) have became important tools for information processing (Haykin, 1993). Much research has been conducted in pursuing new neural network models and adapting the existing ones to solve real life problems, such as those in engineering (Haykin, 1993). ANN are made of arrangements of processing elements called neurons. The artificial neuron model basically consists of a linear combiner followed by an activation function, Figure 1 (left side), given by:

$$y_k = \varphi\left(\sum_{j=1}^{n} w_{kj}x_j + b_k\right) \tag{34}$$

where w_{kj} are the connection weights, b_k is a threshold parameter, x_j is the input vector and y_k is the output of the k_{th} neuron.



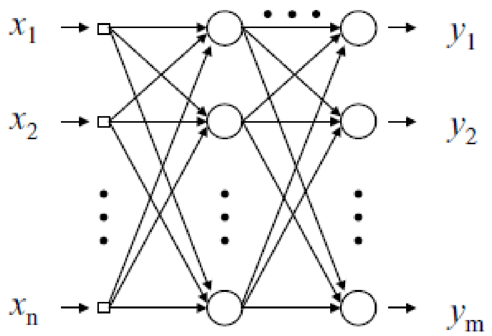


Figure 2: (left side) Single Neuron, (right side) Multilayer Neural Network

Arrangements of such units form the ANN that are characterized by:

- Very simple neuron-like processing elements;
- Weighted connections between the processing elements;
- Highly parallel processing and distributed control;
- Automatic learning of internal representations.

ANN aim to explore the massively parallel network of simple elements in order to yield a result in a very short time slice and, at the same time, with insensitivity to loss and failure of some of the elements of the network. These properties make artificial neural networks appropriate for application in pattern recognition, signal processing, image processing, financing, computer vision, engineering, etc.

There are different architectures of ANN that are dependent upon the learning strategy adopted. This paper briefly describes the Multilayer Perceptron (MLP) with error backpropagation learning. Detailed introductions on ANN can be found in (Haykin, 1993) and (Nadler, 1993). MLP with backpropagation learning algorithm, are feedforward networks

ISSN 2238-1007

composed of an input layer, an output layer, and a number of hidden layers, whose aim is to extract high order statistics from the input data (Haykin, 1993). Figure 1(right side) depicts a multilayer neural network with a hidden layer. Functions $\varphi(.)$ provide the activation for the neuron. Neural networks will solve nonlinear problems, if nonlinear activation functions are used for the hidden and/or the output layers. From several activation functions, the sigmoid are commonly used:

$$\text{bipolar function } \varphi(v) = \frac{1-\exp(-av)}{1+\exp(-av)} \ . \tag{35}$$

A feedforward network is a non-linear mapping to compute the output vector from an input vector. The connections among the several neurons (Figure 2 (right side)) have associated weights that are adjusted during the learning process, thus changing the performance of the network. Two distinct phases can be devised while using ANN: the training phase (learning process) and the run phase (activation of the network). The training phase consists of adjusting the weights for the best performance of the network in establishing the mapping of many input/output vector pairs. Once trained, the weights are fixed and the network can be presented to new inputs for which it calculates the corresponding outputs, based on what it has learned.

4 RESULTS

The linear wave model presented in section 2.1 has been integrated with the FTCS (Forward-Time Central-Space method). The initial conditions and boundary are periodic. The data was assimilated each 10 time step and each 4 space grid point.

The observed data for assimilation are synthetic observations. The observation data were generated from the integration of the model (2), and adding a random noise of variance 0.04. The reference of truth to the assimilation method is the curve obtained from the integration of models without noise.

The architecture for the neural artificial network implemented for this experiment was one hide layer with three neurons, two input data, and one neuron at output layer, according to Figure 3. The algorithm for training the network was the backpropagation (Haykin, 1993).

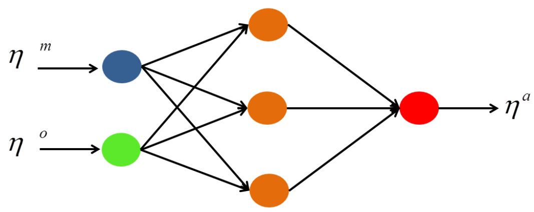


Figure 3: Architecture for the neural artificial network. η^m data model; η^o data observed; η^a data analysis.

Figure 4 shows the result for equation (2) with initial conditions equal zero: only propagation of the external forcing $F = 0.0001$. Figure 5 shows the initial condition used was sine function; in the left side the estimative obtained by representer method, and the right side the estimative obtained by multlayer perceptron. Figure 6 compare the two methods with the truth.

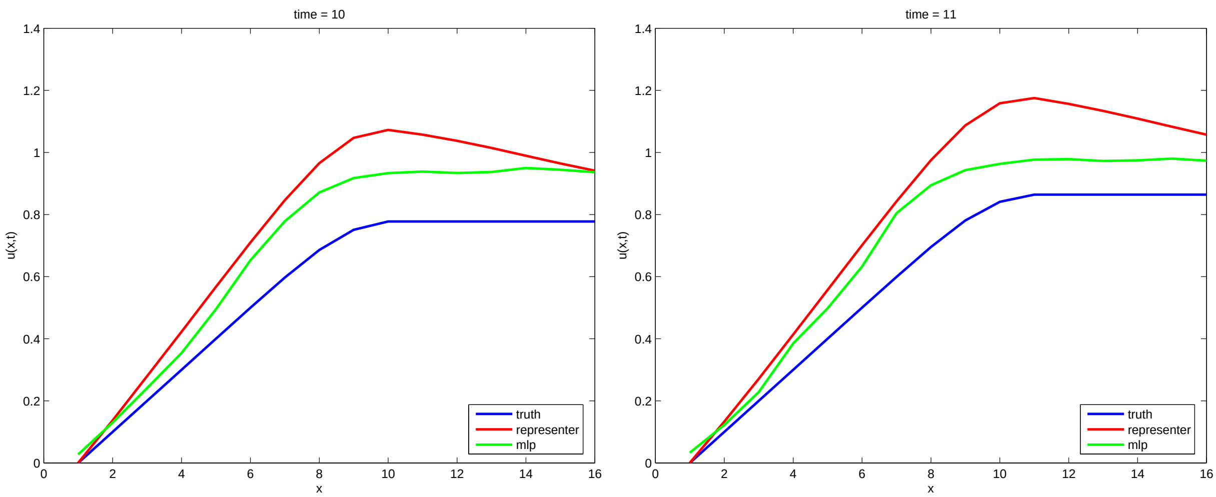


Figure 4: Blue curve: truth; red curve: estimated by representer method; green curve estimated by mlp. (left side) Time serie t = 10; (right side) Time serie t = 11.

ISSN 2238-1007

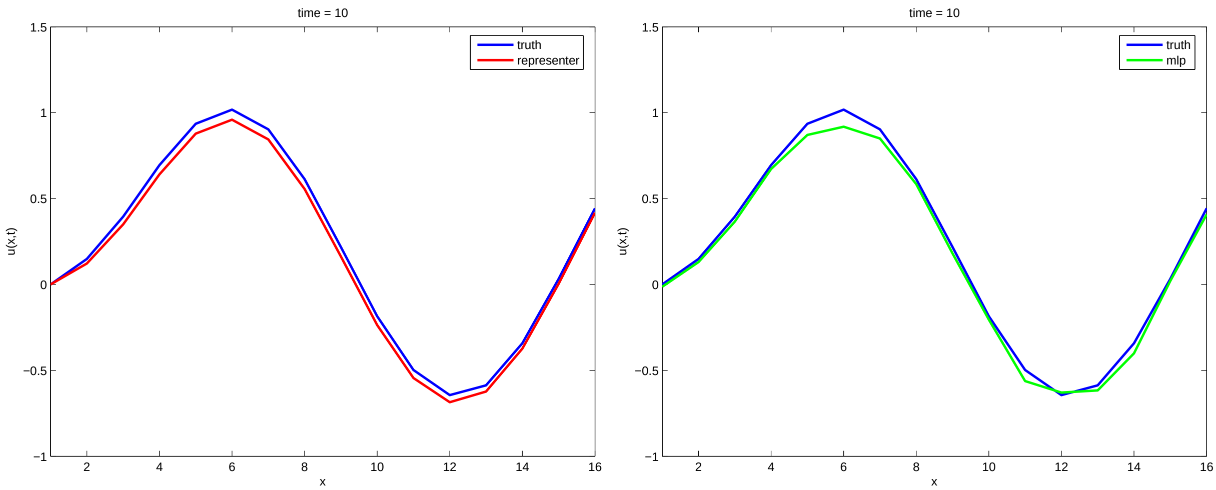


Figure 5: (left side) blue curve: truth; red curve: estimated by representer method; (right side) green curve: estimated by mlp; blue curve: truth

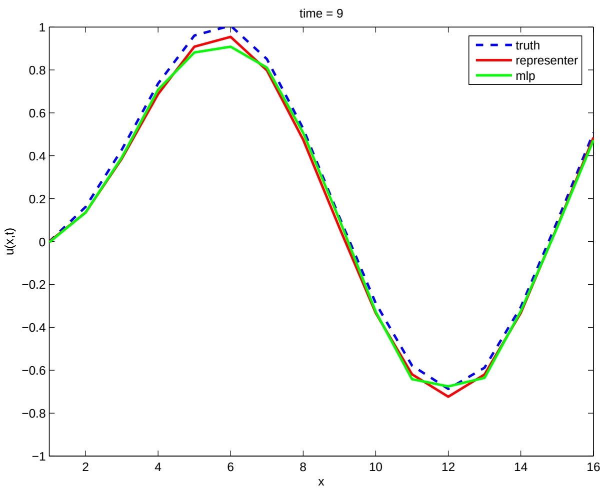


Figure 6: Blue curve: truth; red curve: estimated by representer method; green curve estimated by mlp.

5 CONCLUSION

Neural network can be successfully used as a data assimilation method. The Multilayer Perceptron technique was tested in a linear wave equation. As neural network employed supervised training, where the target of network was determined by a variational method presented section 3.1. This ANN presents a lower computational complexity than the Particle Filter, and variational approach (Furtado, 2008) or Kalman filter (Harter, 2008). The numerical experiments are a new evidence for the neural network is a good method for data assimilation, where the new method presented better estimative.

ACKNOWLEDGEMENTS

The authors wants to thank to the CNPq, CAPES, and FAPESP, Brazilian agencies for research support.

6 REFERENCES

Daley, D. 1993 “Atmospheric Data Analysis” (Cambridge University Press).

Kalnay, E. 2003 “Atmospheric modeling, data assimilation and predictability” (Cambridge University Press).

Bennett, A. F. 2002 “Inverse Modeling of the Ocean and Atmosphere” (Cambridge University Press).

Garner, T. W; Wolf, R. A; Spiro, R. W; Thomsen, M. F. 1999 “First attempt at assimilation data to constrain a magnetospheric model” Journal of Geophysical Research 104, 25145-25152.

Schunk R W, Scherliess L, Sojka J J, Thompson D C, Anderson D N, Codrescu M, Minter C, Fuller-Rowell T J, Heelis R A, Hairston M, Howe B M 2004 “Global assimilation of ionospheric measurements” (GAIM00). Radio Science, 1, RS1S02.

Belyaev, K. P; Tanajura, C. A. S. and O’Brien J. J. 200 “A data assimilation method used with an ocean circulation model

ISSN 2238-1007

and its application to the tropical Atlantic". Applied Mathematical Modelling, v. 25 - 655-670)

C. Grebogi, E. Ott, J. Yorke 1987 "Chaos, strange attractors, and fractal basin boundaries in nonlinear dynamics". Science 238 (585-718).

F. P. Härter, H. F. de Campos Velho "New approach to applying neural network in nonlinear dynamic model". Applied Mathematical Modellin - Science Direct 32 (2008) 2621-2633).

Nakano S, Ueno G, Higuchi T 2007 "Merging particle filter for sequential data assimilation Nonlinear" (Processesin Geophysics 14 395-408).

Chua, B. S.; Bennett, A. F. "An inverse ocean modeling system" (Ocean modeling, v.3, p. 137-165, March 2001).

Jazwinski, A. H. 1970 "Stochastic Processes and Filtering Theory" (Academic Press).

Evensen, G. 2007 "The Ensemble Kalman Filter" (Springer Verlag).

Belyaev, K; Meyers, S. O'Brien J J 2000 "Fokker-Planck equation application to data assimilation into hydrodynamic models" J. Math. Sci., 99(4) 1393-1403.

Belyaev, K. P; Tanajura, C. A; S, O'Brien J J 2001 "A data assimilation method used with an ocean circulation model and its application to the tropical Atlantic Appl. Math. Modelling 25 655-670.

Campos Velho H F, Cintra R S, Furtado H C M. "Introduction to Data Assimilation". Short-course, INPE, São José dos Campos, Brazil (In Portuguese). Available in the internet (Accessed: September 22, 2010): `www.lac.inpe.br/~haroldo/Curso-DataAssimilation/Curso_Assi_Dados-INPE-2.pdf`.

Furtado H C M, Campos Velho H F, Macau E E N "Redes neurais e diferentes métodos de assimilação de dados em dinâmica não linear" (125 p). (INPE-15235-TDI/1322). MSc thesis on Applied Computing – Instituto Nacional de Pesquisas Espaciais, São José dos Campos. 2008. Available in the internet: `http://urlib.net/sid.inpe.br/mtc-m17@80/2008/02.07.10.49`

Nowosad A G, Rios Neto A, Campos Velho H F 2000 "Data assimilation in chaotic dynamics using neural networks". In: International Conference on Nonlinear Dynamics, Chaos, Control and Their Applications in Engineering Sciences 6 212-221 (Associação Brasileira de Ciências Mecânicas).

Härter F P, Campos Velho H F 2005 "Recurrent and feedforward neural networks trained with cross correlation applied to the data assimilation in chaotic dynamic". Brazilian Journal of Meteorology 20 411-420.

Härter F P, Campos Velho H F 2008 "New approach to applying neural network in nonlinear dynamic model" Applied Mathematical Modelling 32 2621-2633 - DOI 10.1016/j.apm.2007.09.006 - ISSN: 0307-904X.

Härter F P, Rempel E L, Campos Velho H F, Chian A 2008 "Application of artificial neural networks in auroral data assimilation". Journal of Atmospheric and Solar – Terrestrial Physics, 70 1243-1250, - DOI 10.1016/j.jastp.2008.03.018.

Lorenz, E. N. 1963 "Deterministic Nonperiodic Flow" Journal of the Atmospheric Sciences 20(2) 130-141.

Lorenz, E. N. 1965 "A study oh the predictability of a 28-variable model" Journal Tellus 17

Haykin, S. 1993 "Neural Networks: A Comprehensive Foundation" (New York: Prentice Hall).

Nadler, M. and Smith, E. P. 1993 "Pattern Recognition Engineering" (New York: John Wiley & Sons)

Furtado, H. C. M. and Campos Velho, H. F. and Macau, E. E. N. and Härter, F. P. 2007 "Diferent data assimilation methods applied to the Lorenz dynamical system", 6th Brazilian Conference on Dynamics, Control and Their Applications (DINCON-2007) May 22-25 (UNESP São José do Rio Preto (SP) Brazil) Proceedings in CD-Rom 1030-1035 - in Portuguese.

RESPONSIBILITY NOTICE

The author are the only responsible for the printed material included in this paper.