Data Assimilation using Artificial Neural Networks: Missing Observation

Rosangela Cintra[1], Haroldo F. de Campos Velho

National Institute for Space Research, São José dos Campos, SP, Brazil

Abstract

Numerical weather prediction (NWP) uses atmospheric general circulation models (AGCMs) to predict the weather future conditions. The process is done inserting observation data into computer model to compute initial conditions – also named "analysis". Such feature is called data assimilation (DA). Several techniques have been developed for DA. Ensemble Kalman filter. Here, a set of artificial neural networks (ANNs) – multi-layer perceptron with back propagation learning – is configured to emulate the Local Ensemble Transform Kalman Filter (LETKF) applied to the SPEEDY model. The novelty is to perform the *analysis* when some observations are missing in a DA cycle, or several cycles. A comparison between analysis produced by the LETKF and ANN is carried out. The numerical experiment was done at January 1985.

Keywords: Data assimilation, atmospheric model, artificial neural network, ensemble Kalman filter.

1. Introduction

The procedure to determine the initial condition in the operational prediction centers is called *data assimilation* (DA), where the background fields are combined with observations for producing the *analysis* – the new initial condition. There are many different types of data assimilation schemes. A useful overview of some of the most common data assimilation methods can be found in texts such as Daley [7] and Kalnay [16].

Two DA schemes are analized here: Local Ensemble Transform Kalman Filter (LETKF) [14], and the approach based on Artificial Neural Networks (NN) to emulate the LETKF [19, 11, 4, 5]. The main advantage to use NN is to improve the computational performance.

Numerical experiment is performed using syntethic observations from surface stations (data at each 6 hours/day) and upper-air soundings (data

[1]E-mail Corresponding Author: rocintra@gmail.com

at each 12 hours/day). The research investigates the behaviour of DA techniques with missing observiations for data assimilation cycles. Each day a sequence of analyses at 00:00, 06:00, 12:00, and 18:00 UTC. The grid of observations reproduces the stations of World Meteorological Organization (WMO) of radiosonde observations.

The atmospheric model is the 3D SPEEDY(Simplified Parameterizations PrimitivE-Equation Dynamics) [2], a model with simplified physics parameterization. We run 44 data assimilation cycles (analysis/model/forecast) for LETKF and ANN approach, where some cycles have missing observations.

2. Data Assimilation

Considering a general nonlinear system with a $n-$dimensional state vector x^f and a $m-$dimensional observation vector y^o evolving according to

$$x^f_{k+1} = f(x^f_k, t_k) + q_k \tag{1}$$

$$y^o_k = h(x^f_k, t_k) + v_k \tag{2}$$

where q_k and v_k are Gaussian noise terms.

Forecasting is a step to predict a state x^f_{k+1} of a system from the last state by numerical weather prediction model. From mathematical point, the assimilating process can be represented by

$$x^a_k = x^f_{k+1} + W[y^o_k - H(x^f_{k+1})] \ , \tag{3}$$

$$W = (HP^fH^T + R) \ . \tag{4}$$

Equation (3) is the analysis step, where H is the observation operator, W is the weighting matrix, computed from the error covariance matrices P^f and R, representing model and observations errors respectively.

The data assimilation approach with ANN uses neural networks to implement the function:

$$x^a = F_{NN}(y^o, x^f) \tag{5}$$

where F_{NN} is the data assimilation process, y^o represents the observations, x^f is a model forecast, often called the first guess, and x^a is the analysis field with innovation that represents the correction to the model.

2.1 Local Ensemble Transform Kalman Filter

Ensemble Kalman filter (EnKF) is a Bayesian approach proposed by Evensen [8]. Several schemes for the EnKF has been developed. The local ensemble transform Kalman filter (LETKF) is one of them [14].

The basic idea for the LETKF is to perform the analysis at each grid point simultaneously using the state variables and all observations in the region centred at given grid point. The local strategy separates groups of neighbouring observations around a central point for a given region of the grid model. Each grid point has a local patch; the number of local vectors is the same as the number of global grid points [17].

3. Atificial Neural Networks

An ANN is composed of simple processing units (*neurons*) for computing certain mathematical functions, and consists of interconnected artificial neurons or nodes. The neurons are connected to others to form an ANN. Each artificial neuron has one or more inputs and outputs The connection among neurons stores a weighted sum, called synaptic weight. In ANN processing, the inputs are multiplied by weights feeding an activation function. This function activates or inhibits the next neuron. Mathematically, we can describe the i-th neuron with the following form:

$$\begin{array}{ll} \text{input summation:} & u_i = \sum_{j=1}^{p} w_{ij} x_i \\ \text{neuron output:} & y_i = \varphi(u_i) \end{array} \tag{6}$$

where $x_1, x_2, \cdots, x_n$ are the inputs; $w_{i1}, \cdots, w_{ip}$ are the synaptic weights; u_i is the output of linear combination; $\varphi(\cdot)$ is the activation function, and y_i is the i-th neuron output, n is number of patterns, p is number of neurons.

A feed-forward network, which processes in one direction from input to output, has a layered structure. The first layer of an NN is called the input layer, the intermediary layers are called hidden layers, and the last layer is called the output layer. Some parameters, as number of layers and the quantity of neurons in each layer, define the neural network topology, but other parameters are also need to be computed, such as learning ratio and momentum. These parameters are determined by the nature of the problem.

There are two distinct phases in using a suvervised NN: the training phase (learning process) and the run phase (activation or generalization). The training phase of the ANN consists of an iterative process for adjusting the weights for the best performance of the NN in establishing the mapping of input and target vector pairs. The learning algorithm is the set of procedures for adjusting the weights. The single pass through the entire training set in one process in called an 'epoch'. The iterative training process continues or stops after defined criteria, which can be minimum error of mapping or a determined number of epochs. Once the ANN is trained the weights are fixed, and the ANN is ready to receive new inputs (different from training inputs) for which it calculates the corresponding outputs.

The multilayer perceptron (MLP) is the ANN topology used in this study; with at least one intermediate layer of neurons (hidden layer) [12]. The supervised learning process, the functional to be minimized is tread as a function of the weights w_{ij} (Eq. 6), instead of the NN inputs. For a given input vector x, the output vector x^a_{NN} is compared to the target answer x^a_{ref}. If the difference is smaller than a required precision, no learning takes place; on the other hand, the weights are adjusted to reduce this difference. The goal is to minimize the error between the actual output y_i (or x^a_{NN}) and the target output (d_i) (or x^a_{ref}) from the training data. The set of procedures to adjust the weights is the learning algorithm *backpropagation*, which is generally used for the MLP training.

4. The SPEEDY Model

The SPEEDY model [18], is an Atmospheric Global Circulation model (AGCM) developed in the International Centre for Theoretical Physics (ICTP) in Triesty-Italy, based on a spectral dynamical core with a simplified set of physical parametrization schemes. The dynamic variables for the primitive meteorological equations are integrated by the spectral method in the horizontal grid at each vertical level – more details in [2].

The SPEEDY is executed on T30L7: horizontal resolution with triangular spectral truncation at total wave-number 30 (T30), with seven levels on the vertical coordinate (100, 200, 300, 500, 700, 850, and 925 hPa). The vertical coordinate are defined on sigma ($\sigma = p/p_s$), where p_s is the surface pressure. The horizontal coordinates are latitude and longitude on Gaussian grid, with a regular grid with 96 zonal points (longitude), and 48 meridian points (latitude). The prognostic variables of input and output model are the absolute temperature (T), surface pressure (p_s), zonal and meridional wind components (u, v), and specific humidity (q).

5. MLP-DA: Data assimilation with missing observation

The ANN configuration for this experiment is a set of multilayer perceptrons, referred to as MLP-DA, defined in the experiment described by Cintra and Campos Velho [6]. The topology has the following characteristics:

1. Two input nodes, one node for the meteorological observation vector and the other for the 6-hours forecast model vector;

2. One output node for the analysis vector results;

3. One hidden layer with eleven neurons;

4. The hyperbolic tangent as the activation function;
5. Learning rate η is defned do each MLP;
6. Training phase stops when the error reaches 10^{-5} or after 5000 epochs, which criterion first occurs.

The input vectors represent the model grid point for a single variable with a correspondent observation-forecast. In the training algorithm, the MLP-DA computes the output and compared it with the analysis vector from the LETKF – the target data. The output vectors represent the analysis values for one grid point.

The MLP-DA scheme were developed with a set of *thirty* NNs: six regions with five prognostic variables (p_s, u, v, T, q). All NNs have one hidden layer, with the same number of neurons for all regions, but with *different* connection weights. The MLP-DA scheme divides the entire globe into six regions: for the Northern Hemisphere, 90^o N and three longitudinal regions of 120^o each; for the Southern Hemisphere, 90^o S and three longitudinal regions of 120^o each. All regions have the same size, but the number of observations is distinct for each region – see Figure 1. This regional division is applied only for the MLP-DA; the LETKF procedures are not modified.

The MLP-DA is designed to emulate the LETKF analysis for SPEEDY initial condition. The LETKF analysis is the average field from an ensemble of analyses.

The upper levels and the surface covariance error matrices to run the LETKF system, as well as the SPEEDY model boundary conditions data and physical parameterizations are the same as those used for Miyoshi's [17] and Cintra's [6] experiments. The so-called "true" (or control) model is generated the SPEED simulation without *noise*. The data assimilation cycles are considered at four times per day (00:00, 06:00, 12:00, 18:00 UTC).

The synthetic observations are generated, reading the "true" SPEEDY model fields and adding a random noise on meteorological variables. The observation localization is on some grid model point. An observation *mask* is designed, adding a positive flag to grid point where the observation should be considered, the locations are similar to the WMO data stations observations from rawinsonde – see Fig. 1. Except for p_s observations, the other observations are also descripted on upper seven levels.

The MLP-DA data assimilation scheme has no error covariance matrices to spread the observation influence. Therefore, it is necessary to capture the influence of observations from the neighbouring region around a grid point. This calculation is based on the distance from the grid point related

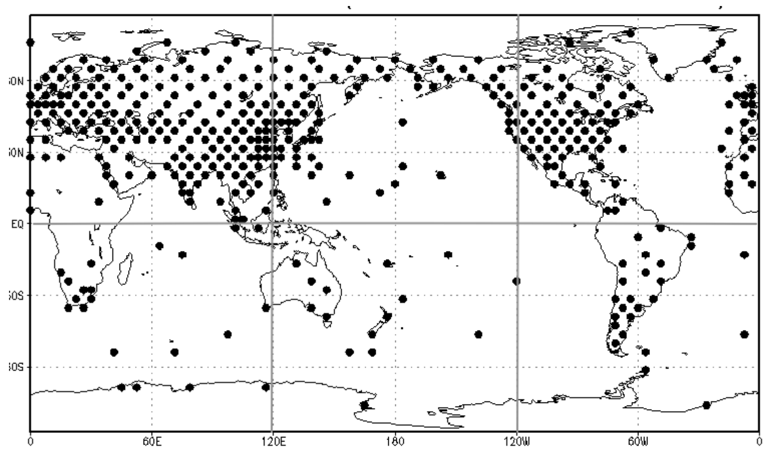

Figure 1: Observations localizations in global area. The dot points represent rawinsonde stations ($\sim$ 415).

to observations inside a determined neighbourhood (initially: $\gamma = 0$)

$$\hat{y}^o_{i\pm m,j\pm m,k\pm m} = \frac{y^o_{ijk}}{(6-\gamma_r)\, r^2_{ijk}} + \sum_{l=1}^{6} \alpha_l \frac{y^o_{i\pm m,j\pm m,k\pm m}}{r^2_{ijk}} \qquad (m = 1, 2, \ldots, M) \tag{7}$$

$$\alpha_l = \begin{cases} 0 & \text{(if there is no observation)} \\ 1 & \text{(if there is observation, and: } \gamma_{r+1} = \gamma_r + 1) \end{cases} \tag{8}$$

where $\hat{y}^o$ is the weighted observation, M is the number of discrete layers considered around observation, $r^2_{ijk} = (x_p - y^o_i)^2 + (y_p - y^o_j)^2 + (z_p - y^o_k)^2$, where (x_p, y_p, z_p) is grid point coordinate, and the (y^o_i, y^o_j, y^o_k) is the observation coordinate, and γ_r is a counter of grid points with observations around (y^o_i, y^o_j, y^o_k). If $\gamma_r = 6$, there is no influence to be considered. Hereafter, the values computed from Eq. (7) are referred as pseudo-observation. grid points considered to LETKF analysis, although these calculations are made without interference on LETKF system.

The training processes was made by the Cintra's experiment [6]. With trained NNs, the system is ready to used as a data assimilation process. The MLP-DA results a global analysis field. The MLP-DA activation has input values (forecast and observations) at each grid point. This experiment of MLP-DA data assimilation is performed for 11 days (44 cycles). It starts at 10 January 1985 00UTC up to 20 January 1985 18UTC.

The DA cycles starting with complete global mask, to five prognostic variable, with four upper-air variables with 7 vertical layers. During three days (12 cycles), SPEEDY analyses were computed from 415 stations for all variables (12035 observations) – phase-1. Next two days (12/Jan and 13/Jan) with eight cycles, there is no observations for U, V, and Q, again with 415 stations with temperature and surface pressure only (3320 observations) – phase-2. For next 5 days (20 cycles), there is no observations points from a set of stations – phase-3.

6. Results and Conclusion

The input and output values of prognostic variables (p_s, u, v, T, q) are processed on grid model points for time integrations. Taking into account the psudo-observation (Eq. 7), with two grid layers ($M = 2$) around a given observation are considered, as carried out by [4]. The results show the comparison of analysis fields, generated by the MLP-DA and the LETKF, and the true model fields.

The differences between analyses for global humidity q observation at 950hPa as presented in Figure 2 of cycle performed at 10 January 06UTC: phase-1. Figure 3 shows the humidity with missing observations (u,v,q) in phase-2 (13 January 12 UTC).

For absolute temperature (950hPa) and suface pressure, Figures 4 and 5 present the phase-2, January 13th 12 UTC, with missing observations (u,v,q), for 415 station with temperature (7 vertical layers) and surface pressure. The temperature field are in Kelvin unit (values above 273.15) with the diferrences up to 10 K in some points. For surface pressure field, there are differences on the entire domain, because there are pseudo-observations on all domain.

In the phase 3, LETKF and MLP-NN are executed with missing observation of some points. Figure 6 shows the surface pressure at 18 January 00UTC, last day of this fail of observation cycles. For phase-3, the observation mask is complete, similar phase-1. The model has the same behavior that previous Cintra's experiment [6] – see Fig. 7 for temperature analysis for 19 January 00UTC.

The analysis tries to be a correction from a forecast. Less observation means a worse analysis. Missing observations could occur in a different assimilation cycles. The MLP-DA was trained with 2 layers for observation influence ($M = 2$ in Eq. (7)). Figures 3, 5, and 7 show a good agreement between the LETKF and MLP-DA analyses. Figures 2, 4, and 6 display some desagreement between the analyses obtained from the two methodolo-

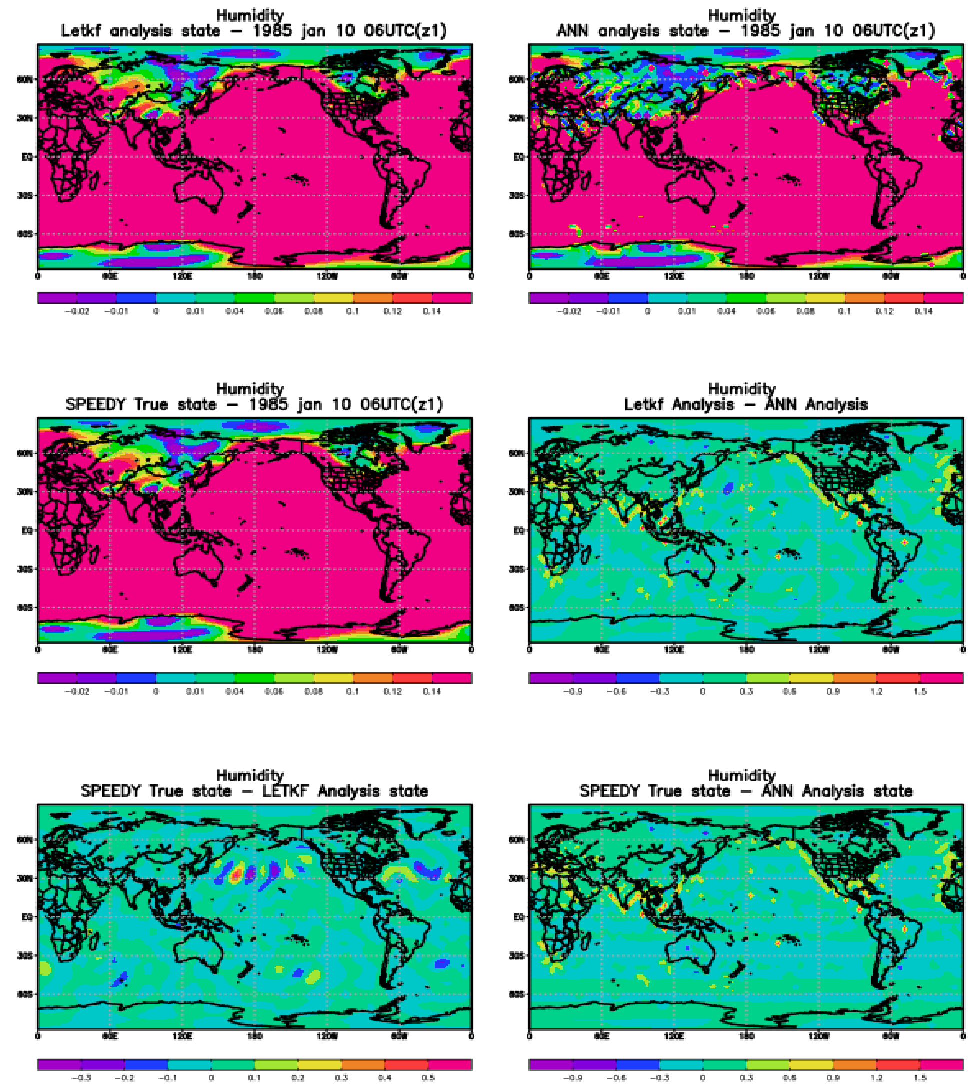

Figure 2: Humidity (kg/kg) for 10 January 06UTC: LETKF analysis, MLP-DA analysis, True model — and diferences: (LETKF and MLP-DA), (True model and LETKF), (True model and MLP-DA).

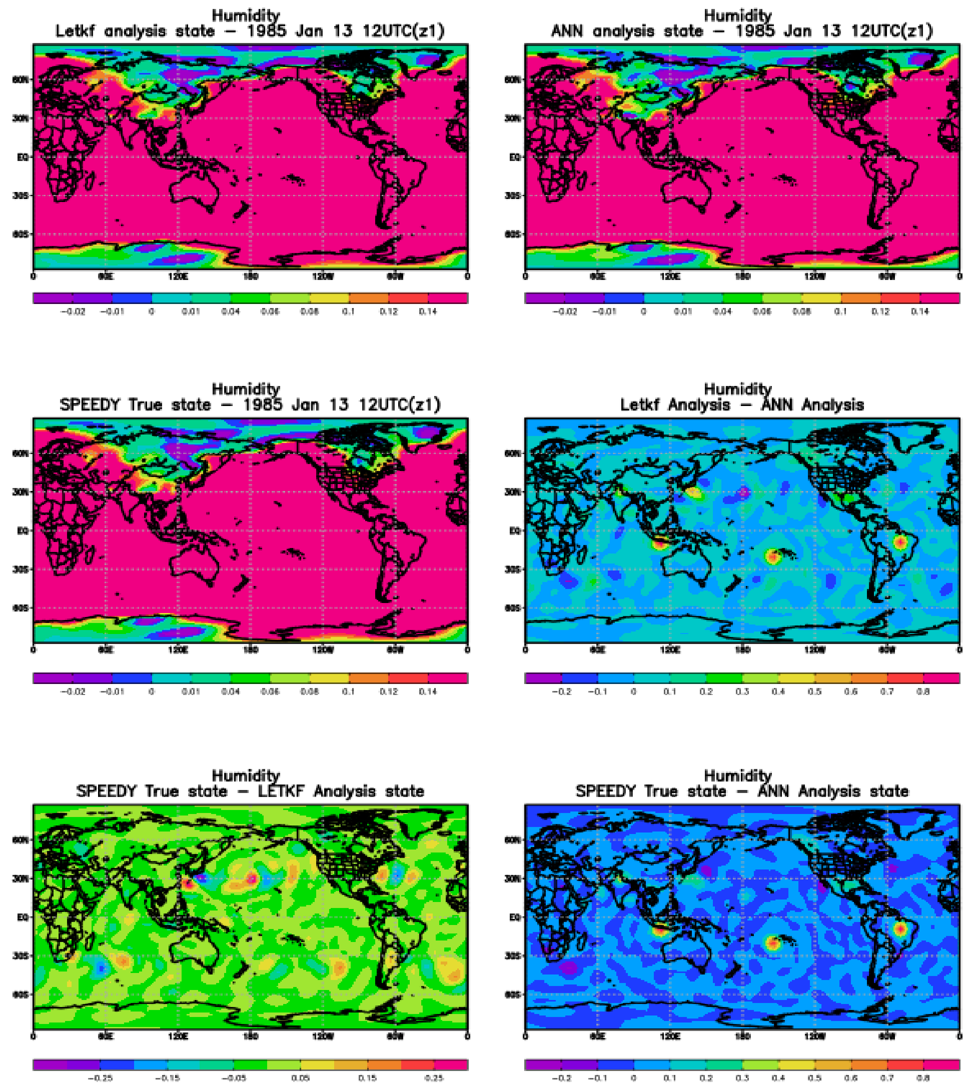

Figure 3: Humidity (kg/kg) for 13 January 12UTC: LETKF analysis, MLP-DA analysis, True model — and diferences: (LETKF and MLP-DA), (True model and LETKF), (True model and MLP-DA).

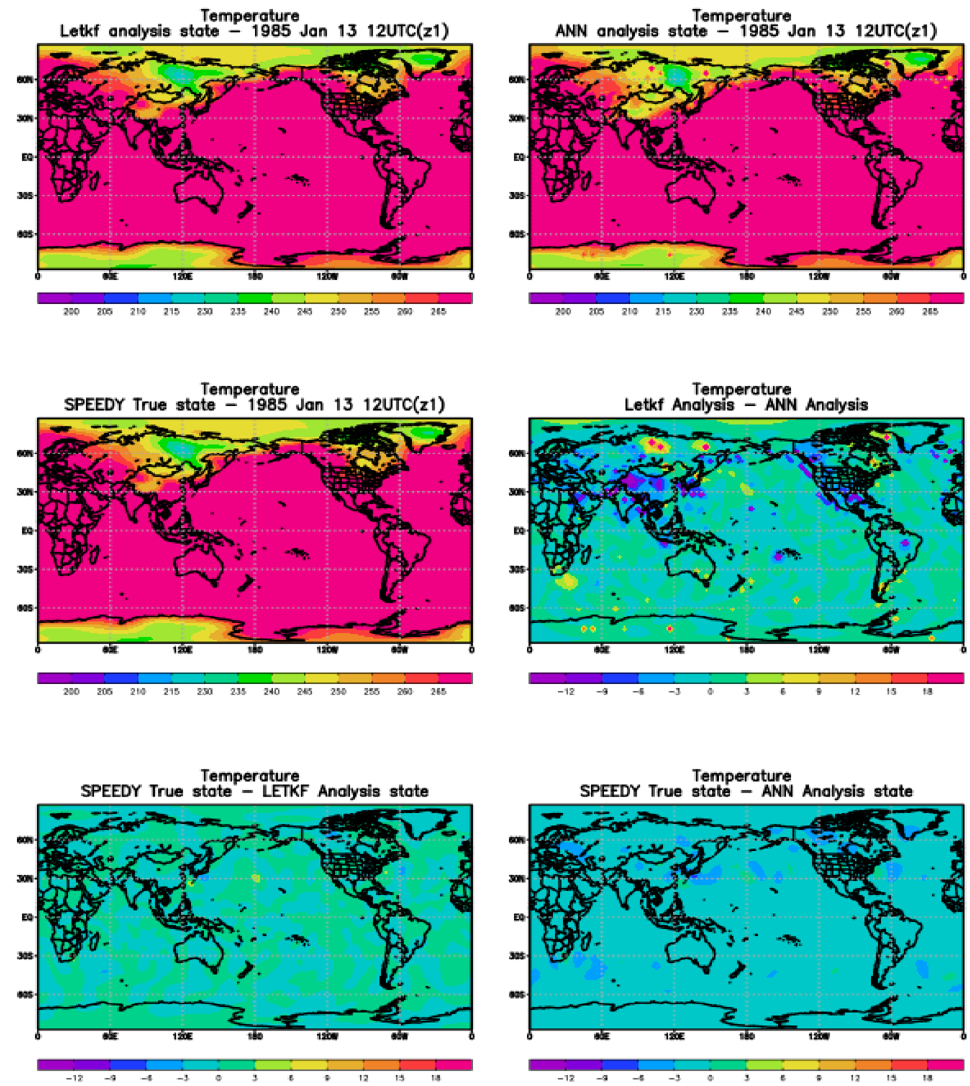

Figure 4: Absolute temperature (K) for 13 January 12UTCLETKF analysis, MLP-DA analysis, True model — and diferences: (LETKF and MLP-DA), (True model and LETKF), (True model and MLP-DA).

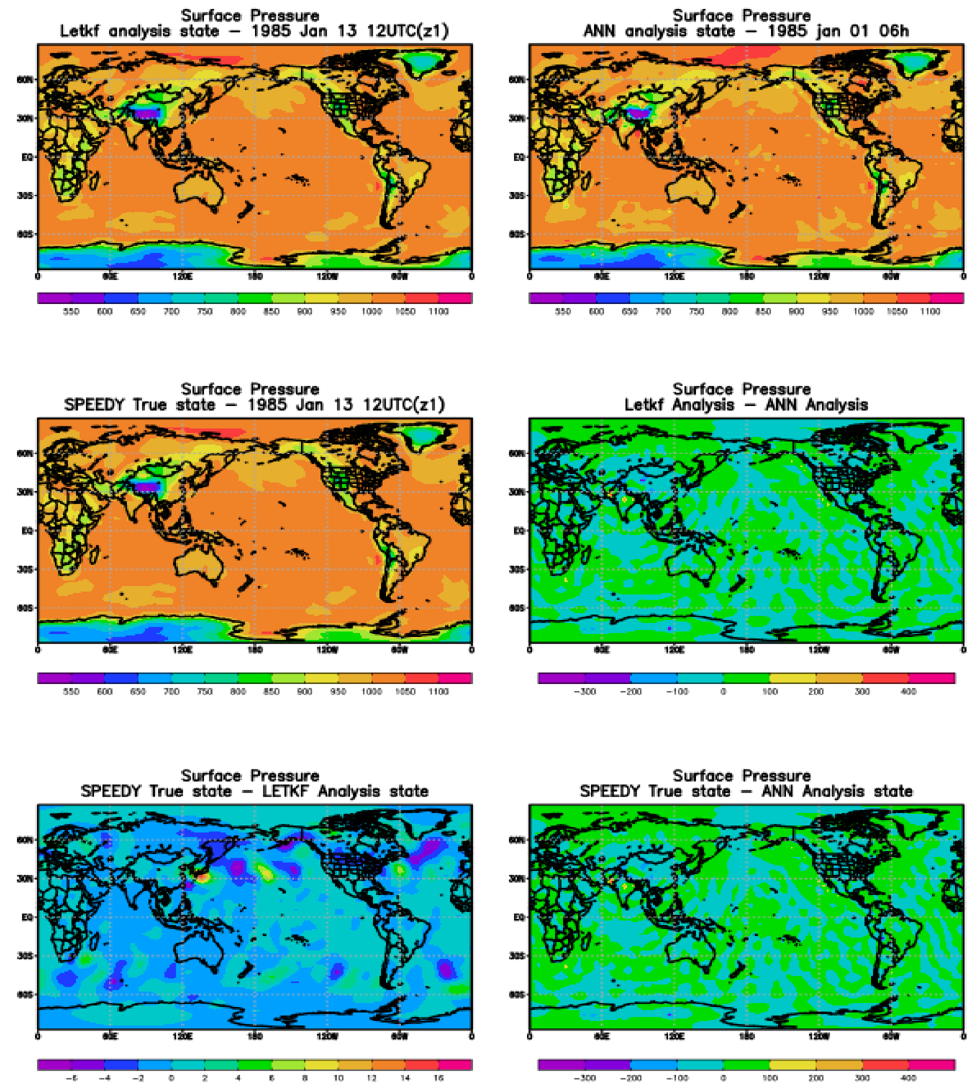

Figure 5: Surface Pressure (hPa) for 13 January 12UTCLETKF analysis, MLP-DA analysis, True model — and diferences: (LETKF and MLP-DA), (True model and LETKF), (True model and MLP-DA).

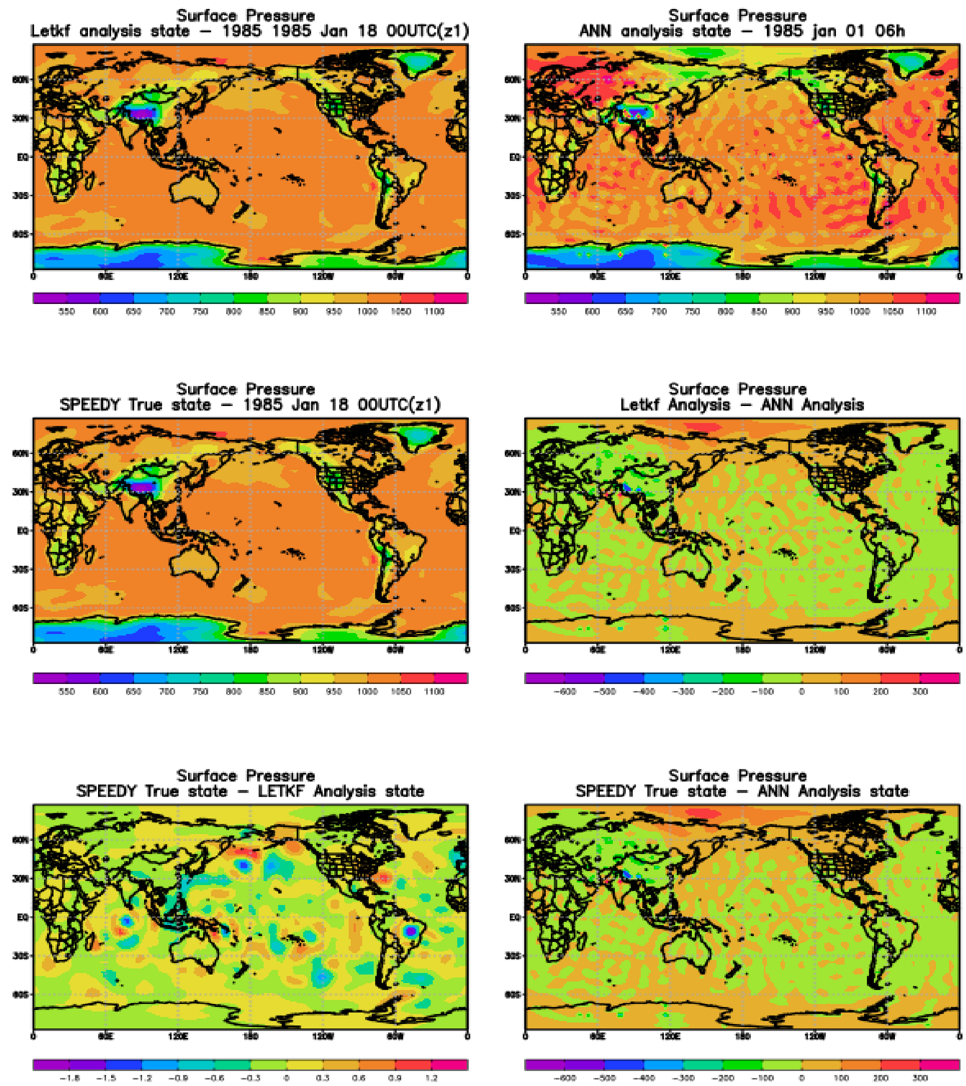

Figure 6: Surface Pressure (hPa) for 18 January 00UTC: LETKF analysis, MLP-DA analysis, True model — and diferences: (LETKF and MLP-DA), (True model and LETKF), (True model and MLP-DA).

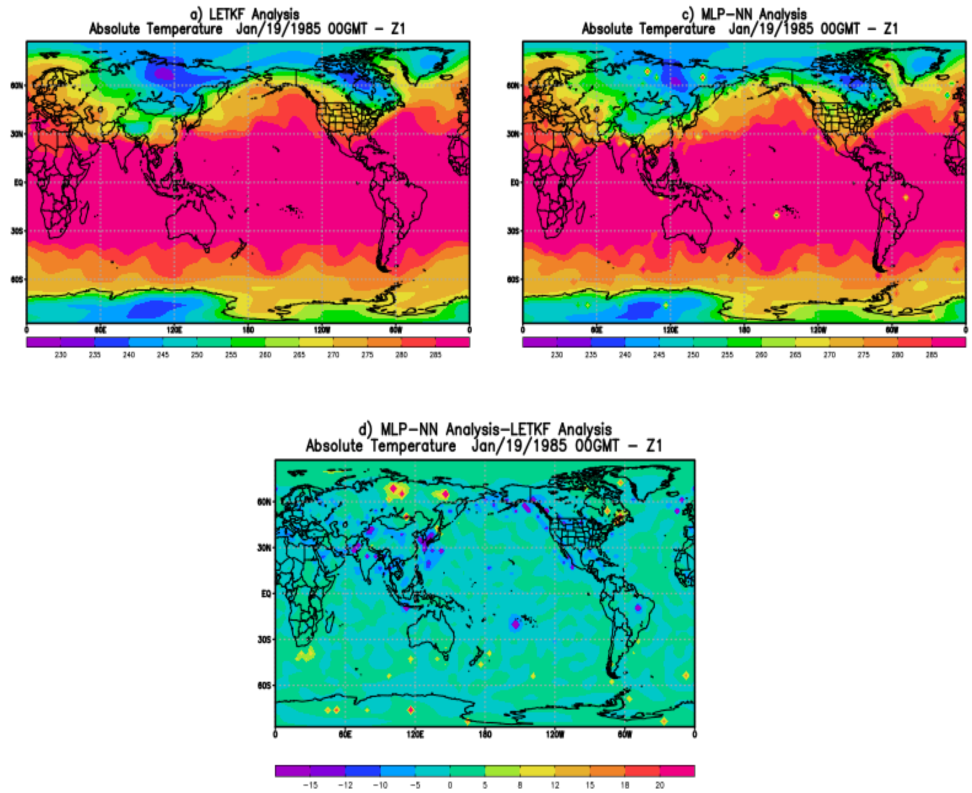

Figure 7: Absolute Temperature for 19 January 00UTC: LETKF analysis, MLP-DA analysis – and diferences: (LETKF and MLP-DA).

gies. Looking at Figure 6, we can realize significant differences on regions over Europe-Asia interface zone, and mainly over the oceans.

There are difference among points on the domain – see Fig. 2 over Alaska and Vladivostok (red dots), indicating the observation influence was not enough effetive by computing pseudo-observation with $M = 2$. One procedure to overcome such issue is to enhance the observation influence ($M \geq 3$).

Acknowledgments: Authors want to thank to Fapesp and CNPq, Brazilian agencies for research support.

References

[1] Bishop, H. C., Etherton, B. J., Majumdar, S. J.: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. Monthly Weather Review, 129, 420–436, 2001.

[2] Bourke,W.: A multilevel spectral model: I. formulation and hemispheric integrations. Monthly Weather Review, 102, 687–701, 1974.

[3] Burgers, G., van Leeuwen, P., Evensen, G.: Analysis scheme in the ensemble Kalman filter. Monthly Weather Review, 126, 1719–1724, 1998.

[4] Cintra, R. S., Campos Velho, H. F.: Global Data Assimilation using Artificial Neural Networks in SPEEDY Model. Proceedings of the International Symposium on Uncertainty Quantification and Stochastic Modelling. pp 648-654, Maresias, SP, Brazil, 2012.

[5] Cintra, R. S. C.; Campos Velho, H. F.; Anochi J.; Cocke S.: Data Assimilation by Artificial Neural Networks for the global FSU atmospheric model: Surface Pressure. 2nd Latin-American Congress on Computational Intelligence (LA-CCI), Curitiba, Brazil. CBIC and LA-CCI 2015.

[6] Cintra, R. S. C.; Campos Velho, H. F.: Data Assimilation by Artificial Neural Networks for an Atmospheric General Circulation Model. In: *Advanced Applications for Artificial Neural Networks* (Org.: Adel El-Shahat), IntechOpen, (2018) – DOI: 10.5772/intechopen.70791.

[7] Daley, R.: Atmospheric Data Analysis. Cambridge University Press, New York, USA, 1991.

[8] Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. Journal of Geophysics Research, 99(10), 143–162, 1994.

[9] Evensen, G.: The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dynamics, 53, 343–367, 2003.

[10] Furtado, H. C. M.; Campos Velho, H. F.; Macau, E. E. N.: Data assimilation: Particle filter and artificial neural networks. Journal of Physics. Conference Series (Online), v. 135, p. 012073, 2008.

[11] Harter, F. P., Campos Velho, H. F: New approach to applying neural network in nonlinear dynamic model. Applied Mathematical modeling, 32(12), 2621–2633, 2008.

[12] Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice Hall, 2nd Edition, 1999.

[13] Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique. Monthly Weather Review, 126, 796–811, 1998.

[14] Hunt, B., Kostelich, E. J., Szunyogh, I.: Efficient data assimilation for spatiotem- poral chaos: a local ensemble transform Kalman filter. Physica D, 230, 112-126. 2007.

[15] Kalman, R. E., Bucy, R. S.: New results in linear filtering and prediction theory. Trans. of the ASMEJournal of Basic Engineering, 83(Series D), 95–108, 1961.

[16] Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability. 2d ed., Cambridge University Press, New York, 2003.

[17] Miyoshi, T.; Yamane, S: Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. Monthly Weather Review, 135, 3841–3861, 2007.

[18] Molteni, F.: Atmospheric simulations using a GCM with simplified physical parametrizations: model climatology and variability in multidecadal experiments. Climate Dynamics, 20, 175–191, 2003.

[19] Nowosad, A., Rios Neto, A., Campos Velho, H.: Data assimilation in chaotic dynamics using neural networks. International Conference on Nonlinear Dynamics, Chaos, Control and Their Applications in Engineering Sciences, 212–221, 2000.

[20] Ott, E., Hunt, B. R., Szyniogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., York, J.: A local ensemble Kalman filter for atmospheric data assimilation. Tellus A, 56, 415–428, 2004.

[21] Widrow, B., Hoff,M. Adaptive switching circuits. IRE WESCON Conv. Record, Pt.4.94,95, 96–104, 1960.

[22] Whitaker, J. S., T. H. Hamill: Ensemble data assimilation without perturbed observations. Monthly Weather Review, 130, 1913–1924, 2002.