XLI CILAMCE
XLI Ibero-Latin American Congress on Computational Methods in Engineering
16-19 November, 2020 FOZ DO IGUAÇU PARANÁ - BRAZIL

NEURAL NETWORKS FOR DATA ASSIMILATION IN THE METROPOLITAN AREA OF RIO DE JANEIRO

Vinícius A. Almeida[1], Nelson F. F. Ebecken[2], Haroldo F. de Campos Velho[3], Gutemberg B. França[1]

[1] *Laboratory for Applied Meteorology, Federal University of Rio de Janeiro*
Rio de Janeiro - RJ, Brazil
vinicius@lma.ufrj.br, gutemberg@lma.ufrj.br
[2] *COPPE, Federal University of Rio de Janeiro*
Rio de Janeiro - RJ, Brazil
nelson@ntt.ufrj.br
[3] *Associate Laboratory for Computing and Applied Mathematics, National Institute for Space Research*
São José dos Campos - SP, Brazil
haroldo.camposvelho@inpe.br

Abstract. The use of neural networks for data assimilation in the terminal area of Rio de Janeiro - emulating the 3D-Var method as implemented in the Weather Research and Forecasting Data Assimilation module - is explored here. Surface and upper-air data (air temperature, relative humidity and wind speed and direction) from airport stations and 6-hour forecast from WRF are used as input for the model and the 3D-Var analysis for each grid point is used as target variable. Periods of 168h from 2014 and 2015 are used with 6h and 12h assimilation cycles for surface and upper-air data, respectively. The neural network model was built using two different approaches: multilayer perceptron with static topology in Weka and a metaheuristic called Multi-Particle Collision Algorithm, where different topologies are tested until the optimum solution is found. Results show that Weka-NN and MPCA-NN are able to emulate the 3D-Var method with negligible differences in comparison of the magnitude of the assimilated data. Also, the neural networks are able to speed up the data assimilation process. In this study, neural network models were able to run from 70 to 100 times faster than the conventional method (3D-Var analysis) under the same hardware configurations. This time reduction enables the execution of data assimilation methods using less CPU time and using personal computers.

Keywords: Neural network, data assimilation, surface data, profile data.

1 Introduction

Numerical weather prediction is an initial-value problem as stated in the beginnings of the twentieth century by Bjerknes [1]. Therefore, given an estimate of the present state of the atmosphere (initial conditions), and appropriate surface and lateral boundary conditions, the model simulates (forecasts) the atmospheric evolution – see Kalnay [2]. Currently, data assimilation methods are used by operational centers to determine as accurately as possible the state of the atmospheric (also for ocean circulation and atmospheric pollutant predictions) flow [3]. The determination of such state is essential, sophisticated, and demands high computational effort [4]. Several methods have been developed since the 1950s to tackle this problem. See Daley [4], Talagrand [3], Zupanski and Kalnay [5], and Kalnay [2] for reviews on the subject.

The traditional approaches applied for data assimilation usually are both computer and time-consuming requiring super computers for daily use in operational centers. Therefore, in the last two decades efforts have been made in order to overcome the resource restrictions of data assimilation speeding-up the process without loss of quality. In this context, neural networks have been applied to emulate several data assimilation methods, such as: Kalman filter (Härter and de Campos Velho [6]), Particle Filter (Furtado et al. [7]) and variational methods (Furtado et al. [8]).

The Weather Research and Forecasting model (WRF) is a community atmospheric modeling system, and its development and capabilities are the result of the contributions of a host of individuals and institutions from over the years (Skamarock et al. [9]). This model contains a limited-area three-dimensional variational data assimilation

(3DVAR) system applicable to both synoptic and mesoscale numerical weather prediction (Barker et al. [10]).

Although neural networks have already been applied for data assimilation in different contexts, the use of machine learning to emulate variational methods in limited-area numeric weather prediction, mainly for the Brazilian aviation sector, has not been explored yet. Therefore, the objective of this work is to evaluate the ability of an artificial neural network to emulate the 3DVAR method for surface and upper-air data in the metropolitan area of Rio de Janeiro, Brazil. The present article is part of a sequence of studies related to nowcasting that have been executed by the Applied Meteorological Laboratory at the Federal University of Rio de Janeiro, following Almeida [11], Silva et al. [12], França et al. [13], França et al. [14] and Almeida et al. [15]. All these studies encompass researches based on artificial intelligence methods for weather forecasts, mainly for high-impacting phenomena for aviation.

2 Data Assimilation Techniques

2.1 Variational method (3D-Var)

Variational method is a well known scheme applied to data assimilation, where a cost function needs to be minimized Kalnay [2], Daley [4]. The WRF model has a module for data assimilation, including the 3D variational (3D-Var) approach Barker et al. [10]. The analysis x^a – the computed initial condition for the numerical weather prediction new cycle – is determined to find a minimum for the cost function (J) given by:

$$J(x) = 1/2\{[y - H(x)]^{\mathrm{T}} R^{-1} [y - H(x)] + [x - x^b]^{\mathrm{T}} B^{-1} [x - x^b]\} \tag{1}$$

where R and B are error variance matrices for the background value x^b and the observation vector y. Assuming the statistical independence for the measurements, the matrix R is a diagonal one. The matrix B is found as a difference between two short predictions: $B \approx [x^b_{12hs} - x^b_{24hs}]^{\mathrm{T}}\,[x^b_{12hs} - x^b_{24hs}]$ – see Barker et al. [10].

2.2 Neural network (MLP-NN)

The multi-layer perceptron (MLP) [16] is probably one of the most neural network (NN) architecture employed. It is a supervised NN configured by produce an output similar to the training dataset – the *target*, where the connection weights are calculated by a learning process. One learning method is the delta rule of the error back-propagation. Two schemes are used to configure the MLP-NN: as an NN architecture solution from a optimization problem – the one is solved by the multi-particle collision algorithm (MPCA) (Pacheco da Luz et al. [17], Luz et al. [18]), and the application of the WEKA (Waikato Environment for Knowledge Analysis) package software for machine learning (Witten et al. [19]).

The MPCA approach to identify the best NN topology is applied to compute the minimum of the objective function (Anochi and de Campos Velho [20]):

$$J(\mathbf{P}) = \Big[c_1\, e^{x^2} + c_2\, y + 1\Big] \times \left[\frac{\rho_1\, E_{\mathrm{train}}(\mathbf{P}) + \rho_2\, E_{\mathrm{gen}}(\mathbf{P})}{\rho_1 + \rho_2}\right] \tag{2}$$

where $\mathbf{P}$ is the unknown vector; E_{train} and E_{gen} are errors for training phase and generalization (after training: the square difference between the NN output and the analysis produced by 3D-Var). The parameters ρ_1 and ρ_2 are employed to describe the balance between the two error types. Finally, x is the number of neurons in the network, y is the number of iterations to reach the convergence in the training phase, and c_1, c_2 are free parameters. The unknown vector $\mathbf{P}$ collects the parameters: number of neurons in the maximum of two hidden layers, type of activation function (logistic, Gaussian, or hyperbolic tangent), and parameters for training phase (learning ratio and momentum). More information about the optimization process can be found in Pacheco da Luz et al. [17] and also on the MPCA GitHub page at `https://github.com/scsr-inpe/mpca-ann`.

The same number of hidden layers (2), and the same number of neurons for each hidden layer (25) were also applied to the WEKA software (Witten et al. [19]). The differences between the two NN formulation are: backpropagation algorithm for weights adjustment, fixed learning rate (0.3), fixed momentum (0.2), number of epochs (500), sigmoid activation function used for all hidden units. In comparison to the static topology used in WEKA experiments, the MPCA final topology (after the optimization process) is presented in Table 1.

Table 1. Final topology for MPCA experiments. The columns represent the number of hidden units in hidden layer one (HL1) and two (HL2), the best activation function (AF), the momentum (alpha) and the learning rate (eta).

Variable	HL1	HL2	AF	Alpha	Eta
Air temperature (°C)	11	24	logistic	0.53	0.80
Relative Humidity (%)	9	0	hyperbolic tangent	0.67	0.81
Zonal wind (m/s)	5	0	logistic	0.10	0.46
Meridional wind (m/s)	21	11	logistic	0.09	0.84

3 Design for Numerical Executions

3.1 Data and study area

The study area is the metropolitan area of Rio de Janeiro and its surroundings (Fig. 1) located approximately at latitude 22°55'44.3"S and longitude 43°24'21.1"W. The most import airports in the region are located in Figure 1 identified by their International Civil Aviation Organization (ICAO) codes: Santos Dumont Airport (SBRJ), Galeão International Airport (SBGL), Santa Cruz Air Force Base (SBSC), Jacarepaguá Airport and Afonsos Air Force Base (SBAF).

Each airport is responsible for local hourly routine and special surface observations of several meteorological parameters as surface wind (direction and speed), visibility, significant weather, cloud cover, air and dewpoint temperature, and station pressure. Besides, the SBGL airport has an upper-air (or sounding) station that produces regularly atmospheric soundings twice a day, the atmospheric profile of pressure, air and dewpoint temperature, relative humidity, and wind (direction and speed), from the surface up to more than 25 km.

Figure 1 also presents the simulation domain where the WRF experiments were performed with 2-km horizontal resolution. It is important to highlight that the use of nested domains was tested for smaller integration periods and the results were similar to single-domain simulations and, thus, here only the latter results are presented due to less computational resources demand.

Here, four variables (air temperature, wind direction and speed, and relative humidity) retrieved at sensor locations at the airports every six and twelve hours from surface and upper air stations, respectively, were evaluated for data assimilation.

Figure 1. Study area and sensor locations. The codes in the zoom represent the ICAO identifications codes for Santos Dumont Airport (SBRJ), Galeão International Airport (SBGL), Santa Cruz Air Force Base (SBSC), Jacarepaguá Airport and Afonsos Air Force Base (SBAF).

3.2 Description of Experiments

The period of the experiments performed in this work was in February. This month is characterized by a peak of atmospheric discharges in Rio de Janeiro (Paulucci et al. [21]) and the development of intense convective events. Lastly, after the end-of-the-year holidays, February has a peak of movements in airports, becoming the period relevant – since our study is a joint research between the Federal University of Rio de Janeiro and the Department of Airspace Control (DECEA), a division of the Brazilian Air Force.

Two 7-day simulations (168 hours), starting on Feb 1st, 2014 and Feb 1st, 2015, were performed using the WRF model. As mentioned above, the data assimilation is carried out every 6 hours for surface variables (air temperature, relative humidity, and wind direction and speed) at the airport locations, and every 12 hours for upper-air variables (air temperature, relative humidity, and wind direction and speed) at SBGL location.

The experiment steps are described as follows: (i) a white-noise perturbation is applied to the background field at the airport locations for surface and upper-air data, generating synthetic observations; (ii) synthetic observations are placed on the exact coordinates where real sensors are located; (iii) new analysis field is generated from synthetic observations and background field using the 3D-Var data assimilation technique and 6-h integration is performed ; (iv) the 6-h forecast output from step (iii) is retrieved and the steps (i)–(iii) are repeated considering surface data assimilation every 6h and upper-air data assimilation every 12h; (v) the impact of the synthetic observations on the surroundings is computed using 5-grid points radius with the value. In grid points under the influence of more than one station, the inverse of the distance is used as a weighting factor; (vi) synthetic observations, background field, and analysis are employed; (vii) a pre-processing is executed for data cleansing and normalization; (viii) a shuffle and split are performed on the dataset defining three subsets for each training software: (a) MPCA: training (60%), generalization (20%) and test (20%) ; (b) Weka: as there is no generalization dataset in Weka, this dataset was provided as a "supplied test set" option and, after the model training, this model was applied to the remaining dataset (the 20% used as test dataset by MPCA); (ix) Training, validation and test are performed using MPCA (Luz et al. [18]) and Weka (Witten et al. [19]); (x) an evaluation is performed comparing the results for the data assimilation process by the 3D-Var and the trained neural networks applied to the test dataset .

4 Results and discussions

This section presents the results of the trained neural networks (MPCA and Weka) applied to the generalization dataset.

Figure 2 presents a comparison between the probability distribution of the 3D-Var approach and neural networks (Weka-NN and MPCA-NN) for air temperature (a and b), relative humidity (c and d), wind u-component (e and f) and wind v-component (g and h). Both MPCA-NN and Weka-NN have difficulties to reproduce values in tails what is expected since this region of the probability distribution is where there is more uncertainty. Neural networks have greater errors for vector variables (components of wind) in comparison to scalar variables (air temperature and relative humidity). However, the Weka-NN is still able to reproduce the behavior of the upper distribution tail. Weka-NN show an overall superior performance compared to MPCA-NN.

A complete evaluation was also performed looking at the maps from the difference fields (not shown here) between the control and the analysis – the estimated initial condition – to investigate the impact of the data assimilation on the simulation domain. Weka-NN and MPCA-NN are able to reduce the perturbation in the observed data showing the noise smoothing behavior as expected for the data assimilation systems at the sensor locations and surroundings.

Table 1 presents a summary of the statistics for the generalization dataset for both Weka-NN and MPCA-NN. The magnitude of the Root mean square error (RMSE) show that for both methods the differences between the 3D-Var approach and the neural networks were smaller than the white-noise perturbation of the assimilated data (around 10% of the variable magnitude) for scalars (air temperature and relative humidity). On the other hand, differences with greater magnitudes were observed for vector variables (u- and v-component of the wind) and remains a challenge for future works.

It is noteworthy to explain why the comparison is performed only between neural networks and the 3D-Var approach, not to the data used in the assimilation process. As presented in the objectives of the present study, the intention is to test the hypothesis that neural networks are able to emulate 3D-Var method more efficiently (with lower computational demand and faster). The evaluation of the 3D-Var method is widely discussed in the related papers in Barker et al. [10].

One of the most important aspects of the results is not presented in Figure 2 and Table 2. The present section up to this point showed that the differences between the traditional approach (3D-Var) and neural networks (Weka-NN and MPCA-NN) are negligible. However, small differences indicate a sort of equivalence but is not sufficient to justify the replacement of the traditional method. The answer for such question relies on the CPU time and infrastructure demanded by each methods. While the 3D-Var method demands computers with great power (many cores and GB of RAM) the neural networks are able to run in personal computers with simple configurations in a wide variety of operational systems. Another aspect worthy of note is that the neural networks (both Weka-NN and MPCA-NN) are able to run from **70 to 100 times faster** than 3D-Var.

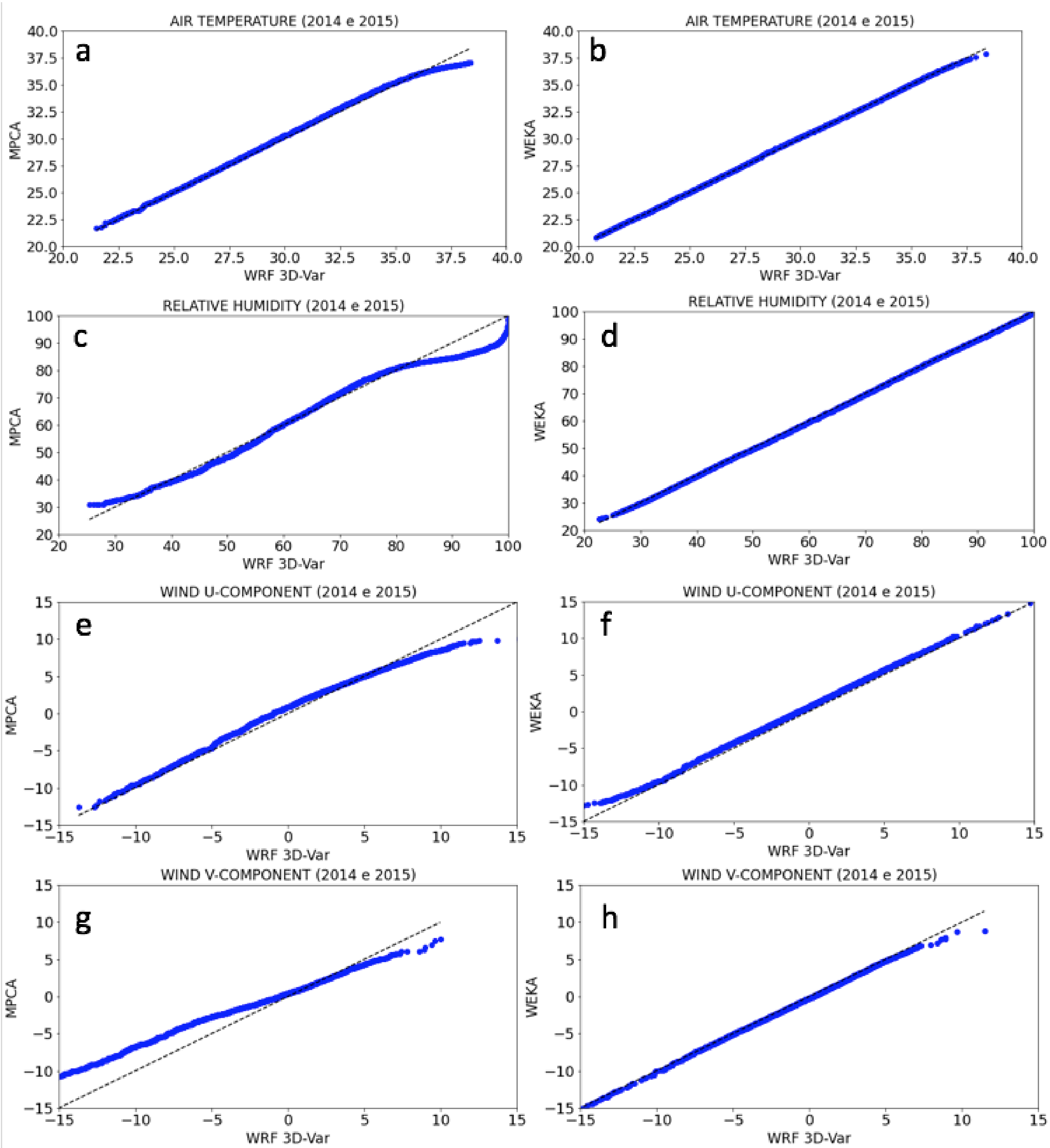

Figure 2. Quantile-quantile plot comparing the probability distribution of the trained neural networks (MPCA and Weka) (y-axis) and the actual values (x-axis) for air temperature (a and b), relative humidity (c and d), wind u-component (e and f) and wind v-component (g and h). The dashed line represents the perfect correspondence (1:1) between the trained neural network and the 3D-Var approach.

Table 2. Statistics for the generalization dataset.

Variable	Software	Root mean square error (RMSE)	Standard Deviation (SD)
Air temperature (^{o}C)	Weka	0.09	0.07
	MPCA	0.37	0.31
Relative Humidity (%)	Weka	0.48	0.42
	MPCA	4.04	3.10
Wind (u-component) (m/s)	Weka	0.32	0.72
	MPCA	1.05	0.63
Wind (v-component) (m/s)	Weka	0.48	0.25
	MPCA	1.56	1.10

5 Conclusions

This paper provides preliminary results of the evaluation of trained neural networks – using static neural network for the Weka framework and a metaheuristic for neural network self-configurated by MPCA – to emulate the 3D-Var method for data assimilation in a limited-area atmospherical model (WRF). The goal was to evaluate data assimilation of surface and upper-air data retrieved from airports in the metropolitan area of the Rio de Janeiro in 168-h simulations in February, a critical month due to intense convective activity in the study area.

Current results include: (i) similar analysis produced by the 3D-Var method and both trained networks; (ii) greater differences for vector variables; (iii) greater standard deviation for almost all MPCA-NN results, excluding the u-component of the wind, in comparison to Weka-NN; (iv) results follow preview reports(Furtado et al. [8], Cintra and Campos Velho [22], Härter and de Campos Velho [23]); (v) after training, the neural networks are able to perform from 70 to 100 times faster than 3D-Var approach.

More computational effort than 3D-Var is verified by 4D-Var method. Using the test case for the native variational schemes in a quad-core computer, the 3D-Var demands 45 seconds while 4D-Var demanded 9550 seconds (approximately 210 times slower than 3D-Var). Future works will investigate the performance of neural networks face on 4D-Var data assimilation.

As a final note, considering the operational centers with a relative short time window to elaborate forecast bulletins, the reduction of CPU-time of order 71 times from a standard method to the assimilation cycle is important for several aspects: possibility of assimilation of a greater amount of data and/or the use of finer model computer resolution.

Acknowledgements. This study is funded by the Department of Airspace Control (DECEA) via the Brazilian Organization for Scientific and Technological Development of Airspace Control (CTCEA) (GRANT: 002-2018/COPPETEC_CTCEA). Authors would also like to thank the National Council for Scientific and Technological Development (CNPq, Brazil) for the research grants: HFCV (CNPq: 312924/2017-8), and GBF (CNPq: 304441/2018-0).

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] Bjerknes, V., 1904. Das problem der wettervorhers-age, betrachtet vom standpunkte der mechanik und der physik. *Meteor. Z.*, vol. 21, pp. 1–7.
[2] Kalnay, E., 2002. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press.
[3] Talagrand, O., 1997. Assimilation of observations, an introduction (gtspecial issueltdata assimilation in meteology and oceanography: Theory and practice). *Journal of the Meteorological Society of Japan. Ser. II*, vol. 75, n. 1B, pp. 191–209.
[4] Daley, R., 1991. *Atmospheric Data Analysis*. Cambridge University Press.

[5] Zupanski, M. & Kalnay, E., 1999. Principles of data assimilation. in global energy and water cycles.

[6] Härter, F. P. & de Campos Velho, H. F., 2010. Multilayer perceptron neural network in a data assimilation scenario. *Engineering Applications of Computational Fluid Mechanics*, vol. 4, n. 2, pp. 237–245.

[7] Furtado, H. C. M., de Campos Velho, H. F., & Macau, E. E. N., 2008. Data assimilation: Particle filter and artificial neural networks. *Journal of Physics: Conference Series*, vol. 135, pp. 012073.

[8] Furtado, H. C. M., de Campos Velho, H. F., & Macau, E. E. N., 2011. Neural networks for emulation variational method for data assimilation in nonlinear dynamics. *Journal of Physics: Conference Series*, vol. 285, pp. 012036.

[9] Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., & Powers, J. G., 2019. A description of the advanced research wrf version 4.

[10] Barker, D., Huang, W., Guo, Y., Bourgeois, A., & Xiao, A., 2004. A three-dimensional variational data assimilation system for mm5: Implementation and initial results. *Monthly Weather Review - MON WEATHER REV*, vol. 132.

[11] Almeida, M. V., 2009. *FORECASTING SHORT-TERM OF VISIBILITY AND CEILING FOR AIRPORT GUARULHOS - SP BY APPLICATION OF TECHNIQUES OF ARTIFICIAL NEURAL NETWORKS*. Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

[12] Silva, W., Albuquerque, F., França, G., & Matschinske, M., 2016. Conceptual model for runway change procedure in guarulhos international airport based on sodar data. *The Aeronautical Journal*, vol. 120, pp. 725–734.

[13] França, G. B., de Almeida, M. V., & Rosette, A. C., 2016. An automated nowcasting model of significant instability events in the flight terminal area of rio de janeiro, brazil. *Atmospheric Measurement Techniques*, vol. 9, n. 5, pp. 2335–2344.

[14] França, G. B., de Almeida, M. V., Bonnet, S. M., & Neto, F. L. A., 2018. Nowcasting model of low wind profile based on neural network using sodar data at guarulhos airport, brazil. *International Journal of Remote Sensing*, vol. 39, n. 8, pp. 2506–2517.

[15] Almeida, V. A., França, G. B., & de Campos Velho, H. F., 2020. Short-range forecasting system for meteorological convective events in rio de janeiro using remote sensing of atmospheric discharges. *International Journal of Remote Sensing*, vol. 41, n. 11, pp. 4372–4388.

[16] Haykin, S., 1999. *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, New Jersy, USA, 2 edition.

[17] Pacheco da Luz, E., Becceneri, J., & Campos Velho, H. F., 2008. A new multi-particle collision algorithm for optimization in a high performance environment. *Journal of Computational Interdisciplinary Sciences*, vol. 1, n. 1, pp. 03–10. URL: http://www.epacis.net/jcis/PDF_JCIS/JCIS11-art.01.pdf.

[18] Luz, E. F. P., Becceneri, J. C., & Campos Velho, H. F., 2011. Multiple particle collision algorithm applied to radiative transference and pollutant localization inverse problemsk. In *IEEE International Symposium on Parallel and Distributed Processing Workshops and PhD Forum*, pp. 347–35. DOI: 10.1109/IPDPS.2011.171.

[19] Witten, I. H., Frank, E., & Hall, M. A., 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann, Amsterdam, 3 edition.

[20] Anochi, J. A. & de Campos Velho, H. F., 2014. Optimization of feedforward neural network by multiple particle collision algorithm. pp. 128–134, Orlando, FL, United States of America.

[21] Paulucci, T., França, G., Libonati, R., & Ramos, A., 2019. Long-term spatial–temporal characterization of cloud-to-ground lightning in the metropolitan region of rio de janeiro. *Pure and Applied Geophysics*, vol. 176, pp. 1–15.

[22] Cintra, R. S. C. & Campos Velho, H. F., 2012. Global data assimilation using artificial neural networks in speedy model. In *Proceedings of the 1st International Symposium on Uncertainty Quantification and Stochastic Modeling*, pp. 648–654, Maresias, são Sebastião (SP), Brazil.

[23] Härter, F. P. & de Campos Velho, H. F., 2012. Data assimilation procedure by recurrent neural network. *Engineering Applications of Computational Fluid Mechanics*, vol. 6, n. 2, pp. 224–233.