proceedings

MDPI

sType of the Paper: Proceedings

Bias correction method based on artificial neural networks for quantitative precipitation forecast

Adrián Fuentes Barrios [1,*] and Maibys Sierra Lorenzo [1,*]

[1] Atmospheric Physics Center, Institute of Meteorology, Havana, Cuba
* CorrespondenceFuentes, A: adrian.fuentes@insmet.cu
Sierra, M: maibys.lorenzo@insmet.cu

† Presented at the title, place, and date.

Abstract: The nowcasting and very short-term prediction system (SisPI, for its acronym in Spanish) is among the tools used by the National Meteorological Service of Cuba, for the quantitative precipitation forecast (QPF). SisPI uses the WRF model as the core of its forecasts and one of the challenges to overcome is to improve the precision of the QPF. With this purpose, in this work we present the results of the application of a bias correction method based on artificial neural networks. The method is applied to the highest resolution domain of SisPI (3km), and the correction is made from the precipitation estimation GPM satellite product. Results shows higher correlation with the artificial neural network model in relation to the values predicted by SisPI (0.76 and 0.34 respectively). The mean square error applying the artificial neural network model is 3.69, improving the performance of SisPI with 6.78. In general, the bias correction has good ability to correct the precipitation forecast provided by SisPI, being less evident in cases where precipitation is reported and SisPI is not capable of forecasting it. In cases of overestimation by SisPI (which happens quite frequently), the correction achieves the best results.

Keywords: QPF; WRF; artificial neural networks; bias correction

Citation: Fuentes, A.; Sierra, M. Title. *Proceedings* **2021**, *65*, x. https://doi.org/10.3390/xxxxx

Received: date
Accepted: date
Published: date

1. Introduction

Artificial neural networks are a mathematical technique inspired by biological neural networks [1, 2]. At present, the development of this tool continues, being widely used in different branches of science including atmospheric sciences and related areas [3-5]. Within meteorology there are many applications, highlighting its use in the detection of cloud patterns, in weather and climate forecasting, and as a method for correcting forecast errors generated by numerical weather forecast models (NWM), among others [5].

This research is focused on this last application of the ANN. The work presented is one more contribution that explores the use of artificial neural networks to increase the precision of the numerical forecast, in particular, the quantitative precipitation forecast (QPF). For this purpose, a mul-tilayer perceptron is used as the network model. It is trained using a configuration obtained in previous studies [5]. As data for the training, the estimation of the satellite precipitation of the GPM product and the precipitation forecast with 3 km of spatial resolution of the nowcasting and very short-term forecast system (SisPI) [6-8] are used.

In the section Model and Configuration of the ANN, the information regarding the ANN model and it's configuration is presented. The details of the data used, as well as the SisPI description can be found in the section Data used, while the results achieved for two case studies are discussed in the section Results and Discussion, followed by a preliminary conclusions.

2. Model and Configuration of the ANN

As was mentioned before, a multi-layer perceptron model (MLP) is used in order to extend a previous result in which the observation of meteorological surface stations were used for training. The MLPs describes an artificial neural network that uses the output to establish a relationship with the input data. In our case, the input data was the QPF values directly taken from SisPI and as output the precipitation estimation of GPM was used. The MLP was configured with 64 neurons in the hidden layer and a sigmoid function activation for the hidden and the output layers. The machine learning platform Tensorflow and the Keras library [9], were used for implementing the MLP. The training , validation and verification steps were development using the available data in the period of 2018-2019.

3. Data used

As observation data for training, the precipitation data from the Global Precipitation Mission was used. In particular we use the GPM_3IMERGHH, which is the GPM IMERG Final Precipitation L3 product (version 06) with temporal resolution of 30 minutes and spatial resolution of 0.1 degree × 0.1 degree. This is a multi-satellite precipitation product with global coverage and it is a Level 3 NASA product that unifies and inter-calibrates data of about some constellation and types of satellites from several space agencies [10].

The correction using the MLP was applied to the forecast of SisPI over the highest spatial resolution domain (3km, see Figure 1). SisPI uses the Weather Research & Forecast (WRF) V3.8.1 model [11], as numerical forecast core initialized with the 0.5 degrees of spatial resolution output of the Global Forecast System (GFS). The main details of the physical configuration are shown in Table 1. This system generates 24 forecast hours updated four times during day at 0000, 0600, 1200 and 1800 UTC, in this investigation is used the 0000 UTC initialized forecast.

Table 1. Physical configuration of the WRF used in SisPI (see [11] for the details of each parameterization).

Parameters	Settings
Spatial resolution	Three nested domains of 27, 9 and 3 km of resolution
Nx	145, 162, 469
Ny	82, 130, 184
Nz	28, 28, 28
Domain center	21.8 N, 79.74 W
Time step	150s
Microphysics	WSM5,WSM5, double moment Morrison
Cumulus	Grell-Freitas, Grell-Freitas, not activated
Long radiation	RRTM scheme: Rapid Radiative Transfer Model
Short radiation	Dudhia scheme: Simple downward integration allowing efficiently for clouds and clear-sky absorption and scattering
Surface physics	Noah Land Surface Model: Unified NCEP/NCAR/AFWA scheme with soil temperature and moisture in four layers, fractional snow cover and frozen soil physics.
Surface layer	Eta similarity: Used in Eta model
PBL	Mellor-Yamada-Janjic, Mellor-Yamada-Janjic, Mellor-Yamada-Janjic

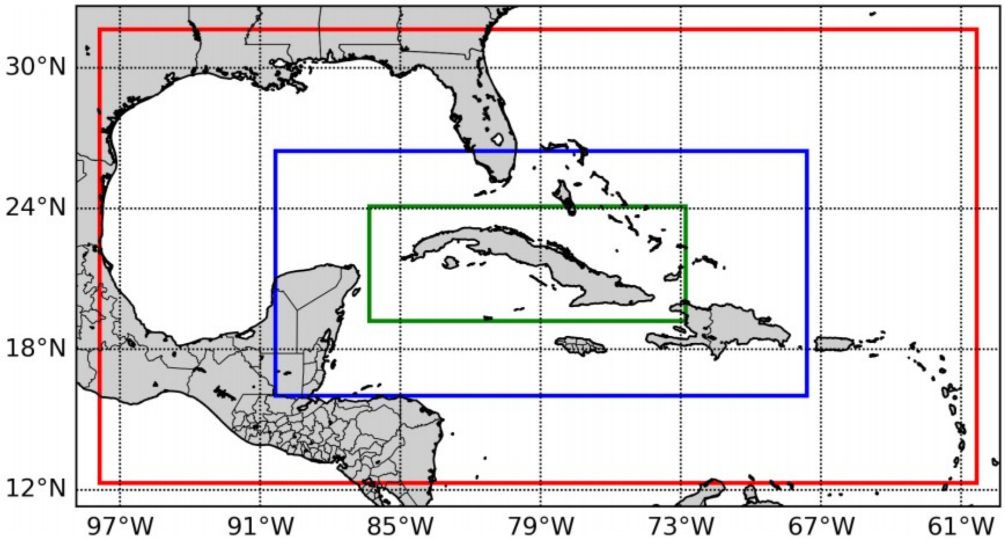

Figure 1. Simulation domains for SisPI. The green square corresponds with 3km resolution domain used in this study.

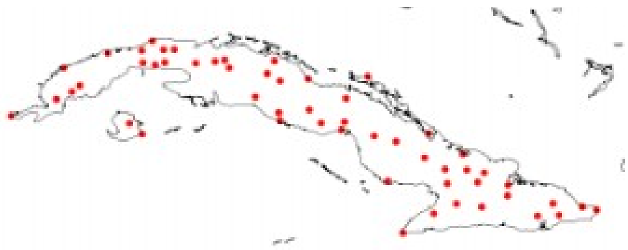

Figure 2. Location of the meteorological surface stations.

Once the training step was complete, the validation and the verification steps were carried out using the observations of the surface meteorological stations. Figure 2 shows the location of the 67 surface weather stations that were included in this study. A conventional verification process was applied in order to analyze the ability of the MLP for correcting the bias. The following statistical metrics were computed: Mean Absolute Error (**mae**), the Mean Square Error (**mse**) and Pearson's Correlation Coefficient (p_{corr}); applying the cell-point verification approach [12].

3. Results and Discussion

In this section we discuss some preliminary results that shows the improvement for the QPF when an ANN model is used as bias correction method. The analysis is done taken two study cases: July 5th and 10th of 2019. Rainfall between July 4th and 10th , 2019 was encouraged by the evening instability as a result of the diurnal warming, the sufficient moisture content in the lower troposphere and conditions in the mesoscale, together with the transit of active tropical waves through the seas at the South of Cuba (http://www.insmet.cu/asp/genesis.asp?TB0=PLANTILLAS&TB1=MES&TB2=/Mes/JULIO2019.HTM&TB3=2019). The reason for selecting these cases is because are one of the situations in which SisPI frequently fails [6,7].

3.1 *Study case July 5th 2019*

The **mae**, **mse** and p_{corr} for all forecast times are shown in Figures 3, 4 and 5 respectively. Notice that for the first 7 forecast hours the correction through the MLP model is greater with an error reduction of 2.0 mm/3h for **mae** and 60 mm/3h for **mse**. In the evening hours, although there is also a decrease in the error values, this decrease is more discreet. This makes sense when one takes into account that when using the SisPI runs initialized at 0000 UTC, in the first 7 hours the WRF is in the spin up period which can extend to the first 12 hours of forecasting. Therefore, for this case study, the MLP model constitutes a tool not only to reduce the error, but also to enhance the quality of the forecast in the WRF spin up period.

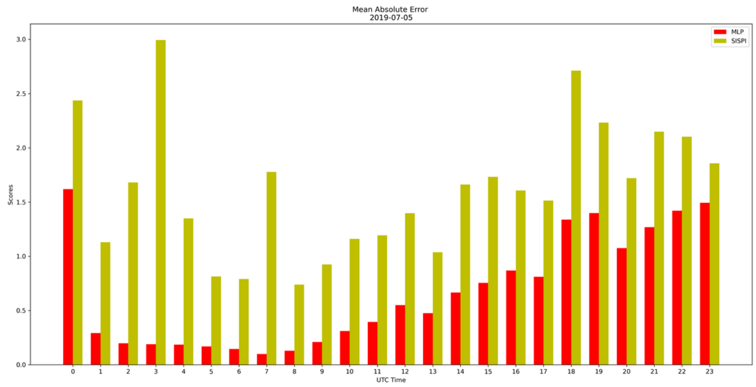

Figure 3. Mean absolute error of the SisPI forecast for July 5th 2019 and the SisPI forecast after bias correction with the MLP model.

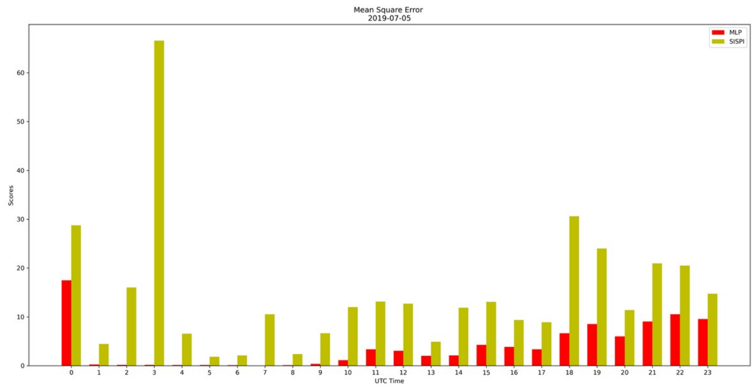

Figure 4. Mean square error of the SisPI forecast for July 5th 2019 and the SisPI forecast after bias correction with the MLP model.

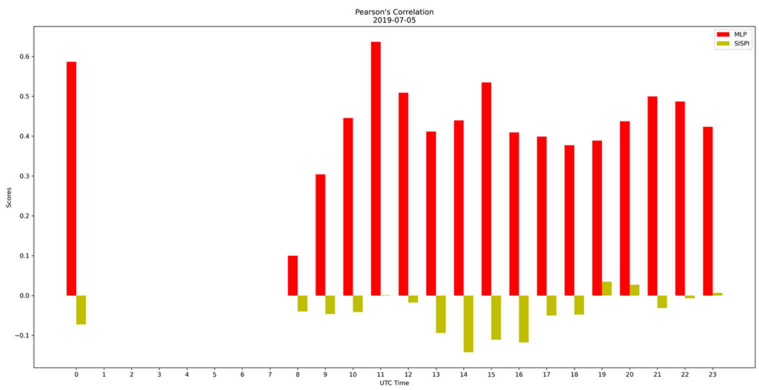

Figure 5. Pearson's correlation of the SisPI forecast for July 5th 2019 and the SisPI forecast after bias correction with the MLP model.

The behavior of the correlation curve indicates that the SisPI forecast behaves contrary to what was recorded in the observations. The negative values of p_{corr} suggest that when precipitation was recorded in the surface meteorological station, the SisPI did not predict rain, or it could be the opposite. However, a more rigorous and in-depth study must be done to be able to affirm this behavior. When it comes to very high resolution forecasts, errors have a double penalty due to position error. Therefore, the fact that SisPI does not predict precipitation at a given point could be due to a position error of the precipitation area. In the case of the correction with the MPL, however, the previous situation is corrected, and the correlation reach values of up to 0.6, being positive for all forecast periods. In the spin up hours, p_{corr} is zero since the mean is zero as well.

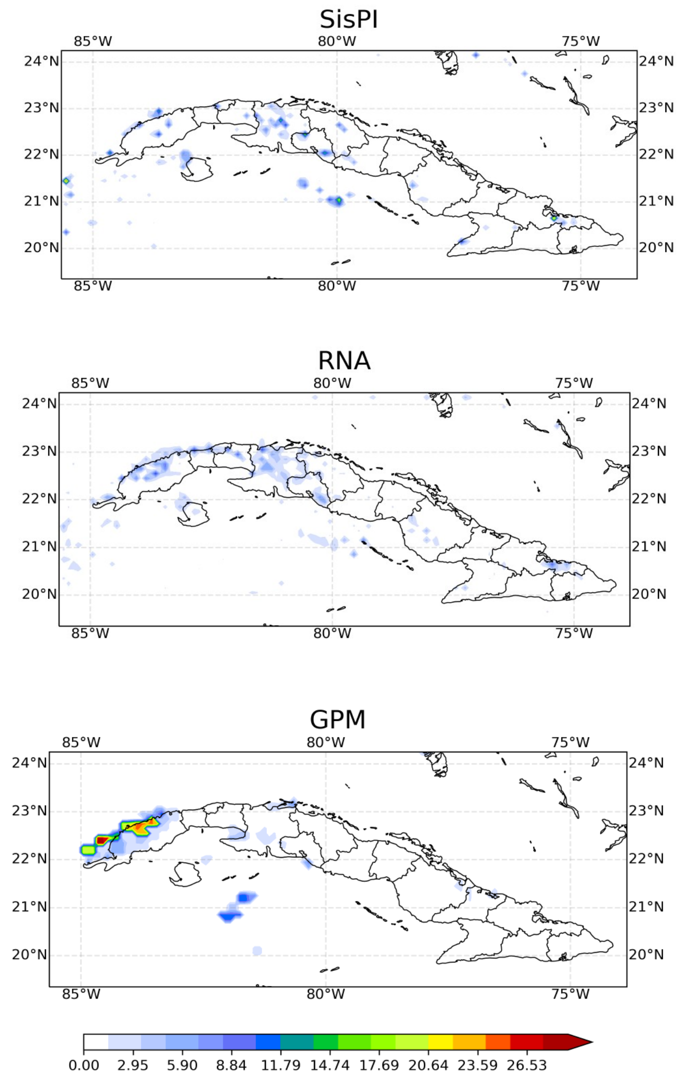

Figure 6. SisPI precipitation forecast (top panel), MLP correction (middle panel) and precipitation estimated by GPM product (bottom panel) for 5th July 2019 at 1800 UTC.

Figure 6, shows the SisPI forecast, the corrected SisPI forecast and the GPM precipitation for 1800 UTC. It can be seen that the SIsPI forecast underestimates the precipitation over the province of Pinar del Río not only quantitatively but also from a spatial point of view. While the GPM estimate presents values of more than 20 mm / 3h, the SisPI barely predicts 11 mm / 3h. On the province of Matanzas the opposite is observed, the SisPI suggests the occurrence of rain practically in the entire province being a false alarm. The correction fails to improve SisPI's prognosis. Over the province of Pinar del Río, the MLP achieves a better spatial representation of precipitation, however, no improvement is obtained in terms of the quantitative forecast of precipitation. Furthermore, the MLP fails to eliminate false alarms.

3.2 Study case *July 10th 2019*

A similar behavior of **mae**, **mse** and p_{corr} is observed for this study case (Figures 7, 8 and 9 respectively). Again during the spin up time, the MLP correction shows its ability to reduce the WRF error.

According to the GPM estimate (Figure 10 bottom panel), at 2200 UTC, non-significant values between 6 mm / 3h and 12 mm / 3h occurred over the provinces of Pinar del Río, Artemisa, Havana, Camaguey and Holguín. SisPI, however, predicts more than 30 mm / 3h in the North region of Matanzas, which is considered a false alarm. In the rest of the country there is also an overestimation of precipitation values. In this forecast period, the application of the MLP failed to correct any of the errors indicated above and increased the spatial overestimation of precipitation. The above indicates that there is still much work to be done, and that other ANN models with more appropriate characteristics for this type of application should be explored, convolutional networks, for example.

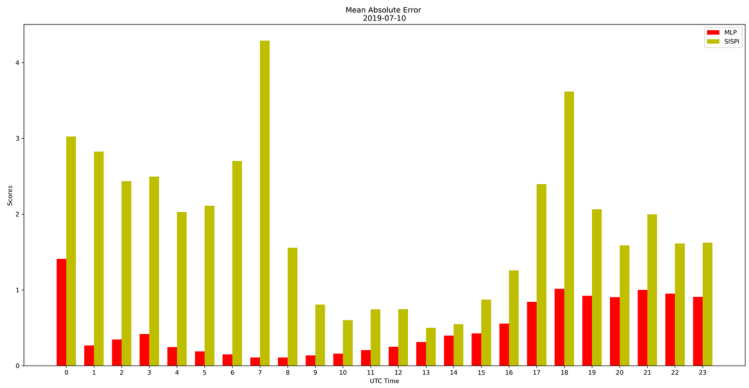

Figure 7. Mean absolute error of the SisPI forecast for July 10th 2019 and the SisPI forecast after bias correction with the MLP model.

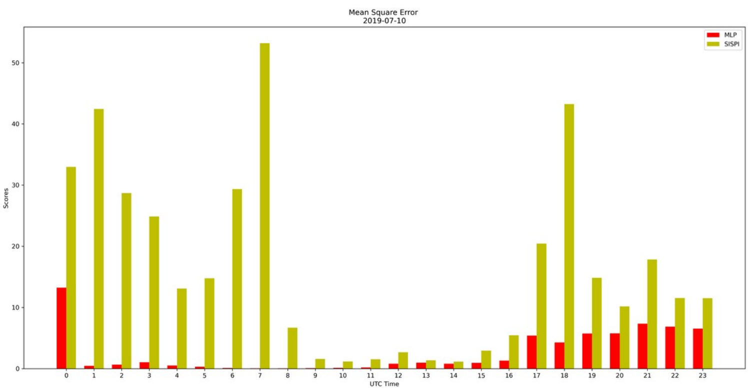

Figure 8. Mean square error of the SisPI forecast for July 10th 2019 and the SisPI forecast after bias correction with the MLP model.

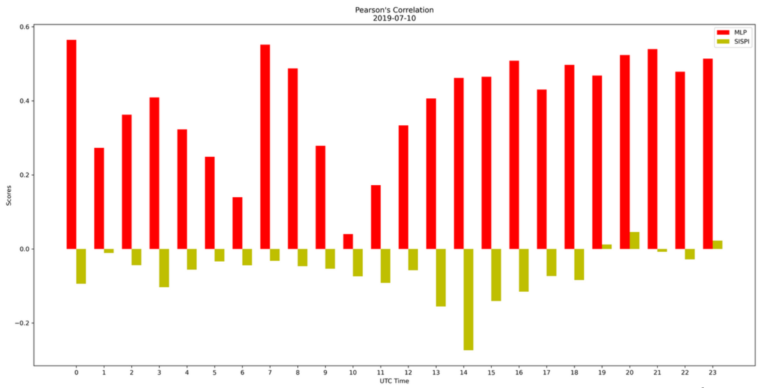

Figure 9. Pearson's correlation of the SisPI forecast for July 10th 2019 and the SisPI forecast after bias correction with the MLP model.

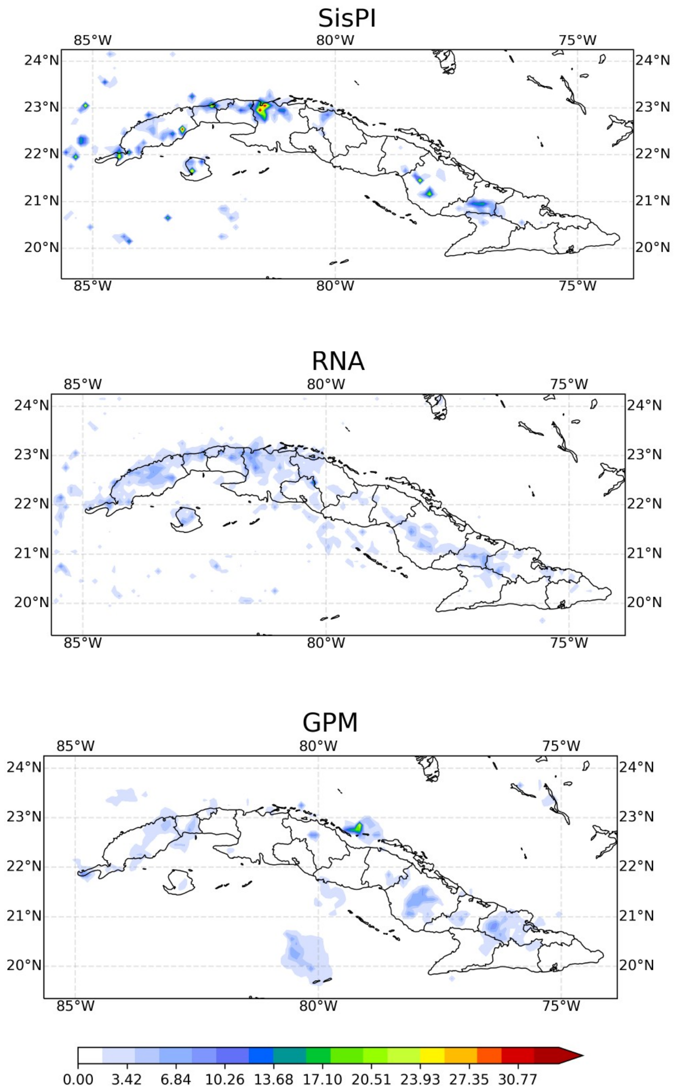

Figure 10. SisPI precipitation forecast (top panel), MLP correction (middle panel) and precipitation estimated by GPM product (bottom panel) for 10th July 2019 at 2200 UTC.

4. Conclusions

The research presented consist in an application of a multi-layer perceptron artificial neural network for bias correction of QPF. Some positive results were obtained in terms of the reduction of the metrics **mae** and **mse**, being better the correction in the spin-up period of the WRF. However, the MLP model was not able to overcome quantitative and positional errors when looking at a given forecast timeframe. It is recommended to extend the experiments to more appropriate ANN models for correcting the quantitative forecast of precipitation, taking into account its value and also the spatial location.

Author Contributions: Conceptualization, M.SL. and A.FB.; methodology, M.SL. and A.FB.; software, A.FB.; validation, A.FB. and M.SL.; formal analysis, A.FB. and M.SL.; investigation, A.FB.; resources, M.SL.; data curation, A.FB.; writing—original draft preparation, A.FB.; writing—review and editing, M.SL.; visualization, A.FB.; supervision, M.SL. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bishop, CM. Neural Networks for Pattern Recognition. **1995**, Oxford: University Press.
2. Haykin, S. Neural Networks: A Comprehensive Foundation. 2. **1998**.
3. Hsieh, WW and Tang, B. Applying neural network models to prediction and data analysis in meteorology and oceanography. **1998**. pp. 1855-1870.
4. Sanaz, M. Bias Correction of Global Circulation Model Outputs Using Artificial Neural Networks. Georgia Institute Technology: sn, **2015**.
5. Fuentes-Barrios, A., Sierra-Lorenzo, M. and Morfa-Ávalos, Y. Corrección del pronóstico cuantitativo de laprecipitación mediante el uso de redes neuronales Revista Cubana de Meteorología, Vol. 26, No. 3, Julio-Septiembre **2020**, ISSN: 2664-0880
6. Sierra-Lorenzo, M., Ferrer-Hernández, A. L., Hernández-Valdés, R., González-Mayor, Y., Cruz-Rodríguez, R. C., Borrajero-Montejo, I. and Rodríguez-Genó, C. F. Sistema automático de predicción a mesoescala de cuatro ciclos diarios. **2014**, Technical Report, Instituto de Meteorología de Cuba.
7. Sierra-Lorenzo, M., Borrajero-Montejo, I., Ferrer-Hernández, A. L., Hernández-Valdés, Morfa-Ávalos, Y., Morejón-Loyola, Y., Hinojosa-Fernández, M. Estudios de sensibilidad del sispi a cambios de la pbl, la cantidad de niveles verticales y, las parametrizaciones de microfisica y cúmulos, a muy alta resolución. **2017**, Technical Report, Instituto de Meteorología de Cuba.
8. SisPI, **2021**, http://modelos.insmet.cu/sispi/
9. Buhigas, J. Everything you need to know about TensorFlow, Google's Artificial Intelligence platform. Digital Bridges. [Online] 2 14, **2018**. https://puentesdigitales.com/author/javierbuhigas/.
10. Hou, A.Y., Kakar, R.K., Neeck, S., Azarbarzin, A.A., Kummerow, C.D., Kojima, M., Oki, R., Nakamura, K. and Iguchi, T. The Global Precipitation Measurement Mission. Bulletin of the American Meteorological Society, **2014,** 95, 701-722. https://doi.org/10.1175/BAMS-D-13-00164.1
11. Mesoescale & Microescale Meteorology Division, ARW Version 3 Modeling System User's Guide. Complementary to the ARW Tech Note, pp.141, NCAR: Boulder, Colorado. **2014**. Available on line: URL http://www2.mmm.ucar.edu/wrf/users/docs/userguideV3/ARWUsersGuideV3.pdf
12. WWRP/WGNE Joint Working Group on Verification. Forecast Verificacion-Issues, Methods and FAQ (pdf version). Available on line: URL http://www.cawcr.gov.au/staff/eee/verif/verif_web_page.html , **2008**. (accessed on may, 15, 2020)