

Comparison between 3DVAR and 3DEnVAR Methods for Fog Forecast Applied to Short-Range Forecast System [†]

Pedro M. González ¹, Maibys Sierra ¹, Adrián L. Ferrer ¹ and Carlos M. González ²

¹ Center for Atmospheric Physics, Institute of Meteorology; pedro.met90@gmail.com (P.M.G.); (M.S.); (A.L.F.)

² Meteorological provincial Center Artemisa-Habana-Mayabeque, Institute of Meteorology; (C.M.G.)

* Correspondence:

[†] Presented at 5th International Electronic Conference on Atmospheric Sciences, 16–31 July 2022; Available online: <https://ecas2022.sciforum.net/>.

Abstract: The research's objective is to compare the 3DVAR and 3DEnVAR's methods skill over fog forecasting applied to the Short-range Forecast System (SisPI). Prebufr and radiances data joint to observations from the KBYX and KBMA radars data are assimilated in a combined way for a first time in Cuba. On the other hand, five different hydrometeors species with the vertical velocity was included as control variables into a traditional covariance matrix CV7 used for these experiments. The evaluation is performed considering the phenomenon as a binary event. Dichotomous analysis uses the present weather code data with the visibility predicted by the model, which is obtained by an empirical algorithm. The results suggest that the hybrid scheme allows a more realistic representation of the environment where the phenomenon develop and leads to more accurate forecasts.

Keywords: SisPI; WRFDA; fog; assimilation

1. Introduction

In Cuba, the observational limitations and the relatively low frequency of these events have resulted in its shallow study. As a result of these studies, fog processes at country are currently forecast based on two fundamental methods. The first is related to fog synoptical patterns persistence, based on the presence of favorable synoptic configurations; while the second is associated with climatology, where an evaluation is made of the frequency and intensity according to the time of year. Neither considers the dynamic processes associated with the phenomenon and, in both cases, more weight is given to the volume of available data, which leads to the fact that the national meteorological service currently issues forecasts of this type with very low frequency.

Actually a numerical weather forecasting project based on the WRF (Weather Research & Forecasting) is currently in operation, whose fundamental objective is short- and very short-term forecasting [3,4], called SisPI (Immediate Forecast System) and which is currently working on incorporating data assimilation techniques into its operational scheme, the purpose is to achieve initial conditions that produce the best possible numerical forecast.

This research evaluates the impact of two data assimilation schemes applied to SisPI in a continuous episode of dense fog that occurred on 30 and 31 December 2019 over a western region of Cuba. For this, the skill with which it reproduces the synoptic environment where the fog develops is evaluated, but an improved algorithm is also applied that allows estimating the horizontal visibility, which is analyzed using the present weather data of the meteorological stations located on study area.

Published: 22 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

2. Materials and Methods

2.1. Study Area

The study area includes Havana, Artemisa and Mayabeque provinces, which have 10 conventional meteorological stations irregularly distributed over their geography (Figure 1b). The fog events occurrence in this zone is linked to the proximity of frontal systems in the Gulf of Mexico, where the subtropical anticyclone imposes a weak flow from the second quadrant (southeast region). They also tend to appear in situations of strong anticyclonic dominance, where the presence of weak pressure gradients is combined with strong nocturnal radiative cooling [1]. Its appearance is less frequent towards the coasts and greater towards the inner region, particularly towards the interior plains of Mayabeque and the south of the capital [2].

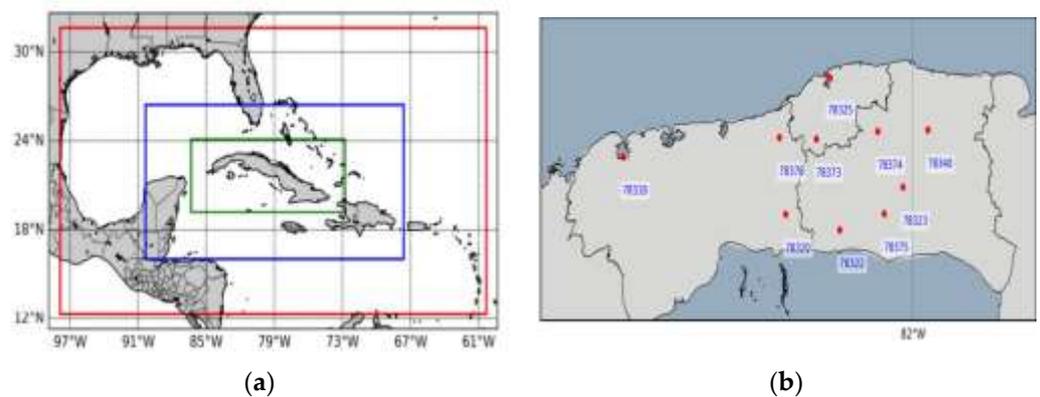


Figure 1. (a) SisPI’s domains. Red box represents the parent domain; blue box represents the 9 km nested domain and the green box the high resolution domain (3 km). (b) location of weather stations over the study area.

2.2. Short-Range Forecast System (SisPI)

To develop the experiments, the WRF model was used with the ARW dynamic core in its version 3.8.1 [3]; which constitutes the main element of SisPI’s project [4,5]. The main objective of this system is very short and short-term forecasting. SisPI’s desing has two way nested domains of 27 and 9 km (kilometers) respectively and a 3 km one way nested domain of grid resolution (Figure 1). The model was initialized from GFS (Global Forecast System) forecast data with 0.5° horizontal resolution, used at SisPI’s operative desing too.

The configuration proposed by SisPI includes 28 vertical levels, the Mellor-Yamada-Nakanidhi and Niino2.5 PBL scheme and the RRTM longwave parameterization for all domains. In the case of low-resolution domains, it contains the microphysics of WSM5, the Grell-Freitas cumulus parameterization, and the shortwave Dudhia scheme. For the high resolution domain, the Morrison double-moment microphysics is used, the cumulus parameterization is deactivated and the Goddard shortwave radiation scheme is used.

2.3. Assimilation Methods

2.3.1. 3-Dimensions Variational (3DVAR)

The 3DVAR method can be summarized as the iterative solution to find the state x that minimizes the cost function (Equation (1)). This solution represents the maximum probability (least variance) estimate of the true state of the atmosphere given two a priori data sources: the background field and the observations.

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(y - H(x))^T R^{-1}(y - H(x)) \quad (1)$$

where $J(x)$ is the cost function; x is the analysis (what you want to find); the background field; B the background error covariance; H is the observational operator included in the

WRFDA itself and finally R is the covariance matrix related to the error of the observations.

2.3.2. Hybrid Assimilation Method (3D $EnVAR$)

Hybrid assimilation schemes employ a combination of the background error applied in the variational methods with the flow-dependent BECs used in the Ensemble Kalman Filter (EnKF). In the 3D $EnVAR$ formulation (Equation (2)) the ensemble is valid only for the analysis time:

$$J(x; \alpha) = B_s \frac{1}{2} (x_1 - x_b)^T B^{-1} (x_1 - x_b) + B_e \frac{1}{2} \sum_{i=1}^n (a_i^T C^{-1} a_i) + \frac{1}{2} [y - H(x_1 + x_e)]^T R^{-1} [y - H(x_1 + x_e)] \tag{2}$$

In this case, B is the weight attributed to the ensemble, with C being the correlation matrix for the effective location of the ensemble’s perturbations.

2.4. Experiments Desing

2.4.1. Background Error Covariance and Outers Loops

For all the experiments, a background error covariance (BEC) generated specifically for the domain in question was used, built from the NMC method (National Meteorological Center; [6]) from the forecasts generated up to 15 previous days at the time of initialization. This BECs were generated with the inclusion of hydrometeors as additional control variables using the `gen_be` version 2 available in the WRFDA package version 4.3.1.

At 3D VAR ’s case, the application of multiple outers loops (OLs) was decided, which, according to authors such as [7], has an influence on the cost function minimization process, which can translates into best results with the method. Following this idea, the multiplicative weight of the static BEC scale and variance functions was varied with each OL, this is equivalent to a variational control [8] giving more weight to the improved background field obtained as a result. of each OL (Table 1).

Table 1. Modification of the multiplicative weight and control variables.

| Control Variables (CV7 Background Error Co- variance) | Modification on Variance Scaling (1ro-2do-3ro OL) | Modification on Length Scaling (1ro-2do-3ro OL) |
|---|--|--|
| U component | 1.5-1.0-0.5 | 1.0-0.5-0.2 |
| V component | 1.5-1.0-0.5 | 1.0-0.5-0.2 |
| Temperature | 1.5-1.0-0.5 | 1.0-0.5-0.2 |
| Pseudo relative humidity | 1.5-1.0-0.5 | 1.0-0.5-0.2 |
| Surface pressure | 1.5-1.0-0.5 | 1.0-0.5-0.2 |

The ensemble designed to obtain the flow-dependent perturbations for the application of the 3D $EnVAR$ method was built using previous SisPI’s runs, with the criterion that they included members from up to 12 h prior to the initialization time, in order to mitigate the possible effects of model spin-up in calculating flow-dependent perturbations. This allowed to obtain a small ensemble with 5 members. For this research, a weight of 75% was arbitrarily assigned to the ensemble contribution and 25% to the static covariance matrix, thus giving greater relevance to flow-dependent errors on the assimilation process [7,9].

2.4.2. Data Employed for Assimilation and Verification

For this research, prepbuf and radiances data were used. The data contained in the prepbuf files includes information of FM12 codes, METAR, ship data, buoys and soundings. Regarding the assimilation of radiances, data in buf format from microwave sensors were used: AMSU-A (NOAA-15/16/18/19), MHS (NOAA-18/19), SSMIS (DMSP-16) and ATMS (Suomi-NPP), in all cases extracted from the site <https://rda.ucar.edu/datasets>. This information was combined with radar data from the KBYX and KBMA stations, which in turn were obtained from the <https://www.ncdc.noaa.gov/nexradinv> site. This last is a novelty into national researches because is the first time that radar data are included inside an assimilation desing for fog's studies.

2.4.3. Visibility Algorithm

An algorithm is used to estimate the horizontal visibility based on the pre-existing Cvis (Equation (3)); [10]. Which is expressed as the minimum between the formulation of [10] based on different hydrometeor species (Equation (4)) and a relationship that considers the effects of dew point depression and relative humidity (Equation (5)).

$$Cvis = \min(SW_{99}; FSL) \quad (3)$$

$$SW_{99} = \frac{-\ln(0.02)}{\beta} \quad (4)$$

$$FSL = 1.609 * 6000 * \left(\frac{T - Td}{RH e^{1.75}} \right) \quad (5)$$

2.5. Forecast Verification

The results evaluation on study area was done on the highest resolution domain through a cell-point verification scheme. Due to the inherent subjectivity in fog measurements at conventional stations, it was decided to use the present weather code, since through it the presence and intensity of the phenomenon can be distinguished. This validation was designed considering the fog cases individually as binary events, for which it was necessary to generate a contingency table. The generation of these tables allowed the calculation of a series of statisticians (Equation (6)), through which it was possible to quantify the forecasting fog skill.

$$H = \frac{a}{a+c} \quad F = \frac{b}{b+d} \quad CSI = \frac{a}{a+b+c} \quad EDI = \frac{\log(F) - \log(H)}{\log(F) + \log(H)} \quad (6)$$

3. Results

3.1. Evaluation of Synoptic Flow Representation

On the days corresponding to 30 and 31 December 2019, the western region was under the weak influence of the subtropical anticyclone, with a cold front advancing through the eastern portion of Gulf of Mexico toward east (figure omitted). The fog manifestations on 30th day began between 06:00 and 09:00 UTC, reaching their greatest intensity around 12:00 UTC, a behavior that is characteristic at study area [12,13]. Towards the southern region, the phenomenon was particularly strong as it persisted until around 15:00 UTC. Already on the 31st, with the cold front proximity, the fog manifestations decreased in intensity and temporal extension compared to previous day. This fact seems to be conditioned by slight increases in wind speed recorded fundamentally towards coastal portions of the study area, with the consequent increase in turbulent mixing processes.

3.2. Synoptic Environment Evaluation Were Fogs Develop

Temperature forecasts (Figure 2a) show a tendency to forecast warmer environments, a fact that assimilation methods fail to improve significantly. The results obtained here coincide with [14], who carried out experiments for radiation fog cases with various

boundary layer configurations, and obtained that MYNN2.5 (Mellor Yamada Nakanishi and Ninno 2.5) tends to generate warmer environments. Regarding the dew point temperature (figure omitted), the results are very similar. It highlights the fact that the values of SisPI and 3DVAR present few differences, while 3DEnVAR presents higher values as a result of predicting more moisture environments.

In relation to relative humidity, the results suggest that the model tends to overestimate its values, however, it underestimates the saturation conditions. Its means that model presents problems to reproduce environments with relative humidity above 95% (Figure 2b). This result may be a response to model difficulties in correctly reproducing the isobaric cooling rate by long wave emission on the one hand, and on the other, to a tendency to overestimate warm advection over the study area. The advective component has also been identified by other authors due to its influence on the night cooling rate, as is the case of [15].

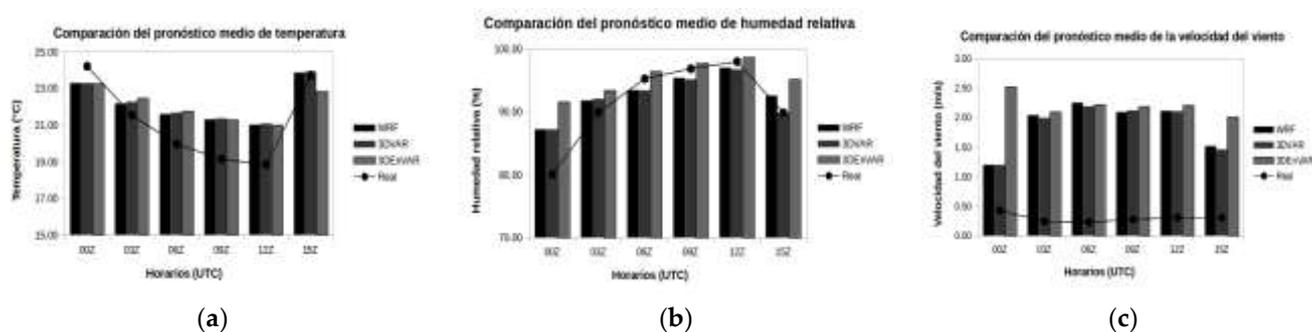


Figure 2. Behavior of SisPI and the assimilation methods designed during the first 15 forecasting hours (including the initialization time); (a) temperature (°C); (b) relative humidity (%); (c) surface wind speed (m/s).

The wind speed forecast shows the greatest deficiencies (Figure 2c). The mean error of this variable is around 1.63 m/s and the assimilation schemes fail to significantly improve this value. It is appreciated that 3DVAR’s solutions are usually close to SisPI’s, an appreciable characteristic in all studied variables. On the other hand, 3DEnVAR reflected the worst initial conditions, something that may be related to the fact that the background field provided by the ensemble obtained from previous SisPI outputs does not adequately reflect the fluxes related to the wind field above the study area. In the critical hours of genesis and development of the phenomenon, the errors of this scheme are comparable to the rest of the experiment.

3.3. Algorithms Performance

A fact that has negative weight is the low model skill to simulate the concentrations of hydrometeors. This is fundamentally reflected in the forecast of liquid water content equal to zero over large portions of the study area (Figure 3). This difficulty was also identified by [16]. These authors obtained predictions of weak and short-lived events due to this effect using the SW99 algorithm, a component of the Cvis algorithm used in this research.

The critical success index obtained by Cvis show that the algorithm fails to forecast fog events that begin to appear at the first forecasting hours with SisPI, it is not until 9 h after the initialization instant that fog episodes are correctly forecast with the index values around 50%. Data assimilation allows earlier detections as a result of improved meteorological fields from the observations included in the model. Despite the fact that 3DVAR produces slight improvements respect to SisPI, its solutions are very close, therefore the main contribution of this method lies in an earlier fog detection, since in the rest of the forecast times, its forecasts are comparable to SisPI.

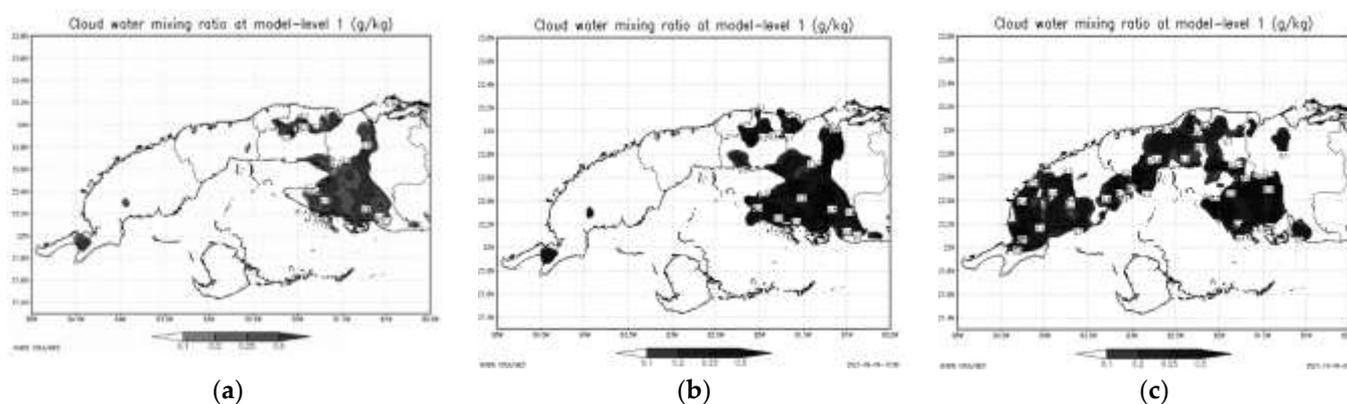


Figure 3. Comparison of liquid water content forecasting over the model level closest to the surface for 30 December at 12:00 UTC. (a) SisPI; (b) 3DVAR; (c) 3DENVAR.

The 3DENVAR method produces significant improvements compared with 3DVAR and SisPI, although its effectiveness tends to decrease around 12:00 UTC due to the forecast of early dissipations in coastal regions due mainly to wind speed overestimation. The dichotomous statisticians analyzed show that SisPI registered values around 62%, while 3DVAR and 3DENVAR favored forecasts that led to correct detections of 64% and 79% respectively. The false alarms volume was comparable between the assimilation schemes with 25% and 26% for 3DVAR and 3DENVAR respectively, in both cases lower than SisPI (35%). As can be expected, the CSI was much higher with 3DENVAR (62%), being closer between SisPI (50%) and 3DVAR (51%) and a similar behavior was shown by the EDI with 83%, 70% and 69% for 3DENVAR, 3DVAR and SISPI respectively.

3.4. Cost Function and Incremental Analysis

A complementary analysis is presented for to allow a better understanding of the cause of the results described previously. At the first place, the 3DENVAR scheme is more efficient minimizing the cost function than 3DVAR, this can be seen in a higher minimization rate achieved with a lower number of iterations (figure omitted). This means that the hybrid scheme manages to propagate the observations more successfully and that the background field obtained by the previous ensemble used as a first guess, contributes to providing a more realistic starting point than the mere initial condition of the model that uses 3DVAR. The application of multiple OLs to 3DVAR does not achieve, through the variational control, a minimization similar to 3DENVAR and its impact can vary from one case to another, so we see how in the case of day 30 this technique leads to a progressive minimization of $J(x)$, while on day 31st the third and fourth OL barely influence the cost function.

3DVAR produces very homogeneous temperature changes throughout the column and lead to a heating, therefore increasing the vapor retention capacity of the air mass, this being the lowest point of the method. As a positive balance, it achieves wind speed reductions greater than 3DENVAR, a fact that allows the method to lead the Cvis algorithm to make more accurate forecasts than those made with SisPI.

3DENVAR method contributed to greater cooling in the superficial layers allowed early detections in relation to SisPI, however this occurred mainly towards the southern region of the study area, in others portions it was not as efficient. On the other hand, the heating of the lower layers and those close to the surface is a characteristic that usually appears in situations of radiation fog and is related to the anticyclone subsidence, that contributes to the formation of a thermal inversion, a process that favors fog appearances. Another characteristic that could be observed with this method is the decrease in vertical wind speeds, with which it is possible to forecast less sheared environments than with SisPI, another factor that constitutes a forecasting improvement.

4. Conclusions

With Cvis algorithm applied to SisPI, successful fog forecasts are achieved over the study area, with CSI values around 50%. However, the detection does not occur until at least the first 9 h after the instant of initialization, presumably due to spin up process of model's mass and momentum fields.

The 3DVAR method manages to improve the CSI values by only 1% and the correct detections by 2%. These discrete values respond to a limited modification of the background field as a consequence of an inadequate dispersion of the impact of the observations on the domain. However, these improvements are mainly due to the obtention of correct forecasts at the first 6 forecasting hours.

The 3DENVAR method generates a more realistic analysis field compared to 3DVAR and achieves a more efficient dispersion of the observations over the domain. This leads to more realistic forecasts in the short term, that allows not only earlier detections with this desing, but also to sustain these advantages over time, this is observed in values of correct detection up to 17% and CSI of 11% higher than SisPI.

Author Contributions: Conceptualization, P.M.G. and M.S.; methodology, M.S.; software, Adrián Ferrer.; validation, P.M.G., M.S. and A.L.F.; formal analysis, C.M.G.; investigation, P.M.G., C.M.G.; resources, M.S.; data curation, A.L.F., C.M.G.; writing—original draft preparation, P.M.G.; writing—review and editing, M.S.; visualization, A.L.F.; supervision, M.S.; project administration, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement:

Data Availability Statement: Not applicable.

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

References

1. Alfonso, A.P.; Florido, A. *El clima de Matanzas*; Editorial Academia; La Habana, Cuba, 1993; p. 113.
2. Álvarez, E.L.; Borrajero, M.I.; Álvarez, M.R.; León, L.A. Estudio de la marcha interanual de la frecuencia de ocurrencia de los fenómenos nieblas y neblinas a partir del código de estado de tiempo presente. *Revista Ciencias de la Tierra y el Espacio* **2011**, *12*, 31–46.
3. Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Wang, W.; Powers, J.G. A Description of the Advanced Research WRF Version 3. NCAR Tech. Note NCAR/TN-4751STR. 2008; p. 125. Available online: http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf (accessed on).
4. Sierra, M.; Ferrer, M.; Hernández, R.; González, Y.; Borrajero, I.; y Rodríguez, R.F. *Sistema automático de predicción a mesoescala de cuatro ciclos diarios*; Informe del Resultado No.1 del Proyecto: Sistema de Predicción a muy corto plazo basado en el Acoplamiento de Modelos de Alta Resolución y Asimilación de Datos; Instituto de Meteorología de Cuba: La Habana, Cuba, 2014.
5. Sierra, M.; Borrajero, I.; Ferrer, A.L.; Morfa, Y.; Morejón, Y.; Hinojosa, M. Estudios de sensibilidad del SisPI a cambios de la PBL, la cantidad de niveles verticales y, las parametrizaciones de microfísica y cúmulos, a muy alta resolución. *Informe de resultado* **2017**.
6. Parrish, D.F.; Derber, J.C. The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Weather. Rev.* **1992**, *120*, 1747–1763.
7. Schwartz, C.S.; Liu, Z.; Huang, X.; Kuo, Y.; Fong, C. Comparing limited-area 3DVAR and hybrid variational-ensemble data assimilation methods for typhoon track forecasts: Sensitivity to outer loops and vortex relocation. *Mon. Weather. Rev.* **2013**, *141*, 4350–4372. <https://doi.org/10.1175/MWR-D-13-00028.1>.
8. Barker, D.M.; Huang, W.; Guo, Y.R.; Xiao, Q.N. A Three-Dimensional (3DVAR) Data Assimilation System For Use With MM5: Implementation and Initial Results. *Mon. Weather. Rev.* **2004**, *132*, 897–914.
9. Malakar, P.; Kesarkar, A.P.; Bate, J.; Deshamukhya, A. Apraisal of data assimilation techniques for dynamical downscaling of the structure and intensity of tropical cyclones. *Earth Space Sci.* **2020**. <https://doi.org/10.1029/2019EA000945>.
10. Bang, C.H.; Lee, J.W.; y Hong, S.Y. Predictability experiments of fog and visibility in local airports over Korea using the WRF model. *Korean Soc. Atmos.* **2009**, *24*, 92–101.
11. Stoelinga, M.T.; Warner, T.T. Nonhydrostatic, Mesobeta-scale model simulations of cloud ceiling and visibility for an east coast winter precipitation event. *J. Appl. Meteorol.* **1999**, *38*, 385–404.

12. Álvarez, E.L.; Borrajero, M.I.; Álvarez, M.R.; León, L.A. Estudio de la marcha interanual de la frecuencia de ocurrencia de los fenómenos nieblas y neblinas a partir del código de estado de tiempo presente. *Revista Ciencias de la Tierra y el Espacio* **2011**, *12*, 31–46.
13. Hernández, J.F.; González, C.M.; González, P.M. *Pronóstico de nieblas en las provincias de Artemisa, La Habana y Mayabeque. IX Congreso Cubano de Meteorología, La Habana, Cuba*; Available: Sociedad Meteorológica de Cuba. Consultado: Marzo 2020; 2017; ISBN 978-959-7167-60-0.
14. Román-Cascón, C.; Yagüe, C.; Sastre, M.; Maqueda, G. Observations and WRF simulations of fog events at the Spanish Northern Plateau. *Adv. Sci. Res.* **2012**, *8*, 11–18. <https://doi.org/10.5194/asr-8-11-2012>.
15. Bergot, T.; Guedalia, D. Numerical Forecasting of Radiation Fog. Part I: Numerical Model and Sensitivity Tests. *Mon. Weather. Rev.* **1993**, *122*, 12181230. <https://doi.org/10.1175/1520-0493>.
16. Ryerson William, R. Toward Improving Short-Range Fog Prediction in Data-Denied Areas Using the Air Force Agency Mesoscale Ensemble. 2012. Available online: <https://calhoun.nps.edu/handle/10945/17454> (accessed on).