

# DATA ASSIMILATION IMPACT ON FOG / HAZE FORECAST APPLIED TO THE SHORT-RANGE FORECAST SYSTEM

## IMPACTO DE LA ASIMILACIÓN DE DATOS SOBRE EL PRONÓSTICO DE NIEBLA/NEBLINA APLICADO AL SISTEMA DE PRONÓSTICO INMEDIATO

P. M. GONZÁLEZ JARDINES<sup>a†</sup>, M. SIERRA LORENZO<sup>a</sup> AND C. M. GONZÁLEZ RAMÍREZ<sup>b</sup>

a) Centro de Física de la Atmósfera, Instituto de Meteorología, La Habana, Cuba; pedro.met90@gmail.cu<sup>†</sup>, maibys.lorenzo@insmet.cu

b) Centro Meteorológico Provincial Habana-Artemisa-Mayabeque, Instituto de Meteorología, La Habana, Cuba; carlosm.ramirez@insmet.cu

<sup>†</sup> corresponding author

Recibido 25/1/2022; Aceptado 25/8/2022

During the early morning of January 4<sup>th</sup>, 2019, a dense fog episode occurred over the western region of Cuba, which affected areas of socioeconomic importance. This research studies three data assimilation methods; 3DVAR and two hybrid variants, 3DnVAR and ETKF-3DVAR, applied to the 00:00 UTC initialization of the WRF model. The way of obtaining the first guess and the flow-dependent perturbations used in the hybrid schemes is modified. Data in prepbufr and radianc formats are assimilated, analyzing the impact of each type of data on assimilation and forecasting. The results obtained indicate that, although all the assimilation techniques studied lead to improved forecasting in the short and very short term, the combination of the 3DVAR method and the data content in prepbufr format is the one that offers the best results.

Durante la madrugada del 4 de enero de 2019 se presentó un episodio de niebla densa sobre la región occidental de Cuba, que afectó zonas de importancia socioeconómica. En esta investigación se estudian tres métodos de asimilación de datos; 3DVAR y dos variantes híbridas, 3DnVAR y ETKF-3DVAR, aplicados a la inicialización a las 00:00 UTC del modelo WRF. Se modifica la forma de obtener la primera aproximación y las perturbaciones dependientes del flujo utilizadas en los esquemas híbridos. Se asimilan datos en formato prepbufr y radianzas analizando el impacto de cada tipo de datos en la asimilación y pronóstico. Los resultados obtenidos indican que, aunque todas las técnicas de asimilación estudiadas conducen a mejorar el pronóstico a corto y muy corto plazo, la combinación entre el método 3DVAR y el contenido de datos en formato prepbufr es la que mejores resultados ofrece.

PACS: Atmospheric optics (óptica atmosférica), 42.68.Ge; Weather analysis and prediction (análisis y predicción del tiempo), 92.60.Wc; Meteorology (meteorología), 92.60.-e.

### I. INTRODUCTION

A large number of investigations have been oriented to the forecast of fog events. A group of researchers has based their work on the implementation of empirical algorithms that allow estimating horizontal visibility [1–6]. The algorithms proposed by these authors based the estimation of the horizontal visibility in the concentration of the different species of hydrometeors and the relative humidity. These algorithms are obtained in most cases as a result of regression analysis, for which they contain empirical coefficients based on measurements made in the United States fundamentally. This can have as a consequence that they do not adequately represent the behavior of the variables involved in their calculation in a certain region.

Other researches have advanced in the sensitivity study with different parameterizations in order to predict favorable conditions for the appearance of fog events, where the use of the WRF (Weather Research and Forecasting) mesoscale model has prevailed [7]. Such works study the impact of the parameterizations of microphysics, planetary boundary layer and radiation in the genesis, development and dissipation of fog events, [8–12]. A detail that appears in almost all of these studies is the influence of the model's spin-up process, that prevents satisfactory results, at least during the first 12

forecasting hours. This creates difficulties for to obtain an efficient short-term fog forecast.

Nowadays in Cuba, a numerical weather forecasting project based on the WRF is in operation, with the fundamental objective of short and very short-term forecasts [12], called Short-range Forecast System (SisPI, acronyms in Spanish). In order to mitigate the effect of spin up the variant proposed in SisPI is the incorporation of data assimilation techniques to the model [12]. There are diverse assimilation methods implemented in software packages like GSI (Grid-point Statistical Interpolation) and WRFDA (WRF-Data Assimilation) that have been widely used, and several ones have been tested for forecasting fog events with satisfactory results.

For instance, [13] applied the 3DVAR (Three-Dimensional Variational) method with satellite data to predict a fog event over the Yellow Sea, finding that better humidity and temperature profiles are obtained on the marine boundary layer. Other authors such as [14] compared the results obtained by the 3DVAR and EnKF (Ensemble Kalman Filter) methods, in two study cases on the Yellow Sea, concluding that the EnKF offers better results than those obtained with 3DVAR. Similar results were achieved in [15], with the application of the EnKF to a one-dimensional model oriented to the fog forecast, obtaining a significant improvement in the

initial conditions of the model, thanks to more realistic profiles of temperature and specific humidity.

These studies introducing a series of questions focused on the application of data assimilation to fog prediction. Those points are around which data is most appropriate to assimilate or which method is the most optimal, taking into account not only the results, but also the computational cost for its application for operational purposes.

This research evaluates the impact of several assimilation methods in the forecast of a fog event that occurred on January 4<sup>th</sup>, 2019 in a sector of the western region of Cuba. It also analyzed the impact of use of prepbufr and radiance data, individually and combined, on the assimilation to obtain the most optimal configuration (method and data). The evaluation is carried out in the initialization corresponding to 00:00 UTC (United Time Coordinated) in order to assess the assimilation impact on the short and very short-term numerical forecast.

## II. MATERIALS AND METHODS

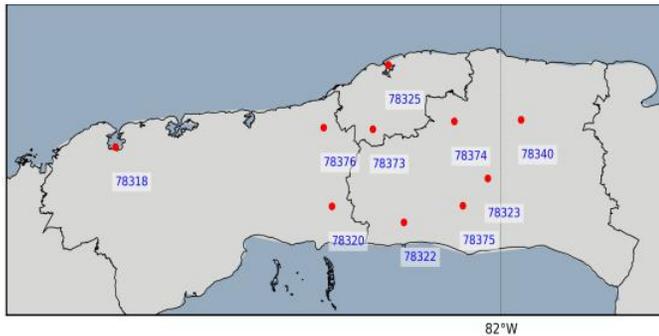


Figure 1. Study area. Geographical location of the meteorological stations.

The study area is framed in the provinces of Artemisa, Havana and Mayabeque which are located in the western region of the country and have ten meteorological stations (Fig. 1). In this region, radiation fog/haze events predominate, and they are related to the proximity of frontal systems in the Gulf of Mexico, where the subtropical anticyclone imposes a weak flow from the second quadrant. They also tend to appear in situations of marked anticyclone influence, where the presence of weak pressure gradients is combined with strong nighttime radiation [16]. They are more common and tend to be more intense in the dry period of the year (November-April). The time of 12:00 UTC is the one with the highest frequency of reports of both phenomena [17].

### II.1. Model Setup

To develop the experiments, the WRF model was used with the dynamic core ARW (Advanced Research WRF) in its 3.8.1 version; which constitutes the fundamental core of SisPI [12]. It has two-way nested domains of 27 (d01: 140 x 78) and 9 km (kilometers) (d02: 199 x 112) of spatial resolution and a domain of 3 km (d03: 421 x 184) of one-way nested solver using the ndown tool (Fig. 2). This design is the same as the operational

configuration of SisPI that currently runs in INSMET (Institute of Meteorology, acronyms in Spanish).

The temporal resolution of 27 and 9 km domains is every three hours, while the 3 km domain provides forecasts every hour. The model was initialized from forecast data of the GFS (Global Forecast System) with 0.5° horizontal resolution.

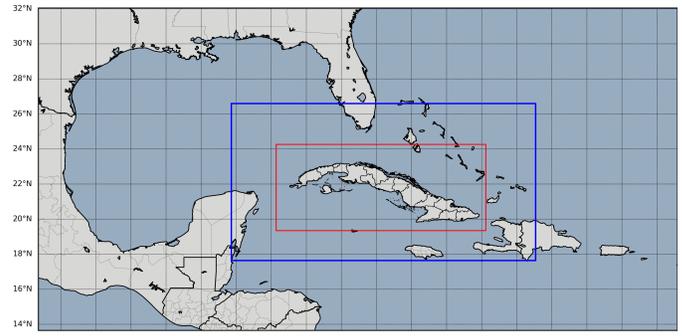


Figure 2. Domains used in SisPI; (external) 27 km, (blue square) 9 km Study area. Geographical location of the meteorological stations.

The physical configuration of the WRF used in SisPI is presented in Table 1:

Table 1. Some physical parametrization schemes used by SisPI for the three domains [12]

Parametrizations	Domain 27 km	Domain 9 km	Domain 3 km
Microphysics	WDM5 (WRF-single moment 5)		Morrison 2-moment
Cumulus	GF (Grekk-freitas)		not used
Boundary layer	MYNN 2.5 (Mellor-Yamada Nakanishi and Niino 2.5)		
Short wave radiation	Dudhia	Goddard	
Long wave radiation	RRTM (Rapid Radiative Transfer Model)		
Surface border	Monin-Obukhov		
Surface	Unified Noah land-surface model		
Vertical levels	28		

### II.2. Design of assimilation experiments

Assimilation was applied only at the highest resolution domain. Two fundamental factors lead to this choice: the SisPI execution philosophy that implies that, the highest resolution domain have initial and boundary conditions adjusted to it and includes the whole island and adjacent seas (Fig. 2), at second place data assimilation only can be applied to one domain individually. The main disadvantage lies in the fact that, being a relatively small spatial coverage nest, the volume of the available data to assimilate can highly variable, especially with the ingestion of radiances, because its come from polar-orbiting satellites.

This differs from other studies that perform assimilation in the parent domain as is the case of the study examples on the Yellow Sea [13, 14]. However, in other researches of mesoscale

phenomena, such as heavy rain, where the WRF execution philosophy is similar to SisPI, the assimilation is also carried out in the domain with the highest resolution [18, 19].

The methods 3DVAR [20] and two hybrid variants are used. The 3DVAR method can be summarized as the iterative solution to find the state  $x$  that minimizes the cost function (1). This solution represents the a posteriori maximum likelihood (minimum variance) estimate of the true state of the atmosphere given the two sources of a priori data: 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 field,  $x_b$  the background field,  $B$  the covariance matrix,  $H$  is the observational operator included into WRFDA and  $R$  is the observational covariance error.

The hybrid method 3DEnVAR (Three Dimensional Ensemble-Variational [21–23], requires a previous ensemble whose average is used as a first approximation. This ensemble is also used to represent the flow-dependent background errors that are combined with the static covariance error (2).

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

where  $\alpha_i$  is the ensemble weight and  $C$  the correlation matrix for the localization of ensemble perturbations.

The other hybrid method used responds to the combination of the ETKF (Ensemble Transform Kalman Filter) and 3DVAR [24], known as ETKF-3DVAR, which will simply be referred as ETKF throughout the paper. This method provides a framework for assimilating observations and also for estimating the effect of observations on forecast error covariance. It differs from other ensemble Kalman filters in that it uses ensemble transformation and a normalization to rapidly obtain the prediction error covariance matrix associated with a particular spread of observation sources.

This last scheme shows the disadvantage that, being based in the EnKF, the localization is carried out in the space of observation, unlike purely variational or 3DEnVAR schemes that use the model space, allowing non-local observations such as radiances to be assimilated. This limits the assimilable data for the application of ETKF to the prepbufr format in this research.

For the methods 3DEnVAR and ETKF the construction of the first guess was changed. In a first variant, a 10-member multiphysics ensemble is used, modifying the microphysics and boundary layer parameterizations that have shown the best performance in the study area, according to sensitivity studies developed by [12, 25]. For this case, the ensemble members building is more computationally

expensive, because required 3 km outputs for to be used like first guess.

The second variant uses a 6-member ensemble using the previous SisPI outputs that contain the initialization time, with a configuration described in the (Table 1). The number of members in the second case responds to the fact that currently the SisPI is initialized every 6 hours (four cycles per day), with a 36 hours forecast extension, which allows up to 6 previous initializations. This desing is less expensive than previous, because it use existing model outputs and is not necessary generate them like desing above.

Through the different assimilation techniques proposed, the data assimilation employed prepbufr and radiance formats, individually and jointly, in order to analyze the impact of the type of data that is assimilated on the fog/haze forecast (Fig. 3). The data contained in the prepbufr files include information in FM12, METAR codes, ship data, buoys and sounding. Regarding the assimilation of radiances, data in the microwave channel bufr format 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>.

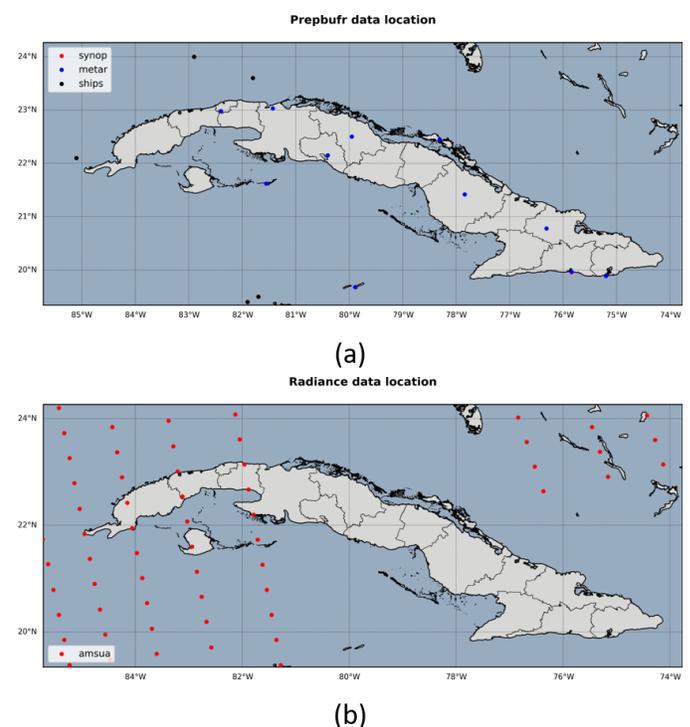


Figure 3. Data types assimilated with proposed methods; (a) prepbufr format, (b) radiances in bufr format.

For the design of the experiments, the recommendations proposed by [26] were followed and which are contained in the presentation about good practices with the WRFDA, available on the site <https://www.mmm.ucar.edu/wrf/users/wrfda>. They include the use of three iterative outliers loops with the 3DVAR method and only one with the hybrid schemes because the impact of the cycles is stronger in 3DVAR [26]. In the hybrid methods cases, was established empirically a

weight of 50% between the generic covariance matrix and the flow-dependent background error obtained from the members of the ensemble.

Due to the fact that the process of generating domain-dependent matrices to obtain the background error is usually computationally expensive, it was decided to use the generic matrix provided by the WRFDA module in these experiments. This matrix can be used as background error for any regional application according to the WRF user manual [12].

Table 2, below, contains the selected nomenclature for the experiments, the types of data that are assimilated and the parameters that vary in each case.

Table 2. Data assimilation experiments.

Assimilation method	Assimilated data	Description
3DVAR.P	prepbufr	3 outer loops
3DVAR.R	radiance	
3DVAR.PR	prepbufr + radiance	
3DEnVAR.P	prepbufr	10 members multiphysics, 1 outer loops
3DEnVAR.R	radiance	
3DEnVAR.PR	prepbufr + radiance	
3DEn2VAR.P	prepbufr	6 members configuration Table 1, 1 outer loops
3DEn2VAR.R	radiance	
3DEn2VAR.PR	prepbufr + radiance	
ETKF	prepbufr	10 members multiphysics, 1 outer loops
ETKF2	prepbufr	6 members configuration Table 1, 1 outer loops

### II.3. Algorithms used to calculate horizontal visibility

With the purpose of generating fog/hazes events forecasts in the study area, an algorithm is used to estimate the horizontal visibility. This algorithm name's is  $C_{vis}$  [5,9] (3), which choose the minimum value among the algorithms proposed by [2]; (3) and [3]; (4).

$$SW_{99} = \frac{-\ln(0.02)}{\beta} \quad (3)$$

where

$$\beta = \beta_{cw} + \beta_{rw} + \beta_{ci} + \beta_{sn} \quad (3.1)$$

$$\beta_{cw} = 144.7(C_{cw})^{0.88} \quad (3.2)$$

$$\beta_{rw} = 1.1(C_{rw})^{0.75} \quad (3.3)$$

$$\beta_{ci} = 163.9(C_{ci})^{1.0} \quad (3.4)$$

$$\beta_{sn} = 10.4(C_{sn})^{0.78} \quad (3.5)$$

$\beta$  is the light extinction coefficient (km<sup>-1</sup>) and  $C$  the concentration of four different species of hydrometeors (water liquid content-cw, rain water-rw, ice-ci and snow-sn); obtained from the corresponding mixing ratios calculated by the model [9].

$$FSL = 1.609 \cdot 6000 \cdot \frac{T - T_d}{rh^{1.75}} \quad (4)$$

This algorithm proposes that visibility can be obtained from a relationship between dew point depression (numerator) and relative humidity (rh). The coefficient 1.609 is used to convert from miles to kilometers [5,9].

Therefore

$$C_{vis} = \min(SW_{99}; FSL) \quad (5)$$

### II.4. Forecast verification

To verify the results of the forecast, a cell-point verification strategy was used in the study area, always on the domain with the highest resolution. The analysis is focused in the variables wind speed, dew point depression and relative humidity, and their real data are obtained from the measurements recorded in the meteorological stations from study area (figure 1). This allowed obtaining the values of the mean absolute errors (6) and root mean square error (7) in each of the experiments (MAE and RMSE respectively).

$$MAE = \frac{1}{n} \sum_{i=1}^n (O - P) \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (O - P)^2} \quad (7)$$

Where  $O$  represents the observation,  $P$  the forecast and  $N$  represents the total data to be compared.

Since assimilation allows correcting the initial background condition (first guess) by using observations, a good technique to measure the impact of assimilation is incremental analysis (Fig. 4). This is obtained by calculating the differences between the first guess and the analysis field a posteriori. This study performs an incremental vertical analysis to evaluate the impact of the different assimilation designs on the background field and how assimilation improves (or not) the model initial condition.

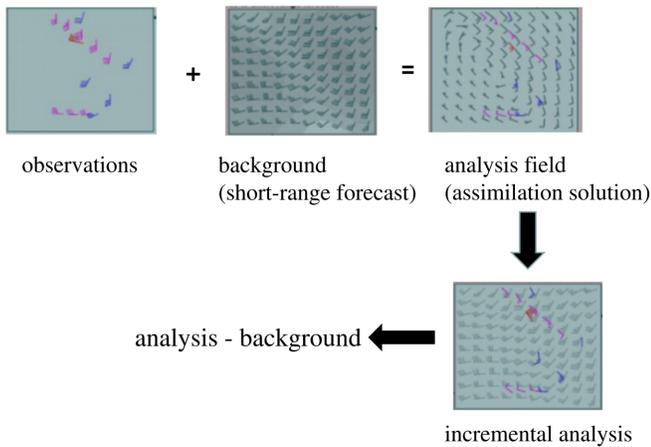


Figure 4. Representative diagram of incremental analysis (modified from COMET www.meted.ucar.edu).

Taking into account that horizontal visibility measurements have a high subjective component since they depend on the observer's skill, a binary analysis of the phenomenon is used as an evaluation methodology, where its occurrence/non-occurrence is verified comparing the model's solutions with the present weather code value registered at weather stations (FM-12 code for weathers stations). For the calculation of the dichotomous statisticians a contingency table is used (Table 3).

The contingency analysis was carried out individually for fog and hazes reports which allow obtaining an approximate idea of the degree of overestimation or underestimation of the intensity of the phenomenon. Due to the need of homogenizing the forecasts for their comparison with the present weather code, the definitions of fog and haze dictated by the World Meteorological Organization [27] are considered.

Taking into account those facts, in the case of fog events, a hit is considered when the predictor algorithm (Cvis) forecasted values less than or equal to 1 km and the present weather code registered values between 40 and 49 (fog); failures when the algorithm forecasted values greater than 1 km and in the present weather code values related to fog were recorded and false alarms when the forecast revealed values equal to or less than 1 km, but in the present weather code no values relative to fog appeared. Regarding the haze, the philosophy described above for them was followed, considering the predicted values, greater than 1 km and less than or equal to 5 km according to present weather records of 10.

Table 3. Contingency table

Forecasted	Observed	
	Yes	No
Yes	a (hit)	b (false alarms)
No	c (miss)	d (correct rejection)

Starting from the results of the contingency analysis, three statistics were calculated that allow the evaluation of the

forecasts (8,9,10).

$$H = \frac{a}{a + c} \quad (8)$$

$$F = \frac{b}{b + d} \quad (9)$$

$$CSI = \frac{a}{a + b + c} \quad (10)$$

### III. RESULTS

#### III.1. Synoptic situation. Characteristics of the events and its representation with the WRF model

Between the night of the 3<sup>rd</sup> and the early morning of January 4<sup>th</sup>, 2019, a weak cold front was over the eastern portion of the Gulf of Mexico. This synoptic pattern generates weak from the second quadrant winds over the study area and a strong radiative cooling favored by the stability conditions imposed by the subtropical anticyclone periphery's influence. This causes the heat accumulated on the surface during the day rises towards the lower layers, limiting its rise due to the subsidence generated by the anticyclone, which leads to a sufficient surface cooling to allow moisture to rise and condense causing the appearance of a fog/haze process by irradiation (Fig. 5a).

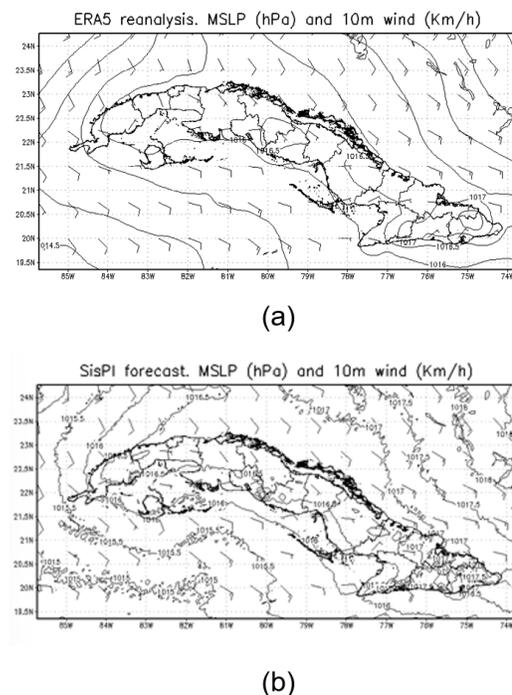


Figure 5. (a) Surface wind and pressure fields obtained from the ERA5 reanalysis (adapted to the highest spatial resolution domain latitude-longitude dimensions) (b) SisPI's forecast. January 4<sup>th</sup>, 12:00 UTC.

In situations like this, the appearance of an inversion of the temperature is common in the lower layers where the heat

transferred by the surface is concentrated. The weakness of the winds at the low levels, generated by this synoptic configuration, diminishes the mixing processes by turbulence allowing the super saturation below the base of the inversion.

SisPI adequately forecast the synoptic flow representing the position and intensity of the synoptic scale systems that day. However, a more westward expansion of the anticyclonic ridge can be seen, coinciding with a slight strengthening of surface wind speeds, mainly toward the west of the ridge, presumably as a result of a slight strengthening of baric gradients in that ridge side at the model solutions (Fig. 5b).

Consequently to these characteristics, the phenomenon begun at the night of the 4<sup>th</sup> by 00:00 UTC (Fig. 6) with hazes reports in the 78320 station, located at southeast of the study area (Fig. 1). The haze event continued strengthening and at 03:00 UTC, 4 stations already reported this phenomenon. By 06:00 UTC, a total of 6 weather stations were reporting fog, while 78340, 78373 and 78376 ones were reporting fog starting in the preceding hour, with estimated visibilities around 500 meters.

As can be appreciated, the 09:00 UTC marked the moment of greatest intensity of the phenomenon according to the present weather reports. A total of 6 stations reported visibilities less than or equal to 300 meters, being the inner region of the study area the most affected. At 12:00 UTC a gradual weakening of the phenomenon began to be experienced, 6 stations reported haze, while the others, all in the inward portion, continued reporting fog with visibilities estimated between 200 and 700 meters. By the 15:00 UTC observation, the phenomenon was completely dissipated.

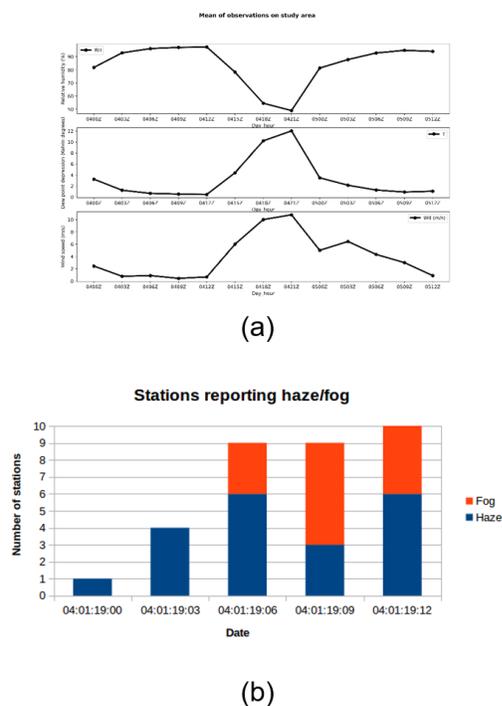


Figure 6. Number of stations in the study area reporting fog or haze (b) and average behavior of relative humidity, wind speed and dew point depression (a) in the period of time between 00:00 and 12:00 UTC on January 4<sup>th</sup>, 2019 in the study area.

Observations indicate that the phenomenon was weaker

towards the north coast where was reported haze only and stronger towards the inner region. This behavior has been described by several Cuban author's who state that these phenomena tend to be weaker towards the coasts, apparently due to the modulating effect of the sea that reduces the radiative cooling rate [16,17,27].

### III.2. Forecast evaluation without data assimilation

The radiative cooling is one of the main physical mechanisms associated with radiation fogs [10,16], as it diminishes the capacity of a given air mass to retain water vapor. This is essential because in tropical regions, even a relatively small cooling could lead to the genesis of fog/haze processes. In the no-assimilation experiment the model predicts an average radiative cooling rate slower than the observations (0.88°K/h [Kelvin degrees per hour] vs 1.66°K/h), consequently the mean absolute errors of temperature over the study area were between -2.5 and -3.5°K.

Warmer environments imply a higher vapor retaining capacity from the surrounding air mass; therefore this solution may correlate with an underestimation of relative humidity. The results obtained in the evaluation of the predictions of this variable corroborate the previous sentence.

In the simulation there was an underestimation of its values with mean errors close to 5 % (Fig. 7). These results coincide with several authors who obtain that the WRF tends to underestimate the relative humidity values, placing them sometimes below the threshold required for fog formation [10,28].

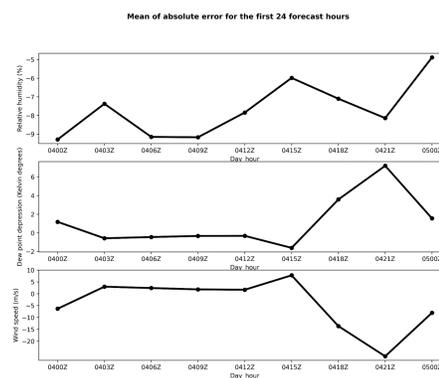


Figure 7. Mean absolute error over the study area for the relative humidity, dew point depression and wind speed from the moment of initialization until 12:00 UTC on January 4<sup>th</sup>, 2019.

However, the radiative cooling and increased humidity may not be sufficient elements to generate and sustain a radiation fog/haze event. For this, it is necessary that low wind speeds reduced the mixing processes due to turbulence allowing in this way, the development of the event [10]. The wind speeds increasing not only contributes to the increase of the turbulent mixing, even to the erosion of the thermal inversion too, that tends to be weak in tropical regions if compared to mid-latitudes making it's, consequently is more susceptible to changes in wind speed.

In the experiment the wind speed overestimation was above  $-2.5$  m/s as average, exceeding  $-3$  m/s between 06:00 and 09:00 UTC when the winds were in calm at almost all weather stations on the study area. (Fig. 7).

In the opposite direction, the others schemes contributes to a somewhat stronger wind regime, occasionally higher than the SisPI solutions, which indicates that the application of the assimilation methods in these cases, as they have been designed, they are not effective.

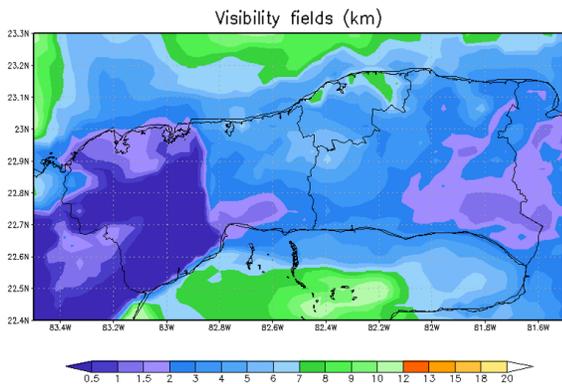


Figure 8. Forecast obtained by the Cvis algorithm for the study area corresponding to 09:00 UTC on January 4th, 2019

This scenario is consistent with the Cvis's results (Fig. 8). The without assimilation model's solution predicts haze events from moderate to weak over the study area, underestimating the fog's intensity that occurred in the early morning on January 4<sup>th</sup>, 2019, as it was not able to forecast fog events that occurred in the interior and southern regions even at times of greatest intensity.

### III.3. Assimilation impact

The assimilation schemes fail to substantially modify the errors in the representation of the synoptic flow (Fig. 9). They continue to represent the ridge extended to westward, which is evidenced at the wind direction from second quadrant. However in some cases like 3DVAR\_P (Fig. 9a), 3DVAR\_PR (omitted figure), 3DEn2VAR\_R (Fig. 9b) and ETKF (omitted figure) it is possible to appreciate a relative weakness in the flow as a result of weaker baric gradients, that leads to more realistic solutions and more a successful forecast.

### III.4. Incremental analysis

At the incremental analysis it is observed that the 3DVAR\_P experiment showed an increase in humidity values at troposphere low levels with respect to without assimilation solution and with a tendency to rapidly decrease generating a drier environment starting at surface of 925 hPa, confining the moisture layer below this level although the increments does not exceed 1%. That means that this experiment build a initial condition slightly more realistic forecasting an air mass with higher water vapor content. The other combinations that include radiance data generated analysis fields with drier environments at all levels, compared to obtained without data assimilation, in a threshold between  $-0.1$  and  $-1.5$  % (Fig. 10a). This solutions are less closer to observations.

The hybrid methods that use the multiphysics ensemble generated an more moist initial condition with increases between 3 and 12 %, presenting a main maximum contribution between 750 and 850 hPa as well as a second maximum in the levels closer to the surface. Those results suggest an increase in the vapor content in the superficial layers and the region corresponding to the boundary layer, elements that are needed for the formation of fogs (Fig. 10b).

On the other hand, the methods that start from the ensemble generated from SisPI previous runs showed an increase in humidity with a maximum around 10 % at the 950 hPa level, showing also a fast decrease in the layer between 900 and 750 hPa, generating an increasing trend above the surface of 700 hPa.

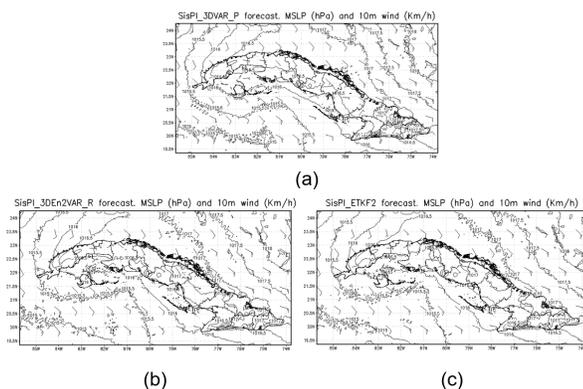


Figure 9. Surface wind and pressure fields forecasted (a) 3DVAR.P (b) 3DEn2VAR.R and (c) ETKF2. January 4<sup>th</sup>, 12:00 UTC.

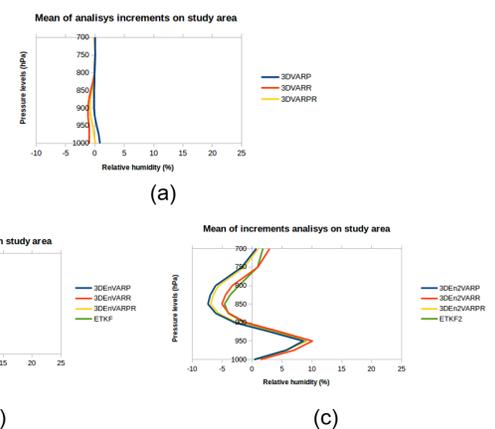


Figure 10. Incremental analysis of relative humidity for the different assimilation schemes applied. (a) 3DVAR; (b) hybrids built using the multiphysics ensemble; (c) hybrids built using SisPI's ensemble of solutions.

The results coincide, in the case of 3DVAR, with the initialization of warmer environments according to the model.

In the experiments developed with 3DVAR the radiances assimilation, individual and combined, led to warmer environments forecasts with increments thresholds between 0.2 and 0.4°K (Fig. 11a). On the contradictory, with an individual assimilation of prebufr data, a slight cooling is observed in the layer below 975 hPa with an equally very limited increase above this level. For the 3DVAR\_P case, the assimilation contribution did not exceed the threshold of  $\pm 0.1^\circ\text{K}$ .

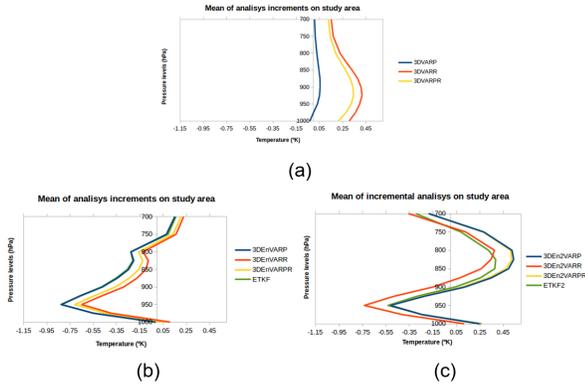


Figure 11. Incremental analysis of temperature for the different assimilation schemes applied. (a) 3DVAR; (b) hybrids built using the multiphysics ensemble; (c) hybrids built using SisPI's suite of solutions.

The hybrid techniques built from the multiphysics ensemble generated a maximum cooling around the surface of 950 hPa, with values around -0.6 and -0.8°K. A poor differentiation can be seen between the analysis field obtained from the different data combinations to be assimilated and even in relation to the method used, since the solutions obtained from 3DVAR and ETKF were very similar. This fact, also observed in the case of relative humidity, suggests a strong influence of the first approximation on the assimilation process. (Fig. 11b).

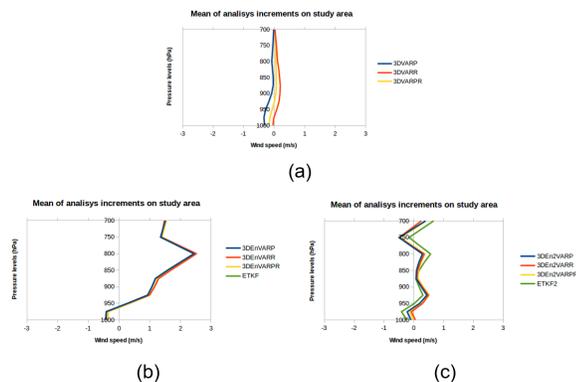


Figure 12. Incremental analysis of wind speed for the different assimilation schemes applied. (a) 3DVAR; (b) hybrids built using the multiphysics ensemble; (c) hybrids built using SisPI's suite of solutions.

The contributions, in relation to wind speed followed the trend appreciated in the previous variables. The 3DVAR\_P combination resulted in a reduction of the wind force in the 1000 to 700 hPa layer with mean values between -0.3 and -0.04 m/s. The 3DVAR\_R and 3DVAR\_PR showed an increase in the surface environment of 950 hPa compared to the run without assimilation, in this sense the 3DVAR\_R variant predicted the

highest increases of wind speed with the height and therefore an initial environment slightly more sheared (Fig. 12a).

The 3DVAR and ETKF designs generated initial environments with lower wind speeds regarding without assimilation run. However, forecasting increases of the speed above the surface of 950 hPa, which reached a maximum of 2.5 m/s, that means a slightly sheared environment. Those results were different in relation to the 3DVAR2 and ETKF2 schemes which led to discrete contributions comparable to 3DVAR, with wind speed decreases below the surface of 950 hPa.

### III.5. Impact on surface forecast

The contributions generated by the different assimilation schemes generated more or less favorable environments for the appearance of the phenomenon object of study, which caused different impacts on the surface forecasts. The proposed assimilation techniques predicted environments less warm than the solution without assimilation, however the observed differences do not exceed 0.5°K (Fig. 13a).

In relation to the dew point depression, the purely variational scheme tends to reduce the mean square errors with the forecasting time advance, slightly improving the solution without assimilation. However, the hybrid systems show a limited impact on the forecast, since after the first three hours the errors are even higher than the run without assimilation.

In relation to relative humidity, the 3DVAR method predicted a more realistic initial environment and, although the differences with respect to solution without assimilation were low, it showed an increase rate of the relative humidity in surface higher than the one obtained with the hybrid methods. The best results were obtained with the 3DVAR\_P experiment. Precisely, the solution of the hybrid schemes in a general sense showed greater precision in the initialization, but can be observed a decreasing performance after the three first forecasting hours with relative humidity underestimations up to close to 8% around 09:00 UTC, time of higher intensity of the episode, as it was already expressed (Fig. 13b).

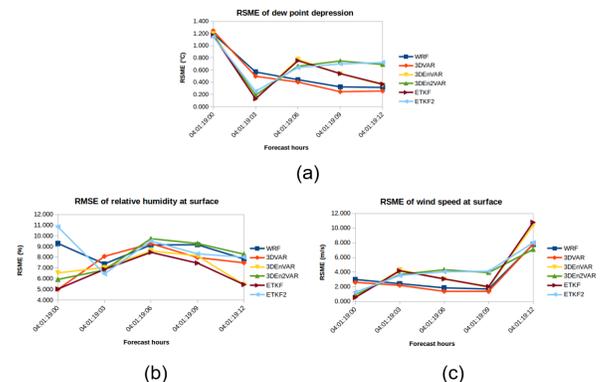


Figure 13. Behavior of (a) temperature forecasts; (b) relative humidity; (c) wind speed between 00:00 and 15:00 UTC for all experiments.

For the surface wind speed case, the 3DVAR\_P experiment shows the most realistic results, although the 3DVAR method

reduced in general the wind speeds on the surface. The hybrid methods built from the multiphysics ensemble achieved a more successful forecast in the initialization stage at surface, with absolute errors at the time of initialization around 0.19 m/s, the lowest. However, during the following hours, the mean absolute errors grew in a 1.5 and 2.5 m/s threshold, top reached at 09:00 UTC. The hybrid methods built from the previous runs were characterized by an overestimation of the mean wind speed at the initialization time and they were even worse than the run without assimilation (Fig. 13c).

### III.6. Binary analysis with Cvis algorithm

The application of the different assimilation techniques led, in some cases, to obtaining fog forecasts over the study area something that, as it was previously explained, did not achieve the without assimilation solution.

The late forecasting of the phenomenon around 06:00 UTC persisted with the 3DVAR technique. It differs from the solution without assimilation as its beginning consisted of a moderate to strong haze episode forecasting the fog event appropriately at 12:00 UTC in most cases. The delay in the fog appearance is related to a weakening of the phenomenon around 09:00 UTC, a result consistent with the characteristics of a more hostile environment previously described. The use of radiances is conducive to an increase in false alarms related to the appearance of ephemeral fog events towards the coasts around 06:00 UTC and the overestimation of the intensity of the event towards the south coast weather stations around 12:00 UTC.

Table 4. CSI's behavior for different assimilation techniques during the first 12 hours of forecast.

Method	Fog	Haze	Fog + Haze
SisPI-NA	0	0.412	0.311
<b>3DVAR.P</b>	<b>0.308</b>	<b>0.483</b>	<b>0.428</b>
3DVAR.R	0.231	0.400	0.349
3DVAR.PR	0.231	0.448	0.381
3DEnVAR.P	0.143	0.355	0.289
3DEnVAR.R	0.231	0.355	0.318
3DEnVAR.PR	0.231	0.375	0.333
3DEn2VAR.P	0.167	0.355	0.302
3Den2VAR.R	0.333	0.448	0.415
3DEn2VAR.PR	0.200	0.333	0.289
ETKF	0.167	0.364	0.311
ETKF2	0.077	0.387	0.289

Hybrid methods that use the multiphysics ensemble also showed difficulties, mainly when forecasting in the stations near the north coast (78325-78376-78318) generating miss values higher than 3DVAR and the solution without assimilation. Those methods showed more difficulties than 3DVAR to detect fog events.

Finally, the methods which use like first guess the previous runs of SisPI exhibited better results than the multiphysics ensemble but its were not result in higher performance than 3DVAR. In this case, the prepbufr data combinations leads to worst forecast, being the radiance data individually

assimilated the one that obtained better detection rates of fog, thanks to a more accurate forecast towards the north coast. On the other hand, the forecast over inner region of the study area did not achieve relevant results.

Since the CSI is a value that summarizes the relationship between successful detections, failures, and false alarms, the (Table 4) also summarizes the values of this index for all assimilation schemes considering individually the fogs and hazes as well the combination of both phenomena.

## IV. CONCLUSIONS

In a conclusive way it can be stated that the assimilation of data improves the forecast of episodes of fog and/or haze in short and very short terms. The greatest impacts were observed in fog forecast. However, the applied methods not able to significantly improve the tendency to underestimate relative humidity and overestimate temperature and wind speed.

The schemes that used the 3DVAR method led to a minor modification of the initial field in relation to the hybrids. It could be observed in the multiphysics ensemble experiments that the multiphysics first guess has a determinant impact in the assimilation results, as a consequence, the influence of the type of data or even the method turn up irrelevant because in those cases at 3DEnVAR and ETKF techniques the error curves and incremental analysis are superimposed.

Hybrid methods in general show a limited temporary assimilation impact of approximately to around 3 hours, with rapid transitions towards the solution without assimilation after this time, contrary to 3DVAR, which extends its influence a few hours more over forecasting time. Here the influence of the generic covariance error can be detrimental to hybrid schemes, which suggests exploring the use of domain-dependent static matrices (which increases the computational cost of the method) and evaluating the cost-benefit ratio.

The individual use of prepbufr data seems to be sufficient to improve the fog and/or haze forecast in short and very short terms being the 3DVAR method the most appropriate selection, as the impact of assimilation seems to spread more over time than hybrid methods and it is less computationally expensive.

An increase in the study cases and modifications in the configuration of the assimilation methods, such as the generation of a domain-dependent covariance matrix, or modifications of the weight of the dependent flow errors with in respect to static background in the case of the hybrid schemes, could provide more concrete results in order to achieve a more effective short and very short terms fog/haze forecast.

## REFERENCES

- [1] B. A. Kunkel, J. Clim. Appl. Meteorol. **23**, 34 (1984).
- [2] A. Talamo, J. Nucl. Mater. **373**, 407 (2008).
- [3] J.A. Doran, P.J. Roohr, D.J. Beberwyk et al., The MM5 at the Air Force Weather Agency — New products

- to support military operations. The 8th Conference on Aviation, Range, and Aerospace Meteorology, NOAA / NWS, Dallas, Texas (1999).
- [4] I. Gultepe, M. D. Müller, Z. Boybeyi, J. Appl. Meteorol. Climatol. **45** (2006).
- [5] C. H. Bang, J. W. Lee, and S. Y. Hong, Korean Soc. Atmos. **24**, 92 (2009).
- [6] G. Creighton, E. Kuchera, J. McCormick, S. Rentschler, B. Wickard, AFWA Diagnostics in WRF, (2014) (Available online at <https://www2.mmm.ucar.edu/wrf/users/docs/>).
- [7] W. C. Skamarock et al., A description of the advanced research WRF version 3, NCAR Tech., 125 (2008) (Available online at [http://www.mmm.ucar.edu/wrf/users/docs/arw\\_v3.pdf](http://www.mmm.ucar.edu/wrf/users/docs/arw_v3.pdf)).
- [8] C. Román-Cascón, C. Yagüe, M. Sastre, G. Maqueda, Adv. Sci. Res. **8**, 11 (2012), DOI:10.5194/asr-8-11-2012.
- [9] C. Lin, Z. Zhang, Z. Pu and F. Wang, Numerical Simulations of an Advection Fog Event over Shanghai Pudong International Airport with the WRF model **31**, 874 (2017), (Available online at <https://www.researchgate.net/publication/320820726>)
- [10] N. Franciscatto, F. Puhales, V. Anabor, E. Dal Piva, E. Nascimento, Revista do Centro de Ciências Naturais e Exatas **37**, 613 (2015), DOI:10.5902/2179-460X17277.
- [11] C. Chen, M. Zhang, W. Perrie, R. Chang, X. Chen, P. Duplessis and M. Wheeler, Earth and Space Science **7**, (2020), DOI:10.1029/2019EA000703
- [12] M. Sierra, A. Ferrer, R. Hernández, Y. González, R. Cruz, I. Borrajero, C. Rodríguez, Sistema de Predicción a muy corto plazo basado en el Acoplamiento de Modelos de Alta Resolución y Asimilación de Datos, Informe de resultado, Programa: "Meteorología y Desarrollo Sostenible del País", (2014), DOI:10.13140/RG.2.1.2888.1127.
- [13] F. Wang et al, Assimilating MTSAT-Derived Humidity in Nowcasting Sea Fog over the Yellow Sea. (2013), DOI:10.1175/WAF-D-12-00123.1.
- [14] X. Gao, S. Gao, Y. Yang, Preprints, (2018), DOI:10.20944/preprints201807.0577.v1.
- [15] S. Remy and T. Bergot, Meteorol. Weather Res. **138**, (2009), DOI:10.1175/2009MWR3110.1.
- [16] A. P. Alfonso, A. Florido, El clima de Matanzas, (Editorial Academia, La Habana, 1993) pp. 113.
- [17] E. L. Álvarez, M. I. Borrajero, M. R. Álvarez and L. A. León, Rev. Cienc. Tierra y Espacio **12**, 31 (2011).
- [18] I. Maiello, R. Ferretti, S. Gentile, M. Montopoli, E. Picciotti, F. S. Marzano and C. Faccani, Atmospheric Measurements Techniques, (2014), DOI:10.5194/amt-7-2919-2014.
- [19] V. Mazarella, I. Maiello, V. Capozzi, G. Budillon and R. Ferretti, Adv. Sci. Res. **14**, 271 (2017), DOI:10.5194/asr-14-271-2017.
- [20] D. Barker et al, Meteor. Soc., **93**, 831 (2012).
- [21] X. Wang, D. M. Barker, C. Snyder and T. M. Hamill, Mon. Wea. Rev. **136**, 5116 (2008).
- [22] X. Wang, D. M. Barker, C. Snyder and T. M. Hamill, Mon. Wea. Rev. **136**, 5132 (2008).
- [23] M. Tong, J. A. Sippel, V. Tallapragada, E. Liu, C. Kieu, I. Kwon, W. Wang, Q. Liu, Y. Ling and B. Zhang, Mon. Weather Rev. **146** 4155 (2018).
- [24] C. Bishop, B. J. Etherton, S. J. Majumdar, Mon. Weather Rev. **129**, (2000).
- [25] A. Pérez, I. Mitrani and O. Díaz, Sistema de Predicción Numérica Océano-Atmósfera para la República de Cuba, (Informe de Resultado Científico, Instituto de Meteorología, Centro de Física de la Atmósfera, 2014), (Available online at <https://modelos.insmet.cu/static/models/docs/>).
- [26] C. Schwartz, Z. Liu, X. Huang, Y. Kuo and C. Fong, Comparing limited-area 3DVAR and hybrid variational-ensemble data assimilation methods for typhoon track forecasts: sensitivity to outer loops and vortex relocation, (2013), DOI:10.1175/MWR-D-13-00028.1.
- [27] OMM, Vocabulario Meteorológico Mundial **182**, 784 (1992).
- [28] G. Acosta, Caracterización de las nieblas para cinco estaciones del occidente de Cuba en el período 2000-2009, Tesina del Diplomado de Física de la Atmósfera, UDICT, Instituto de Meteorología, (2011).

This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0, <http://creativecommons.org/licenses/by-nc/4.0>) license.

