Atmospheric Pollution Research 4 (2013) 64-74

Atmospheric Pollution Research

www.atmospolres.com

Assessment of the Weather Research and Forecasting model implementation in Cuba addressed to diagnostic air quality modeling

Leonor Turtos Carbonell [1], Gil Capote Mastrapa [1], Yasser Fonseca Rodriguez [1], Lourdes Alvarez Escudero [2], Madeleine Sanchez Gacita [3], Arnoldo Bezanilla Morlot [2], Israel Borrajero Montejo [2], Elieza Meneses Ruiz [1], Saturnino Pire Rivas [4]

[1] *Centro de Gestión de la Información y Desarrollo de la Energía, CUBAENERGIA, Calle 20, No.4111 e/18a y 47, Miramar, Playa, La Habana, Cuba*
[2] *Instituto de Meteorología de Cuba, INSMET, Cuba*
[3] *Centro de Previsão de Tempo e Estudos Climáticos – CPTEC/INPE, Brazil*
[4] *Instituto Superior Politécnico José Antonio Echeverría, Cuba*

ABSTRACT

This paper evaluates the implementation of the Weather Research and Forecasting model, WRF, for its use as the meteorological pre–processor for diagnostic air quality modeling in Cuba. The implementation of the WRF involved two studies: the first one was aimed at defining which global meteorological data is more suited for Cuba; the second one consisted of an analysis of the results for long–term runs on two domains, with the specific objective of assessing the general performance of the model. The results of the model were compared with the observations of the National Weather Service surface stations. The comparisons showed good performance for temperature and acceptable performance for prediction of wind tendencies. On average, the wind speed is overestimated in the model and the wind direction deviations exceed 30 degrees for several of the meteorological stations. These deviations are related to nearby topography and the low–wind speed. Some additional studies must be conducted in order to clarify and reduce the wind deviations. The research concludes that the WRF output is able to provide realistic meteorological patterns for air quality models, which require high–resolution three–dimensional (3D) meteorological data. The WRF–fsl tool was developed to use WRF to feed the local models as AERMOD when upper air data is not available. This tool takes the WRF output and gets the upper air data, in the fsl radiosonde format. The WRF–fsl results were compared to other solution, which incorporates a surface data parameterization. The conclusion is that the efforts, to run WRF for long periods, are not justified with the improvement in the results for regulatory purposes. However, as the differences in convective mixing height could be significant, this solution would be very useful for other kind of studies.

Keywords: *WRF model, AERMOD model, air quality modeling, statistical analysis, photochemical modeling*

doi: 10.5094/APR.2013.007

Corresponding Author:
Leonor Turtos Carbonell
☎ : *+53-7-202-7527*
🖷 : *+53-7-204-1188*
✉ : *leonorturtos@gmail.com*

Article History:
Received: 25 May 2012
Revised: 30 October 2012
Accepted: 31 October 2012

1. Introduction

Anthropogenic activities; in particular, energy production, the transport sector and industrial facilities; may cause significant air pollution at local, regional and global scales. This can be from both direct emissions of primary pollutants, and by the formation of secondary harmful species from the primary ones. The assessment of these activities is vital in order to understand the damages they cause to the environment. Minimizing them to achieve medium to long–term sustainable development must be an everyday goal.

Air pollution may be defined as a situation in which substances that result from anthropogenic activities are present at concentrations sufficiently high above their normal ambient levels to produce a measurable and undesirable effect on humans, animals, vegetation, or materials (Seinfeld and Pandis, 2006). To evaluate the air pollution from a source, it is imperative to consider both their emissions (concentration, temperature and flow rate of the exhaust gas streams, release height, etc.) as well as the contribution of these emissions on air quality (concentration of pollutants in the air).

Both emissions and air quality impacts can be: (1) measured and/or (2) estimated through models and calculation programs. The use of models is more cost–effective and quicker than other methods and recently they have proven to be very realistic in many situations and at different scales: local, regional and global. In Cuba, there have been important advances in studies at local scale (Carbonell et al., 2007a; Carbonell et al., 2007b; Carbonell et al., 2010b; Carbonell et al., 2011; Herrera et al., 2011), using screening models, like SCREEN (EPA, 1995a) and Berlyand (NC, 1999), but also using more refined one, like ISCST3 (EPA, 1995b) and AERMOD (EPA, 2004). As these latest models require upper air meteorological data, which is not available in Cuba, parameterizations from surface data were developed for their implementation. The solution for AERMOD is described in Carbonell et al. (2010a).

AERMOD introduces state–of–the–art modeling concepts into the EPA's local air quality models because it incorporates air dispersion based on planetary boundary layer turbulence structure and scaling concepts, including treatment of both surface and elevated sources, and both simple and complex terrain. AERMOD uses AERMET as the meteorological preprocessor.

The importance of the evaluation of air pollution at regional scale has been evident in recent years. Many studies have shown that the major impacts on human health from many of the primary pollutants such as sulfur and nitrogen oxides are not caused directly, but by the sulfate and nitrate aerosols in which they are

transformed during their dispersal at regional scale (Spadaro, 1999). Conversely, models as AERMOD are not capable of simulating pollutant transport and diffusion within spatially variable meteorological fields (Klausmann et al., 2003).

For the above reasons, it is required to introduce other models, for example a puff model such as CALPUFF (Scire et al., 2000a). CALPUFF model was defined by the U.S. Environmental Protection Agency as the regulatory model for regional transport of pollutants (EPA, 2003), between 50 and 300 km, although it is also proposed for those applications involving complex wind regimes at local scale. CALPUFF requires CALMET (Scire et al., 2000b) as meteorological pre–processor to display its full potential. CALMET also requires upper air meteorological data, which is not available in Cuba. Parameterizations made for the AERMOD model in Cuba (Carbonell et al., 2010a) have not been introduced in CALMET due to the complexity of the three–dimensional meteorological grid. CALMET–CALPUFF handles many options, one of which involves the use of mesoscale models for the preparation of meteorological data, such as MM5 (Anthes and Warner, 1978; Dudhia et al., 2005) or WRF (PSU/NCAR, 2010). The latter feature involved the use of interface software such as CALMM5 (TRC, 2008a) and CALWRF (TRC, 2008b), respectively; or the recently released MMIF (Brashers and Emery, 2012). Previous studies on the application of MM5 have been carried out by the Cuban Meteorology Institute (INSMET) (Mitrani et al., 2003) and CUBAENERGIA, but there has been no earlier experience with the WRF implementation in the country.

Current researches also include the incorporation and evaluation of the mesoscale models to provide upper meteorological data to local models; the MMIF interface is an example of these efforts. This matter is also discussed in this paper, in particular the use of WRF results to feed AERMOD.

Photochemical models are typically used in regulatory or policy assessments to simulate the impacts from all sources by estimating the pollutant concentrations and deposition of both inert and chemically reactive pollutants over large spatial scales. When they are not on–line or coupled to atmospheric–chemistry models, such as CMAQ (Community Multiscale Air Quality) or CHIMERE (CNRS, 2007), they also need the high–resolution 3D meteorological data as an input, which could be provided directly by WRF.

The central objective of this research is the implementation of the WRF model (WRF–ARW V3.1), aimed at assessing the pollutant dispersion rather than weather forecasting, which is its main use. In this way several modeling options should be evaluated to take into consideration the work with historical data and therefore the possibility of improved results based on the assimilation of the meteorological data observations.

The WRF model is designed for both operational and scientific purposes. It features a flexible and efficient code, with parameterizations that reflect the state of the art in the fields of physics and atmospheric dynamics, thanks to the experience of a wide scientific community. The introduction of WRF in developing countries is limited by its high computational requirements. This model requires the use of parallel processing, either through clustering or through GRID technology in distributed systems with independent workloads.

The WRF implementation is completed through three numerical experiments or case studies. The first one, described in the Section 3 and identified as the *early* case, aims at evaluating different global meteorological data as boundary and initial conditions. The other ones, described in the Section 4, consist of two nested cases, Case 1 and Case 2, with three domains each, were used for *long–term* simulations and corresponding validation.

The use of the WRF results to feed local air quality models like AERMOD, when upper air data is not available is achieved with the development of an interface module, WRF–fsl. The evaluation of this solution is described in the Section 5 through a numerical experiment, which includes the comparison of the AERMET and AERMOD results for several energy facilities located in the respective inner domains of the above–mentioned Case 1 and 2, with respect to a previously implemented solution, which incorporates a surface data parameterization.

2. Statistical Procedures to Use in the WRF Evaluation

A main step in the WRF implementation was to identify the methodology required to evaluate the results. This included identifying the statistical functions, variables to be used as indicators and their respective reference values, and the verification procedures.

2.1. Statistical functions

Statistical analysis using observations is the most common method for determining model uncertainty. The model output is compared directly to observations, statistically assessed using a number of metrics, and statements concerning the quality of the model are provided. In many ways, this procedure follows the methodologies linked to validation, but the aim of the assessment is intended to provide information on how uncertain a model is with regard to the observations. To the best of our knowledge, a protocol for the assessment of the performance of weather forecasting models has not been developed. Most scientific studies carried out a qualitative assessment, and those with a quantitative approach generally used simple statistical functions (Seaman, 2000; Titov et al., 2005; Han et al., 2008; Hanna et al., 2010; EEA, 2011). Finally, the following functions were selected for comparison in this study: root mean square error, mean absolute error, bias and index of agreement.

Mean Absolute Error (*MAE*):

$$MAE = \sum_{i=1}^{N} \frac{|M_i - O_i|}{N} \quad (1)$$

where M_i is the modeled value for cell *i,* O_i is the observed value for cell *i*, and *N* is the number of values analyzed.

Root Mean Square Error (*RMSE*) is similar to *MAE* but more sensitive to occasional large errors due to its quadratic term:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(M_i - O_i)^2}{N}} \quad (2)$$

Bias (*BIAS*) provides information on the trend of the model to overestimate or underestimate a variable, quantifies the systematic error of the model. Pielke (1984) defines *BIAS* as:

$$BIAS = \sum_{i=1}^{N} \frac{(M_i - O_i)}{N} \quad (3)$$

BIAS is intended primarily for scalar magnitudes as it calculates the tendency of the model to overestimate or underestimate the variable.

Index of Agreement (*IoA*) provides further insight into the behavior of the model for scalar magnitudes. It ranges from 0 to 1, ($0<IoA<1$) and is calculated by:

$$IoA = 1 - \frac{\sum_{i=1}^{N}(M_i - O_i)^2}{\sum_{i=1}^{N}(|M_i - M_{mean}| + |O_i - O_{mean}|)^2} \quad (4)$$

where M_{mean} is the average modeled value and O_{mean} is the average observed value.

2.2. Indicators and references values

Three meteorological variables were selected as indicators in this comparison: surface temperature, wind speed and wind direction at 10 m. The latter two were estimated from *U* and *V* wind vector components calculated by the model. *BIAS* and *IoA* were used for temperature and wind speed.

Although there is much uncertainty in this area, reference values were chosen for evaluating the model performance (Russell and Dennis, 2000; Borge et al., 2008):

- *MAE* and *RMSE*, ≤2°C for temperature, ≤2 m s^{-1} for wind speed and ≤30 degrees for wind direction,
- *BIAS* absolute value, ≤0.5°C for temperature and ≤0.5 m s^{-1} for wind speed,
- *IoA*, ≥0.8 for temperature and ≥0.6 for wind speed.

These references should not be interpreted as if they are definitive numbers. The performance measures will vary depending on the situation. There is a minimum *RMSE*, of about 1 m s^{-1} for near–surface wind speed, which cannot be improved upon due to inherent uncertainty. In addition, the wind direction *RMSE*, is found to be a function of mean wind speed (approximately inversely proportional) (Hanna and Yang, 2001).

2.3. Verification procedures: cell–cell and cell–point

Statistical methods in model grids are applied in two ways: cell–cell and cell–point verification (Pielke, 1984). The cell–cell testing consists of comparing the model results with spatial analysis data calculated from intermediate models that can average observations over grids around the world. The advantage associated with this method is the simplicity in the computation, since all points of the modeled and observed values coincide. However, some authors have noted the tendency of this methodology to produce a bias in favor of the results with lower resolutions (Stenger, 2000). This also makes the comparisons dependent on the model used for the averaging process.

The other methodology used is the cell–point verification. In this case, observations are compared with the values of the corresponding grid cells to the site of these observations. In this study, a cell–point methodology has been used. The model results were compared with surface observations, choosing the model data corresponding to the cell closest to the location of the observation. There has been no interpolation of model data to fit the specific point of observation.

3. Selecting the Global Meteorological Input Data

An important goal of this study is defining which global meteorological data is more suited as the boundary and initial condition, to run WRF for Cuba. In order to do so, an initial implementation; from now on identified as the *early case*; was completed. Several meteorological data input sources for the model were analyzed and the most currently used three sources were considered in the assessment:

- NCEP Final Analysis (FNL from GFS) (ds083.2) GFS with spatial and temporal resolutions of 1 degree and 6 hours respectively,
- NCEP/NCAR Reanalysis (ds090.0) NNRP with spatial and temporal resolutions of 2.5 degrees and 6 hours,
- NCEP Eta/NAM (ds609.2) NAM with spatial and temporal resolutions of 40 km and 6 hours.

A test run with each data set was performed for a common domain with identical physical parameterizations.

3.1. Physics options in the early case

The mesoscale meteorological model WRF offers multiple physical and dynamical options that can be combined in several ways. The options typically range from simple and efficient to sophisticated and more computationally costly and from newly developed schemes to well tried schemes such as those in current operational models.

For this research, these options were analyzed taking into account the final objective of providing input data for air quality modeling. The selection of the physical parameterizations for the *early* case considered several aspects: Cuban specific meteorological and weather conditions; when doable, use of settings already tested in the country for the MM5 model (Mitrani et al., 2003) and WRF configurations used in other countries in the region, e.g. Venezuela (CvM, 2009). These settings are listed below:

- Microphysics: WSM3, WRF Single–Moment 3–class scheme. A simple efficient scheme with ice and snow processes (mp_physics=3)
- Cumulus Parameterization: Grell–Devenyi ensemble scheme for coarse grids. Not necessary for domains with cells lower than 4 km (cu_physics=3)
- Shortwave Radiation: Dudhia Scheme. Scheme with simple downward integration allowing efficiently for clouds and clear–sky absorption and scattering (ra_sw_physics=1)
- Long–wave Radiation: RRTM scheme (Rapid Radiative Transfer Model). An accurate scheme using look–up tables for efficiency. Accounts for multiple bands, trace gases, and microphysics species (ra_lw_physics=1)
- Surface Layer: MM5 similarity, based on Monin–Obukhov with Carslon–Boland viscous sub–layer and standard similarity functions from look–up tables (sf_sfclay_physics=1)
- Land Surface: 5–layer thermal diffusion. Soil temperature only scheme, using five layers (sf_surface_physics=1)
- Planetary Boundary Layer (PBL): Yonsei University scheme. Non–local–K scheme with explicit entrainment layer and parabolic K profile in unstable mixed layer (bl_pbl_physics=1)

The Cuban implementation of MM5 used the following options: the simple ice as explicit moisture schemes; the Grell scheme for cumulus parameterization; the cloud–radiation scheme, sophisticated enough to account for long–wave and shortwave interactions and the Burk–Thompson PBL scheme.

3.2. Location of the study domains in the *early case*

The *early case* consisted of two nested squared domains of 45 and 49 cells with 9 and 3 km grid sizes respectively and 50 isobaric levels (eta) in height, with a common center located at 23.1 N and 82.35 W, solved in two–way nesting. The case uses Lambert Conic Conformal (LCC) projection. The nested grid included the provinces of Artemisa, Mayabeque and Havana (see Figure 1).

3.3. Comparison with observations in the *early case*

The WRF hourly outputs were matched with the corresponding observations in eight surface stations of the National Meteorological Service (NMS). The stations are located within the inner modeling domain and shown in Figure 1. At NMS, variables are recorded every three hours.

The ID code of the stations and their coordinates, in Latitude (N) and Longitude (W), are indicated below:

(1) Casablanca, 78325, 23.144 N, 82.342 W
(2) Bauta, 78376, 22.97 N, 82.53 W
(3) Bainoa, 78340, 23.03 N, 81.92 W
(4) Batabano, 78322, 22.72 N, 82.28 W
(5) Guines, 78323, 22.85 N, 82.03 W
(6) Melena, 78375, 22.77 N, 82.13 W
(7) Tapaste, 78374, 23.02 N, 82.13 W
(8) Santiago de Las Vegas, 78373, 22.97 N, 82.38 W

Due to limitations in computing capacities, the modeling period was one week (07/10/2008 to 14/10/2008), selected as a representative week of the year according to a cluster statistical analysis of the surface meteorological data. Average statistical results of the analysis, including *IoA*, are presented in Table 1, at the top. In the case of temperature, the average of the *MAE, RMSE* and *BIAS* during the week in the eight meteorological stations are minimum using GFS data, by around one, with the highest *IoA*, being 0.92. For NNRP data, the statistical functions *MAE, RMSE* and *BIAS* were slightly higher than GFS, but the *IoA* was slightly lower. In the case of NAM data, the WRF´s results were significantly lower than the observations, as is evident by the large negative value of the average temperature *BIAS*.

In the case of wind speed, from the three global input data, the modeled values are overestimated, mainly due to poor management of the calm phenomena in WRF (Zawar–Reza et al., 2005; de Meij et al., 2009), a common limitation to all current models of this type. The best performance so far is using NAM data. A detailed explanation about the performance of the mesoscale models in the case of low–wind speed is included in the following section. For wind direction, the *MAE* ranges from 31 degrees when GFS data is used to 43 degrees when NAM data is used. Episodic large errors are evidenced from the marked difference between the values of *MAE* and *RMSE*.

In summary, the overall performance of the model using GFS and NNRP data is similar, with GFS having better results for temperature and wind direction. The WRF´s results using NAM data show large differences with the observed temperature. These results supported the use of GFS data in *long–term* simulations.

4. Model Performance in *Long–Term* Simulation Cases

Once the *early* case was solved and the GFS data set was selected for the boundary and initial conditions, two *long–term* simulation cases were completed. The results of these cases are the focus of the present research.

4.1. The domains and physical configuration in *long–term* simulation cases

Thanks to a partnership between the Center for Information Management and Energy Development (CUBAENERGIA) of Cuba and the Center for Energy, Environment and Technology (CIEMAT) in Spain, temporally consecutive weekly runs of the WRF model were conducted for two cases (Figure 2) during one year, 2009. Each case included three domains solved in two–way nesting. The outer domain is common in both case studies and it covers the entire Cuba Island. It contains 45 x 30 cells of 27 km in Lambert Conformal Conic projection center in 22.19 N, 79.52 W.

Case 1 contains the western zone of the island as the second domain and Havana´s counties as the inner domain. In Case 2, the medium domain covers the central part of the island and the inner one, the province of Cienfuegos. In both cases, the domains 2 and 3 contain 34 x 34 cells of 9 and 3 km, respectively. The center of both internal domains is located at 23.1 N, 82.35 W and 22.19 N, 80.52 W for Case 1 and 2 respectively.

Meteorological input data type GFS was used with 1 degree resolution every 6 hours and 28 eta levels in height for all domains used.

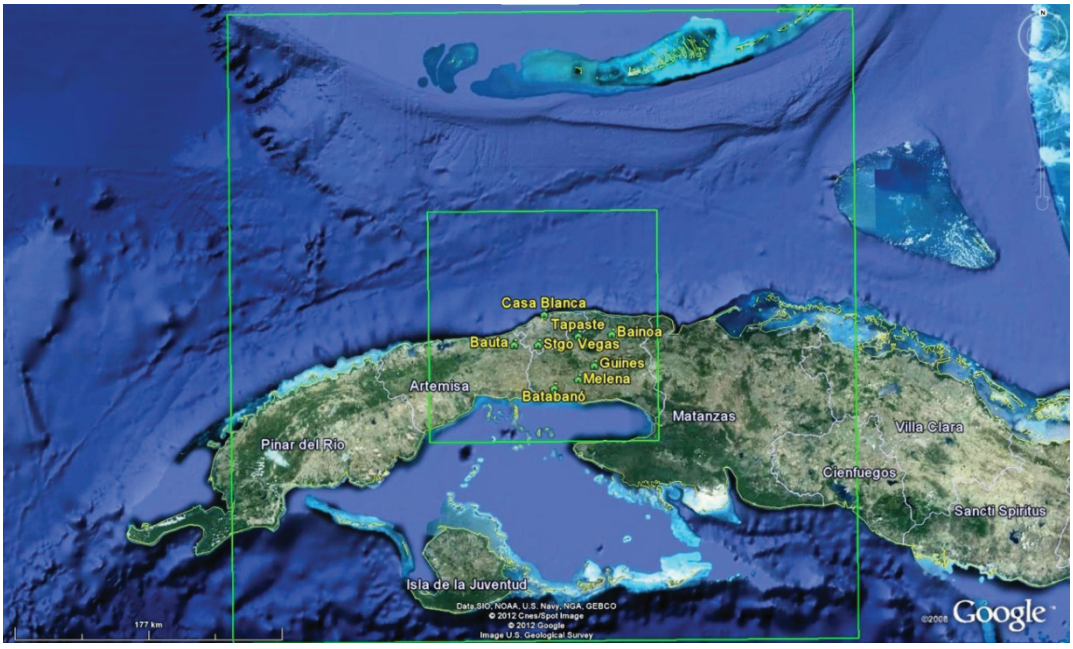

Figure 1. *Location of the early case domains and surface meteorological stations used for the evaluation of WRF results.*

Table 1. *Average of statistical functions for temperature, wind speed and direction in the WRF case studies: the early case and the long-term simulation cases*

EARLY CASE				
Temperature (°C)	**MAE**	**RMSE**	**BIAS**	**IoA**
GFS	1.26	1.6	1.05	0.92
NNRP	1.49	1.94	1.31	0.9
NAM	5.63	6.61	-5.46	0.59
Wind speed (m s^{-1})				
GFS	2.43	2.78	2.07	0.51
NNRP	2.31	2.71	1.93	0.51
NAM	1.75	2.13	1.12	0.58
Wind direction, degrees				
GFS	31.02	43.09		
NNRP	37.08	51.65		
NAM	43.2	55.34		
LONG-TERM SIMULATION CASES				
Temperature (°C)	***MAE***	***RMSE***	***BIAS***	
Casablanca	1.25	1.66	0.22	
Batabano	1.54	2.08	0.89	
Bauta	1.31	1.81	0.37	
Melena	1.6	2.12	0.16	
Santiago de la Vegas	1.24	1.76	0.26	
Tapaste	1.38	1.95	0.68	
Average in Case 1	1.39	1.9	0.43	
Cienfuegos	1.24	1.77	0.42	
Aguada de Pasajeros	1.4	1.9	0.78	
Average in Case 2	1.32	1.83	0.6	
Wind speed (m s^{-1})	**MAE**	**RMSE**	**BIAS**	**Calm (%)**
Casablanca	1.46	1.89	0.33	4.76
Batabano	2.3	2.72	1.92	26.03
Bauta	2.66	3.02	2.54	39.38
Melena	1.93	2.33	0.73	6.47
Santiago de la Vegas	2.11	2.51	1.85	12.81
Tapaste	3.15	3.49	3.07	27.53
Average in Case 1	2.27	2.66	1.74	
Cienfuegos	2.09	2.53	1.74	16.44
Aguada de Pasajeros	2.76	3.13	2.7	37.60
Average in Case 2	2.43	2.83	2.22	
Wind direction (degrees)	**WRF**	**Obs.**	**Diff.**	
Casablanca	86.31	72.66	13.65	
Batabano	67.3	36	31.3	
Bauta	89.61	18.34	71.27	
Melena	83.93	54.15	29.78	
Santiago de la Vegas	91.44	55.39	36.05	
Tapaste	91.91	56.79	35.12	
Average in Case 1	85.08	48.89	36.20	
Cienfuegos	79.56	27.45	52.11	
Aguada de Pasajeros	81.07	37.33	43.74	
Average in Case 2	80.31	32.39	47.92	

Both cases used the same physical options. The high computational capacities in CIEMAT, available for the calculation allowed substitution of the microphysical options, the WSM3 for WSM5. WSM5 is a slightly more sophisticated version of WSM3, allows for mixed–phase processes and super–cooled water, (mp_physics=4). In addition, RRTMG, a shortwave and long–wave radiation scheme with Montecarlo Integrated Column Approach (MCICA) method of random cloud overlap (ra_sw_physics=4, ra_lw_physics=4) was used in the Domain 1.

Rapid Radiative Transfer Model (RRTM) for long–wave radiation (ra_lw_physics=1) and Dudhia Scheme for shortwave radiation (ra_sw_physics=1) were used in Domains 2 and 3. Domain 3 did not use any Cumulus Parameterization because it is not necessary as its cells are lower than 4 km.

4.2. Statistical results in long–term simulation cases

The previous methodology of analysis was applied for the innermost domain in both cases. In Case 1, six surface stations, listed above, are located within the domain (Casablanca, Bauta, Batabano, Melena, Tapaste and Santiago de las Vegas); in Case 2, the following two surface stations are located within the domain:

(1) Cienfuegos, 78344, 22.186 N, 80.445 W
(2) Aguada de Pasajeros, 78335, 22.383 N, 80.85 W

Statistical functions were calculated for temperature, wind speed and wind direction (see Table 1). For temperature, *RMSE* and *MAE* were lower than 2 K and *BIAS* less than 0.5 K in most stations.

For wind speed, the model tends to overestimate measurements in most stations with values above 2 m s^{-1} for both *RMSE* and *MAE*. This behavior was further strengthened by the average *BIAS* obtained, fulfilling the reference threshold of ±0.5 m s^{-1} only for the Casablanca station. It is important to point out that Casablanca is the reference station of the National Meteorological Service, located in the headquarters of the Cuban Meteorological Institute in Havana, therefore the quality of the observations in this station is higher than of any others. This behavior confirms the early case results about the poor management of the calm phenomena in WRF. For wind speed in the *long–term* simulation cases, Table 1 includes an additional column with the percent of hours with calm conditions in each meteorological station. The correlation factor between *BIAS* and the prevalence of calms is 0.86. This value increases to 0.91 when Tapaste station is not considered in the analysis. Tapaste station shows the highest values of *RMSE*, *MAE* and *BIAS* for wind speed but this overestimation must be related to nearby topography, showed in the left section of Figure 3. The station is located beyond a hill in the predominant wind directions.

It should also be noted that in the model results, the wind speed range is smaller than actual observations, an indication of the model´s predisposition to smooth values of this variable.

To deepen the understanding of the ability of WRF to reproduce the wind speeds under different seasons/months and day/night, Figure 4 shows the *BIAS* wind speed, between the values modeled by WRF and observed on Casa Blanca meteorological station in different months and hours of the day. Although the average *BIAS* is 0.33 (the WRF overestimates the observed wind speed), during April, the month with highest average wind speed (5.4 m s^{-1}), the wind speed obtained by WRF is significantly underestimated, except for the hours before dawn. In March, May, June and July, WRF underestimated wind speed during the daytime but at nighttime the *BIAS* is positive, indicating that the wind speed is overestimated.

As for the wind direction, in most of the stations reviewed in Case 1 the difference between the model resultant vector and the station measurement's is approximately 30°, with borderline cases for the Casablanca station with 13.6 and Bauta with 71.3 degrees. For Case 2 both stations' resultant vectors from the model and the observations overcome this value. Figure 5 shows this behavior for

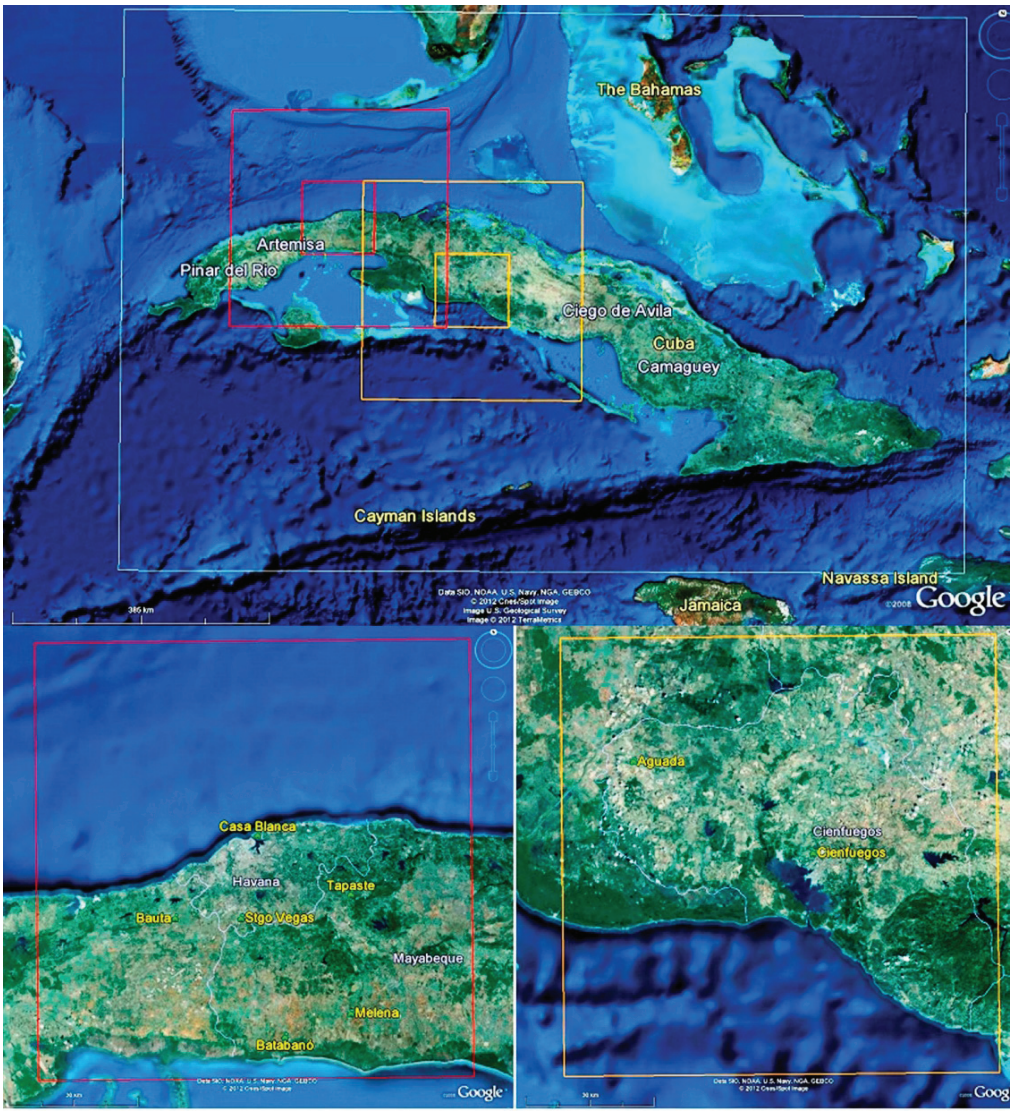

Figure 2. *Long-term case study domains (top); Case 1 (in red), Case 2 (in orange). Location of the surface meteorological stations in most inner domain (bottom).*

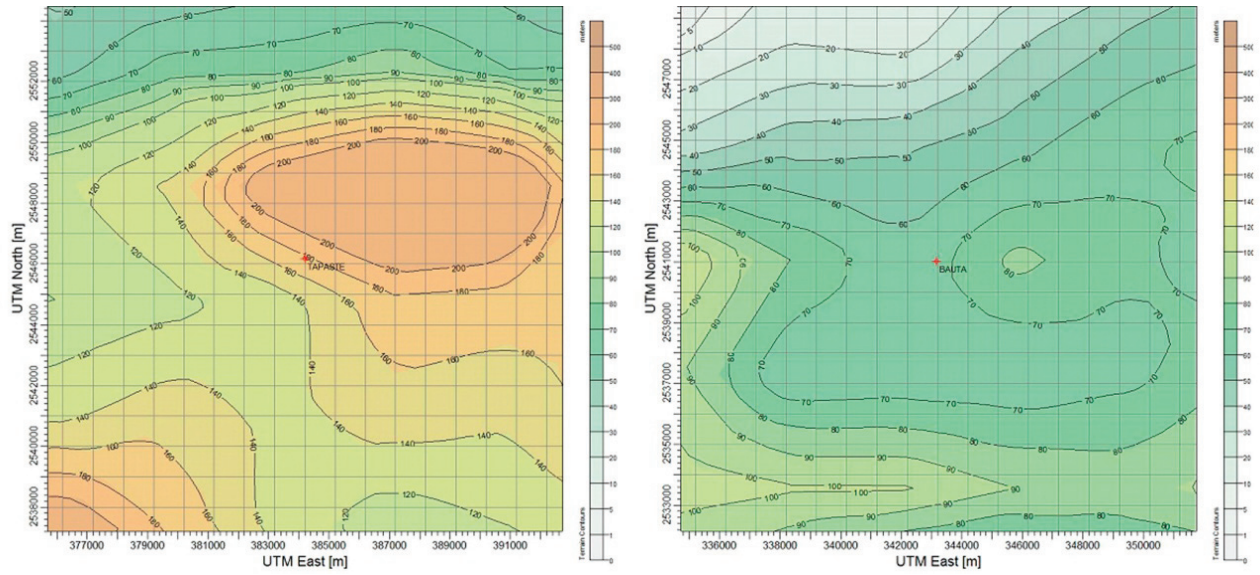

Figure 3. *Nearby topography around Tapaste and Bauta surface meteorological stations.*

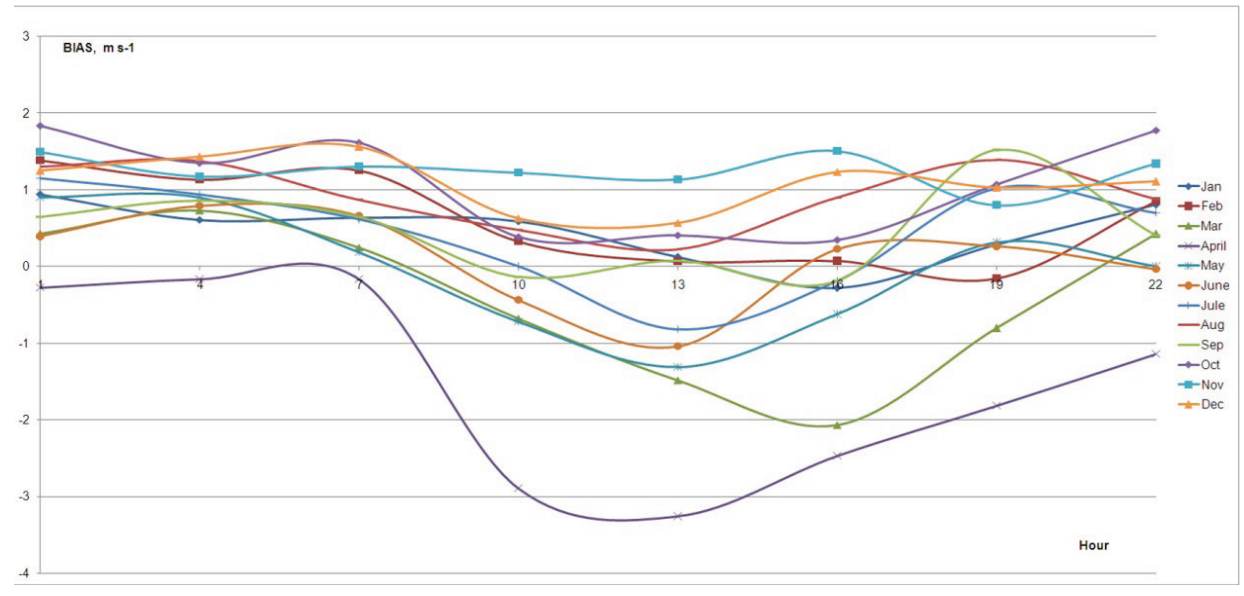

Figure 4. *BIAS in wind speed, between the values modeled by WRF and observed on Casa Blanca meteorological station in different months and hours of the day.*

some representative stations in Case 1. The differences in the wind roses are likely to be due to nearby topography, especially for Bauta, where the highest differences were found. In Bauta, the WRF is simulating the dominant synoptic easterlies, while the observations are indicating dominant winds from the north sector, due to a channeling by terrain, as Figure 3 (right section) shows.

This analysis could indicate the necessity to increase the horizontal resolution of the model at least to 1 km, in order to provide a better representation of the surface heterogeneity. The WRF limitations with high resolutions, at around 1 km and beyond, must be taken into consideration in order to reach an optimal resolution according to the specific case study. In addition, increasing the topographic dataset resolution should be evaluated. WRF used topographic data with resolution of 30", but there are freely available topographic data in Internet, with higher resolution (3") all over the world, like SRTM2 files (Rodriguez et al., 2005; Farr, 2007).

The deviations in Batabano are related with the poor management of the calm phenomena in WRF. The percent of hours with calm conditions in this station is 26%.

It should be noted that comparable results were obtained in other studies around the world for wind speed and direction (Hanna and Yang, 2001; Jimenez et al., 2005; Perez et al., 2006; Galeas, 2009; Kusaka et al., 2009). The significant deviations in wind variables do not indicate a poor ability of the model to reproduce wind patterns. Some additional studies must be conducted in order to improve the implementation.

5. Feeding AERMOD with the WRF Outputs

In Cuba and in other countries, upper air soundings are not performed at all or they are not available with the necessary frequency (twice daily). The simplest solution was presented in Carbonell et al., 2009; Carbonell et al., 2010a, the MPPBL module of AERMET was expanded and a new version, AERMET+, was obtained. AERMET+ does not require the upper meteorological data and it estimates the convective mixing heights, the convective velocity scale and the potential temperature gradient above the mixing height based on surface meteorological data. AERMET+ is available for latest two version of AERMET, 06341 and 11059.

Another more complicated solution is to use the WRF results to feed AERMOD. To implement this solution, an interface module between WRF and AERMET was developed. The following sections compare the AERMET and AERMOD results using these two solutions.

5.1. *WRF–fsl* tool

The study started from the analysis of how the WRF results can be used by the AERMOD modeling system:

(1) WRF can directly feed the surface and upper meteorological data to AERMET.

(2) WRF results are directly fed to AERMOD because all variables required by the AERMET output (input for AERMOD) are contained in the WRF output or they can be estimated through simple processing.

(3) WRF only feeds AERMET with the upper air data, the surface data is extracted from surface local stations.

The first and second options should only be used if the WRF assimilates local data. Other studies, which solved the same problem, were reviewed (Randolph, 2002; Brode, 2008; Davis et al., 2008; Myers–Cook et al., 2010). The option selected was the third, which can be used to run both AERMOD and another local model that requires sounding data in *fsl* radiosonde format (NOAA, 2012), such as ISCST3. It is also the simplest to implement. In this way, WRF provides a file, which replaces the upper air sounding. The *WRF–fsl* tool was developed with this objective and it could be used to create radiosonde files, both in original or new *fsl* format.

5.2. AERMET+ vs. *WRF–fsl*→AERMET

For evaluating of the implemented solution, another numerical experiment was conducted. This experiment includes two case studies, matching with the cases described in Section 3.1 for WRF. In Case 1 and 2 of this experiment, the emission facilities are located in 23.11 N, 82.35 W and in 22.19 N, 80.52 W, almost in the middle of the inner domains of the WRF long–term simulation, Case 1 and Case 2 respectively.

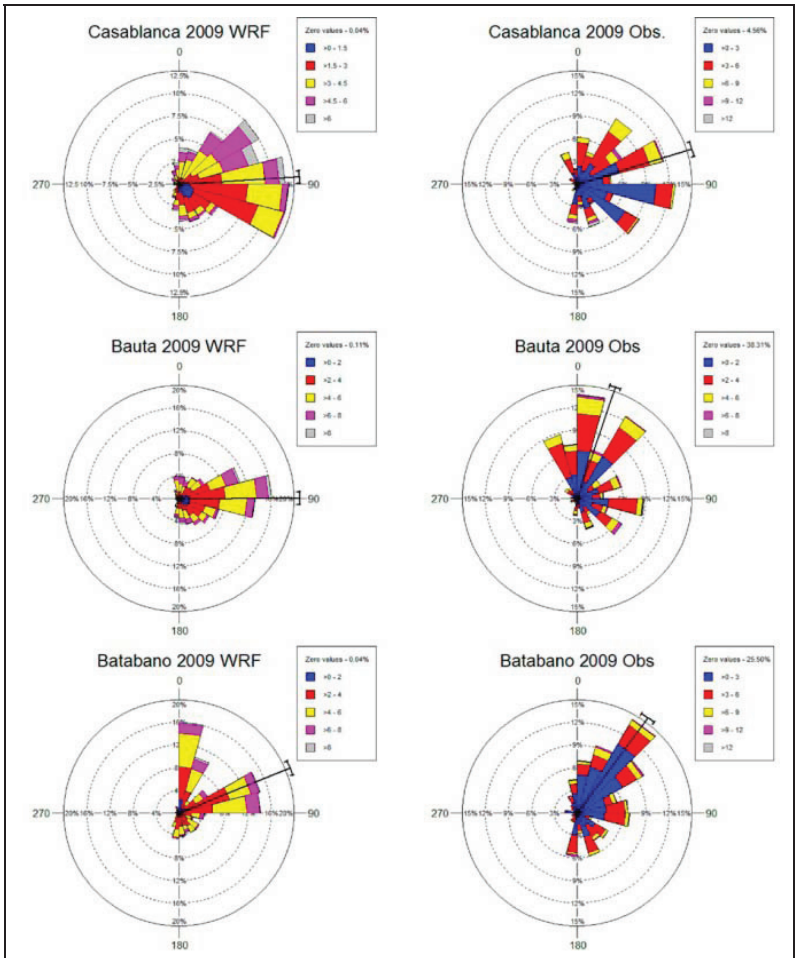

Figure 5. *Wind roses WRF vs. observations in representative stations in Case 1.*

Figure 6 compares the convective mixing height, *Zc,* calculated with AERMET+ (X–axis) and *WRF–fsl*→AERMET (Y–axis) for both cases. Therefore, the figure is comparing two models simulations of convective mixing height and there is no comparison with observations. Additionally, a histogram was plotted for Case 1, in which the larger deviations are observed. The upper part of the histogram shows the *Zc* differences (classes) using AERMET+ and *WRF–fsl*→AERMET versus frequency (m) and cumulative (%), sorted by frequency, up to a cumulative 99%. The classes are represented by the average value of each range, in this case of ±33 m. The right part of the histogram shows the classes symmetrically distributed around zero deviation. As 8 760 hours were evaluated, it is appreciated that more than 50% of the time the deviation is less than 50 m.

It can be concluded, that the results obtained using each version of the AERMET pre–processor, are comparable but with significant differences, especially in Case 1, where the linear relationship for the convective mixing height is 0.828. In both cases, AERMET+ estimates bigger *Zc* values than *WRF–fsl*→AERMET.

In Figure 6 for Case 1, there seems to be two groups of points. One group follows the line of good agreement but the second group has a slope significantly lower than one, showing that *Zc* values estimated by AERMET+ are higher than the value estimated by WRF–fsl–AERMET. This happens for high wind speeds (10 m s^{-1} and higher) because in the algorithm implemented in AERMET+, the *Zc* depends directly on the friction velocity. In Case 1, for 112 hours, the wind speed is higher than 10 m s^{-1}. The situation is different in Case 2, where the highest wind speed is around 9 m s^{-1}. This is the reason why there is a better correspondence in Case 1 than Case 2.

5.3. Comparing AERMOD results

To evaluate the influence of the considered options in AERMOD results, Case 1 was chosen; in which the larger deviations are observed in AERMET results. AERMOD estimated the environmental incremental concentrations of SO_2 and PM_{10} due to the emissions from the most representative energy facilities in the country: gas turbines, power plants with oil steam boilers and generations set with internal combustions engines. In addition, a flare was also considered due to the high impact of these technologies into the air pollution. The main source characteristics are included in Table 2. At the local scale, the convective boundary layer is a key parameter in the impacts on the sources, in particular when stack height is low, as is the case for gas turbines, flares and generation sets.

The analysis of the results included the maximum and average SO_2 and PM_{10} concentrations estimated by AERMOD using AERMET+ and WRF–fsl→AERMET for different averaging periods; hourly, daily and annually (Table 2). In spite of the significant differences in convective mixing height calculated by AERMET; for both pollutants there is a coincidence in the highest and average concentrations for the entire averaging period considered. The hourly averages show a small variation: the SO_2 concentrations range from 227 using AERMET+, to 256 µg m^{-3} using WRF–fsl→AERMET and for PM_{10} concentrations from 21 to 22 µg m^{-3} respectively.

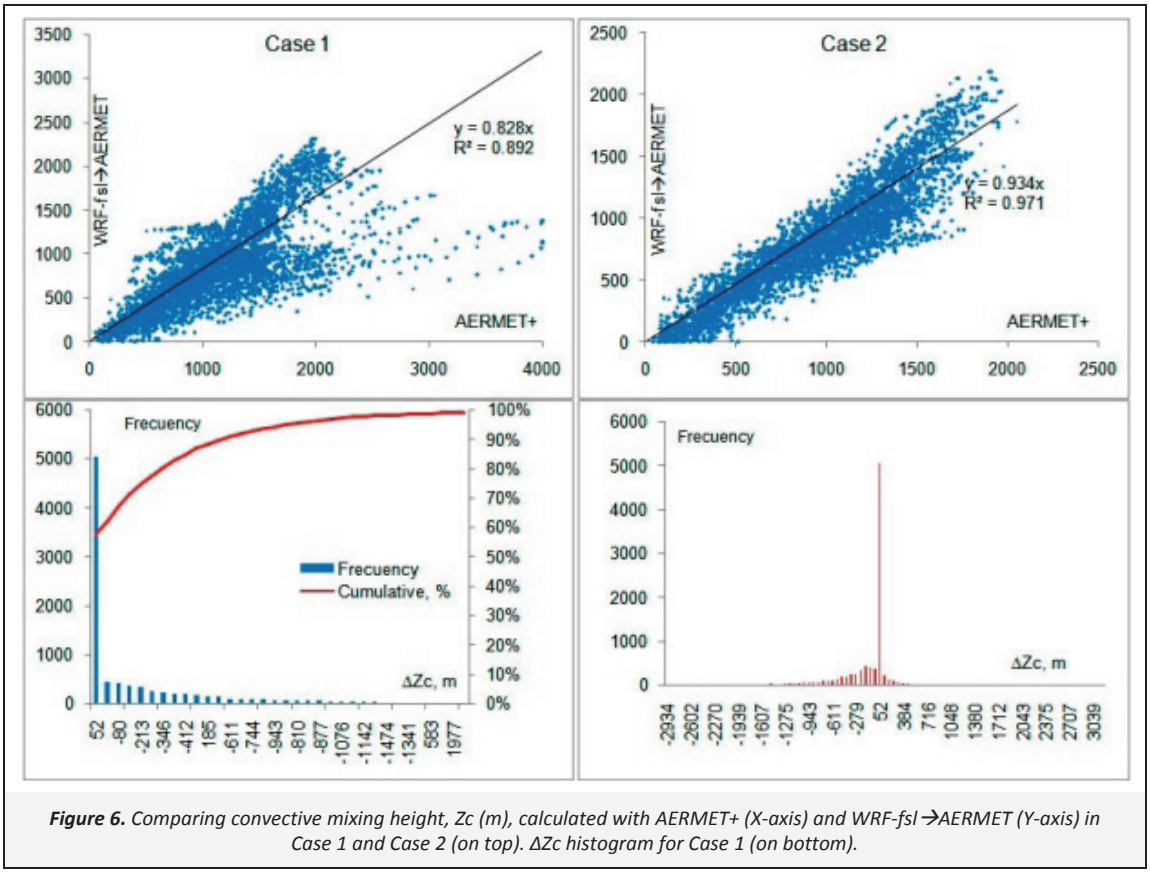

Figure 6. *Comparing convective mixing height, Zc (m), calculated with AERMET+ (X-axis) and WRF-fsl→AERMET (Y-axis) in Case 1 and Case 2 (on top). ΔZc histogram for Case 1 (on bottom).*

Table 2. *Main characteristics of sources used in AERMOD modeling. Maximum and average SO_2 and PM_{10} concentrations estimated by AERMOD using the AERMET+ and WRF-fsl→AERMET results*

Source characteristics	Flare	Gas turbine	Oil steam boiler	Internal combustion engine
Stack Height (m)	65.4	12	100	37.5
Stack Diameter (m)	2.3	3	6	1.2
Flue gas speed (m s^{-1})	6.1	40.4	7	15
Flue gas temperature (K)	1 273	823	423	520
SO_2 emissions (g s^{-1})	157	0.2	1 000	13
PM_{10} emissions (g s^{-1})	10	0.4	50	1
Incremental concentrations	**Using AERMET+**		**Using *WRF-fsl*→AERMET**	
			SO_2 concentrations (μg m^{-3})	
Averaging period	**Maximum**	**Average**	**Maximum**	**Average**
1 hour	1 010.8	226.9	1 010.8	256.3
24 hours	339.2	30.2	340.6	30.5
Annual	63.6	2.6	62.0	2.7
			PM_{10} concentrations (μg m^{-3})	
1 hour	45.0	20.8	45.0	22.4
24 hours	18.8	2.5	18.8	2.5
Annual	3.6	0.2	3.5	0.2

Significant deviations in the *Zc* estimation, explained in the previous section, correspond to the hours with high wind speeds. Generally, these deviations are not considered in AERMOD because at these hours, the mixing height in the Convective Boundary Layer (CBL) is equal to mechanical mixing height. Note that AERMET estimates the mixing height in the CBL, taking into account its dependence on both mechanical and convective processes. Then, the mixing height is calculated during the day, from the larger of the convective and the mechanical mixing height.

For regulatory purposes, the use of the WRF results as the input to the AERMOD system is not justified, as it requires unquestionably greater resources, storage, computing time, etc. than the alternative, AERMET+, which is sufficient to use. However, as the differences in convective mixing height could be significant, it would be very useful for other studies.

6. Conclusions

This research was undertaken to assess the possibility of using the WRF mesoscale model as the meteorological preprocessor for

air quality modeling in Cuba. Three input boundary conditions were analyzed for the *early* case to determine the best overall performance, and model verification methods were reviewed. Two different Cuban scenarios were run with the same configuration, and their results were compared with the available surface meteorological data from stations located within each domain.

The correspondence of the modeled variables and observations is consistent with the statistical reference limits for the case of temperature, but not as well matched for wind speed and wind direction, in which the reference values in almost all stations are exceeded. Variable winds and calm phenomena, very common in Cuban climate, contribute to this variance. This fact was verified by the high correlation (0.91) between the wind speed *BIAS* and the prevalence of calm conditions at weather stations. In Tapaste and Bauta stations, which show the highest values of *RMSE*, *MAE* and *BIAS* for wind speed and wind direction respectively, the analysis concluded that the main reason for the deviations is the nearby topography. The deviation in wind speed has great relevance since higher values of wind speed favors the dispersion processes, therefore an overestimation of the speed can lead to significant errors in air quality modeling.

In spite of the above comments, it is significant to point out that in Casablanca, the reference station of National Meteorological Service, located in the headquarters of the Cuban Meteorological Institute in Havana, the wind speed and wind direction deviations were less than the threshold–established values.

It is essential to implement and validate this model for both air quality studies and weather forecast. Given its high computational requirements in both processing speed and storage, which require parallel processing, an integrated national strategy, that includes the use and expansion of existing resources, is needed.

As a first attempt of WRF implementation in the country, the results provide recommendations for future studies more than conclusions. In this regard, it may be required to:

- Increase the horizontal resolution of the model at least to 1 km, taking into consideration the WRF limitations with high resolutions, at around 1 km and beyond, in order to reach an optimal resolution according to the specific case study.
- Increase the resolution of the topography data used to run WRF from 30” to 3”.
- Study more specific model settings for Cuban weather conditions, given the importance of accurate predictions of wind in air quality modeling.
- Conduct further assessment with data from gradient wind towers installed in the country.
- Install at least one upper air station in the country, to validate model results at different levels of the atmosphere, aimed at evaluating uncertainties introduced by the model in air quality studies.

The paper also evaluated the use of the WRF outputs to feed the AERMET using the *WRF–fsl* tool, comparing the AERMOD results in two case studies, with other previously implemented solution where upper air meteorological data is not available: the parameterization of the surface meteorological data with AERMET+. Even in the Havana case study, where the adjustment of convective mixing height is worse, 0.828 with a correlation of 0.892, the maximum and average concentrations of the modeled species with AERMOD reach values almost identical to most of the periods evaluated. These results indicate that the use of WRF does not justify the unquestionably great effort required, the use of resources, storage, computing time, etc. and that the alternative, AERMET+, is sufficient for regulatory purposes.

Acknowledgements

We thank Fernando Martin, Marta Garcia and Inmaculada Palomino from Atmospheric Pollution Modeling Department of the CIEMAT for their support in WRF runs. We also thank the International Atomic Energy Agency, IAEA, which funded the project CUB7007 “Assessing the Atmospheric Pollution of Energy Facilities for Supporting Energy Policy Decisions”, including the fellowships of CUBAENERGIA’s researchers in the CIEMAT. We also appreciate the support of the National Research Programs: Environmental Protection and Sustainable Development from Environmental Agency and Cuban Nuclear Program. The editing of earlier versions of this paper by Mario Alberto Arrastia Avila and Francesca Boughey is gratefully acknowledged.

References

Anthes, R.A., Warner, T.T., 1978. Development of hydrodynamic models suitable for air pollution and other mesometerological studies. *Monthly Weather Review* 106, 1045-1078.

Borge, R., Alexandrov, V., del Vas, J.J., Lumbreras, J., Rodriguez, E., 2008. A comprehensive sensitivity analysis of the WRF model for air quality applications over the Iberian Peninsula. *Atmospheric Environment* 42, 8560-8574.

Brashers, B., Emery, C., 2012. The Mesoscale Model Interface Program (MMIF), EPA Contract No. EP-D-07-102, Novato, 41 pp.

Brode, R.W., 2008. MM5-AERMOD tool, *Proceedings of 9th Conference on Air Quality Modeling*, October 9, 2008, North Carolina, USA, 22 pages.

Carbonell, L., Alonso, D., Capote G., Paz, E., Meneses, E, Fonseca, Y., Ricardo, H., Lopez, D., Molina, E., Rodriguez, A., Rodriguez, M., 2011. Impact study of the atmosphere pollution due to Camilo Cienfuegos refinery. Final report. Contract SCT No.8/2010, CUBAENERGÍA-CUVENPETROL S.A.

Carbonell, L.M.T., Gacita, M.S., Oliva, J.D.J.R., Garea, L.C., Rivero, N.D., Ruiz, E.M., 2010a. Methodological guide for implementation of the AERMOD system with incomplete local data. *Atmospheric Pollution Research* 1, 102-111.

Carbonell, L., Ruiz, E.M., Carrera, D.W., Zucchetti, M., 2010b. Global and local atmospheric pollution evaluation and control. challenges for a small island and for developing countries. *Fresenius Environmental Bulletin* 19, 2354-2360.

Carbonell, L.T., Rivero, J.R., Garea, L.C., Gacita, M.S., Ruiz, E.M., Diaz N., 2009. Method for the estimation of the convective mixing height aimed to atmospheric local dispersion modeling, in *Environmental Impact Assessments*, edited by Fridian, Y.T., Halley, G.T., Nova Science Publishers, Hauppauge, 12 pages.

Carbonell, L.T., Ruiz, E.M., Gacita, M.S., Oliva, J.R., Rivero, N.D., 2007a. Assessment of the impacts on health due to the emissions of Cuban power plants that use fossil fuel oils with high content of sulfur. estimation of external costs. *Atmospheric Environment* 41, 2202-2213.

Carbonell L., Gacita, M. S., Diaz, N., Meneses, E., Rivero, J, Curbelo, L, Molina, E, Fernández, M., 2007b. Local dispersion study local of gases emitted by Moa Nickel S.A. – Pedro Soto Alba. Inal Report, Contract SCT 03/2007, CUBAENERGÍA-Moa Nickel S.A.

CNRS (French National Centre for Scientific Research), 2007. http://www.lmd.polytechnique.fr/chimere/, accessed in May 2012.

CvM (Centro virtual de Meteorología), 2009. http://met.ivic.gob.ve/cvm, accessed in May 2009.

Davis, N., Arunachalam, S., Brode, R., 2008. MCIP2AERMOD: a prototype tool for preparing meteorological inputs for AERMOD. *Proceedings of 7th Annual Models-3 CMAS Users Conference*, October 6-8, 2008, Chapel Hill, 21 pages.

Dudhia, J., Gill, D., Manning, K., Wang, W., Bruyere C., 2005. PSU/NCAR Mesoscale Modeling System Tutorial Class Notes and User's Guide: MM5 Modeling System Version 3, Boulder, CO, 366 pages.

EEA (European Environment Agency), 2011. The Application of Models under the European Union's Air Quality Directive: A Technical Reference Guide, Technical Report No. 10/2011, Luxembourg, 69 pp.

EPA (Environmental Protection Agency), 2004. User's Guide for the AMS/EPA Regulatory Model – AERMOD, EPA-454/B-03-001, North Carolina, 216 pages.

EPA (Environmental Protection Agency), 2003. Revision to the Guideline on Air Quality Models: Adoption of a Preferred Long Range Transport Model and Other Revisions, 40 CFR Part 51, USA.

EPA (Environmental Protection Agency), 1995a. SCREEN3, Model User's Guide, EPA-454/B-95-004, North Carolina, 60 pp.

EPA (Environmental Protection Agency), 1995b. User's Guide for the ISC3 Dispersion Models, Volume I, EPA-454/B-95-003a, North Carolina, 390 pages.

de Meij, A., Gzella, A., Thunis, P., Cuvelier, C., Bessagnet, B., Vinuesa, J.F., Menut, L., 2009. The impact of MM5 and WRF meteorology over complex terrain on CHIMERE model calculations. *Atmospheric Chemistry and Physics Discussions* 9, 2319-2380.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The shuttle radar topography mission. *Reviews of Geophysics* 45, art. no. Rg2004.

Galeas, G., 2009. Valuation of the MM5 meteorological model for photochemical studies. Grade thesis, Engineer College, Santiago University.

Han, Z.W., Ueda, H., An, J.L., 2008. Evaluation and intercomparison of meteorological predictions by five MM5-PBL parameterizations in combination with three land-surface models. *Atmospheric Environment* 42, 233-249.

Hanna, S.R., Reen, B., Hendrick, E., Santos, L., Stauffer, D., Deng, A., McQueen, J., Tsidulko, M., Janjic, Z., Jovic, D., Sykes, R.I., 2010. Comparison of observed, MM5 and WRF-NMM model-simulated and HPAC-assumed boundary-layer meteorological variables for 3 days during the IHOP field experiment. *Boundary-Layer Meteorology* 134, 285-306.

Hanna, S.R., Yang, R., 2001. Evaluations of mesoscale models' simulations of near-surface winds, temperature gradients, and mixing depths. *Journal of Applied Meteorology* 40, 1095-1104.

Herrera, I., Garcia, Y., Martínez, R., Ocana, V., Nunez, V., Ocana, Y., Ruyck, J., 2011. Impact assessment of the emissions related to fuel oil power generation plants based on set of internal combustions engines in Santa Clara city, Cuba. IV Workshop Atmospheric Pollution vs. Sustainable Development.

Jimenez, P., Jorba, O., Parra, R., Baldasano, J.M., 2005. Influence of high-model grid resolution on photochemical modelling in very complex terrains. *International Journal of Environment and Pollution* 24, 180-200.

Klausmann, A., Phadnis, M., Scire, J., 2003. The application of MM5/WRF models to air quality assessments. *Proceedings of 13th PSU/NCAR Mesoscale Model Users' Workshop*, June 10-11, 2003, USA, 124-127.

Kusaka, H., Chen, F., Mukul, T., Duda, M., Dudhia, J., Miya, Y., Akimoto, Y., 2009. Performance of the WRF model as a high-resolution regional climate model: model intercomparison study. *Proceedings of 7th International Conference on Urban Climate*, June 29 - July 3, 2009, Yokohama, Japan, 4 pages.

Mitrani, I., Alvarez, L., Borrajero, I., 2003. Implementation and optimization of the MM5V3 on Cuba, using personal computers. *Cuban Meteorology Journal* 10, 84-94.

Myers-Cook, T., Mallard, J., Mao, Q., 2010. Development of a WRF-AERMOD tool for use in regulatory applications. *Proceedings of 16th Conference on Air Pollution Meteorology (TVA*), January 17-21, 2010, Atlanta, USA.

NC, 1999. NC 93-02-202/1987: CUBAN STANDARDS. ATMOSPHERE. Health and Sanitary Requirements: Maximum admit table conditions, expulsion minimum height and sanitary protection zones. NC 39:1999: Air quality, Health and Sanitary Requirements, AMENDMENT 1.

NOAA (National Oceanic and Atmospheric Administration), 2012. http://www.esrl.noaa.gov/raobs/intl/fsl_format–new.cgi, accessed in May 2012.

Perez, C., Jimenez, P., Jorba, O., Sicard, M., Baldasano, J.M., 2006. Influence of the PBL scheme on high-resolution photochemical simulations in an urban coastal area over the Western Mediterranean. *Atmospheric Environment* 40, 5274-5297.

Pielke, R.A., 1984. *Mesoscale Meteorological Modeling.* Academic Press, Orlando, 612 pages.

PSU/NCAR, 2010. User's Guide for the Advanced Research WRF (ARW) Modeling System Version 3.1, Boulder, 312 pages.

Randolph, E., 2002. Using mesoscale model data as input to AERMOD. *Proceedings of 12th Joint Conference on the Applications of Air Pollution Meteorology with the Air and Waste Management Association*, May 19-23, 2002, USA, 2 pages.

Russell, A., Dennis, R., 2000. NARSTO critical review of photochemical models and modeling. *Atmospheric Environment* 34, 2283-2324.

Rodriguez, E., Morris, C.S., Belz, J.E., Chapin, E.C., Martin, J.M., Daffer, W., Hensley, S., 2005. An Assessment of the SRTM Topographic Products. Technical Report JPL D-31639, Pasadena, California, 143 pp.

Scire, J.S., Strimaitis, D.G., Yamartino, R.J., 2000a. User's Guide for the CALPUFF Dispersion Model (Version 5.0), Concord, 521 pages.

Scire, J.S., Robe, F.R., Fernau, M.E., Yamartino, R.J., 2000b. User's Guide for the CALMET Meteorological Model (Version 5.0), Concord, 332 pages.

Seaman, N.L., 2000. Meteorological modeling for air-quality assessments. *Atmospheric Environment* 34, 2231-2259.

Seinfeld, J.H., Pandis, S.N., 2006. *Atmospheric Chemistry and Physics: from Air Pollution to Climate Change*, Wiley, New York, 1225 pages.

Spadaro, J.V., 1999. *Quantifying the Damages of Airborne Pollution: Impact Models, Sensitivity Analyses and Applications*. Ph.D. Thesis, Ecole des Mines de Paris, Centre d'Energetique, France, 130 pages.

Stenger, R.A, 2000. *Sensitivity Studies on a Limited Area Mesoscale Model: An Examination of Lateral Boundary Placement, Grid Resolution and Nesting Type.* MSc Thesis, Air Force Institute of Technology, Ohio, 193 pp.

Titov, M., Sturman, A., Zawar-Reza, P., 2005. Application of MM5 to study of air pollution in Christchurch, New Zealand – some problems of using MM5 with global analysis data. *Proceedings of 6th WRF/ 15th MM5 User´s Workshop*, June 27-30, 2005, USA.

TRC, 2008a. CALMM5 software. TRC Environmental Corporation.

TRC, 2008b. CALWRF software. TRC Environmental Corporation.

Zawar-Reza, P., Kingham, S., Pearce, J., 2005. Evaluation of a year-long dispersion modelling of PM_{10} using the mesoscale model TAPM for Christchurch, New Zealand. *Science of the Total Environment* 349, 249-259.