Multi-criteria selection of an Air Quality Model configuration based on quantitative and linguistic evaluations

V.H. Almanza *, I. Batyrshin, G. Sosa

Instituto Mexicano del Petróleo, 07730 México D.F., Mexico

A R T I C L E   I N F O

Keywords:
Multi-criteria evaluation
Pareto Fronts
Air quality
WRF-Chem

A B S T R A C T

This study presents the application of multi-criteria evaluation in the selection of an optimal configuration for an Air Quality Model. The simulation domains focus on the Mexico City Metropolitan Area. A set of 10 different configurations were considered as alternatives. These configurations included convective parameterization, 6th order diffusion and exclusion of data assimilation within the Planetary Boundary Layer. In addition, model integration in a continuous setup and in a segmented setup was also considered. The modeling variables were surface temperature, wind speed, wind direction, and sulfur dioxide. The performance of the meteorological fields was evaluated with statistical metrics together with the Local Trend Association measure and further used as criteria. The air pollution field was evaluated qualitatively with five expert-based linguistic criteria and further converted into numerical terms. Meteorological variables and sulfur dioxide were aggregated into two single arrays, one for representing meteorology and the other for representing transport of a large industrial plume. Pareto Fronts were constructed for these two arrays under different weights scenarios. Results suggested that a model with no parameterizations in continuous integration setup and a model with segmented integrations using 6th order diffusion were the optimal configurations to conduct future air quality studies.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

It has long been recognized the negative impacts of air pollution on human health and ecosystems. The role of anthropogenic emissions is at the forefront of these impacts (WMO, 2012). To effectively address the air pollution at several spatial scales, air quality modeling is of utmost relevance. Air quality forecasting uses source and receptor models to study the degree of pollution in future events. There are several approaches to model air pollution, and all of them rely on the scientific goals under study which include the spatial scale of interest. In this respect, expert-based approaches are important in air quality studies. For example: in forecasting air pollution with fuzzy time series methods (Domboska & Wojtlyak, 2012); classifying concentrations of criteria pollutants in Mexico (Barrón-Adame, Cortina-Januchs, Vega-Corona, & Andina, 2012); for controlling emissions (Zhou, Huang, & Chan, 2004), identification of geographical distribution of particle pollution (Li & Shue, 2004); or in suggesting a system for air quality management at regional scale based on interval programming with stochastic variables (Cao, Huang, & He, 2011).

Another approach consists in the application of deterministic models which solves the conservation equations of different physical and chemical processes (Zhang, Bocquet, Mallet, Seigneur, & Baklanov, 2012). These models are known as Air Quality Models (AQM) or Chemical Transport Models (CTM). This kind of models combines in a systematic approach current knowledge of meteorology, emissions and atmospheric chemistry to make forecasts of ambient concentrations (Molina et al., 2004). In addition, they provide scientific understanding of pollutant processes; can address and track, both in space and time, the long range transport of air pollutants; and can help in formulating scenarios involving changes in meteorology or emissions (Zhang et al., 2012). In Mexico, air quality studies using this type of models have been conducted in recent years focusing on different research objectives and on different kind of pollutants: sensitivity of ozone formation to emission changes (Lei, de Foy, Zavala, Volkamer, & Molina, 2007); assessment of possible underestimation of volatile organic compounds in the official National Emissions Inventory (West et al., 2004) and estimation of source contribution of large sulfur dioxide plumes (de Foy et al., 2009) among others.

Meteorological fields, including wind, hourly temperature, mixing depth and solar insolation fields are an important input for any modeling exercise with Air Quality Models (Russell & Dennis, 2000). These fields can have great uncertainty which contribute...
in mispredicting airborne chemical species, aerosols and particulate matter (Seaman, 2000). Thus, accurate meteorological fields in a CTM are of utmost importance. They will lead to reliable forecasts of air pollution events.

Multi-criteria evaluation is suitable for addressing complex problems involving high uncertainty, conflicting objectives and different data among other features (Wang, Jing, Zhang, & Zhao, 2009). It selects the best choice in a set of potential alternatives (Dalalah, Hayajneh, & Batieha, 2011). Since it is unlikely to achieve optimal values in all objectives at the same time, the final solution is a compromise of the initial alternatives in the selection process (Behnamian, Ghomi, & Zandieh, 2009). In environmental sciences it has been used in the development of a support system to help planning water resources management in the Haihe river (China) (Weng, Huang, & Li, 2010) and in the estimation of the optimum number of parameters for the coupled system of a multi-layer perceptron and the Numerical Weather Prediction (NWP) model HIRLAM (Niska et al., 2005). In this respect, the best configuration for either a NWP or a CTM cannot be known a priori, even though there are recommended parameterizations which produce reliable results. Therefore, sensitivity experiments are conducted in order to choose the best configuration among different physical parameterizations (Borge, Alexandrov, del Vas, Lumbreras, & Rodríguez, 2008; Lo, Yang, & Pielke, 2008). However, not all the possible combinations are feasible to be evaluated since some parameters are specific for some modeling options. Although sensitivity studies usually focus on the meteorological fields, there are works which additionally assess the potential transport of pollutants for each configuration. Trajectory calculations are computed in order to evaluate the meteorological fields (Godowitch, Gilliam, & Rao, 2011; Ngan et al., 2012; Seaman, 2000). This allows depicting the degree of uncertainty in the meteorological fields. Nevertheless, the plume transport is not considered in these works as an extra criterion to evaluate model performance.

The main objective of this study is to select an optimum configuration for an AQM based on multi-criteria evaluation. The main motivation is to have a trade-off between meteorology and plume transport in order to conduct a regional air quality simulation focusing on the Mexico City Metropolitan Area (MCMA). Section 2 presents the AQM and the area of study. The methodology is presented in Section 3. Results and discussion are given in Section 4, and conclusions of this study are summarized in Section 5.

2. Air Quality Model and area of study

2.1. WRF-Chem

Detailed Eulerian numerical Air Quality Models have been developed for research purposes and to support emissions-control policy decisions (Seaman, 2000). An example is the WRF-Chem model. It is an on-line chemistry model fully coupled to the Weather Research and Forecasting (WRF) model (Skamarock et al., 2005) with lead developments at the National Oceanic and Atmospheric Administration (NOAA) and the Pacific Northwest National Laboratory (Fast et al., 2006; Grell et al., 2005). In this respect, the WRF model is a state-of-the-art mesoscale system designed for both short- and long-term weather and climate simulations. It is a non-hydrostatic model, with different physical parameterizations, including cumulus parameterizations, Planetary Boundary Layer, land surface models, long-wave and short-wave radiation, and microphysics models. It also includes multi-scale data assimilation. Thus, the WRF-Chem preserves the transport, grid and physics schemes of WRF's meteorological component when solving the transport in sub-grid scales. It has different chemical mechanisms, as well as several schemes of aerosol and photolysis (WRF-Chem, 2010; WRF, 2010). WRF-Chem version 3.2.1 is used in the present study.

2.2. Tula region and MCMA

The city of Tula is located in the Mezquital Valley, in southwest Hidalgo with a total population of nearly 94,000 inhabitants and more than 140 industries. The region is semi-arid with average temperatures of 17 °C and precipitation ranging from 432 to 647 mm, increasing from north to south. In this region, the Tula Industrial Complex is settled in an area of 400 km². The major industries of the city are located within this region, including the Miguel Hidalgo Refinery, the Francisco Perez Rios power plant, several cement plants and limestone quarries. Other minor industries include metal manufacturing, processed food, chemical and incineration of industrial waste. The emission of pollutants from combustion processes of these industries impacts the regional air quality. In addition, the inflow of untreated sewage water from Mexico City promotes severe pollution problems to soil and water resources (Cifuentes, Blumenthal, Ruiz-Palacios, Bennett, & Peasey, 1994; Vazquez-Alarcon, Justin-Jajuste, Siebe-Grabach, Alcantar-Gonzalez, & de la Isla-de-Bauer, 2001). According to current environmental regulations, this region is classified as a critical area due to the high emissions of SO₂ and particulate matter (SEMARNAT-INE, 2006).

The MCMA is the largest megalcity in North America, and the third largest urban agglomeration, with nearly 22 million inhabitants, after Tokyo (Japan) and Delhi (India) (UN, 2012). It is located in the subtropics within an elevated U-shaped basin surrounded by mountain ridges which border the west, east and south regions of the city. The metropolitan area covers 1500 km² on the southwestern side (Parrish, Singh, Molina, & Madronich, 2011; Williams, Brown, Cruz, Sosa, & Streit, 1995). This topography acts like a barrier to large-scale circulations and isolates the basin from the winds of synoptic weather systems at low levels (Zhang & Dubey, 2009). Previous work has shown that emissions from Tula Industrial Complex can impinge into the megacity, exerting an influence on sulfur dioxide pollution levels in the city (de Foy et al., 2009; Williams et al., 1995; Almanza, Molina, & Sosa, 2012).

3. Air Quality Model configurations

A set of 10 different WRF-Chem configurations were considered in this study. Each model run encompass a 6-day simulation period, from 00:00 UTC 22 March to 00:00 UTC 28 March of 2006, using three domains with horizontal resolution of 27, 9 and 3 km, and 35 vertical levels (Fig. 1a). All of the simulations performed Multiscale Four Dimensional Data Assimilation (FDDA) to improve accuracy in the meteorological fields and keep the same physics to maintain consistency. Operational meteorological data, Mexico City Air Quality Network (RAMA) surface data; and radar wind profilers, radiosondes and surface data from MILAGRO field campaign were used in the FDDA assimilation process. MILAGRO field campaign was a major international collaborative project to examine the behavior and export of atmospheric emissions from the Mexico megacity (Molina et al., 2010).

The simulations considered two of the main approaches to run an AQM. The first one of them is a single continuous integration requiring an initialization of meteorology and chemistry just at the beginning of the run. The second one is a multiple integration of overlapping short segments requiring multiple initializations at the beginning of each segment run. Even though they appear similar, final results may differ. In addition, computing time and disk storage are important features to take into account when running the AQM in either of these approaches. Segmented runs tend to require relatively more disk space than a single continuous run.
Previous work obtained reliable results of air quality simulations for the MCMA using some of the most common modeling options (Almanza et al., 2012; Rivera et al., 2009). Thus, rather than performing an exhaustive sensitivity test of all the modeling options, the present study just considered (see Table 1) the inclusion of convective parameterization in the third domain (d03), 6th order horizontal diffusion in the first two domains (d01 and d02), and inclusion of grid nudging of the wind field within the Planetary Boundary Layer (PBL) also in the first two domains with the purpose of increasing the accuracy of future air quality simulations for this specific modeling period. Grid nudging is a kind of Newtonian relaxation which is ideal for assimilating synoptic data covering most or all the model domain (Seaman, 2000). M1 to M10 denote the corresponding set of the 10 WRF configurations. For example, M2 included convective parameterization under a continuous integration setup; whilst M7 included convective parameterization under a segmented integration setup. Both M1 and M6 did not include any of the studied parameterizations and just differ in how they were integrated.

4. Model performance

4.1. Quantitative evaluation

The performance of the meteorological fields obtained with WRF-Chem was assessed by means of the following measures: Root Mean Squared Error (RMSE):

\[ m_1(P,O) = \left( \frac{1}{n} \sum_{i=1}^{n} (P_i - O_i)^2 \right)^{1/2} \]

the mean bias (BIAS):

\[ m_2(P,O) = \frac{1}{n} \sum_{i=1}^{n} (P_i - O_i) \]

and the Index of Agreement (IOA) which is designed to better handling differences in predicted and observed means and variances (Willmott et al., 1985)

\[ m_3(P,O) = 1 - \frac{\sum_{i=1}^{n} (P_i - O_i)^2}{\sum_{i=1}^{n} (P_i - \bar{O})^2 + \sum_{i=1}^{n} (O_i - \bar{O})^2} \]

where \( P_i \) denotes model predictions, \( O_i \) experimental observations and \( \bar{O} \) denotes the average of observations for the total simulation data \( n \).

The meteorological observations from 15 monitoring stations including the Mexico City RAMA monitoring network and the surface measurements from MILAGRO campaign were used to compare model results. The model surface variables considered for this purpose were the temperature at 2 m above ground level (T), wind speed (WS) at 10 m above ground level and wind direction (WD). With respect to the wind field, the RMSE of the vector wind difference (RMSEvec) was calculated. This statistic considers both speed and direction errors (Fast, 1995).
\[ m_4(P, O) = \left[ (u_P - u_0)^2 + (v_P - v_0)^2 \right]^{1/2} \]

where \( u_P, v_P \) are the predicted and \( u_0, v_0 \) are the observed zonal and meridional wind components respectively. In this study the air quality simulations were run just in the third domain (d03) to save computing time. Thus, the performance was evaluated only for the innermost domain.

Additionally to the statistical evaluation methods the Local Trend Association (LTA) measure based on Moving Averaging Transformation (MAT) ([Batyrsyn, Herrera-Avelar, Sheremetov, & Panova, 2007]) was used for the performance evaluation of the model configurations. Unlike statistical evaluation measures LTA measures a similarity in dynamics of time dependent variables given in measurements and in models. MAT transforms a time series \( y = (y_1, \ldots, y_n) \) into a sequence \( \text{MAT}_i(y) = (a_1, \ldots, a_{n-k+1}) \) of slopes of least squares linear regressions \( f_i = a_i t + b_i \) of time series \( (y_1, y_2, \ldots, y_{r+k-1}) \) in a sliding window of size \( k \in \{2, \ldots, n\} \). These slope values are calculated by Batyrsyn et al. (2007):

\[
a_i = \frac{6(2 - k + 1) y_{i+k}}{k(k^2 - 1)}, \quad i = 1, \ldots, n-k+1.
\]

Local Trend Association between time series \( P \) and \( O \) is calculated for window size \( k \) by

\[
\text{LTA}_k(P, O) = \cos(\text{MAT}_i(P), \text{MAT}_i(O)).
\]

This measure was used in Almanza and Batyrsyn (2011) for analysis of associations between atmospheric pollutants and meteorological variables in MCMA. In the present study, we used an aggregation of LTA obtained for window sizes \( k = 2 \) and \( k = 4 \):

\[
M_5(P, O) = \max(\text{LTA}_2(P, O), \text{LTA}_4(P, O)).
\]

### 4.2. Qualitative evaluation

Unlike the meteorological fields, the model performance of the regional plume transport for each configuration was assessed qualitatively in terms of SO2. Five features were considered for this purpose: the dynamics on 22 March (d22), the timing of the concentration peak on 23 March (p23), the magnitude of the peak on 23 March (m23), the concentration on the early morning on 24 March (mo24), and the peak of late night and early morning of 24 and 25 March (p2425) respectively. Each model of the set of 10 configurations was assessed on the basis of previous results and with expected behavior of the model ([Almanza et al., 2012; de Foy et al., 2009; Rivera et al., 2009]). One of the most rigorous tests of model performance is the reproduction of the peak concentration in the grid cell containing the measurement location at the same hour as it is measured, as well as the phasing of the peak, which would be within ±1 h ([Seinfeld, 1988]). For this purpose, observations from 25 monitoring stations including RAMA and MILAGRO were used to analyze plume transport from Tula to the MCMA. It should be noted that some stations which monitor meteorological variables do not monitor sulfur dioxide and vice versa (Fig. 1b). At this stage, just the emissions from Tula are considered in order to simplify the analysis. The aim is to select an optimum configuration for future air quality simulations.

### 5. Multi-criteria evaluation

#### 5.1. Quantitative criteria: meteorology

The evaluation of a set of model configurations taking the statistical quantities (RMSE, IOA, BIAS, RMSEvec, LTA) as the criteria \( j = 1, \ldots, 5 \) and the model configurations as the alternatives \( \{M_1, \ldots, M_{10}\} \) ([Ng, 2008]) was performed for the aggregated meteorological variables T, WS and WD. The aggregation consisted in taking the arithmetic average of measurements from 15 monitoring stations previously mentioned in Section 4.1. The median and the lexicographical minimum were considered as aggregation methods but they tended to exclude the outliers (not shown). Each model configuration \( M_{i-1}, \ldots, M_{10} \) was evaluated by converting the statistical quantities under all criteria into a single performance measure for each meteorological variable \( PM_{ijkl} \). The matrix \( MP_{ij} \) \( \{i = 1, \ldots, 10; j = 1, \ldots, 5\} \) represents the performance of \( M_{ij} \), that is the model configuration \( M_i \) under criterion \( j \), normalized in a 0–1 scale. The normalization used the following linear transformation \( MP_{ij} = \frac{\text{max}_{i=1,10}(M_{i1}) - \text{min}_{i=1,10}(M_{i1})}{\text{max}_{i=1,10}(M_{i1}) - \text{min}_{i=1,10}(M_{i1})} \). Thus, the performance measure of a model configuration for each meteorological variable was expressed as the weighted sum \( PM_{ik} = \sum_{j=1}^{5} \omega_j MP_{ij} \), where \( \omega_j \) is the weight of criterion \( j \) of model configuration \( M_i \) such that \( \sum_{j=1}^{5} \omega_j = 1 \). Subsequently, the performance measures of all meteorological variables were further aggregated into a single performance array \( MET_i \) by taking their weighted sum \( \sum_{k=1}^{3} W_k PM_{ik} \), where \( W_k \) is the weight of meteorological variable \( k \) for performance measure \( PM_{ik} \), such that \( \sum_{k=1}^{3} W_k = 1 \). In this study, wind direction is relatively more important than wind speed and temperature. Thus, a slightly higher weight was assigned to wind direction. If ozone simulations were to be conducted, temperature would be as important as wind direction and similar weights could be assigned. In this work, however, the analysis just considers the gas phase chemistry of sulfur dioxide emissions released by large point sources.

In order to examine the influence of the statistical criteria in the configuration selection, a sensitivity test on criteria weights was conducted. It consisted in assigning equal weights to all statistical quantities (S1), moderately higher weights to some of them (S2), relatively equal weights but one higher than the rest (S3–S6) and considering just one of the statistical quantities (S7–S10) as shown in Table 2. The purpose was to represent the relative importance of each criterion on the selection process and to further investigate how the selection could be influenced if considering just one statistical performance measure. Thus, a single performance array \( MET_i \) was obtained for each sensitivity test.

<table>
<thead>
<tr>
<th>Statistical measure</th>
<th>Sensitivity test</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE/RMSEvec</td>
<td>S1</td>
</tr>
<tr>
<td>IOA</td>
<td>0.25</td>
</tr>
<tr>
<td>BIAS</td>
<td>0.25</td>
</tr>
<tr>
<td>LTA</td>
<td>0.25</td>
</tr>
</tbody>
</table>
5.2. Linguistic criteria: SO2 plume transport

With respect to sulfur dioxide, the model performance was evaluated subjectively by means of linguistic measures. The criteria in this case consisted in the five model features previously mentioned in Section 4.2. The linguistic attributes used for this purpose were {Excellent, Very Good, Good, Fair, Poor}, according to how the model performed for each feature. The subjective evaluation was preferred since at this stage only one emission source, the Tula Industrial Complex, was considered. Thus, the emissions from Mexico City, the Popocatepetl volcano and surrounding regions were excluded, and a quantitative evaluation would not be representative for the performance of air pollutants. Thus, the evaluation of model configurations consisted in taking the model features \{d_{22}, 23, m_{23}, m_{24}, p_{2425}\}, as the criteria and the model configurations as the alternatives \{M_1, ..., M_{10}\}. Similar to the meteorological variables, each model configuration \(M_{i = 1,...,10}\) was evaluated by converting the linguistic attributes \(l_i\) under all model features \(j_i\) into a single performance measure for sulfur dioxide \(PM_{v_{1,...,10}}\). For this purpose, the linguistic attributes \(l_i\) were converted into numerical values in a 0–100 scale. The procedure to assign numerical values considered the priority among the model features. In the present study, the most important model features were the timing of the peak on 23 March (p23) and the magnitude of early morning concentration on 24 March (mo24). Thus, these model features were ranked with higher (H) priority, whilst the other ones with low (L) priority. In order to maintain consistency between high and low priority model features, the numerical assignment was split into two sets, one for each type of priority. This is presented in Table 3. For instance, the highest value (100) was assigned to the linguistic attribute \{Excellent\} with high priority features \(EH\); whilst a similar value (75) was assigned to both the linguistic attribute \{Good\} with high priority features and to the linguistic attribute \{Excellent\} with low priority features.

The numerical assignment was done for each of the 25 monitoring stations. Like for the meteorological variables, the weighted performance of SO2 for all stations was further aggregated into the single performance measure array SD by taking the arithmetic average of the resulting numerical assignments of the monitoring stations and subsequently normalized into a 0–1 scale. With these
two performance arrays \( MET_i \) and \( SD_i \), one for the meteorology and one for sulfur dioxide. The Pareto Front of model configurations was constructed. The best two models were selected by counting the number of dominated models for each point in the Pareto Front. This procedure was performed for each case of the weights sensitivity test \( S_1, \ldots, S_{10} \). The configurations with higher dominance were selected as the best model candidates.

6. Results and discussion

6.1. Pareto Fronts: optimal configurations

The Pareto Fronts of the three sensitivity cases \( S_1, S_5 \) and \( S_{10} \) are presented in Fig. 2. It shows the best models when the quantitative evaluation of the meteorological variables were assumed of equal importance (upper panel); when considering that the BIAS was slightly more important for assessing performance (mid panel) and the case when assuming that LTA was the most important performance criterion (lower panel). In all cases, model configurations \( M_1, M_2 \) and \( M_8 \) were suggested as reliable candidates to conduct an air quality simulation. In the first case \( S_1 \), the configuration \( M_1 \) dominated over four configurations. It is denoted \( M_1 \sim \{M_3, M_5, M_6, M_{10}\} \). The configuration \( M_8 \) also dominated over four configurations \( M_8 \sim \{M_3, M_4, M_5, M_{10}\} \), and the configurations \( M_2, M_7 \) and \( M_9 \) only dominated over two configurations. However, if giving the bias slightly more priority (case \( S_5 \)) similar results to case \( S_1 \) were obtained, except that configuration \( M_8 \) only dominated over three models instead of four as in case \( S_1 \). When comparing \( M_8 \) with respect to \( MET_i \), for \( S_1 \) and \( S_5 \), it can be seen that in \( S_5 \) case, \( M_8 \) presented slightly lower performance in meteorology, implying slightly higher bias in wind direction with respect to \( M_1 \).

According to LTA, the configurations which better represented the dynamics of the observed time series were \( M_1 \) and \( M_2 \), each with four dominated configurations. An interesting finding regarding configuration \( M_2 \) is that even though it presented slightly lower performance in meteorology with respect to the statistical measures, it had the best performance in terms of plume transport. In addition, it represented relatively well the dynamics of both meteorology and transport of sulfur dioxide from Tula region. This suggests that the dynamics were better reproduced with the continuous integration rather than with multiple integrations of segmented runs. Table 4 shows the number of dominated models for the remaining cases of the sensitivity analysis. It only shows model configurations with the highest degree of dominance, so that a zero means that for the respective sensitivity case, those particular configurations were not the highest ones. It is clear that configurations \( M_1 \) and \( M_8 \) were the best candidates for conducting an air quality simulation, since they had the greatest dominance among all the model configurations. \( M_8 \) in particular. However, \( M_2 \) could give reliable results given its performance in representing the dynamics of plume transport.

---

**Fig. 3.** Regional transport of SO\(_2\) plume form the Tula Industrial Complex for the model configurations \( M_1 \) (yellow) and \( M_8 \) (purple dashed) on 23 March at 11:00 LST (left panel), and 24 March at 05:00 LST (right panel). The dots depict the location of the monitoring stations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

**Fig. 4.** Sulfur dioxide time series at Tultitlan (TLI) monitoring station of model configurations \( M_1 \) (blue), \( M_2 \) (orange) and \( M_8 \) (cyan). All times are in Local Standard Time (LST). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
6.2. Regional transport: influence of model configuration

Fig. 3 shows the spatial distribution of the SO2 plum from the Tula Industrial Complex. This Figure depicts the subtle but important differences in the regional transport between the best model candidates. On 23 March at 11:00 LST the modeled plume transport is very similar for configuration M1 (yellow) and configuration M8 (purple dashed), but with a slight delay and lower spread for M1 (Fig. 3, left). In contrast, on 24 March at 05:00 LST configuration M1 presented a wider penetration in the northwest region of the basin even reaching part of the southwest: whilst M8 prevailed mostly on the western ridge of the basin suggesting no plume impingement at this hour (Fig. 3, right).

The time series at Tultitlan monitoring station (TLU) of these model configurations are presented in Fig. 4 to better depict the contribution of model parameterizations in the final plume transport. It also includes the results obtained with M2 configuration. Although M8 overpredicted the magnitude of SO2 concentration on 22 March it reproduced the concentration in the early morning on 24 March better than M1. In addition M1 and M8 reproduced the timing of the peak on 23 March, whilst M2 predicted it one hour earlier. Moreover, the early morning peak on 25 March was better reproduced with M8 in both magnitude and timing, since M1 and M2 presented higher overprediction.

In order to support these results, the best two models of this study for each sensitivity case were compared with the best models obtained with the method of Ng (2008). This method ranks the criteria in descending order of importance. In this study we ranked the modeling variables as $r_v = [SO_2, WD, WS, T]$ in that order. This resulted in a matrix containing the aggregated performance measures of each variable $[PM_{E_{x,i}(SO_2)}, PM_{E_{x,i}(WD)}, PM_{E_{x,i}(WS)}, PM_{E_{x,i}(T)}]$. Results are presented in Table 5.

Ng method showed a great similarity with the suggested modeling configurations, in particular M1 and M8. Nevertheless, when all statistical quantities have the same importance (S1), Ng method could not give representative results. The main differences resulted when RMSE (S3) and LTA (S6) were slightly more important in the evaluation of model performance. In this approach, however, M2 was reproduced relatively well the dynamics of the observed time series (S10). Thus, M1 and M8 were selected as the best alternatives for conducting a full air quality simulation in this specific simulation domain. They correspond to the simplest case (M1) and to a segmented integration with 6th order diffusion (M8). This suggests that convective parameterization would not be required. However, M2 included this parameterization and performed relatively well with respect to the dynamics. Ignoring data assimilation within the Planetary Boundary Layer resulted in the worst configurations in terms of both meteorology and plume transport (M2 and M10). Another aspect that is important in practical terms concerns to the portability, disk storage and scripting between M1 (continuous integration) and M8 (segmented integration). The results suggests that a single run could be as accurate as a segmented run, which would require relatively less disk storage, less scripting, increased portability, and slightly lower computing time. Scripting refers to the required code to pre-process, automate, and post-process the model inputs and outputs. For instance, in their numerical experiments (Lo et al., 2008) analyzed the influence of applying nudging in short-segments re-initializations for regional climate downscaling, and found that their results were the same as a continuous run with nudging of their experiments. They suggested that it could improve the portability of their experiments. Even though these practical aspects were not included in the procedure to select the best model configurations, it would be feasible to also weight model configurations with these practical criteria.

7. Conclusions

The selection of the optimal model configurations for an Air Quality Model using Multi-Criteria Evaluation was investigated. The main motivation to apply this approach is because a preliminary simulation gave relatively good results so that a more accurate run could be used in future studies. For this purpose some of the most widely used statistical metrics for evaluating model performance were used as the criteria. A set of 10 different model configurations were used as the alternatives. In addition to the statistical metrics, an approach based on Local Trend Associations was also applied with the aim of evaluating the dynamics of the model configurations. The results suggested that the most simple model configuration in continuous integration (M1) and the model configuration with 6th order diffusion in multiple-integrations (M8) were the optimal configurations to conduct future air quality studies for the simulation period from 22 March 2006 to 27 March 2006. These configurations were Pareto optimal and prevailed after performing a sensitivity test to the weights of the criteria. In addition the application of the method of Ng (2008) suggested similar configurations. The criterion using Local Trend Associations provided insights of which configurations better reproduced the dynamics of meteorological variables, providing supporting information for selecting potential model configurations probably overlooked. Multi-criteria evaluation provided a more rigorous approach in the selection process of an optimum configuration for an Air Quality Model in a regional scale.

Acknowledgments

The authors thank Dr. Ed Scott Wilson Garcia for his help with MPICH. In addition, the experimental data of the MILAGRO campaign is greatly acknowledged.

References


