# How bias-correction can improve air quality forecast over Portugal

C. Borrego[1], A. Monteiro[1*], M. T. Pay[2], I. Ribeiro[1], A.I. Miranda[1], S. Basart[2], J. M. Baldasano[2,3]

[1]CESAM, Department of Environment and Planning, University of Aveiro, 3810-193 Aveiro, Portugal.

[2]Earth Science Department, Barcelona Supercomputing Center, Jordi Girona 29, Edificio Nexus II, Barcelona, Spain

[3]Environmental Modeling Laboratory, Technical University of Catalonia, Barcelona, Spain

*Corresponding author: alexandra.monteiro@ua.pt, Tel: +351 234370220, Fax: +351 234 370309

## Abstract

Currently three air quality modelling systems routinely operate with high resolution over mainland Portugal for forecasting purposes, namely MM5-CHIMERE, MM5-EURAD and CALIOPE. Each of one operates daily using different horizontal resolutions (10 km x 10 km; 5 km x 5 km and 4 km x 4 km, respectively), specific physical and chemical parameterizations, and their own emission pre-processors (with common EMEP emission database source, but different spatial disaggregation methodologies). The operational BSC-DREAM8b model is offline coupled within the aforementioned air quality systems to provide Saharan dust contribution to particulate matter. Bias-correction studies have demonstrated the benefit of using past observational data to reduce systematic model forecast errors. The present contribution aims to evaluate the application of two bias-correction techniques - multiplicative ratio and Kalman filter in order to improve the air quality forecast over Portugal. Both techniques are applied to the three modelling systems over the full year 2010. Raw and unbiased model results for the main atmospheric pollutants ($O_3$, $NO_2$, $SO_2$, PM10 and PM2.5) are analysed and compared against 18 monitoring stations distributed within inland Portugal in an hourly basis. Statistical analysis shows that both bias-correction techniques improve the raw forecasts skills (for all the modelling systems and pollutants). In the case of $O_3$ max-8h, correlation coefficients improve in 19-45 %, from 0.56-0.81 (raw models) to 0.78-0.86 (corrected models). PM2.5 also present significant improvements, e.g., correlation coefficients increase in more than 50% (both techniques) reaching values between 0.50-0.64. The corrected primary pollutant $NO_2$ and $SO_2$ demonstrate significant relative improvements compared to $O_3$, mostly because the original modelling system skills are lower for those species. Despite the applied techniques have different mathematic formulation and complexity level, there are comparable answers for all of the forecasting systems. Analysis performed over specific situations, such as air quality episodes, not-validated or missing data reveals different behaviour of the bias-correction techniques under study. The results confirm the advantage of the application of bias-correction techniques for air quality forecast. Both techniques can be applied

routinely in an operational forecast system and they will be useful to alerts for the population about accurate exceedances.

**KEYWORDS:** air quality forecast, modelling systems, bias correction, multiplicative ratio, Kalman filter.

## 1. INTRODUCTION

Air quality forecasting is both a challenge and a scientific problem, which has recently emerged as a major priority in many urbanised and industrialised countries due to the increasing consciousness of the effect, on health and environment, caused by airborne pollutant emissions. Furthermore, is one of the requirements of the Air Quality Framework Directive (2008/50/EC) and a key issue of the Clean Air for Europe (CAFE) Programme (Cuvelier et al., 2007). The goals of reliable air quality forecasts are obvious: population exposure can be more efficiently reduced and protected by means of information and short-term action plans.

For that, European legislation settled ambient air quality standards for acceptable levels of air pollutants (like $O_3$, $NO_2$, $SO_2$, PM2.5 and PM10) and also recommended the use of modelling tools to assess and to forecast the air quality, in order to develop emission abatement plans and alert the population when health-related issues occur. In some European member states, like Portugal, air pollution limit values, namely for PM10 and ground-level $O_3$, are being exceeded every year and during long-term periods (Monteiro et al., 2007a; Carvalho et al., 2010; EEA, 2010).

Several operational air quality forecasting systems already exist over Europe (see http://gems.ecmwf.int or http://www.chemicalweather.eu). Some of them forecast at the national level as in Portugal. In particular the MM5-CHIMERE (Monteiro et al., 2005), the MM5-EURAD-IM (Elbern et al., 2007; Strunk et al., 2010) and the CALIOPE (Baldasano et al., 2008a) forecasting systems are advancing our understanding of atmospheric dynamics in Portugal as follows. First, they are applied with a higher resolution over Portugal. Meanwhile most European models use a horizontal resolution of at least 25 x 25 $km^2$, the MM5-CHIMERE, the MM5-EURAD-IM and the CALIOPE systems use horizontal resolution of 10x10 $km^2$, 5x5 $km^2$ and 4x4 $km^2$, respectively. Second, they include contribution of Saharan dust emissions on an hourly basis from the BSC-DREAM8b model. Third, there are several evaluation studies that support the confidence on the three selected systems (MM5-CHIMERE in Monteiro et al. 2007a,b; MM5-EURAD-IM in Monteiro et al., 2011; and CALIOPE in Baldasano et al., 2008a, 2011 and Pay et al., 2011).

Air quality forecast modelling, which rely not only on the meteorological prediction but also on a chemical-transport modelling and on highly uncertain emission inventories, are likely to have significant (systematic) model errors (Borrego et al., 2003, 2008; Chang and Hanna, 2004). In order to improve each model forecast skill, different bias-correction techniques have been recently applied and examined (McKeen et al., 2005; Wilczak et al., 2006; Pagowski et al., 2006; van Loon et al., 2007; Djalalova et al., 2010; Sicardi et al., 2011).

The objective of the present study is to examine the efficacy of two bias-correction techniques, multiplicative ratio and Kalman filter methods, to improve the air quality forecasts (ground-based concentrations of $O_3$, $NO_2$, $SO_2$, PM10 and PM2.5) calculated from the three operational modelling systems available at high resolution over Portugal mainland domain. The model evaluation exercise covers the full year 2010 and observation from 18 air quality monitoring stations.

The present work is organized as follows. Section 2 describes the different forecast modelling systems. Section 3 presents the observational dataset selected and used within this study. The applied bias techniques are described in Section 4 and the analysis and discussion of the results are presented in Section 5. In Section 6, classical/categorical statistics are addressed to investigate the forecast skills after bias correction. Finally, the conclusions are drawn in Section 7.

## 2. THE AIR QUALITY FORECASTING SYSTEMS

There are three air quality forecasting systems operating over Portugal with high resolution. Both MM5-CHIMERE (Monteiro et al., 2005) and MM5-EURAD-IM (Elbern et al., 2007; Strunk et al., 2010) modelling systems are being applied by the University of Aveiro's research group using an European/Iberian Peninsula coarse domain as boundary and initial conditions for the nested domain over Portugal with a 10x10 $km^2$ and a 5x5 $km^2$ horizontal resolution, respectively. The MM5-CHIMERE modelling system is operational with daily forecasts available since 2007: http://adamastor.dao.ua.pt/previsao_qar/. The MM5-EURAD-IM is only operational for Portugal since 2010, with also daily forecasts in an hourly basis, as a result of a scientific collaboration between the University of Aveiro and the Rhenish Institute for Environmental Research at the University of Cologne. The CALIOPE system (Baldasano et al., 2008a) provides high-resolution air quality forecast for Spain. CALIOPE system encompasses a set of models: WRF-ARW meteorological model, the High-Elective Resolution Modelling Emission System (HERMES, Baldasano et al., 2008b) and the chemical transport model CMAQ. CALIOPE is applied over Iberian Peninsula with a 4 x 4 $km^2$ horizontal resolution and also with an hourly basis (Baldasano et al., 2011). Current forecasts and near real-time evaluation are available through the CALIOPE system website (http://www.bsc.es/caliope).

CMAQ, CHIMERE and EURAD-IM are all regional-scale three-dimensional chemical transport models (CTM) designed for short-term and long-term simulations of oxidants and aerosol formation. Both CHIMERE and EURAD-IM CTM are forced by the MM5 meteorological fields (Dudhia, 1993), meanwhile CMAQ uses the outputs of the WRF-ARW model (Michalakes et al., 2004). Both MM5 and WRF-ARW are non-hydrostatic models.

The three modelling system have different degrees of complexity and spatial resolution. A summary of their key features, including emissions and boundary conditions, is listed in Table 1. Additional descriptions can be consulted on the online Model Documentation System (http://pandora.meng.auth.gr/mds/mds.php). CALIOPE configurations in both European and Iberian Peninsula domains are described in detail in Pay et al. (2010) and Baldasano et al. (2011), respectively.

Table 1

Since episodic natural of dust outbreaks are frequently observed over all Iberian Peninsula (Rodríguez et al., 2001; Silva et al., 2003; Basart et al., 2009), and because the representation of these events cannot be well simulated with solely the information of aerosol boundary conditions (Vautard et al., 2005, Jiménez-Guerrero et al., 2008; Menut and Bessagnet, 2010), the long-range transport of mineral dust from Sahara desert is modelled on an hourly basis by the BSC-DREAM8b model (Nickovic et al., 2001; Pérez et al., 2006a,b). The BSC-DREAM8b is fully embedded within the NCEP/Eta meteorological driver (Janjic, 1994). Dust aerosols are represented by 8 bins size distribution within the 0.1-10 µm radius range. Dust-radiation interactions are calculated online. The modelled domain in this study comprises Northern Africa, the Mediterranean basin, Europe and Middle East. It is applied with a 0.3º x 0.3º horizontal resolution using 24 vertical layers extending up to 15 km. In the present study the BSC-DREAM8b model is offline coupled within the hourly forecast PM10 and PM2.5 concentrations from CALIOPE, MM5-CHIMERE and MM5-EURAD-IM.

## 3. MONITORING DATA

The air quality monitoring network of mainland Portugal (http://www.qualar.org/) includes 68 stations of which 42 are background, 19 traffic and 7 industrial, following the classification in Garber et al. (2002). The spatial coverage, together with the background influence and a minimum data collection efficiency of 75% are three of the criteria used for the stations selection. A fourth criterion is related with the measured pollutants. Stations that measure $O_3$ also measure $NO_2$ and the stations that measure PM10 also do it for PM2.5. As a result, a total of 18 stations (8 rural, 5 urban and 5 suburban) are selected for the present study, 13 stations for $O_3/NO_2$, 9 for $SO_2$ and 6 for PM10/PM2.5. Figure 1 shows the location and main characteristics of the selected stations over the study domain. Note that the measured data are in an hourly basis and the data are not validated since they refer to year 2010.

Figure 1

Despite the spatial coverage criteria, there is an evident concentration of monitoring stations over the coastal area and the two metropolitan areas of Porto and Lisbon (see Figure 1). Nevertheless, all the regions of Portugal are covered by at least one rural background station. In terms of topography, the mountainous regions are not so well represented by monitoring sites. The majority of the stations, which are located near/over the coast, have altitudes lower than 300 meters.

## 4. BIAS-CORRECTION TECHNIQUES

As discussed in previous works, the applied forecast systems are found to have significant biases (Monteiro et al., 2007a; Baldasano et al., 2010) that could be removed through bias-correction techniques. There are several techniques by which bias correction can be applied as mean subtraction (McKeen et al., 2005; Wilczak et al., 2006), multiplicative ratio adjustment (McKeen et al., 2005), hybrid forecast (Kang et al., 2008) and Kalman filter (Delle Monache et al., 2006; Kang et al., 2008; Djalalova et al., 2010), model ensembles (van Loon et al., 2007; Wilczak et al., 2006; Djalalova et al., 2010) among others. The bias correction does not try to gain additional insight into model deficiencies or performance neither to correct them artificially, but intends to remove potential errors intrinsic to each model formulation or input data. In the present study two post-processing methods are used to correct the bias of the three forecasting system for all the considered pollutants: a multiplicative ratio correction (McKeen et al., 2005) and the Kalman filter method (Delle Monache et al., 2006; Kang et al., 2008, 2010). Both techniques are site-specific approaches since they use past ground-based measurements and simulated data at each monitoring site to revise and improve the current hourly forecasts for the entire year of 2010.

### 4.1 The multiplicative ratio correction

The multiplicative ratio correction (RAT, McKeen et al., 2005) is a simple approach that can be mathematically expressed by equation 1.

$$C_{\text{model}}^{corrected}(h, day) = \frac{\sum_{day=1}^{n} C^{obs}(h, day)}{\sum_{day=1}^{n} C_{\text{model}}^{raw}(h, day)} \times C_{\text{model}}^{raw}(h, day) \tag{1}$$

The corrected concentration with RAT ($C_{\text{model}}^{corrected}$) is estimated based on the application of a correction factor to the raw modelled concentration ($C_{\text{model}}^{raw}$). The correction factor is calculated as the quotient between the additions of observed ($C^{obs}$) and modelled $C_{\text{model}}^{raw}$ concentrations at a particular hour (h) of the n previous days. To estimate the number of previous days (*n*), Monteiro et al. (2011) tested different training periods and chosen a 4 day training period as a compromise between having a sufficiently long period to gather adequate statistics, but not too long to mask seasonal variations (for $O_3$, for e.g.). According to Stull (1988) and also Tchepel and Borrego (2010), synoptic conditions are characterized by a 3-4 day period, which supports the chosen training period. Thus, the current multiplicative ratio correction approached was applied with a 4 day period (RAT04).

### 4.2. Kalman filter

The Kalman filter (KF) is a recursive, linear, and adaptive method that has been used recently to improve air quality forecast of ground-based $O_3$ (Delle Monache et al., 2006, 2008; Kang et al., 2008; Djalalova, et al., 2010; Sicardi et al., 2011) and PM2.5 (Dajalalova, et al., 2010; Kang et al., 2010). KF performance is sensitive to the error ratio ($\sigma^2_\eta/\sigma^2_\varepsilon$) which indicates the way in which the KF responds to the variations in biases at prior steps. There exists an optimal error ratio to generate the best forecast given the forecast modelling system and the dynamic of the study area. We follow the methodology of Kang et al. (2008) for estimating the optimal error ratio which consists in minimizing the root mean square error and maximizing the correlation coefficient for all the stations. Therefore, optimal errors ratios are selected for each modelling system and for all the selected stations over the year 2010. Only in the case of $O_3$, optimal errors ratios are selected seasonality because it was found that corrected $O_3$ simulation improved when using seasonally varying values. Table 2 presents the optimal error ratios selected for each pollutant.

Table 2

## 5. BIAS-CORRECTION ASSESSMENT

The evaluation of the different bias-correction approaches applied to the three modelling system is carried out using classical statistical indicators (Tilmes et al., 2002; Borrego et al., 2008; Denby et al., 2010; Dennis et al., 2010). The global skills of the bias-correction approaches are represented by means Taylor diagrams. Additionally, this evaluation is complemented with analysis of the most important critical points of each bias-correction technique find on the air quality forecast of the three modelling systems under study.

The Taylor diagram (Taylor, 2001) is a powerful tool frequently used in model evaluation studies (Cuvelier et al., 2007; Denby et al., 2010; Dennis et al., 2010) for the simultaneous visualization of three statistical indicators, in the present study we present the observed and modelled standard deviation (SD), the centred root mean square error (CRMSE) and the correlation coefficient (R) in a single point. Together these statistical parameters provide a quick outline of the degree of pattern correspondence among the raw and the unbiased simulated values of each forecasting system and the observed data.

Figure 2 shows the Taylor diagrams for each pollutant. $O_3$ is expressed in maximum daily concentration ($O_3$ max-1h) (Figure 2a) and in maximum daily eight-hour running average ($O_3$ max-8h) (Figure 2b) following the current 2008/50/EC European directive (European Commission, 2008). $NO_2$, $SO_2$, PM10 and PM2.5 are expressed in daily mean concentrations (Figure 2c-f, respectively). Each Taylor diagram shows the annual performance of the two bias-correction techniques, KF and RAT04, applied to the three forecasting systems and the corresponding raw modelling systems over all the studied stations.

Figure 2

Visualization of every single polar plots shows that the application of both KF and RAT04 techniques improve the raw forecasts for all the modelling systems and pollutants, bringing unbiased SD closer to the observed SD than raw modelled SD, reducing errors and increasing correlation coefficients close to the unit. For $O_3$ max-1h the improvements in annual performance is significant after applying bias-correction techniques. The maximum variability increases with KF (SD = 25.2-28.0 $\mu g.m^{-3}$) and RAT04 (SD = 26.4-29.5 $\mu g.m^{-3}$) falling closer to the observed SD (29.3 $\mu g.m^{-3}$) than raw modelled SD (22.1-26.9 $\mu g.m^{-3}$), which means that techniques adjust high $O_3$ peaks although they are still slightly underestimated. Annually, unbiased error decreased in 21-22% (KF) and 16-26%, from 20.4-25.2 µg.m$^{-3}$ (raw model) to 16.2-19.6 µg.m$^{-3}$ (KF) and 17.1-19.2 µg.m$^{-3}$ (RAT04); and correlation coefficient increase in 16-34% (KF) and 13-37% reaching 0.75-0.84 (KF) and 0.77-0.83 (RAT04). The same tendency, but with slight better skills, is found in the case of $O_3$ max-8h. Although the variability is improved, the unbiased standard deviations are usually smaller than their observed field. CRMSE is reduced in 25-26% (KF) and 25-33% (RAT04) and correlation coefficient range between 0.78-0.85 and 0.81-0.86, respectively.

In the case of $NO_2$ daily mean, after applying bias-correction techniques unbiased concentration increase the daily variability getting closer to the observed SD (14.0 µg.m$^{-3}$) from 7.7-11.2 µg.m$^{-3}$ (raw model) to 12.4-12.9 µg.m$^{-3}$ (KF) and 14.1-14.6 µg.m$^{-3}$ (RAT04) showing slightly more daily variability with RAT04. CRMSE decreases from 10.6-12.2 µg.m$^{-3}$ (raw model) to 7.2-7.3 µg.m$^{-3}$ (KF) and 5.9-6.7 µg.m$^{-3}$ (RAT04); and temporal correlations increase from 0.55-0.66 (raw model) to 0.85-0.86 (KF) and 0.89-0.91 (RAT04).

As for $NO_2$ primary pollutant, annual unbiased modelled $SO_2$ daily means present higher skills than raw modelled concentrations. Raw modelled $SO_2$ concentrations present higher daily variability (SD = 4.1-12.7 µg.m$^{-3}$) than observed field (2.4 µg.m$^{-3}$). In this sense, both bias-correction techniques get to deduced raw modelled SD till 3.3-7.3 µg.m$^{-3}$ (KF) and 3.1-7.3 µg.m$^{-3}$ (RAT04) which means that high $SO_2$ peaks have been reduced and decreased the daily concentration. Annual CRMSE are reduced in 34-75% after applying bias-correction techniques in the range of 1.6-5.8 µg.m$^{-3}$ (KF) and 2.1-5.7 µg.m$^{-3}$ (RAT04). Unbiased models also improve temporal annual correlation in more than 100%, reaching 0.17-0.50 and 0.14-0.59 with KF and RAT04, respectively.

Raw modelled PM present higher daily variability than observations which is reduced after applying bias-correction techniques. For PM10, raw modelled SD are reduced from 13.1-22.3 µg.m$^{-3}$ to 1.3-18.3 µg.m$^{-3}$ (KF) and 14.1-16 µg.m$^{-3}$ (RAT04) closer to 13.6 µg.m$^{-3}$ (observed PM10 SD). PM2.5 daily mean presents the same tendency, raw modelled concentrations are reduced from 8.0-13.7 µg.m$^{-3}$ to 7.1-10.2 µg.m$^{-3}$ (KF) and 6.8-8.9 µg.m$^{-3}$ (RAT04) closer to 7.5 µg.m$^{-3}$ (observed PM2.5 SD). The higher variability observed with PM10 and PM2.5, even after applying bias-correction techniques, is deviated by the high overestimation urban stations such as CAM. Temporal variability improves for PM10, in the range of 7-20 % (KF) and 12-33% (RAT), reaching correlation of 0.49-0.58 (KF) and 0.58-0.61 (RAT04). Improvements are higher with PM2.5 for temporal variability, (>50% with both KF and RAT04) reaching R in the range of 0.50-0.64 (KF) and 0.57-0.62 (RAT04).

Based on annual statistics indicator, the biggest percentage of improvement after applying bias-correction techniques are found for $SO_2$ daily concentration, R increase in more than 100% and error (CRMSE) decrease in the range of 34-51 % (for both KF and RAT04), following by $NO_2$ daily concentration where R increase in 30-65% and error decrease in 32-51% (for both KF and RAT04). The percentage of improvement is smaller for $O_3$ max-1h and max-8h, although with significant impact in correlation that reach 0.78-0.86 (both KF and RAT04) in the case of $O_3$ max-8h, since the raw modelled present high skills. Note that to get high skills after applying bias-correction techniques modelling systems has to demonstrate their relative accuracy. Overall, Taylor diagrams (Figure 2) point out that despite the applied techniques have different mathematic formulation and complexity level, there is comparable answers for all of the forecasting systems (see e.g. Figure 2c). There is a slightly superiority of RAT04 technique over Kalman filter in terms of statistical indicator and graphical representation of Taylor diagrams. However the aforementioned evaluation has the limitation that it is done over all the stations in annual basin and it gives no information whether the unbiased concentrations are correct for the right or wrong reason. Therefore, in order to go more in detail on the skills of bias-correction techniques specific examples of the successes/failures of both techniques are illustrated following, since is important to know how RAT04 and KF behave in specific situations, such as air quality episodes, not-validated or missing data, in order to choose the most convenient bias-correction technique to apply on air quality forecast over Portugal.

In Figure 3 (top) where the hourly observed $O_3$ concentrations (red points) at the CAL station is presented along with the raw CALIOPE outputs (blue) and the post-processed KF and RAT04 values (orange and green, respectively) during a summer period (June month). This example demonstrates how both KF and RAT04 techniques improve the forecasted $O_3$ daily cycles, since they agree with the observed hourly variability in both diurnal maximum and night minimum, reducing the persistent overestimation with respect to measurements (Figure 3, bottom). Hourly statistical analyses (not shown here) quantify that maximum and minimum annual bias are in the range of ±5 µg.m$^{-3}$ after post-processing with both KF and RAT04. That means a bias improvement of more than 80% in the maximum overestimation (from 40-20 µg.m$^{-3}$ to less than 5 µg.m$^{-3}$) for all the system.

Figure 3

Figure 4 shows PM10 time series at FUN station during an air quality episode in August 2010. In the first part of time series, from August 7$^{th}$ to 10$^{th}$, a desert dust outbreaks arrives to Portugal due to a North Africa advection (Figure 4 c). The raw CALIOPE system reproduces the event thanks to the contribution of the BSC-DREAM8b model (Figure 4b) although the concentrations are slightly underestimated. After applying bias-correction techniques, unbiased outputs are closer to the hourly observed concentrations. In the second part, from August 10th to 13th, the wind changes the trajectory to northwest (see Figure 4c) and the observed concentrations reach ~170 µg.m$^{-3}$ According to the Portuguese Forest Authority (Autoridade Florestal Nacional, 2010) nine forest fires occurred during this period in a radium of 100 km from FUN station where more than 10,000 ha were burned. In the described fire episode both bias-correction techniques do not reproduce the event since the raw CALIOPE modelling system, as MM5-CHIMERE and MM5-EURAD-IM systems, does not include forest fire emissions. The high bias estimated for this episode generates that both techniques overestimate observed concentration four days later after the fire is finished. KF gets closer to the observations faster than RAT04 since KF gradually spreads the error and RAT04 present high sensitivity to the magnitude of the modelled values.

Figure 4

Frequent problems in the forecast of $SO_2$ are associated to high underestimations of $SO_2$ peaks. The main activity sources of $SO_2$ emissions are related to power plants and transformation/manufacturing industry (source: http://www.emep.int/). Besides a high level of control of the $SO_2$ emissions, these point sources can episodically generate large plumes of high-$SO_2$ content affecting the air quality in urban and regional scales downwind the sources. Accurate $SO_2$ forecasts depend on the accuracy in the meteorological patterns, the variability on the sub-grid scale with respect to measured data (Stern et al., 2008; Baldasano et al., 2011), and the accurate representation of emissions sources.

The Figure 5 illustrates an episode of high $SO_2$ concentrations at the CHA station, on March 27th from 6:00 to 12:00 where any of the forecast systems were able to predict the observed event (only MM5-EURAD-IM is shown in Figure 5). This example demonstrates that both KF and RAT04 produce an error due to high concentrations observed on March 27th which is propagated to the same hour during the days after. The propagated error is higher for RAT04 than KF since RAT04 is a simple technique by which simulated and observed data have the same weight. RAT04 applies a correction on the same hour of the next days and if there is no other high concentration during 4 days, the hourly correction factor error will not be reproduced on the 5th day after. On the other hand, the optimal ratio of KF to MM5-EURAD-IM is low (0.04, see Table 2) which means that KF has more confidence on model simulations than observations data. In this sense, the propagated error by KF is less than RAT04 error. In addition, if no other high concentration is recorded, KF error will decrease over the next days, meaning that corresponding bias will be getting closer to 0. The propagation of an error produced by model simulations or observations data (both by a high recorded concentration and by not validated data) is a common characteristic of both techniques. This example illustrates that despite RAT04 has a better performance in general terms, KF can generate a correction with less error in these specific situations.

Figure 5

The Figure 6 shows an episode registered in October 25th - 30th at the MVE station where the raw CALIOPE system forecasted high $SO_2$ concentrations that actually did not occur. The same behaviour was obtained with the raw MM5-CHIMERE and MM5-EURAD-IM forecasting systems (not shown here). The figure demonstrates the limitations of the KF technique against high overestimation of the models. Meanwhile, the RAT04 technique (green) corrects the raw forecast following the hourly observation with a bias reduction of 80%. This poor performance of KF is related with two facts. First, $SO_2$ optimal error ratio ($\sigma^2_\eta/\sigma^2_\varepsilon$) for the three models result between 0.13 and 0.20, higher compared to the other pollutants ratios (see Table 2). When ratio is high, the forecast-error white-noise variance ($\sigma^2_\varepsilon$) will be relatively small compared to the true forecast-bias white-noise variance ($\sigma^2_\eta$). Therefore, the filter will put excessive confidence on the previous forecast and the predicted bias will respond very quickly to previous forecast errors. Second, KF bias-adjustment is a linear and recursive algorithm. KF predicts the future bias with a linear relationship given by the previous bias estimate plus a quantity proportional to the difference between the present forecast error and the previous bias estimates. Therefore KF is unable to correct large bias due to model overestimations when all the biases for the past few days have been small.

Figure 6

The absence of monitoring data is frequently a problem for data assimilation or bias-correction procedures. In case of the RAT04 approach, if there are no measurements, the unbiased outputs will be equal to the raw modelled data. On the other hand, KF has a capacity to learn the behaviour of simulations data relatively to monitoring data, which means that KF is designed to apply the same correction as that estimated for the previous days. Figure 7 illustrates this problem with an example of two different periods of absence of measurement data registered at the CAL station, from April 10th to May 1st 2010. Once all of the forecast systems presented the same behaviour, just the MM5-EURAD-IM simulation is shown here. In the first half period (from April 10$^{th}$ to the half of April 14$^{th}$) KF and RAT04 produce a reasonable corrections with bias values closer to 0 (Figure 6, bottom). During the periods of April 14$^{th}$ - 18$^{th}$and April 23$^{rd}$ - 25$^{th}$, there are no monitoring data, In this case, KF applies the same correction from previous days and RAT04 does not correct the simulated data, taking the same raw modelled outputs. When data start to be available, KF continues to apply the bias correction base on previous days and after four days the recent measurement have an effective effect on bias correction (observed and simulated data). With the RAT04 technique the simulated data is only possible to be corrected after 4 days of monitoring data availability. In future work RAT04 can be improved/designed in order to minimize this problem, applying the previous correction to the hour without observed data, as KF does.

Figure 7

Both techniques are sensitive to not validated data which is a frequent problem for time forecasting mode working. Figure 8 shows an example of not validated data, specifically, when the station presents a calibration problem. Time series of hourly $SO_2$ concentrations (red) at the MVE station present two clear tendencies in Figure 8. In the first part, $SO_2$ measured concentrations present a background level ~ 8 $\mu g.m^{-3}$, and on June 22$^{sd}$ at 12:00 observed data decrease sharply 7 $\mu g.m^{-3}$ to be oscillating around ~ 1 $\mu g.m^{-3}$ the rest of the year. This suggests that MVE station registered/exhibited a calibration problem in the first part of the time series that is corrected in the second part. In this situation both KF and RAT04 correct the raw forecast to agree with observations in the both aforementioned situations. On one hand KF presents a robust response against a systematic bias. KF gives more confidence to the observations based on persistent systematic bias, and adjusts the background levels to ~ 8 $\mu g.m^{-3}$ in the first part, and to ~ 1 $\mu g.m^{-3}$ in the second part, with a transition period of 4 days till the bias are reduce to 0 (orange line, Figure 8 bottom). On the other hand RAT04 tries to adjust background levels in both situations, but produces overestimations during these periods. These instabilities show its sensitivity to high gradient of concentrations, and it is a limitation of multiplicative techniques (Wilczak et al., 2006).

Figure 8

## 6. FORECAST MODELS PERFORMANCE

The categorical statistical skills (Kang et al, 2005; Eder et al., 2006) are computed in order to evaluate how the two bias correction techniques improve the three air quality forecasts daily produced over Portugal in terms of exceedances and non-exceedances events. Exceedances analysis is based on a comparison with a fixed threshold concentration (T). The present work uses as thresholds those established by the European directive 2008/50/EC on air quality (European Commission, 2008). Only $O_3$ and PM10 are evaluated in terms of categorical statistics because neither PM2.5, $NO_2$ nor $SO_2$ exceeded the European limit values at the selected stations in 2010. The 2008/50/EC directive sets an information threshold of 180 $\mu g.m^{-3}$ for maximum daily concentrations (max-1h) and a target value of 120 $\mu g.m^{-3}$ for maximum daily eight-hour running average (max-8h) not to be exceeded on more than 25 days per year. In the case of PM10, it establishes a limit value of 50 $\mu g.m^{-3}$ for daily average (Mean-24h) not to be exceeded more than 35 times per year. Table 3 shows the annual categorical parameters for all the selected Portuguese stations. The calculated statistics are the accuracy (A), the bias (B), the probability of detection (POD), the false alarm ratio (FAR) and the critical success index (CSI). Kang et al. (2005) shows the formulas of the aforementioned categorical statistics.

Table 3

The percentages of the 2010 exceedances that are actually forecasted are estimated with the value of POD. For $O_3$ max-1h, a total of 51 exceedances of the information threshold (T = 180 $\mu g.m^{-3}$) are observed over the 13 Portuguese stations in 2010 (1/3*(b+d), in Table 3). The bias-correction techniques increase the POD from 3% (raw models) to 10 % in KF and 31% in RAT04. In the case of $O_3$ max-8h, a total of 297 exceedances of the target value (T = 120 $\mu g.m^{-3}$) were observed over all the stations. The POD also improves when bias-correction techniques are applied from 27% (raw) to 48% (KF) - 54% (RAT04). Overall, POD improves strongly after the post-processing techniques for both $O_3$ max-1h and max-8h, reaching an improvement of more than 100% and 50% for max-1h and max-8h, respectively. This means that by means the application of bias-correction techniques the forecast alerts for the population about exceedances would be significant accurate.

For PM10 daily mean a total of 68 exceedances of the daily limit value (T = 50 $\mu g.m^{-3}$) were measured. The bias-correction techniques increase the POD from 32% (raw) to 34% (KF) - 45% (RAT04). However the improvement percentage of POD is less than 50%, lower than for $O_3$, due to the no significant increase of the number of hits (b) (from 65 (raw) to 70 (KF) and to 92 (RAT04)).

The accuracy (A) measures the percentage of simulations that correctly reproduce an exceedance or no-exceedance (ideally 100%). Actually, A is already high for the raw models for three variables (A > 90%), and there are no significant improvements after post-processing. In the present study, careful must be done in the interpretation of the A since the number of the observed exceedances (b+d) is little respect to the total pair of data (a+b+c+d). The categorical bias (B) indicates if the forecasts fail by overestimating (false positive) or underestimating (correct negative) exceedances (ideally 1). For $O_3$ max-1h and max-8h, B remains below 1 before and after post-processing, which indicates that errors by missing of observed exceedances are not totally resolved (d>a). The better performance is found for $O_3$ max-8h, where B improves from 0.7 (raw) to 0.8 (KF and RAT04). Low B performance in $O_3$ max-1h is due to the poor capability to reproduce maximum hourly $O_3$ concentrations. On the other hand, for PM10 daily mean categorical bias originally presents problem with false alarms (B>1). B is significant reduced after post-processing, from 2.2 (raw) to 1.7 (KF) – 1.5 (RAT04). Nevertheless corrected models still present problems with false alarms.

FAR is useful to quantify the fails by simulating exceedances that actually did not occur (ideally 0%). Application of the post-processing techniques reduces of almost the half the value of the FAR for the max-8h. This indicates the ability of the KF and RAT04 techniques to reduces the number of projected false alarms from 371(raw) to 254 (KF) – 266 (RAT04). For the $O_3$ max-1h the false alarms (b) does not improve significantly after the post-processing. However FAR improves owing to the improving of hits detections (b) from 5 (raw) to 16-48 (KF and RAT04, respectively). For PM10 daily mean, FAR improves less than 20% with both post-processing techniques, since the bias-correction techniques do not reduce significantly modelled false alarms (a) for PM10 daily mean. The CSI indicates how well both forecast exceedances and actual exceedances are predicted (ideally 100%). For the three analysed variables CSI improves when both KF and RAT04 techniques are applied. Unlike the POD and the FAR, the CSI takes into account both false alarms and missed events, and it is thus a more balanced score.

Results demonstrate that both techniques improve modelling skills to reproduce exceedances established by the European directive 2008/50/EC for PM10 daily mean and $O_3$ max-1h and max-8h. Better skills are found with RAT04 than for KF in most cases. Nevertheless, it must be taken into account that the categorical statistics only evaluate the model in terms of exceedances; therefore caution is needed when interpreted.

## 7. SUMMARY AND CONCLUSIONS

The current work performs an exhaustive examination of two different bias-correction techniques, the Kalman filter method (KF) and a multiplicative ratio with a 4 days training period (RAT04), within their application inland Portuguese domain. Both approaches have been applied to the three advanced forecasting systems operated routinely over Portugal in 2010 – CALIOPE, MM5-CHIMERE and MM5-EURAD-IM. The evaluation is carried on in terms of ground-based concentrations of gas-phase ($O_3$, $NO_2$, and $SO_2$) and particulate matter (PM10 and PM2.5) pollutants. Statistical parameters were used (classical and categorical) and graphical techniques (Taylor diagram and temporal series) in order to quantify the abilities of the two post-processing techniques to improve the air quality forecast over Portugal.

Comparative statistical analysis, based on Taylor diagram, show that both KF and RAT04 techniques improve the raw forecasts skills (for all the modelling systems and pollutants), bringing unbiased SD closer to the observed SD than raw modelled SD, reducing errors and increasing correlation coefficients close to the unit. In the case of $O_3$ max-8h, temporal variability improves in 19-45 % from 0.56-0.81 (raw models) to 0.78-86 (KF and RAT04, respectively). Similar tendency is found for $O_3$ max-1h. The primary pollutant $NO_2$ and $SO_2$ daily concentrations, demonstrate significant relative improvements compared to $O_3$, mostly because the original modelling system skills are lower for those species. $NO_2$ correlation coefficients improve between 30-65% and more than 100% for $SO_2$ (for both KF and RAT04); and errors decrease also in both cases in ~30-40% (for both KF and RAT04). For PM, improvement after applying both KF and RAT04 are higher with PM2.5 where correlation coefficients increase in more than 50% (both techniques) reaching values between 0.50 – 0.64. Note that to get high skills after applying bias-correction techniques modelling systems has to demonstrate their relative accuracy.

Despite the applied techniques have different mathematic formulation and complexity level, there are comparable answers for all of the forecasting systems. There is a slightly superiority of RAT04 technique over KF in terms of statistical indicator and graphical representation of Taylor diagrams. However the analysis performed over specific situations, such as air quality episodes, not-validated or missing data reveals different behaviour for KF and RAT04. In the case of hourly $O_3$ concentrations, both bias-correction techniques are efficient tools to improve simulated $O_3$ daily cycle remaining bias in the range of ±5 $\mu g.m^{-3}$. Under desert dust advection from North Africa, KF and RAT04 are able to correct PM10 bias within slightly overestimation of RAT04. Nevertheless, under missed pollution events of short-life (< 2 days), as shown with forest fire or high $SO_2$ peaks, KF and RAT04 have no efficient corrections of that large bias.

RAT04 applies a correction on the same hour of the next days and if there is no other high concentration during 4 days, the hourly correction factor error will not be reproduced on 5$^{th}$ day after. In the other hand, the propagation of error in KF is less sharp than for RAT04, since give more confidence to previous persistent bias. This is an advantage of KF under not validated data or missing data since the capability of response is higher than RAT04. One evident disadvantage of KF against RAT04 is when the modelling system presents high overestimations (as shown with hourly $SO_2$ peaks). KF is unable to correct large bias due to model overestimations since the filter puts excessive confidence on modelled forecast. Note that both techniques are sensitive to not validated data.

The improvements of the discussed critical points will conduct to a better unbiased model performance which will be reflected on a higher accuracy of episodes forecasted. Beyond the discussed weaknesses of the both bias-correction approaches, there is a critical point that is common to KF and RAT04: both are site-specific dependents. We are currently working to solve this problem, developing a spatial approach for the bias correction on the overall domain.

Categorical analysis has been performed over air quality pollutant that exceed threshold and limit values establish by the European legislation on air quality which are $O_3$ max-1h (threshold = 180 µg.m$^{-3}$), $O_3$ max-8h (threshold =120 µg.m$^{-3}$) and PM10 daily mean (limit value=50 µg.m$^{-3}$). Results indicate that the probability of detection (POD) of both techniques improve in more than 100% for $O_3$ max-8h and 50% for $O_3$ max-1h with a total increase from 27% to 48% (KF) and 54% (RAT04) in the case of $O_3$ max-8h. However, the improvement percentage of POD is less than 50%, lower than for $O_3$, due to the no significant increase of the number of hits (b) (from 65 (raw) to 70 (KF) and to 92 (RAT04)), may be related with the fact that some missing sources (such as forest fires) are not includes in the raw modelling systems.

These above results confirm the advantage of the application of RAT04 and KF bias-correction techniques for air quality forecast. Both techniques can be applied routinely in an operational forecast system and they will be useful to alerts for the population about accurate exceedances.

## ACKNOWLEDGEMENTS

The authors acknowledge the CRUP by the support of the Integrated Action E 122-10 and Integrated Action PT2009-0029 from the Ministerio de Ciencia e Innovación. Thanks are extended to the Portuguese ‘Ministério da Ciência, da Tecnologia e do Ensino Superior’ for the financing of BIOGAIR (PTDC/AAC-AMB/103866/2008) project, for the PhD grant of Isabel Ribeiro (SFRH/ BD/60370/2009) and the post doc grant of Alexandra Monteiro (SFRH/BPD/63796/2009). The Spanish Ministry of Science and Innovation is also thanked for the Formación de Personal Investigador (FPI) doctoral fellowship held by María Teresa Pay

(CGL2006-08903). COST ES0602 is also acknowledged. The authors wish to thank Luca Delle Monache and Ronald B. Stull for providing the Kalman filter algorithm used in this study. The computation with CALIOPE system has been done at the MareNostrum supercomputer hosted by the Barcelona Supercomputing Center-Centro Nacional de Supercomputación.

## REFERENCES


Autoridade Florestal Nacional, 2010. Relatório provisório de incêndios florestais. Technical Report RP9/2010. Lisboa, Portugal. 15 pp. Available at http://www.afn.min-agricultura.pt/portal/dudf/relatorios/resource/ficheiros/2010/20100815_AFN_RP6.pdf.

Baldasano J.M, Jiménez-Guerrero P., Jorba O., Pérez C., López E., Güereca P., Martin F., García-Vivanco M., Palomino I., Querol X., Pandolfi M., Sanz M.J., Diéguez J.J., 2008a. CALIOPE: An operational air quality forecasting system for the Iberian Peninsula, Balearic Islands and Canary Islands- First annual evaluation and ongoing developments. Adv. Sci. Res. 2, 89-98.

Baldasano J.M., Güereca L. P., López E., Gassó S., Jimenez-Guerrero P., 2008b. Development of a high-resolution emission model for Spain: the High-Elective Resolution Modelling Emission System (HERMES). Atmos. Environ. 42 (31), 7215-7233.

Baldasano J.M., Pay M. T., Jorba O., Ortiz J., Gonçalves M., BasarT S., Gassó S., Jiménez-Guerrero P., 2010. Evaluation of the Spanish operational air quality forecasting system: diagnostic and near real time. In: International Workshop on Air Quality Forecasting Research. Québec, Canada, 16-18, November.

Baldasano J.M., Pay M.T., Jorba O., Gassó, S., Jiménez-Guerrero P., 2011. An annual assessment of air quality with the CALIOPE modeling system over Spain. Sci. Total Environ. 409, 2163-2178. doi:10.1016/j.scitotenv.2011.01.041.

Basart S., Pérez C., Cuevas E., Baldasano J.M., Gobbi P., 2009. Aerosol characterization in Northern Africa, Northeastern Atlantic, Mediterranean Basin and Middle East from direct-sun AERONET observations. Atmos. Chem. Phys., 9, 8265-8282.

Bessagnet B., Hodzic A., Vautard R., Beekmann M., Cheinet S., Honoré C., Liousse C., Rouïl L., 2004. Aerosol modeling with CHIMERE-preliminary evaluation at the continental scale. Atmos. Environ., 38, 2803-2817.

Bessagnet B., Hodzic A., Blanchard O., Lattuati M., Le Bihan O., Marfaing H., Rouil L., 2005. Origin of particulate matter pollution episodes in wintertime over the Paris Basin. Atmos. Environ., 39 6159-6174.

Binkowski F. S., Roselle S. J., 2003. Models-3 Community Multiscale Air Quality (CMAQ) model aerosol component: 1. Model description. J. Geophys. Res. 108(D16), 4183.

Borrego C., Schatzmann M., Galmarini S., 2003. Quality assurance of air pollution models. In: Moussiopoulos N (ed.), SATURN – Studying air pollution in urban areas – EUROTRAC-2 Subproject final report, Springer Verlag, Heidelberg, Germany, Chapter 7, 155-183.

Borrego C., Monteiro A., Ferreira J., Miranda A.I., Costa A.M., Carvalho A.C., Lopes M., 2008. Procedures for estimation of modelling uncertainty in air quality assessment. Environ. Int. 34, 613-620.

Bott A., 1989. A positive definite advection scheme obtained by non-linear renormalization of the advective fluxes, Mon. Wea. Rev., 117, 1006-1015.

Carvalho A., Monteiro A., Ribeiro I., Tchepel O., Miranda A.I., Borrego C., Saavedra S., Souto J.A., Casares J.J. (2010). High ozone levels in the Northeast of Portugal: analysis and characterization. Atmos. Environ. 44, 8, 1020-1031. doi:10.1016/j.atmosenv.2009.12.020.

Cuvelier C., Thunis P., Vautard R., Amann M., Bessagnet B., Bedogni M., et al., 2007. CityDelta: A model intercomparison study to explore the impact of emission reductions in European cities in 2010. Atmos. Environ. 41, 189-207.

Delle Monache L., Nipen T., Deng X., Zhou Y., Stull R., 2006. Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction. J. Geosphys. Res. 111(D05308). doi: 10.1029/2005JD006311.

Delle Monache L., Wilczak J., McKeen S., Grell G., Pagowski M., Peckham S., Stull R., McHenry J., McQueen J., 2008. A Kalman-filter bias correction method applied to deterministic, ensemble averaged, and probabilistic forecast of surface ozone. Tellus Ser. B, 60, 238-249. doi: 10.1111/j.1600-0889.2007.00332.x.

Denby B., Larssen S., Guerreiro C., Li L., Douros J., Moussiopoulos N., et al., 2010. Guidance on the use of models for the European Air Quality Directive. A working document of the Forum for Air Quality Modelling in Europe. FAIRMODE. In: Denby, B., editor. Technical Report Version 4.2ETC/ACC.

Dennis R., Fox T., Fuentes M., Gilliland A., Hanna S., Hogrofe C., Irwin J., Trivikrama R., Scheffe R., Schere K., Steyn D., Venkatram, A., 2010. A framework for evaluating regional-scale numerical photochemical modeling systems. Environ. Fluid Mech. doi: 10.1007/s10652-009-9163-2.

Djalalova I., Wilczak J., McKeen S., Grell G., Peckhama S., Pagowski M., DelleMonache L., McQueen J., Tang Y., Leeg P., McHenry J., Gong W., Bouchet V., Mathur R., 2010. Ensemble and bias-correction techniques for air quality model forecasts of surface $O_3$ and PM2.5 during the TEXAQS-II experiment of 2006. Atmos. Environ. 44, 455-467.

Dudhia J., 1993. A nonhydrostatic version of the PennState/NCAR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front. Mon. Weather Rev. 121, 1493-1513.

Eder B., Kang D., Mathur R., Yu S., Schere K., 2006. An operational evaluation of the Eta-CMAQ air quality forecast model, Atmos. Environ. 40, 4894–4905, doi:10.1016/j.atmosenv.2005.12.062.

Elbern H., Strunk A., Schmidt H., Talagrand O., 2007. Emission Rate and Chemical State Estimation by 4-Dimensional Variational Inversion. Atmos. Chem. Phys. 7, 3749-3769.

European Commission, 2008. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe, Technical Report 2008/50/EC, L152, Off. J. Eur. Comm, 2008.

EEA, 2010. The European Environment. State and outlook 2010. Air pollution. Luxembourg, Publication Office of the European Union. ISBN 978-92-9213-152-4. doi:10.2800/57792. 46 pp.Garber W., Colosio J., Grittner S., Larssen S., Rasse D., Schneider J., et al., 2002. Guidance on the Annexes to Decision 97/101/EC on Exchange of Information as revised by Decision 2001/752/EC. Technical Report. European Commission, DG Environment; 2002.

Geiger H., Barnes I., Bejan I., Benter T., Spittler M., 2003. The tropospheric degradation of isoprene: an updated module for the regional atmospheric chemistry mechanism. Atmos. Environ., 37, 1503-1519.

Gery M.W., Whitten G.Z., Killus J.P., Dodge M.C., 1989. A photochemical kinetics mechanism for urban and regional scale computer modeling. J. Geophys. Res., 94 (D10), 12925-12956.

Guenther A.B., Hewitt C.N., Erickson D., Fall R., Geron C., Graedel T., et al., 1995. A global model of natural volatile organic compound emissions. J. Geophys. Res. 100, 8873-8892.

Chang J.C., Hanna S.R., 2004. Air quality model performance evaluation. Meteorol. Atmos. Phys., 87, 167-196.

Janjic Z.I., 1994. The step-mountain ETA coordinate model: Further developments of the convection, viscous sublayer and turbulence closure schemes. Mon. Weather Rev. 122, 927-945.

Jiménez-Guerrero P., Pérez C., Jorba O., Baldasano J. M., 2008. Contribution of Saharan dust in an integrated air quality system and its on-line assessment. Geophys. Res. Lett. 35, L03814, doi:10.1029/2007GL031580.

Kang D., Eder B.K., Stein A.F., Grell G.A., Peckham S.E., McHenry J., 2005. The new England air quality forecasting pilot program: development of an evaluation protocol and performance benchmark. J. Air Waste Mange. Assoc. 55, 1782-1796.

Kang D., Mathur R., Rao S.T., Yu S., 2008. Bias adjustment techniques for improving ozone air quality forecasts. J. Geophys. Res. 113 (D23308). doi: 10.1029/2008JD010151.

Kang D., Mathur R., Rao S.T., 2010. Assessment of bias-adjusted PM2.5 air quality forecasts over the continental United States during 2007. Geosci. Model Dev. 3, 309-320. doi: 10.5194/fmd-3-309-2010.

McKeen S., Wilczak J., Grell G., Djalalova I., Peckham S., Hsie E.-Y., Gong W., Bouchet V., Menard S., Moffet R., McHenry J., McQueen J., Tang Y., Carmichael G. R., Pagowski M., Chan A., Dye T., Frost G., Lee P., Mathur R. 2005. Assessment of an ensemble of seven real-time ozone forecasts over eastern North America during the summer of 2004. J. Geophys. Res. 110, D21307, doi:10.1029/2005JD005858.

Menut L. Bessagnet B., 2010. Atmospheric composition forecasting in Europe. Ann. Geophys. 28, 61–74.

Michalakes J., Dudhia J., Gill D., Henderson T., Klemp J., Skamarock W.,Wang W., 2004. The weather research and forecast model: software architecture and performance. In: Mozdzynski, E.G. (Ed.), To Appear in Proceeding of the Eleventh ECMWF Workshop on the Use of High Performance Computing in Meteorology, 25-29 October 2004, Reading, U.K, pp. 117-124.

Monteiro A., Vautard R., Lopes M., Miranda A.I., Borrego C., 2005. Air Pollution Forecast in Portugal: a demand from the new Air Quality Framework Directive. Int.l J. Environ. Pollut. 25, No 2, 4-15.

Monteiro A., Miranda A.I., Borrego C., Vautard R., 2007a. Air quality assessment for Portugal. Sci. Total Environ. 373, 22-31.

Monteiro A., Borrego C., Miranda A.I., Gois V., Torres P., Perez A.T., 2007b. Can air quality modelling improve emission inventories?. In: Proceedings of the 6th International Conference on Urban Air Quality, 26-30 March, Limassol, Cyprus, 13-14.

Monteiro A., Ribeiro I., Techepel O., Sá E., Ferreira J., Carvalho A., Martins V., Strunk A., Galmarini S., Elbern H., Schaap M., Builtjes P., Miranda A.I., Borrego C., 2011. Bias correction techniques to improve air quality ensemble prediction: focus on $O_3$ and PM over Portugal. Submitted to Environ. Modell. Assess.

Nickovic S., Kallos G., Papadopoulos A., Kakaliagou O., 2001. Model for prediction of desert dust cycle in the atmosphere. J. Geophys. Res. 106(D16), 18113-18129, doi:10.1029/2000JD900794.

Pagowski M., Grell G.A., Devenyi D., Peckham S., McKeen S.A., Gong W., Delle Monache L., McHenry J.N., McQueen J., Lee P., 2006. Application of dynamic linear regression to improve the skill of ensemble-based deterministic ozone forecasts. Atmos. Environ. 40, 3240-3250. doi:10.1016/j.atmosenv.2006.02.006.

Parra, R., Gassó, S., Baldasano, J.M., 2004. Estimating the biogenic emissions of non-methane volatile organic compounds from the North western Mediterranean vegetation of Catalonia, Spain. Sci. Total Environ., 329, 241-259.

Pay M.T., Piot M., Jorba O., Gassó S., Gonçalves M., Basart S., Dabdub D., Jiménez-Guerrero P., Baldasano J.M., 2010. A full year evaluation of the CALIOPE-EU air quality modeling system over Europe for 2004. Atmos. Environ. 44, 3322-3342.

Pérez C., Nickovic S., Baldasano J.M., Sicard M., Rocadenbosch F., Cachorro V.E., 2006a. A long Saharan dust event over the western Mediterranean: Lidar, sun photometer observations, and regional dust modeling. J. Geophys. Res. 111, D15214, 1-16, doi:10.1029/2005JD006579.

Pérez C., Nickovic S., Pejanovic G., Baldasano J.M., Ozsoy E., 2006b. Interactive dust-radiation modeling: A step to improve weather forecast. J. Geophys. Res. 111, D16206, doi:10.1029/2005JD006717.

Rodríguez S., Querol X., Alastuey A., Kallos C., Kakaliagou O., 2001. Saharan dust contribution to $PM_{10}$ and TSP levels in Southern and Eastern Spain. Atmos. Environ., 35, 2433-2447.

Schell, B., Ackermann, I.J., Hass, H., Binkowski, F.S., Ebel, A., 2001. Modeling the formation of secondary organic aerosol within a comprehensive air quality model system. J. Geophys. Res., 106 (D22), 28275-28293. Doi:10.1029/ 2001JD000384.

Sicardi V., Ortiz J., Rincón A., Jorba O., Pay M.T., Gassó S., Baldasano J.M., 2011. Ground-level ozone concentration over Spain: an application of Kalman Filter post-processing to reduce model uncertainties. Geosci. Model Dev. Discuss. 4, 343-384.

Silva A., Costa M., Elias T., Formenti P., Belo N., Pereira S., 2003. Ground-based aerosol monitoring at Évora, Portugal. Glob. Change NewsLetter, 56, December.

Simpson D., Winiwarter W., Börjesson G., Cinderby S., Ferreiro A., Guenther A., Hewitt C., Janson R., Khalil M., Owen S., Pierce T., Puxbaum H., Shearer M., Skiba U., Steinbrecher R., Tarrason, L., Öquist, M., 1999. Inventorying emissions from nature in Europe. J. Geophys. Res. 104 (D7), 8113-8152.

Stern R., Builtjes P., Schaap M., Timmermans R., Vautard R., Hodzinc A., et al., 2008. A model inter-comparison study focussing on episodes with elevated PM10 concentration. Amos. Environ. 42, 4567-4588.

Strunk A., Ebel A., Elbern H., Friese E., Goris N., Nieradzik L.P., 2010. Four-dimensional variational assimilation of atmospheric chemical data - application to regional modelling of air quality, in: Lecture Notes in Computer Science (LNCS), 5910, 222–229, Springer.

Stull R.B., 1988. An Introduction to Boundary-Layer Meteorology, Kluwer, 666 pp.

Taylor K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. 106 (D7), 7183–7192.

Tchepel O., Borrego C., 2010. Frequency analysis of air quality time series for traffic related pollutants. J. Environ. Monitor. 12, 544 – 550. doi:: 10.1039/b913797a.

Tilmes S., Brandt J., Flatoy F., Bergrstrom R., Flemming J., Langner J., Christensen J.505 H., Frohn L.M., Hov O., Jacobsen I., Reimer E., Stern R., Zimmermann J., 2002. Comparison of five Eulerian air pollution forecasting systems for the summer of 1999 using the German ozone monitoring data. J. Atmos. Chem. 42, 91–121.

van Loon M., Vautard R., Schaap M., Bergström R., Bessagnet B., Brandt J., Builtjs P., Christensen J., Cuvelier C., Graff A., Jonson J., Krol M., Langner J., Roberts P., Rouil L., Stern R., Tarrasón L., Thunis P., Vignati E., White L., Winda P., 2007. Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble. Atmos. Environ. 41 (10), 2083-2097.

Vautard R., Bessagnet B., Chin M., Menut L., 2005. On the contribution of natural Aeolian sources to particulate matter concentrations in Europe: Testing hypotheses with a modelling approach, Atmos. Environ. 39, 3291–3303.

Wilczak J., McKeen S.A., Djalalova I., et al., 2006. Bias-corrected ensemble and probabilistic forecasts of surface ozone over eastern North America during the summer of 2004. J. Geophys. Res. 111, D23S28. doi:10.1029/2006JD007598.

## Figures captions

Figure 1: Location and main characteristics of the selected stations from the mainland Portuguese air quality monitoring network for 2010. (a) Station environments according to Garber et al. (2002) and the terrain elevation (in m). (b) Measured pollutants in each station.

Figure 2: Taylor diagram for each air quality system (CALIOPE, MM5-CHIMERE and MM5-EURAD-IM) and for each bias correction technique (KF and RAT04) over all selected monitoring stations. (a) $O_3$ max-1h; (b) $O_3$ max-8h (c) $NO_2$ daily mean; (d) $SO_2$ daily mean, and (e) PM10 daily mean; (f) PM2.5 daily mean. Black dots represent the reference point (observed data). The radial distances from the origin (0, 0) to the points are proportional to the standard deviations (in $\mu.m^{-3}$). Azimuthal positions give the correlation coefficients. The distances between single points and reference point give the centred root mean square error (in $\mu.m^{-3}$).

Figure 3: (Top) hourly $O_3$ time series ($\mu g.m^{-3}$) at the CAL station, estimated by the CALIOPE forecasting system (blue) and after the two bias correction techniques KF (orange) and RAT04 (green) from June 9$^{th}$ to 30$^{th}$, 2010. (Bottom) hourly bias evolution ($\mu g.m^{-3}$) corresponding to CALIOPE forecasting system, KF and RAT04.

Figure 4: (a) hourly PM10 time series ($\mu g.m^{-3}$) at the FUN station for the CALIOPE forecasting system (blue line) and the two bias correction techniques KF (orange) and RAT04 (green) from August 5$^{th}$ to 16$^{th}$, 2010. Area plot shows the modelled desert dust contribution (DD, light blue area) and anthropogenic contribution (CALIOPE-DD, dark blue area). (b) Desert dust concentration ($\mu g.m^{-3}$) forecast with the BSC-DREAM8b at 12h August 9$^{th}$ (available at http://www.bsc.es/plantillaH.php?cat_id=521). (c) 5 day HYSPLIT back-trajectories ending at FUN station at different levels (500, 1000, 1500 m a.g.l.) for August 10$^{th}$. (d) 3 day HYSPLIT back-trajectories ending at FUN station at different levels (500, 1000, 1500 m a.g.l.) for August 12$^{th}$.

Figure 5: (Top) hourly $SO_2$ time series at the CHA station, measured and estimated with the MM5-EURAD-IM forecasting system (blue) values and applying the two bias correction techniques KF (orange) and RAT04 (green), from March 26$^{th}$ to April 1$^{st}$ 2010. (Bottom) bias evolution ($\mu g.m^{-3}$) corresponding to MM5-EURAD-IM forecasting system, KF and RAT04.

Figure 6: (Top) hourly $SO_2$ time series (µg.m$^{-3}$) at the MVE station for the CALIOPE forecasting system (blue) and the two bias correction techniques KF (orange) and RAT04 (green) from October 25$^{th}$ to 30$^{th}$. (Bottom) hourly bias evolution (µg.m$^{-3}$) corresponding to CALIOPE forecasting system, KF and RAT04.

Figure 7: (Top) hourly $O_3$ time series at the CAL station for MM5-EURAD-IM forecasting system (blue) and the two bias correction techniques KF (orange) and RAT04 (green), from April 10$^{th}$ to May 1$^{st}$, 2010. (Bottom) hourly bias evolution (µg.m$^{-3}$) corresponding to MM5-EURAD-IM forecasting system, KF and RAT04.

Figure 8: (Top) hourly $SO_2$ time series (µg.m$^{-3}$) at the MVE station for the MM5-CHIMERE system (blue) and the two bias correction techniques KF (orange) and RAT04 (green) from June 3$^{th}$ to July 3$^{th}$. (Bottom) hourly bias evolution (µg.m$^{-3}$) corresponding to MM5-CHIMERE forecasting system, KF and RAT04.

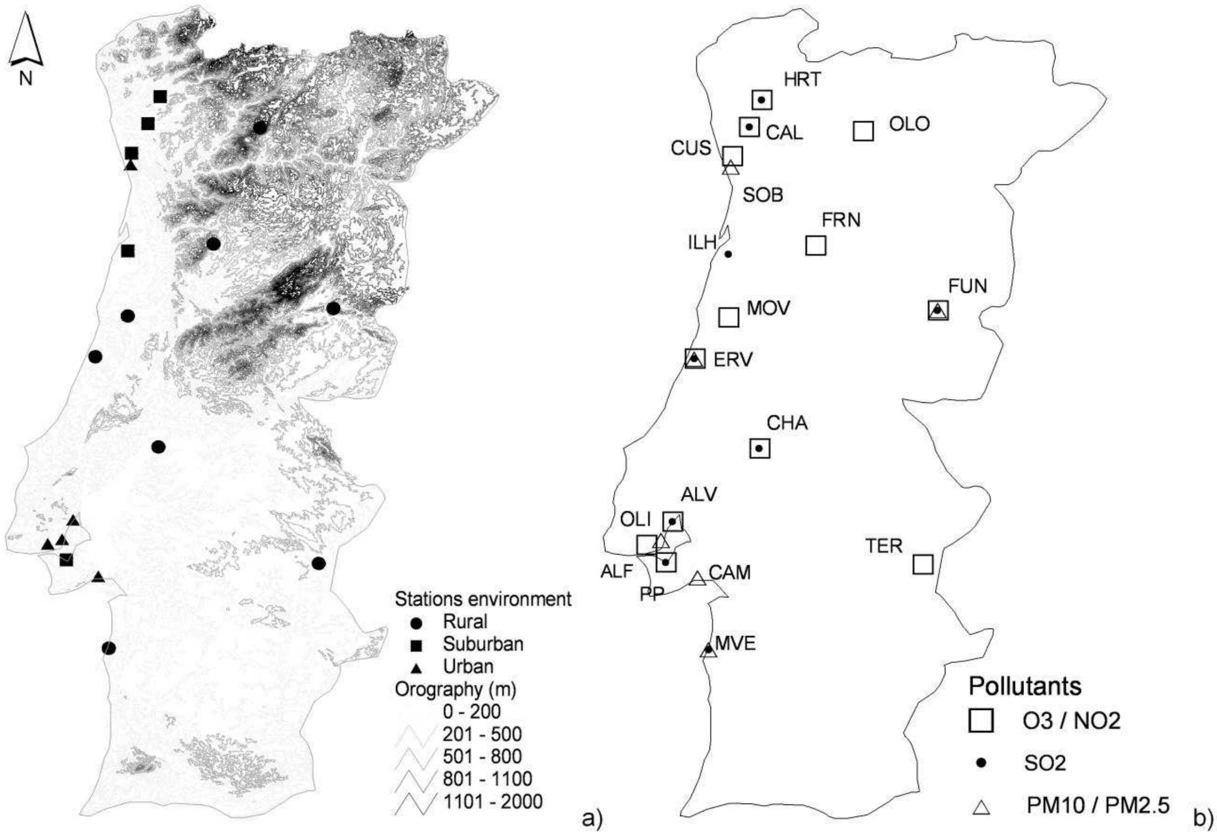

N
Stations environment
Rural
Suburban
Urban
Orography (m)
0 - 200
201 - 500
501 - 800
801 - 1100
1101 - 2000
a)
HRT
CAL
OLO
CUS
SOB
FRN
ILH
FUN
MOV
ERV
CHA
ALV
OLI
ALF
PP
CAM
TER
MVE
Pollutants
O3 / NO2
SO2
PM10 / PM2.5
b)

Figure 1.

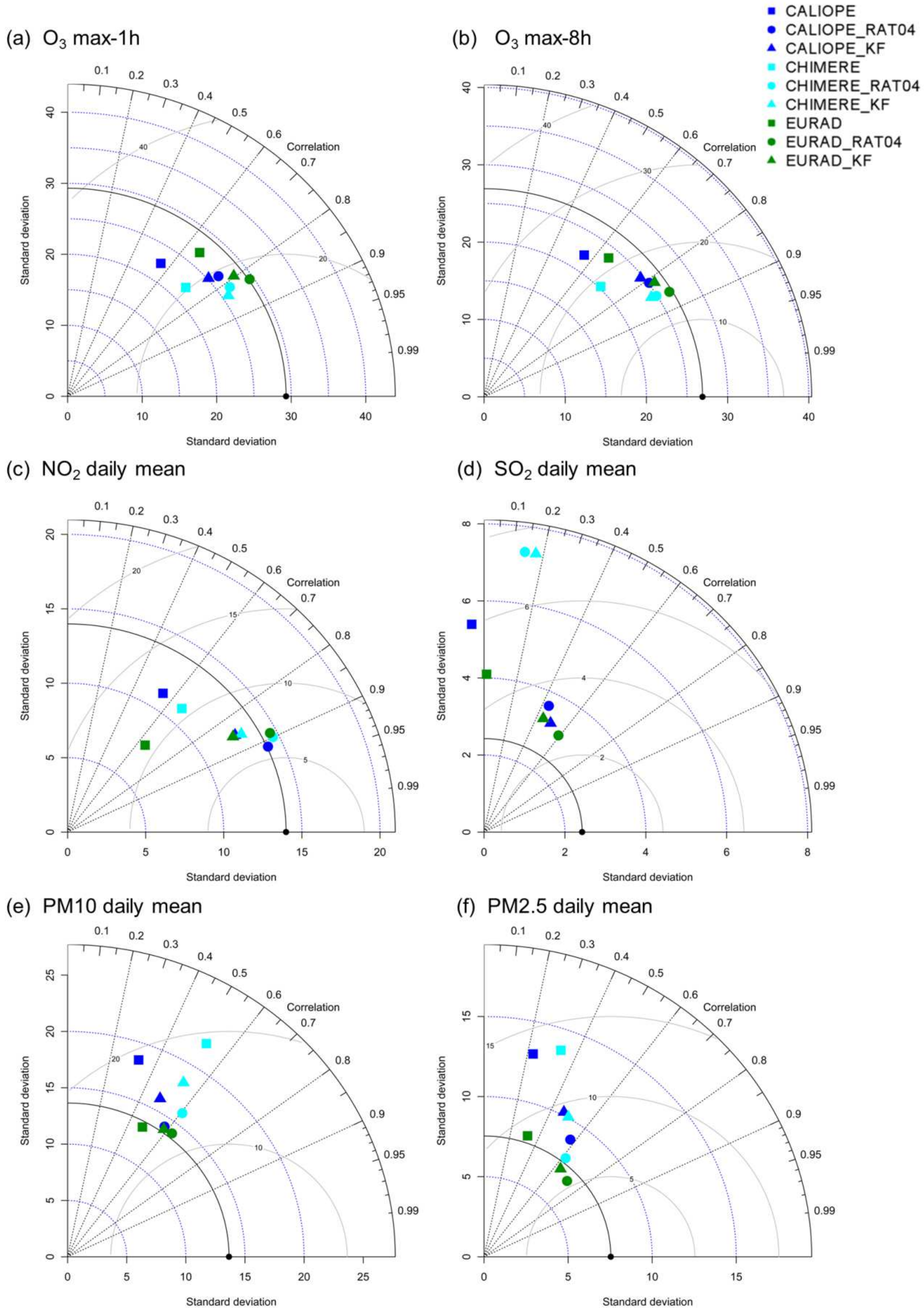
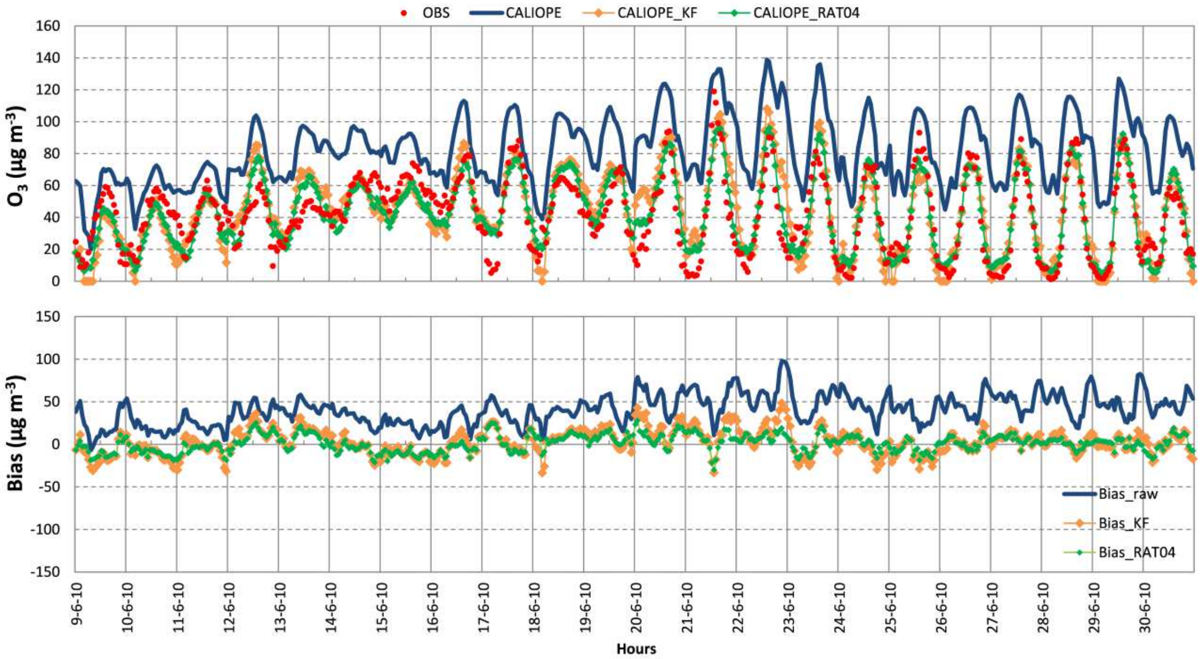

CALIOPE
CALIOPE_RAT04
CALIOPE_KF
CHIMERE
CHIMERE_RAT04
CHIMERE_KF
EURAD
EURAD_RAT04
EURAD_KF
(a) $O_3$ max-1h
(b) $O_3$ max-8h
(c) $NO_2$ daily mean
(d) $SO_2$ daily mean
(e) PM10 daily mean
(f) PM2.5 daily mean
Correlation
Standard deviation

Figure 2.



OBS
CALIOPE
CALIOPE_KF
CALIOPE_RAT04
O3 (µg m-3)
Bias (µg m-3)
Bias_raw
Bias_KF
Bias_RAT04
Hours
9-6-10
10-6-10
11-6-10
12-6-10
13-6-10
14-6-10
15-6-10
16-6-10
17-6-10
18-6-10
19-6-10
20-6-10
21-6-10
22-6-10
23-6-10
24-6-10
25-6-10
26-6-10
27-6-10
28-6-10
29-6-10
30-6-10


Figure 3.

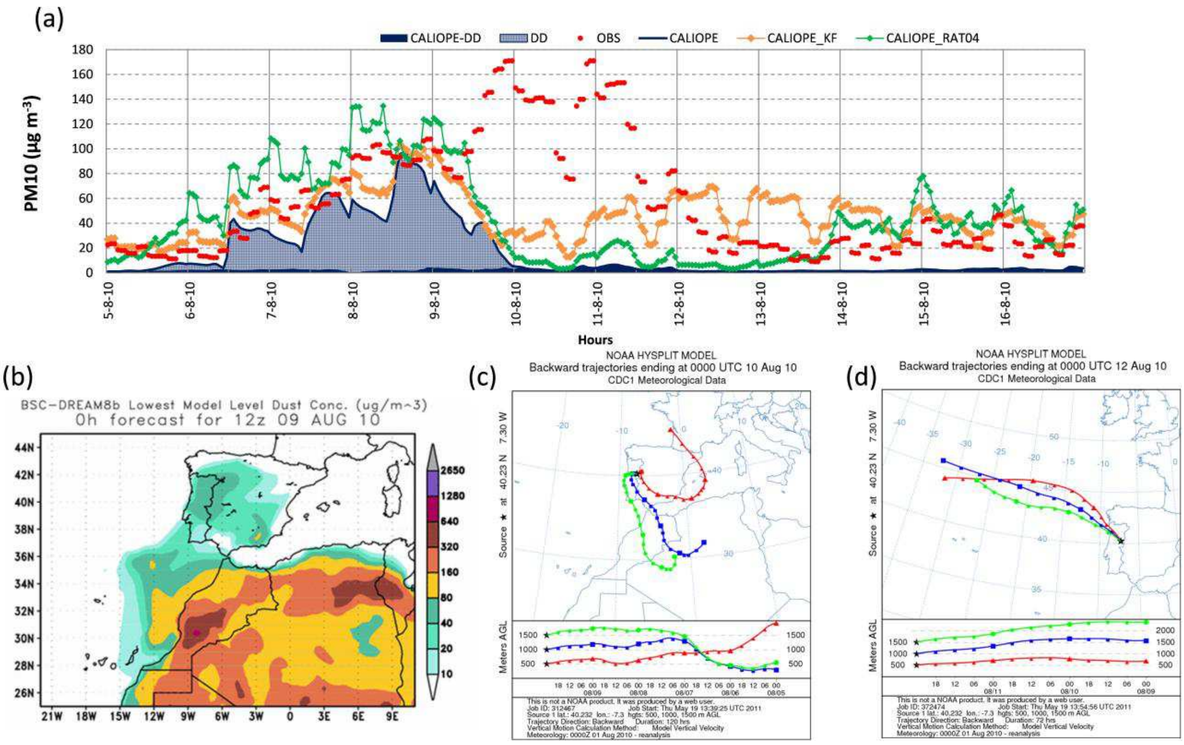

(a)
CALIOPE-DD
DD
OBS
CALIOPE
CALIOPE_KF
CALIOPE_RAT04
PM10 (µg m-3)
Hours
(b)
BSC-DREAM8b Lowest Model Level Dust Conc. (ug/m^3)
0h forecast for 12z 09 AUG 10
(c)
NOAA HYSPLIT MODEL
Backward trajectories ending at 0000 UTC 10 Aug 10
CDC1 Meteorological Data
(d)
NOAA HYSPLIT MODEL
Backward trajectories ending at 0000 UTC 12 Aug 10
CDC1 Meteorological Data

Figure 4.

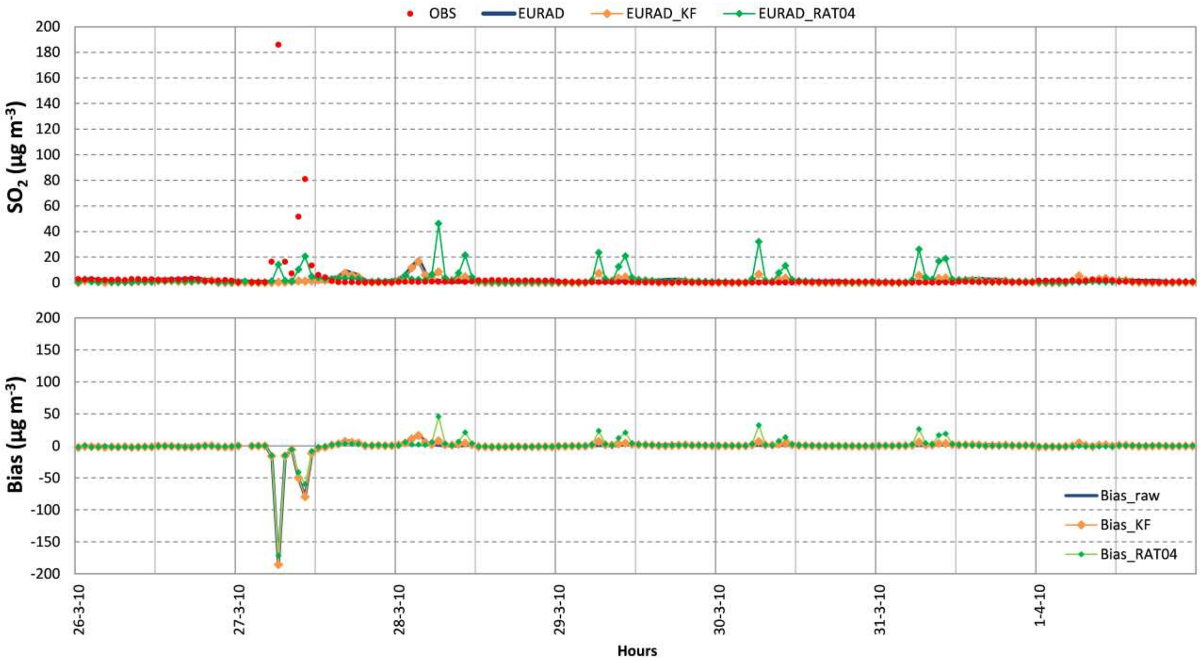

OBS
EURAD
EURAD_KF
EURAD_RAT04
SO2 (μg m-3)
Bias (μg m-3)
Bias_raw
Bias_KF
Bias_RAT04
26-3-10
27-3-10
28-3-10
29-3-10
30-3-10
31-3-10
1-4-10
Hours

Figure 5.

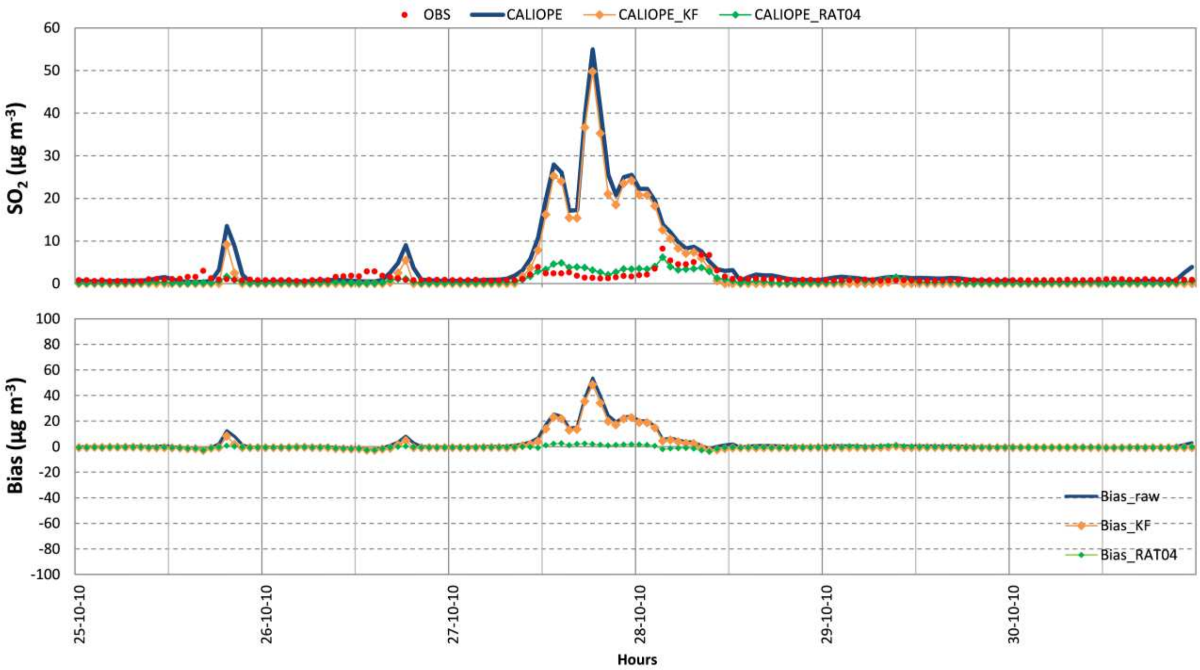
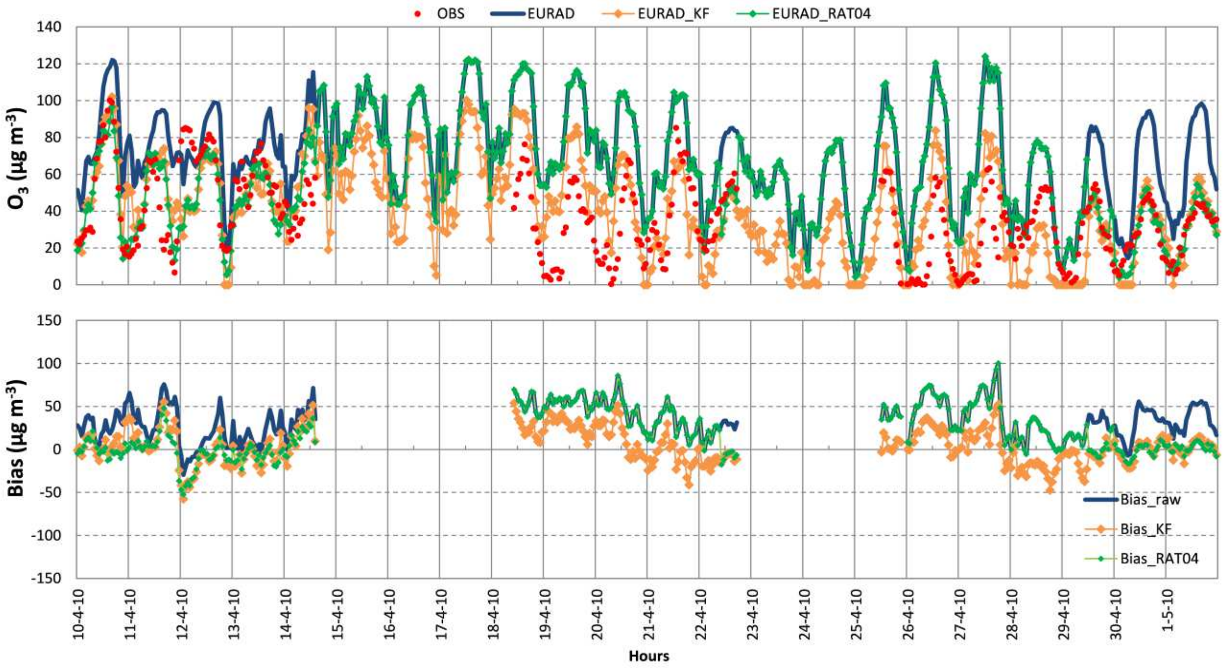
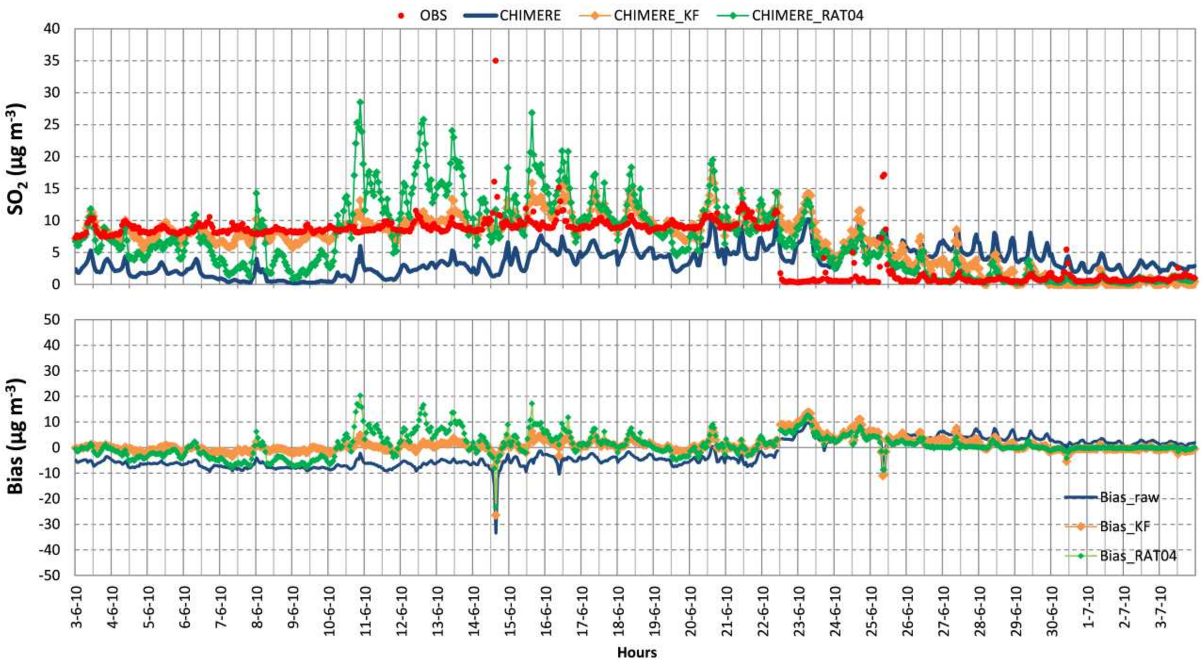

OBS
CALIOPE
CALIOPE_KF
CALIOPE_RAT04
SO2 (µg m-3)
Bias (µg m-3)
Bias_raw
Bias_KF
Bias_RAT04
25-10-10
26-10-10
27-10-10
28-10-10
29-10-10
30-10-10
Hours

Figure 6.



OBS
EURAD
EURAD_KF
EURAD_RAT04
O3 (µg m-3)
Bias (µg m-3)
Bias_raw
Bias_KF
Bias_RAT04
Hours


Figure 7.



OBS
CHIMERE
CHIMERE_KF
CHIMERE_RAT04
SO2 (µg m-3)
Bias (µg m-3)
Bias_raw
Bias_KF
Bias_RAT04
Hours


Figure 8.

Table 1: Configurations of the high-resolution air quality forecasting systems which routinely operate over mainland Portugal.

| | | CALIOPE | MM5-CHIMERE | MM5-EURAD-IM |
|---|---|---|---|---|
| | **Domains** | **Iberian Peninsula** | **Portugal** | **Portugal** |
| **Meteorology** | Model, version | WRF-ARW v3.0.1.1 | MM5v3.7 | MM5 v3.7 |
| | Horizontal resolution | 4 km x 4 km | 10 km x 10 km | 5 km x 5 km |
| | Nx, Ny ,Nz | 400, 400, 38 | 35, 70, 32 | 64, 121, 32 |
| | Mycrophysics | WSM-3 class | Relsner graupel | Relsner graupel |
| | Radiation | RRTM Dudhia scheme | RRTM Dudhia scheme | RRTM Dudhia scheme |
| | PBL | YSU | MRF PBL | MRF PBL |
| | LSM | Noah LSM | Five-layer LSM | Five-layer LSM |
| | Cumulus | Kain-Fritsch | Grell | Grell |
| | Initialization and boundary conditions | Nested from European forecast (NCEP-GFS) | Nested from European forecast | Nested from Iberian Peninsula forecast |
| **Emissions** | Database source (year) | HERMES+EMEP (2004)* | EMEP (2005)** | EMEP (2005)** |
| | Biogenic emissions | Offline<br>Parra et al. (2004) | Online<br>(Simpson et al., 1999) | Online<br>(Guenther et al. 1995). |
| **Chemistry** | Model, version | CMAQ v4.5 | CHIMERE 2006 | EURAD v4.2 |
| | Horizontal resolution | 4 km x 4 km | 10 km x 10 km | 5 km x 5 km |
| | Nx, Ny, Nz | 397, 397, 15 | 29, 58, 8 | 64, 121, 23 |
| | Chemical mechanism | CBM-IV<br>(Gery et al., 1989) | Reduced MELCHIOR<br>(Bessagnet et al., 2004) | RACM-MIM<br>(Geiger et al., 2003) |
| | Aerosol size distribution | Three modes | Eight bins | Three modes |
| | Inorganic aerosol | Thermodynamic ISORROPIA | Thermodynamic ISORROPIA | Thermodynamic |
| | Organic aerosol | Simplified SOA<br>Scheme<br>(Schell et al., 2001) | Simplified SOA scheme<br>(Bessagnet et al., 2005) | APC<br>SORGAM model |
| | Initialization and boundary condition | Nested from Europe (LMDz-INCA) | Nested from Europe | Nested from Iberian Peninsula |
| **Natural dust transport** | Model, version | BSC-DREAM8b | BSC-DREAM8b | BSC-DREAM8b |

*Emissions for Portugal and France are estimated following a top-down methodology from EMEP database. Emissions in Spain are calculated with a bottom-up approach (Baldasano et al., 2008b).

** Emissions for Portugal are estimated with a top-down desegregation methodology (Monteiro et al., 2007a)

Table 2: Estimated optimal error ratios for Kalman filter technique for $O_3$, $NO_2$, $SO_2$, PM10 and PM2.5 for the selected stations in the Portuguese mainland domain for 2010.

| **Pollutant** | **Period** | **CALIOPE** | **MM5-EURAD-IM** | **MM5-CHIMERE** |
|---|---|---|---|---|
| **$O_3$** | Winter | 0.07 | 0.03 | 0.05 |
| | Spring | 0.12 | 0.03 | 0.13 |
| | Summer | 0.12 | 0.03 | 0.09 |
| | Autumn | 0.12 | 0.05 | 0.09 |
| **$NO_2$** | Annual | 0.04 | 0.04 | 0.04 |
| **$SO_2$** | Annual | 0.20 | 0.14 | 0.13 |
| **PM10** | Annual | 0.08 | 0.04 | 0.17 |
| **PM2.5** | Annual | 0.07 | 0.02 | 0.08 |

Table 3: Annual categorical statistics for the three modelling system (CALIOPE, MM5-CHIMERE and MM5-EURAD-IM) (raw models) and for the two bias correction techniques, Kalman filter (KF) and multiplicative ratio (RAT04). The calculated statistics are the accuracy (A), the critical success index (CSI), the probability of detection (POD), the bias (B) and the false alarm ratio (FAR). The number in parentheses next to the statistic indicates the perfect score. Note that A, CSI, POD and FAR are in %. The thresholds (T) used to compute the statistics are chosen from the current European directive (2008/50/EC).

| | | **Raw models** | **KF** | **RAT04** |
|---|---|---|---|---|
| $O_3$ max-1h (T = 180 µg.m$^{-3}$) | | | | |
| (13 stations) | b(hits) | 5 | 16 | 48 |
| | a(false alarm) | 24 | 28 | 36 |
| | d(misses) | 148 | 137 | 105 |
| | c(correct negative) | 13326 | 13322 | 13314 |
| | A (100%) | 99 | 99 | 99 |
| | CSI (100%) | 3.0 | 9.0 | 25 |
| | POD (100%) | 3.0 | 10 | 31 |
| | B (1) | 0.2 | 0.3 | 0.6 |
| | FAR (0%) | 83 | 64 | 43 |
| $O_3$ max-8h (T = 120 µg.m$^{-3}$) | | | | |
| (13 stations) | b(hits) | 240 | 425 | 479 |
| | a(false alarm) | 371 | 254 | 266 |
| | d(misses) | 651 | 466 | 412 |
| | c(correct negative) | 12562 | 12679 | 12667 |
| | A (100%) | 93 | 95 | 95 |
| | CSI (100%) | 19 | 37 | 41 |
| | POD (100%) | 27 | 48 | 54 |
| | B (1) | 0.7 | 0.8 | 0.8 |
| | FAR (0%) | 61 | 37 | 36 |
| PM10 daily mean (T = 50 µg.m$^{-3}$) | | | | |
| (6 stations) | b(hits) | 65 | 70 | 92 |
| | a(false alarm) | 388 | 283 | 214 |
| | d(misses) | 139 | 134 | 112 |
| | c(correct negative) | 5282 | 5387 | 5456 |
| | A (100%) | 91 | 93 | 94 |
| | CSI (100%) | 11 | 14 | 22 |
| | POD (100%) | 32 | 34 | 45 |
| | B (1) | 2.2 | 1.7 | 1.5 |
| | FAR (0%) | 86 | 80 | 70 |