

# Assessment of Kalman filter bias-adjustment technique to improve the simulation of ground-level Ozone over Spain

V. Sicardi<sup>1\*</sup>, J. Ortiz<sup>1</sup>, A. Rincón<sup>1</sup>, O. Jorba<sup>1</sup>, M.T. Pay<sup>1</sup>, S. Gassó<sup>1,2</sup>, J.M. Baldasano<sup>1,2</sup>

[1] {Barcelona Supercomputing Center – Centro Nacional de Supercomputación}

[2] {Laboratory of Environmental Modeling, Universitat Politècnica de Catalunya (LMA-UPC)}

\*Correspondence to:

Valentina Sicardi

Earth Science Department, Barcelona Supercomputing-Center-Centro Nacional de Supercomputación (BSC-CNS). Jordi Girona 29, Edificio Nexus II, 08034 Barcelona, Spain.  
Tel: +34 934137724; Fax: +34 934016646.

[vsicardi@bsc.es](mailto:vsicardi@bsc.es)

## Abstract

The CALIOPE air quality modelling system has been used to diagnose ground level O<sub>3</sub> concentration for the year 2004, over the Iberian Peninsula. We investigate the improvement in the simulation of daily O<sub>3</sub> maximum by the use of a post-processing such as the Kalman filter bias-adjustment technique. The Kalman filter bias-adjustment technique is a recursive algorithm to optimally estimate bias-adjustment terms from previous measurements and model results. The bias-adjustment technique improved the simulation of daily O<sub>3</sub> maximum for the entire year and the all the stations considered over the whole domain. The corrected

simulation presents improvements in statistical indicators such as correlation, root mean square error, mean bias, and gross error. After the post-processing the exceedances of O<sub>3</sub> concentration limits, as established by the European Directive 2008/50/CE, are better reproduced and the uncertainty of the modelling system, as established by the European Directive 2008/50/CE, is reduced from 20% to 7.5%. Such uncertainty in the model results is under the established EU limit of the 50%. Significant improvements in the O<sub>3</sub> timing and amplitude of the daily cycle are also observed after the post-processing. The systematic improvements in the O<sub>3</sub> maximum simulations suggest that the Kalman filter post-processing method is a suitable technique to reproduce accurate estimate of ground-level O<sub>3</sub> concentration. The spatial and temporal characteristics of the adjusted O<sub>3</sub> concentrations are to be considered as a preliminary analysis towards the application of the bias adjustment technique to real near time O<sub>3</sub> forecasts.

**Keywords:** air quality, modelling system, ozone maximum, Kalman filter.

## 1. Introduction

Ozone pollution is one of the main concerns in Europe and in particular in the Mediterranean area (Baldasano et al., 1994, 2003; Millán et al., 1997; Gangoiti et al. 2001; Gerasopoulos et al. 2005; Jiménez et al., 2005a, 2006; Cristofanelli and Bonasoni, 2009). The Iberian Peninsula is characterized by a very complex topography (mountains, plane, desert zones, and large forest areas) and has a peculiar location between the Atlantic Ocean and the Mediterranean Sea. Dynamics of pollutant during summer and primary emission sources along the eastern coast and the central plateau of the IP determinate the location of highest level of ozone over the Iberian peninsula (Gangoiti et al., 2001; Jiménez et al. 2004; Stein et al., 2005; Jiménez et al., 2006; Baldasano et al, 2011). In the South of the Iberian Peninsula the frequent shipping

1 traffic and the high density of industry in the area generate important NO<sub>x</sub> emissions, but  
2 VOC concentrations are not high enough to produce O<sub>3</sub> (Marmer et al., 2009). The opposite  
3 happens over the western Mediterranean. Here, the complex coastal topography and  
4 surrounding mountain favors that the Mediterranean Sea acts as a reservoir of aged pollutants.  
5 The complexity of the terrain generates complex flow structure that leads to local and  
6 mesoscalar circulation phenomena. The mountains lead to complex vertical structural flow,  
7 with downslope warm wind and dry winds (Jimenez et al. 2008). The coastal zones are  
8 characterized by sea breeze flows mostly during summer. The on-shore winds advect  
9 pollutants from the high populated coasts inland, injecting at different altitude (Jorba et al.  
10 2003). In summer the meteorological condition (high pressure, stability, clear sky and high  
11 solar radiation intensity) enhances photochemical processes and emissions of biogenic  
12 volatile organic compounds to the atmosphere (NO<sub>x</sub> limited regime). These phenomena  
13 together with long-range transport of European air toward the Mediterranean basin (Lelieveld  
14 et al., 2002) could be important causes of the O<sub>3</sub> high concentration (Millán et al., 2002;  
15 Jiménez and Baldasano, 2004; Jiménez et al 2005b, 2008; Baldasano et al., 2011).

16

17 Elevated concentrations of tropospheric O<sub>3</sub> may lead to adverse effects on human health,  
18 agricultural crops, forests and materials (Brauer et al. 1997; West et al, 2007; WHO, 2008;  
19 Finlayson-Pitts, 2010).

20

21 The European air quality Directive 2008/50/EC defines target values and long-term objectives  
22 for the protection of human health and vegetation. The objective target value for human  
23 health protection is 120 µg/m<sup>3</sup> (calculated as daily maximum averaged over 8 hours running  
24 mean) and is not to be exceeded on more than 25 days per year averaged over 3 years. The

information threshold must be given to the population when hourly means exceed  $180 \mu\text{g}/\text{m}^3$ , and the alert threshold should be issued if hourly means exceed  $240 \mu\text{g}/\text{m}^3$ . Modelling techniques are valid and recognized means to monitor and predict the air quality and the reliability of such models is essential.

The CALIOPE air quality modeling system, namely WRF-ARW/HERMES-EMEP/CMAQ/BSC-DREAM8b, operatively applied under the Spanish government founded project CALIOPE (Baldasano et al., 2008a) has been used to diagnose the concentration of ground-level  $\text{O}_3$  over Spain for the year 2004. The CALIOPE modelling system has been used in previous studies to assess the air quality over Europe and Spain (Pay et al., 2010; Baldasano et al., 2010). Comparisons of  $\text{O}_3$  model results with observations have revealed that even though the temporal variability in  $\text{O}_3$  is well simulated, further improvement in the  $\text{O}_3$  simulations are still needed.

In order to produce more accurate simulations, that is model results that fit satisfactorily the observations, we post-process the model results with a bias-adjustment technique based on the Kalman filter (KF) (Kang et al., 2008). We carry out the simulation of ground level  $\text{O}_3$  concentration over the Iberian Peninsula (IP) domain, for the year 2004, and analyze the results produced by the modelling system before (hereafter model results) and after the application of the bias-adjustment technique (hereafter KF-output). The Kalman filter (Kalman, 1960), already applied in previous studies of atmospheric pollution modelling, reduces the error in the model results. The KF bias-adjustment technique, in this work, is applied to point stations only when observations are available. Both the model results and KF-output are tested over the available stations located throughout the whole domain.

Models always have uncertainties due to the data limitations and incomplete representation of the physical/chemical mechanisms; this introduces errors in the model results (Borrego, 2003; Chang and Hanna, 2004; Flemming and Stern, 2007). We compute the model uncertainty according to the European Directive 2008/50/CE and to the Spanish Real Decreto 102/2011 related to the air quality in Spain, to verify the accomplishment of this Directive and the improvement in this sense achieved by the post-process. An analysis of the main statistical parameters is also carried out together with an analysis of the daily cycle and their improvement achieved by the application of the post-processing.

In this work we investigate the reliability of the CALIOPE air quality system in reproducing O<sub>3</sub> daily maximum, and the improvement in the O<sub>3</sub> simulation by the application of a Kalman Filter based post-process. This study is an assessment of KF post-processing technique to improve the modelled O<sub>3</sub> concentration. It is a preliminary study to test the robustness of the post-process in order to apply it to our air quality forecast system CALIOPE (<http://www.bsc.es/caliope/>).

## **2 Methods**

### **2.1 Modelling System**

The CALIOPE air quality modelling system is a state-of-the-art modelling framework ([www.bsc.es/caliope](http://www.bsc.es/caliope)). It is a complex system that integrates the meteorological model: WRF-ARW; the emission model: HERMES; the chemical transport model: CMAQ; and the mineral dust atmospheric model: BSC-DREAM8b offline coupled in an air quality forecasting system (Baldasano et al., 2008a).

The Advanced Research Weather Research and Forecasting (WRF-ARW) model v3.0.1.1 (Michalakes et al., 2004; Skamarock et al., 2005; Skamarock and Klemp, 2008) provides the meteorology conditions. For the Iberian Peninsula (IP) domain WRF-ARW is configured with a grid of 397 x 397 points corresponding to a 4 km x 4 km horizontal resolution and 38  $\sigma$  vertical levels with 11 characterizing the planetary boundary layer (PBL). The model top is defined at 50 hPa to resolve the troposphere-stratosphere exchanges.

The Models-3 Community Multiscale Air Quality Modelling System (Models-3/CMAQ, Byun and Ching, 1999; Binkowski, 1999; Byun and Schere, 2006), v4.5 is used to study the behavior of air pollutants from regional to local scales. It includes gas, aerosol and heterogeneous chemistry.

The CMAQ horizontal grid resolution corresponds to that of WRF. Its vertical structure is obtained by a collapse from the 38 WRF layers to a total of 15 layers steadily increasing from the surface up to 50 hPa with a stronger density within the PBL. In order to provide adequate boundary and initial conditions to the IP domain the CALIOPE modelling system is initially run on a regional scale (12 km x 12 km in space and 1 hour in time) to model the European domain (mother domain). Chemical boundary conditions for this domain are provided by the global climate chemistry model LMDz-INCA2 (Hauglustaine et al., 2004; Folberth et al., 2006). A one-way nesting is then performed to retrieve the meteorological and chemical conditions for the Iberian Peninsula domain (Fig. 1). The resolution used for the IP domain is of 4km x 4km. Such high resolution is a key factor to accurately simulate air pollution issues,

1 especially over complex topography (Jiménez et al., 2006) and meteorology patterns  
2 (Baldasano et al., 1994; Millán et al., 2002).

3  
4 The HERMES model (Baldasano et al., 2008b) uses information and state-of-the-art  
5 methodologies for emission estimations. It calculates emissions by sector-specific sources or  
6 by individual installations and stacks. Emissions used for Spain are derived from the  
7 aggregation in space from 1 km x 1 km dataset to 4 km x 4 km. Raw emission data are  
8 processed by HERMES in order to provide a comprehensive description of the emissions to  
9 the air quality model. In this study the emissions are expressed in CBM-IV speciation.  
10 Regarding to biogenic emissions, HERMES calculates the biogenic volatile organic  
11 compounds (bVOC) from vegetation. Three categories of bVOC are estimated according to  
12 their reactivity: isoprene, monoterpenes and other volatile organic compounds (OVOC). The  
13 model considers the influence of temperature and photosynthetically active radiation (PAR)  
14 by Guenther et al. (1995) algorithms, according to Parra et al. (2004, 2006). Emission factors  
15 for each individual vegetal species are associated with emitter land-use categories. The  
16 HERMES model uses 22 land-use categories; these land use categories for each grid cell are  
17 obtained from CORINE Land Cover 2000 map (44 land-use categories) according to Arévalo  
18 et al. (2004). In the updated version of HERMES model used in the present work, the  
19 influence of seasonality in the emission of bVOC is introduced through an environmental  
20 correction factor following Staudt et al. (2000) and Steinbrecher et al. (2009).

21  
22 The CALIOPE system comprises as well The Dust REgional Atmospheric Model (BSC-  
23 DREAM8b), designed to simulate and predict the atmospheric cycle of mineral dust  
24 (Nickovic et al., 2001; Pérez et al., 2006a,b).

The CALIOPE modelling system has been evaluated in depth; a detailed evaluation of the European domain is presented in Pay et al. (2010), and for the Iberian Peninsula domain in Baldasano et al (2010). For a detailed description of the evaluation of this system we refer to these studies. The simulation used in this study has been carried out for the full year 2004 over the Iberian Peninsula domain, hereafter referred as IP. The statistical model skills have been analyzed for the whole year 2004. For the daily cycle analysis, only the data from the ozone campaign (April to September) have been analyzed. The ground level O<sub>3</sub> concentration has been taken into account as representative of the surface concentration and compared with observations.

## 2.2 Observations

The model simulations are tested against the hourly observations from a network of 82 stations (hereafter referred as RedESP) covering the entire Iberian Peninsula domain (Fig. 1). The hourly measurements provided by “Centro de Estudios Ambientales del Mediterraneo (CEAM)”, were subjected to a preliminary quality control. Monitoring data were available for the full year 2004, but only stations with temporal coverage of 85% were taken into account and compared with the model results. The temporal minimum data coverage of 90% as recommended in the Directive 2008/50/EC, refers to the data without calibration and maintenance, thus it reduces to minimum data coverage of 85 % after the quality control (Garber et al., 2002). The uncertainty of the measurements is within the limit established by the European Directive 2008/50/EC equal to 15%.

The air quality monitoring stations are classified as urban, suburban and rural according to their locations (Garber et al., 2002; Annex III of the Directive 2008/50/EC). To extract the model results corresponding to the stations considered we apply two procedures: for the urban



and suburban stations the mean value of the corresponding 4 km x 4 km grid cell has been taken. For the rural stations it has been done a bilinear interpolation with the closest cells. One limitation of this method is that, for example, if two urban stations are located in the same grid cell, the modelled values for these stations will be the same. Nevertheless, the accuracy related to the way of retrieving the modelled data, does not affect the main results presented in our work, as we are looking at the stations globally.

We compute the daily maximum of the hourly data (max 1-hr) and the daily maximum of the 8 hours running average (max 8-hr) for each station over the full year and compare them with the observations. When computing the daily maximum, for both hourly and 8-hr averaged data, only the days with a minimum of 75% of hourly data have been taken into account (according to the Spanish national law for air quality, Real Decreto 102/2011). Following this criterion 24 days of the entire year have been eliminated in the case of the max-1hr calculation and 27 days in the case of max 8-hr calculation. In this work no focus on the type of stations has been done. We analyze the ensemble of stations, without going into details in the stations type. For a detailed analysis of the performance of different kind of stations of the RedESP for the year 2004, the reader can refer to Baldasano et al. 2011.

### **3 Post - processing and definition of the uncertainty**

#### **3.1 The Kalman filter**

The model results have been post-processed by the application of a bias-adjustment technique based on Kalman filter. The Kalman filter (KF) is a post-processing method that uses recent estimates and measurements to revise and improve the current estimate (Kalman, 1960). The

Kalman filter is a linear, adaptive, recursive and optimal algorithm; it works by a mechanism of prediction and correction of the bias (between the model results and the observations) at each time steps. The filter estimates the systematic component of the simulation bias, which is present in model results (Delle Monache et al., 2006). Once the future bias has been estimated, this bias is removed from the model results to produce improved model results, which are more similar to the observations.

The application of the KF bias adjustment technique involves two main steps:

- Estimation of the bias at the next time step, using all available data (model and observed) at the current time step.
- Correction of the modeled estimation at the next time step with the recent bias found in the previous step

The Kalman filter bias-adjustment is a well known and widely used technique. However, for the sake of completeness a description of the method is reported. We follow Delle Monache et al., 2006 for the description of the filter.

The state of the unknown process at time  $t$ , for example the bias between the simulation and the true concentration of a variable, is related to the state at prior time  $(t - \Delta t)$  through the following equation:

$$x_{t|t} = x_{t-\Delta t|t-2\Delta t} + \eta_{t-\Delta t}$$

Eq. 1

where  $\eta$  is a white noise term and assumed to be uncorrelated in time, and is normally distributed with zero mean and variance  $\sigma_\eta^2$ ,  $\Delta t$  is a time lag, and  $t/t - \Delta t$  implies dependency of the variable at time  $t$  on values at time  $t - \Delta t$ .

The bias  $x_t$  is not observable, but is related to the measurable bias  $y_t$ , which is the differences between forecasts and observations. Still, it is to be considered that  $y_t$  is corrupted from the true bias  $x_t$  by a random error  $\varepsilon_t$ . Therefore  $y_t$  is related to  $x_t$  by:

$$y_t = x_t + \varepsilon_t = x_{t-\Delta t} + \eta_{t-\Delta t} + \varepsilon_t \quad \text{Eq. 2}$$

where  $\varepsilon_t$  represents the measurement noise and is assumed to be uncorrelated in time and normally distributed with zero mean and variance  $\sigma_\varepsilon^2$ ;  $\eta_{t-\Delta t}$  represents the process noise.

According to Kalman, 1960, the optimal recursive predictor of  $x_t$  (derived by minimizing the expected mean square error) can be expressed as:

$$\hat{x}_{t+\Delta t|t} = \hat{x}_{t|t-\Delta t} + \beta_{t|t-\Delta t} (y_t - \hat{x}_{t|t-\Delta t}) \quad \text{Eq. 3}$$

where the hat (^) indicates the estimate of the variable and  $\beta$  is the weighting factor, called the Kalman gain, which is recursively computed as follows:

$$\beta = \frac{p_{t-\Delta t|t-2\Delta t} + \sigma_\eta^2}{p_{t-\Delta t|2\Delta t} + \sigma_\eta^2 + \sigma_\varepsilon^2} \quad \text{Eq. 4}$$

With  $p$  the expected mean square error:

$$p_{t|t-\Delta t} = (p_{t-\Delta t|t-2\Delta t} + \sigma_{\eta}^2)(1 - \beta_{t|t-\Delta t}) \quad \text{Eq. 5}$$

Given the forecast and observation time series, the estimates of  $\sigma_{\eta}^2$  and  $\sigma_{\epsilon}^2$ , and the initial estimate of state  $\hat{x}$  and  $p$  at time zero, KF can recursively generate  $\hat{x}_{t+\Delta t}$  (Delle Monache et al., 2006). After that the bias  $\hat{x}_{t+\Delta t}$  is calculated, the new simulation results, “KF-output”, can be simply retrieved by:

$$KF_{t+\Delta t} = z_{t+\Delta t} - \hat{x}_{t+\Delta t|t} \quad \text{Eq. 6}$$

Where  $z_{t+\Delta t}$  is the mode result for the next time step.

The Kalman Filter has been applied in previous O<sub>3</sub> studies (van Loon, 1997, 2000; Sagers, 2002; Hanea et al., 2004; Eben et al., 2005; Delle Monache et al., 2006, 2008; Kang et al., 2008). In this work we apply it following Delle Monache, 2006. The Kalman Filter is applied only to discrete points (our 82 monitoring stations) to improve the simulation at those points. In other words: given the model simulations at those 82 points stations and the observations at the same points, the calculation of the bias is carried out each hour and the simulation of the next hour is recursively corrected. For the correction of the O<sub>3</sub> concentration of the day  $n$ , past observations and model results up the day  $n$  are used. In such way we improve our modelled O<sub>3</sub> concentrations on day  $n$ . The KF has been applied to the original hourly O<sub>3</sub> values to develop a KF-adjusted hourly O<sub>3</sub> time series, which is then used to compute the daily maximum 8-hr and 1-hr O<sub>3</sub> from the adjusted hourly time series. The mean of the adjusted O<sub>3</sub> concentrations for all the stations, is then considered as representative of the whole domain.

1 The ratio  $\frac{\sigma_{\eta}^2}{\sigma_{\epsilon}^2}$  represents the so called “error ratio”. In Delle Monache (2006 and 2008) is  
2 pointed out that the error ratio is a crucial parameter to be calculated in the application of this  
3 post-process. It indicates the relative weighting of the observed and simulated ozone values.  
4 If the ratio is too high, the error white-noise variance ( $\sigma_{\epsilon}^2$ ) will be relatively small compared  
5 to the true bias white-noise variance ( $\sigma_{\eta}^2$ ). Therefore, the filter will put excessive confidence  
6 on the previous simulated values, and the predicted bias will respond very quickly to previous  
7 calculated errors. On the other hand, if the ratio is too low, the predicted bias will change too  
8 slowly over time. Consequently, there exists an optimal value for the ratio, which can be  
9 estimated by evaluating the filter performance in different situations.

10  
11 To test which could be the best error ratio for our O<sub>3</sub> simulation, error ratios ranging from  
12 0.01 to 2 have been selected for all the stations considered over the entire year 2004. We  
13 consider that, in the case of O<sub>3</sub>, due to its dynamics, it is important to take into account the  
14 variations of the ratio over the seasons. Therefore, we calculate root mean square error and  
15 correlation coefficient values over the four seasons to gauge the impact of different error ratio  
16 values on the model performance (Fig. 2). The optimal error ratio will minimize the root  
17 mean square error and maximize the correlation coefficient for all the stations. Based on this  
18 approach we finally use one optimal value varying seasonally for all the stations (Table 1).  
19 According to Delle Monache et al., 2006; Kang et al., 2008; 2010, O<sub>3</sub> simulation over  
20 different areas (e.g. rural versus urban), or for different model results may have different  
21 optimal ratio value. However, we assume valid the hypothesis of spatial uniformity as in  
22 Kang et al, 2008, and consider that the error ratio at the single station does not have a  
23 significant impact on the global results.

**Comment [v1]:** This part has been largely changed

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22

**3.2 Uncertainty according to European Directive 2008/50/CE**

The model uncertainties can be associated with model formulation regarding misrepresentation of atmospheric dynamics and chemistry, numerical solutions, choice of modelling domain and grid structure; with the model input regarding emissions, meteorological data; or with the stochastic processes that are not known (Borrego et al., 2008; Chang and Hanna, 2004). Reducing the uncertainties is fundamental in order to obtain high quality model results.

In this work we refer to “model uncertainty” as defined in the European Directive 2008/50/CE: “the uncertainty for modelling is defined as the maximum deviation of the measured and calculated concentration levels for 90% of individual monitoring points, over the period considered, by the limit value (or target value in the case of ozone) without taking into account the timing of the events....”. The “without timing” in the above definition, implies that the time factor is not taken into consideration, nor the sequences of the events (e.g. when an exceedance of concentration limit occurs), which is fundamental in the case of air quality modelling systems.

The Guidance on the use of models for the European air quality directive proposes the statistical calculations needed to calculate the uncertainty according to the Directive and to validate the air quality models (Denby et al., 2010). The model uncertainty calculated for a single station is defined mathematically as the “Relative Directive Error” (RDE) (Equation 7):

$$RDE = \frac{|O_{LV} - M_{LV}|}{LV} \quad \text{Eq.7}$$

Where  $O_{LV}$  is the closest observed concentration to the limit value concentration (LV) and  $M_{LV}$  is the corresponding ranked modelled concentration. The maximum of this value found at 90% of the available stations is then the Maximum Relative Directive Error (MRDE). The MRDE represents the uncertainty of the model in the evaluation of the air quality. The Directive considers that the model uncertainty related to  $O_3$  simulation must be  $\leq 50\%$ ; therefore models with uncertainty minor to 50% will meet the EU directive requirements. The Directive also requires an uncertainty  $\leq 15\%$  for measurements.

By the calculation of the Maximum Relative Directive Error (MRDE), we verify that our model system accomplishes the uncertainty limits according to the European Directive 2008/50/CE and according to the Spanish Real Decreto 102/2011, related to the air quality in Spain.

### 3.3 Statistics

The model skills are evaluated over the maximum hourly (max 1-hr) and 8-hr running mean  $O_3$  concentration (max 8-hr) in order to verify the ability of the model to reproduce the day-by-day maximum variation. The model skills are also evaluated over the mean daily cycle, as average of all the concentration of all the available days for the 24 hours, in order to verify the ability to capture the day to night variations (Appel et al., 2007; van Loon et al., 2007; Vautard et al., 2007). In addition, the exceedance of the threshold concentration has been

considered by the analysis of a contingency table. All the statistics are calculated over the full year 2004; the daily cycle is analyzed only for the O<sub>3</sub> campaign (April to September).

Since the statistical parameters both for the max 1-hr and max 8-hr consistently show similar results for stations located in urban, suburban and rural areas, we don't analyze the statistics for the different stations types, but we focus rather on the general level of improvement due to the post-process. Specific results for the different type of stations are here omitted. However, for details on the performance of the model for different stations type, refer to Baldasano, 2011.

The model evaluation is carried out using classical statistical indicators for the ground level daily maximum 1-hr and 8-hr O<sub>3</sub> concentration (Dennis et al., 2010). Namely the statistic metrics used are: Mean Bias (MB), the Root Mean Square Error (RMSE), the correlation coefficient (COR). Additionally we compute the Mean Normalize Bias Error (MNBE) and the Mean Normalized Gross Error (MNGE) according to the model evaluation objectives suggested by the United States Environmental Protection Agency (US-EPA). The US-EPA establishes that these parameters should be:  $MNBE \leq 15 \%$  and  $MNGE \leq 35 \%$  (US-EPA, 2007, 2009).

The categorical statistical skills are also evaluated (Kang et al., 2005, Eder et al., 2006), by calculating the Accuracy (A), which measures the fraction of exceedances and no-exceedances correctly predicted; the Bias (B), which measures if there are under-predictions or over-predictions; the Probability of Detection (POD), which measures what fraction of the



exceedances are correctly predicted; the False Alarm Ratio (FAR), which measures what fraction of the predicted exceedances did not occur; the Probability of False Detection (POFD), which measures what fraction of the observed no-exceedances are incorrectly predicted; and the Critical Success Index (CSI), which measures how well both model exceedances and observed exceedances are predicted.

## **4 Results and Discussion**

### **4.1 General Performance**

The CALIOPE air quality system reproduces the temporal variability of O<sub>3</sub> properly, as in previous studies (Gonçalves et al., 2009; Pay et al., 2010; Baldasano et al., 2010). In Baldasano et al., 2011, the highest mean concentrations are showed to be located in the open Mediterranean Sea (up to 90 µg/m<sup>3</sup>) and the Spanish Mediterranean coast (~80 µg/m<sup>3</sup>). Such concentrations are favoured by the prevailing intense photochemistry in the region, the local formation and transport the persistent subsidence over the region, and the low O<sub>3</sub> dry deposition over sea (EEA, 2005; Lelieveld et al., 2002; Vautard et al., 2005b; Gerasopoulos et al., 2005; Cristofanelli and Bonasoni, 2009). The Spanish oceanic region in the north and north-western Spain, characterized by high frequency of precipitation show lower O<sub>3</sub> levels than the Spanish arid and Mediterranean areas. O<sub>3</sub> is found lowest (~50 µg/m<sup>3</sup>) in either regions of low precursor emissions (northern and southern plateaus) or in areas affected by large NO-to-NO<sub>2</sub> concentration ratios (e.g., zones of intense on-road and ship traffic), such as the major Spanish metropolitan cities (i.e., Madrid, Barcelona, Valencia, Sevilla), highways of high traffic flow and the strait of Gibraltar (Baldasano et al., 2011). Considering the totality of 82 stations of the RedESP Baldasano et al., 2011, found that the mean O<sub>3</sub> concentration for

2004 was  $57.7 \mu\text{g}/\text{m}^3$ . The correlation coefficient respect to the results obtained with CALIOPE modelling system was 0.75 and the RMSE  $24.1 \mu\text{g}/\text{m}^3$ .

In Figure 3 the average of all the observations (red line) is overall well represented by both the model results (dark blue) and the KF-output (bright blue) for max-1hr and max-8hr. The model results tend to underestimate the max 1-hr concentration mostly in winter/autumn months (January to April and September to December). It is a known problem of CMAQ to misrepresent  $\text{O}_3$  variability in winter months due to the difficulties in the reproduction of the cross stratosphere-troposphere exchanges (Lam and Fu, 2009; Pay et al., 2010).

The KF is able to correct the simulated concentrations, independently on the magnitude of the bias. The filter adjusts satisfactorily the model results even when the most severe deviation of the model results from the measurements is observed, like during the months March – May. The filter is able to adjust different magnitudes of bias, without the necessity of recalibration; it shows ability to correct both periodic and episodic events.

We have to take into account that the observed bias is not solely due to the measurement or the model, but also to the limitation in the comparison methodology. This is what the filter method is called to mitigate. This is particularly true for stations type such as the urban and the industrial, for which the model result is representative of an area of  $4 \times 4 \text{ km}$ , while in the reality the measurement value is representative of a couple of kilometers. Nevertheless, the improvement of the simulation after the application of the KF, for these types of stations, is of the same order of magnitude than for the rural stations.

1  
2 The results in fig. 3 show that the O<sub>3</sub> chemistry is well represented in summer, and the overall  
3 performance improves notably after applying the Kalman filter. The overall O<sub>3</sub> concentration  
4 shows improvement after the application of the post-processing (bright-blue line): the O<sub>3</sub>  
5 concentration improves in reproducing the observed values even for those months in which  
6 the model results fail to reproduce the observed concentration. In the same figure are plotted  
7 the mean bias, for the model results (dark blue line with triangle) and for the KF-output (light  
8 blue line with triangle). It is notable that in the KF-output the bias is reduced mostly for the  
9 winter months, while for the summer months the bias is already quite low due to the good  
10 representation of the summer O<sub>3</sub> behaviour.

11  
12 The main statistics are summarized in Table 2. The statistical parameters both for the max 1-  
13 hr and max 8-hr consistently show similar results for stations located in urban, suburban and  
14 rural areas. We don't analyze the statistics for the different stations types, but we focus rather  
15 on the level of improvement due to the post-process. The level of improvement in the results  
16 has been found to be independent on the type of stations. Specific results for the different type  
17 of stations are here omitted. However, for details on the performance of the model for  
18 different stations type, refer to Baldasano et al., 2011.

19  
20 All the considered statistics show improvements after the application of the KF, both for max  
21 1-hr and max 8-hr. The RMSE improves of ~30% for max 1hr and ~40% for max 8-hr. A  
22 minor degree of improvement is observed for the correlation coefficients, which are anyway  
23 considerable high already for the model results: improvement of 30% for the max 1-hr and of  
24 ~32% for the max 8-hr. Taking into account the threshold limits established by the US-EPA

1 (MNBE  $\leq$  15% and MNGE  $\leq$  35 %) we see that the model results would meet the US-EPA  
2 target for MNBE for the max 1-hr, but not for the max 8-hr. After the application of the KF  
3 this US-EPA target would be accomplished also for the max 8-hr. In the same way the MNGE  
4 target for the max 1-hr would be accomplished, but not for the max 8-hr. After the application  
5 of the KF, the MNGE target would be met also for the max 8-hr. The improvements in the  
6 statistical metrics are easily visible when looking at their spatial distribution (Fig. 4a and Fig.  
7 4b). All the stations, independently on the type of station, show better statistical parameters  
8 for the KF-output. Generally speaking this finding is true for both the max 1-hr and max 8-hr.  
9 In general no spatial patterns in the model simulation's ability are observed. For all the  
10 available stations, independently on their locations, the degree of improvement after the  
11 application of the KF is comparable. This is due to the fact that the post-process is applied to  
12 the single stations, independently on their location.

13  
14 The found statistics are comparable with the results of similar studies. At the present only  
15 one other study has been conducted with the same modelling system CALIOPE and the same  
16 KF bias adjustment technique, but on a different domain, Portugal. In Borrego et al., 2011,  
17 after the application of the KF to 13 monitoring stations in Portugal, the correlation for all the  
18 stations improves from 0.75 to 0.85 for max 1-hr (improvement of ~12%) and 0.76 to 0.86  
19 (improvement of ~13%) for max 8-hr. In the same study the improvement in the RMSE is of  
20 ~30%. The results of other two studies have been compared with ours since the same KF  
21 algorithm has been applied. In Kang et al., 2008 (over 1000 stations over the US domain) the  
22 improvement in RMSE after the application of the KF is from  $\sim 27 \mu\text{g}/\text{m}^3$  to  $\sim 19 \mu\text{g}/\text{m}^3$   
23 (improvement of ~30%) for the max 8-hr. In Delle Monache et al., 2008, (358 monitoring  
24 stations along the East of US domain) the COR improves from  $\sim 0.62$  to  $\sim 0.75$  (improvement

of ~20%) for the daily max, and the RMSE improves from ~39  $\mu\text{g}/\text{m}^3$  to ~27  $\mu\text{g}/\text{m}^3$  (improvement of ~30%). It is easily understood that when the statistics are globally calculated, the number of the stations is crucial. In the same way, the improvements for the max 8-hr are always higher than for the max 1-hr.

We summarize the improvements in statistics between model results and KF-output using the Taylor diagrams (Taylor, 2001). The Taylor diagram allows us to gauge the improvements between the model results and the KF-output by the means of the visualization in a single polar plot of RMSE, correlation coefficient, and standard deviation for all the stations (Fig. 5). The standard deviation is not normalized to avoid masking the difference between station types. Almost all the stations for the model results (dark blue symbols) have a correlation coefficient in the range 0.6 to 0.7, which becomes 0.7 to 0.9 after applying the KF (light blue symbols). This result is valid for all the stations, independently on their type and geographical location. Also, the standard deviation improves for the KF-output: high correlation coefficient and low standard deviation indicates that the observed variability is well captured. These findings are valid both for the max 1-hr and max 8-hr.

Improvements by the application of the KF are much noticeable by looking at the scatter plot of models versus observations for all the stations over the whole year (Fig. 6). The KF- output (left side of Fig. 6) respect to the model results (right side of Fig. 6) fit more satisfactorily with the observations, as reflected by the reduction of the scattered area in Figure 6. On the same figure are reported the  $r^2$  of the best fit line, the slope and the intercept. After the application of the KF the all the points fit adequately the regression line.

## 4.2 Categorical performance

We perform a categorical analysis of our results in order to highlight the ability of the modelling system to detect the  $O_3$  concentrations that exceed the air quality target for  $O_3$  as established in the EU Directive 2008/50/EC. A visual evaluation of the model performance regarding the exceedance limit is provided in Fig. 6. The figure represents the categorical evaluation of the model results and KF-output, for the max 1-hr and max 8-hr for all the 82 stations. The letters on the plots represent the variable used to formulate the categorical metrics, where  $a$  are the exceedances that did occur and were simulated by the model (hits);  $b$  are the exceedances that did not occur but were simulated by the model (false alarms);  $c$  are the exceedances that did occur but were not simulated by the model (misses), and  $d$  are the exceedances that did not occur and were not simulated by the model (correct negatives). These variables, together with some categorical metrics (Table 3), help to enlighten the improvements carried by the application of the Kalman filter.

In Fig. 6 the higher the aggregation of points, the more the model simulation matches the observations. For the KF-output (light blue points on Fig. 6) most of the points are grouped around the line, indicating better correspondence between the model results and the observed values. The number of hits ( $a$  in the Fig. 6) increases substantially after the KF post-processing for both the max 1-hr and max 8-hr. The improvement in the *hits* detection improves up ~35% for max 8-hr, comparable with the results of Kang et al., 2008, in which the improvement is ~30%. In case of the max 1-hr the improvement is higher, due to the fact that the model results give a very poor hits performance. The false alarms decrease ( $b$  in Fig. 6) after the KF post-processing for the max 8-hr, while for the max 1-hr they remain in the same order of magnitude. The misses ( $c$  in Fig. 6) decrease after the application of the KF

both for the max 1-hr and max 8-hr. The correct negatives ( $d$  in Fig. 6) are well represented by the model and improve slightly after the application of the KF post-processing, see Table 3.

We must take into account that the categorical statistics depend on the number of exceedances or no-exceedances captured by the model; therefore caution is needed when interpreted. In particular in cases as ours, in which the number of exceedances are about the 10%, we must have caution when analyzing the Accuracy (A) (ideally 1) that measures the percentage of simulations that correctly reproduce an exceedance or no-exceedance. The Accuracy is already very high for the model results and no improvements are observed after the post-processing; this is due to the few exceedances observed, respect to the total. The Bias (ideally 1) indicates if our hindcasts are over-predicted (false positive) or under-predicted (correct negative) and improves after the KF post-processing. The Bias improves hardly after the application of the KF. We attribute the low improvement in the Bias in the KF-output to a poorer performance in detecting the exceedances in already in the model results.

To evaluate how many times the model simulates the exceedances, which actually did not occur, we look at the value of False Alarm Ratio (FAR). The application of the KF reduces of almost the half the value of the FAR for the max 1-hr. This finding indicates the ability of the post-processing to reduce the number of projected false alarms. The FAR index decreases for the max 8-hr as well, even if at a less important rate. The Probability of Detection (POD) improves strongly after the KF post-processing for both the max 1-hr and max 8-hr. This means that by the application of the KF the exceedances would be captured by the modelling system with significantly certainty. The Probability of False Detection (POFD) improves after

post-processing the data and in this way the false alerts for the population would be reduced. Finally the Critical Success Index (CSI), which indicates how well both the observed exceedances and the false exceedances are projected, improves after the application of the KF. Unlike the POD and the FAR, the CSI takes into account both false alarms and missed events, and it is therefore a more balanced score.

For the model results, while the mean variability of  $O_3$  has been satisfactorily well represented, in case of episodic conditions, the  $O_3$  levels are not so well represented. However, the KF-bias adjustment brings a significant improvement in terms of CSI. The model, without the KF correction would not be suitable for operational purpose. On the contrary the KF adjusted results would be adequate for operational episodic  $O_3$  event. Therefore the KF-output are adequate as operational means to inform and alert the population, as indicated in the European directive 2008/50/EC.

#### 4.3 Temporal analysis

To detect common periodicities in our time series, a standard method of time-series analysis has been used: computing the Fourier transform and plotting the power density spectrum over the frequency. We compute the spectral analysis of the hourly data in order to catch the signal of characteristic periods as done in previous studies (Hies et al., 2000; Sebald et al., 2000; Marr et al. 2002). We use the periodogram method as estimate of the spectral density calculated in its simplest form: the squared amplitude in function of frequency (Alter, 1937). The time series is decomposed into spectral bands representing the different time scales: intra-day, diurnal, synoptic, seasonal and long term fluctuations. The highest amplitudes of a



spectrum indicate the main periodicities of the underlying processes (Fig. 7). The periodogram in Fig. 7 reveals the ability of the modelling system to capture the variability associated with synoptic to long-term scales. The annual cycle in the O<sub>3</sub> time series is not visible in this periodogram because of the limited data range, since the length of the record should be 10 times as long as the longest significant period (Hies et al., 2000). The results show poor model ability in representing the observed daily and intra-day variability. The modelling system shows a tendency to underestimate the high frequency variability (intra-day and hour-to-hour variability), a feature which is corrected by the post-processing. Such behaviour is further investigated by the temporal analysis of the daily cycle.

We compute the average daily cycle of observed and simulated hourly ozone concentrations for both the model results and the KF-output (Fig. 8). The box plot helps to visualize the distribution, its central value, and spread of the represented values. The lower and upper quartiles and the median are also shown in Fig. 8. The mean is over-plotted for completeness. The model results show poor performance when representing the daily cycle amplitude and the extreme values. In the model results the amplitude of the daily cycle is underestimated (dimension of the boxes in Fig. 8), both the night minimum and the daytime maximum are underestimated while the overall mean value (overlying line on the boxes) is overestimated. The simulation improves after applying the Kalman filter: the amplitude of the cycle becomes well represented and the extreme values are better estimated (Fig. 8).

A known problem of the air quality models is to reproduce the night minimum. The possible explanations of such behaviour are examined by several previous studies (Appel et al., 2007; van Loon et al., 2007; Vautard et al., 2007; Chemel et al., 2010). It is probable that the model chemistry does not represent properly ozone nighttime's titration; therefore the model not

only does not simulate production of  $O_3$ , but also it omits the mechanism of  $O_3$  consumption. Possible causes of the observed mismatch are the bad representation of the nocturnal boundary layer or the height of the emissions injection. In stable surface layers, such as night-time, a misrepresentation of the height of concentration in the model can lead to a great disagreement with the observations. At the same time a misrepresentation of the emission height injection can affect the surface layer ozone titration (Appel et al., 2007; van Loon et al., 2007; Vautard et al., 2007; Chemel et al., 2010).

In Pay et al, 2010, and Baldasano et al., 2011, the highest uncertainties are found in the reproduction of  $O_3$  levels are related to  $NO_2$ -limited regime. Under this regime, corresponding to background conditions, the modeled  $O_3$  overestimates the observed values, specifically during nighttime. At the same time, under no  $NO_2$ -limited regime, the  $NO_2/O_3$  ratio is better represented in the model. This behaviour highlights the need to better characterize the emission inventory in either rural or urban areas (Baldasano et al., 2011). Nevertheless in our case we observe high improvements by the application of the post process for all the kind of stations.

#### **4.4 Model Uncertainty**

The European Directive 2008/50/CE establishes that numerical models have to meet certain model quality, namely have a certain modelling uncertainty, to be considered suited for air

quality assessment. The level of uncertainty associated with the air quality modelling is a crucial issue to take into account when evaluating an air quality performance.

We compute the model uncertainty as defined in the above mentioned European Directive and expressed as MRDE, before and after the Kalman filter post-processing to verify any improvements. In both cases the computed uncertainty is within the limit established by the European Directive 2008/50/CE that is 50%. The uncertainty of the model is 22% and 20% for the max-1h and max-8hr respectively, well under the uncertainty limit. This finding gives us confidence in the ability of CALIOPE system to simulate the O<sub>3</sub> concentrations, already before the post-processing. Nevertheless, the improvement of the uncertainty after the application of the KF post-processing is considerable: the uncertainty is reduced from 22% to 19% and from 20% to 7.5% for the max 1-hr and max 8-hr respectively. The stronger improvement observed for the max 8-hr is due to the fact that the model results for the max 8-hr are already highly accurate and comparable with the observations. The difference  $O_{LV} - M_{LV}$  in Equation 7 is smaller in the case of max 8-hr than in the case of max 1-hr.

The uncertainty, calculated in this way, indicates only whether the model is reliable or not. It is important to note that this definition of uncertainty does not take into account the temporal information related to the formation/destruction of ozone. Furthermore, the number of the exceedances is independent of the sequence of the events. This approach is in contrast to a time-related process, in which the temporal correspondence is an important evaluation parameter, as it is in air quality studies.

## 5 Conclusions

We use the CALIOPE air quality modelling system to diagnose the daily maximum of  $O_3$  ground level concentration over Spain for the full year 2004. To the hourly  $O_3$  values has been applied a bias-adjustment technique based on the Kalman filter to check whether the post-processed results reproduce satisfactorily the observations. From the result of the post-process the daily maximum 1-hr and 8-hr  $O_3$  have been computed, following the EU directive 2008/50/EC. The Kalman filter is applied only to discrete points (82 monitoring stations) and then an average of the concentration of all the points (stations) is presented as representative of the whole domain. Therefore, the improvements of the post-process are limited only to the points where data are available. The complexity of the domain, the limit number of stations over the territory, and the difference in emissions sources over the domain suggest that further research is needed to extend the benefit of the post-process correction to the whole domain.

The model results have been evaluated over the 82 stations of the RedESP. For all the stations, independently on the stations' type, the application of the KF-bias adjustment technique leads to statistical improvements. Both the classical and categorical statistical metrics improve up to a 30% from the initial value after the application of the post-process. In particular, the analysis of the Critical Success Index (CSI) and the Probability of Detection (POD), which are the more suitable parameters since they take into account the misses of the model, shows that by the application of the KF bias adjustment technique the CALIOPE system results to be high reliable for prediction of alert threshold. This result confirms the potential for operational use of the presented methodology for real time simulations. This is a remarkable result, indicating that the bias- adjusted results of the CALIOPE system are a

valid instrument in the relation to the measures of the EU Directive 2008/50/EC on alert/information thresholds.

This article is the result of the developing work done within the operational CALIOPE system, which aims at establishing an air quality forecasting frame for Spain. We carried out a diagnosis of the system in the prospective of applying the post-processing to the near real time forecasts. The presented bias-adjustment method is now being applied on the operational air quality forecast system CALIOPE ([www.bsc.es/caliope](http://www.bsc.es/caliope)), and the extension of the method to correct areas where no observations are available is currently under development.

## Acknowledgments

The authors wish to thank Luca Delle Monache and Ronald B. Stull for providing the Kalman filter algorithm used in this study. All the computation has been done at the MareNostrum supercomputer hosted by the Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS). This work is funded by the CALIOPE project of the Spanish Ministry of the Environment (441/2006/3-12.1, A357/2007/2-12.1, 157/PC08/3-12.0).

## References

- Alter D. A simple form of periodogram. *Ann Mat Stat* 1937; 8 (2): 121-126.
- Appel KW, Bhave PV, Gilliland AB, Sarwar G, and Roselle SJ. Evaluation of the community multiscale air quality (CMAQ) model version 4.5: Sensitivities impacting model performance; Part II-particulate matter, *Atmos Environ* 2008; 42: 6057– 6066. doi:10.1016/j.atmosenv.2008.03.036.
- Arévalo G, Salvador R., Gassó S, Millán M, Baldasano JM. Application of a high-resolution emission model in Valencia Community (Spain). *Air Pollution* 2004. WIT Press, Rhodes, Greece, p. 31–40.
- Baldasano JM, Cremades L, Soriano C. Circulation of air pollutants over the Barcelona geographical area in summer. *Proceedings of Sixth European Symposium Physic-Chemical Behavior of Atmospheric Pollutants*. Varese (Italy), 18-22 October, 1993. Report EUR 15609/EN: 474-479; 1994.
- Baldasano JM, Valera E. and Jiménez P. Air Quality Data from Large Cities. *Sci. Total Environ* 2003; 307 (1-3): 141-165.
- Baldasano JM, Jiménez-Guerrero P, Jorba O, Pérez C, López E, Güereca LP, et al. 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* 2008a; 2: 89-98.
- Baldasano JM, Güereca LP, Lopez E, Gassó S, Jiménez-Guerrero P. Development of a high-resolution (1km x 1km, 1h) emission model for Spain: The High-Effective Resolution Modelling Emission System (HERMES). *Atmos Environ* 2008b; 42 (31): 7215–7233.

1

2 Baldasano JM, Pay MT, Jorba O, Jiménez P. An annual assessment of air quality with the  
 3 CALIOPE modeling system over Spain. Submitted to Sci Total Environ 2011; 409: 2163–  
 4 2178. doi: 10.1016/j.scitotenv.2011.01.041.

5

6 Binkowski FS. Aerosols in models-3 CMAQ, in: Byun DW, Ching JKS. (Eds.), Science  
 7 Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modelling  
 8 System. 1999; EPA. pp. 10-23.

9

10 BOE, Real Decreto 102/2011 (BOE, 2011), de 28 de enero, relativo a la mejora de la calidad  
 11 del aire. Ministerio de la Presidencia. BOE numero 25. BOE-A-2011-1645.53.

12

13 Borrego C, Tchepel O, Costa AM, Amorim JH, Miranda AI. Emission and dispersion  
 14 modelling of Lisbon air quality at local scale. Atmos. Environ 2003; 37: 5197–5205.  
 15 doi:10.1016/j.atmosenv.2003.09.004.

16

17 Borrego C, Monteiro A, Ferreira J, Miranda AI, Costa AM, Carvalho AC, Lopes M.  
 18 Procedures for estimation of modelling uncertainty in air quality assessment. Environ Int  
 19 2008; 34 (5): 613–20. doi:10.1016/j.envint.2007.12.005.

20

21 Borrego C, Monteiro A, Pay MT, Ribeiro I, Miranda AI, Basart S, et al. How bias-  
 22 correction can improve air quality forecasts over Portugal. Atmos Environ 2011; 45: 6629-  
 23 6641. doi:10.1016/j.atmosenv.2011.09.006.

24

25 Brauer M, Brook JR. Ozone personal exposures and health effects for selected groups  
 26 residing in the Freser Valley. Atmos Environ 1997; 31 (14): 2113-2121.

27

1 Byun DW, Ching JKS. Science Algorithms of the EPA Models-3 Community Multiscale Air  
2 Quality (CMAQ) Modelling System. 1999; EPA/600/R-99/030, US EPA. National Exposure  
3 Research Laboratory, Research Triangle Park, NC.

4

5 Byun DW, Schere K L. Review of the governing equations, computational algorithms, and  
6 other components of the Models-3 Community Multiscale Air Quality (CMAQ) modelling  
7 system. *Appl Mech Rev* 2006; 59: 51–77.

8

9 Chang JC, Hanna, S.R. Air quality model performance evaluation. *Meteorol Atmos Phys*  
10 2004; 87: 167–196, DOI 10.1007/s00703-003-0070-7.

11

12 Chemel C, Sokhi R, Yu Y, Hayman G, Vincent K, Dore AJ, et al. Evaluation of a CMAQ  
13 simulation at high resolution over the UK for the calendar year 2003. *Atmos Environ* 2010;  
14 44 (24): 2927–2939. doi:10.1016/j.atmosenv.2010.03.029.

15

16 Cristofanelli P, Bonasoni P. Background Ozone in the southern Europe and Mediterranean  
17 area: Influence of the transport processes. *Environ Pollut* 2009; 157(5): 1399-1406,  
18 doi:10.1016/j.envpol.2008.09.017.

19

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

23

24 Delle Monache L, Wilczak J, Mckeen S, Grell G, Pagowski M, Peckham S, et al. A Kalman-  
25 filter bias correction method applied to deterministic, ensemble averaged, and probabilistic  
26 forecasts of surface ozone, *Tellus Ser B* 2008; 60: 238–249. doi: 10.1111/j.1600-  
27 0889.2007.00332.x.

28



1 Denby B, Georgieva E, Larssen S, Guerreiro C, Li L, Douros J, Moussiopoulos N, et. al.  
2 Guidance on the use of models for the European Air Quality Directive. A working document  
3 of the Forum for Air Quality Modelling in Europe FAIRMODE. Technical Report Version  
4 6.1, Editor B. Denby 1-99. 2010.

5

6 Dennis R, Fox T, Fuentes M, Gilliland A, Hanna S, Hogrefe C, et al. A framework for  
7 evaluating regional-scale numerical photochemical modelling systems. *Environ Fluid Mech*  
8 2010; DOI 10.1007/s10652-009-9163-2.

9

10 Eben K, Jurus P, Resler J, Belda M, Pelikan E, Kruger BC, Keder J. An ensemble Kalman  
11 filter for short-term forecasting of tropospheric ozone concentrations. *Q J R Meteorol Soc*  
12 2005; 131, pp. 3313–3322.

13

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

17

18 European Commission, 2008. Directive 2008/50/EC of the European Parliament and of the  
19 Council of 21 May 2008 on ambient air quality and cleaner air for Europe. Technical Report  
20 2008/50/EC, L152. Off. J. Eur. Comm.

21

22 Finlayson-Pitts BJ. Atmospheric Chemistry 2010; PNAS, 107 (15): 6566–6567.

23

24 Flemming J, Stern R. Testing model accuracy measures according to the EU directives-  
25 examples using the chemical transport model REM-CALGRID. *Atmos Environ* 2007; 41:  
26 9206–9216. doi:10.1016/j.atmosenv.2007.07.050.

27

28 Folberth G, Hauglustaine D, Lathiere J, Brocheton F. Interactive chemistry in the Laboratoire  
29 de Meteorologie Dynamique general circulation model: model description and impact

1 analysis of biogenic hydrocarbons on tropospheric chemistry. *Atmos Chem Phys* 2006; 6:  
2 2273–2319. [www.atmos-chem-phys.net/6/2273/2006/](http://www.atmos-chem-phys.net/6/2273/2006/).  
3  
4 Gangoiti G, Millán MM, Salvador R, Mantilla E. Long-range transport and re-circulation of  
5 pollutants in the western Mediterranean during the project Regional Cycles of Air Pollution in  
6 the West-Central Mediterranean Area. *Atmos Environ* 2001; 35: 6267–6276.  
7  
8 Garber W, Colosio J, Grittner S, Larssen S, Rasse D, Schneider J, et al. Guidance on the  
9 Annexes to Decision 97/101/EC on Exchange of Information as revised by Decision  
10 2001/752/EC. Technical Report European Commission, DG Environment 1-71. 2002.  
11  
12 Gerasopoulos E, Kouvarakis G, Vrekoussis M, Kanakidou M, Mihalopoulos N. Ozone  
13 variability in the marine boundary layer of the eastern Mediterranean based on 7-year  
14 observations. *J Geophys Res* 2005; 110: D15309, doi:10.1029/2005JD005991.  
15  
16 Gonçalves M, Jiménez P, Baldasano JM. Contribution of atmospheric processes affecting the  
17 dynamics of air pollution in South-Western Europe during a typical summertime  
18 photochemical episode. *Atmos Chem Phys* 2009; 9 (3): 849–864.  
19  
20 Guenther AB, Hewitt CN, Erickson D, Fall R, Geron C, Graedel T, et al. A global model of  
21 natural volatile organic compound emissions. *J Geophys Res* 1995; 100: 8873-8892.  
22  
23 Hanea RG, Velders GJM, Heemink A. Data assimilation of ground-level ozone in Europe  
24 with a Kalman filter and chemistry transport model. *J Geophys Res* 2004; 109: D10302.  
25 doi:10.1029/2003JD004283.  
26  
27 Hauglustaine DA, Hourdin F, Jourdain L, Filiberti MA, Walters S, Lamarque JF, et al.  
28 Interactive chemistry in the Laboratoire de Meteorologie Dynamique general circulation

1 model: Description and background tropospheric chemistry evaluation. *J Geophys Res* 2004;  
2 109: 1–44.

3

4 Hies T, Treffeisen R, Sebald L, Reimer E. Spectral analysis of air pollutants. Part 1: elemental  
5 carbon time series. *Atmos Environ* 2000; 34 (21): 3495-3502.

6

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

9

10 Jiménez P, Baldasano JM. Ozone response to precursor controls in very complex terrains: Use  
11 of photochemical indicators to assess O<sub>3</sub>-NO<sub>x</sub>-VOC sensitivity in the northeastern Iberian  
12 Peninsula. *J Geophys Res* 2004; 109:D20309. doi: 10.1029/2004JD004985.

13

14 Jiménez P, Parra R, Gassó S, Baldasano JM. Modelling the ozone weekend effect in very  
15 complex terrains: a case study in the north eastern Iberian Peninsula. *Atmos Environ* 2005a;  
16 39: 429–444.

17

18 Jiménez P, Jorba O, Parra R, Baldasano JM. Influence of high-model grid resolution on  
19 photochemical modelling in very complex terrains. *Int J Environ Pollut* 2005b ; 24: 180–200.

20

21 Jiménez P, Lelieveld J, Baldasano JM. Multi-scale modelling of air pollutants Dynamics in  
22 the North-Western Mediterranean Basin during a typical summertime episode. *J Geophys Res*  
23 2006; 111: D18306. doi:10.1029/2005JD006516.

24

25 Jiménez P, Jorba O, Baldasano, JM, Gassó S. The Use of a Modelling System as a Tool for  
26 Air Quality Management: Annual High-Resolution Simulations and Evaluation. *Sci Tot Env*  
27 2008; 390: 323-340. doi: 10.1016/j.scitotenv.2007.10.025.

28

1 Jorba O, Pérez C, Rocadenbosch F, Baldasano JM. Cluster analysis of 4-day back trajectories  
2 arriving in the Barcelona area (Spain) from 1997 to 2002. *J. Appl. Meteorol.* 2004; 43(6):  
3 887–901.

4

5 Kalman RE. A new approach to linear filtering and prediction problems. *J Basic Eng* 1960;  
6 82: 35–45.

7

8 Kang D, Eder BK, Stein AF, Grell GA, Peckham SE, McHenry J. The New England air  
9 quality forecasting pilot program: Development of an evaluation protocol and performance  
10 benchmark. *J Air Waste Manage Assoc* 2005; 55: 1782–1796

11

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

14

15 Kang D, Mathur R, Rao ST. Assessment of bias-adjusted PM<sub>2.5</sub> air quality forecasts over the  
16 continental United States during 2007. *Geosci Model Dev* 2010; 3: 309–320. [www.geosci-](http://www.geosci-model-dev.net/3/309/2010/)  
17 [model-dev.net/3/309/2010/](http://www.geosci-model-dev.net/3/309/2010/).

18

19 Lam YF, Fu JS. A novel downscaling technique for the linkage of global and regional air  
20 quality modelling. *Atmos Chem Phys* 2009; 9: 9169-9185.

21

22 Marmer E, Dentener F, Aardenne J, Cavalli F, Vignati E, Velchev K, et al. What can we learn  
23 about ship emission inventories from measurements of air pollutants over the Mediterranean  
24 Sea? *Atmos Chem Phys Discuss* 2009; 9:7155–211.

25

26 Marr LC, Harley RA. Spectral analysis of weekday–weekend differences in ambient ozone,  
27 nitrogen oxide, and non-methane hydrocarbon time series in California. *Atmos Environ* 2002;  
28 36 (14): 2327-2335. doi:10.1016/S1352-2310(02)00188-7.

29

1 Michalakes J, Dudhia J, Gill D, Henderson T, Klemp J, Skamarock W, et al. The weather  
2 research and forecast model: Software architecture and performance, in: Mozdzynski, E.G.  
3 (Ed.), To appear in proceeding of the Eleventh ECMWF Workshop on the Use of High  
4 Performance Computing in Meteorology, 2529, Reading, U.K., p. 117-124. 2004.

5  
6 Millán M, Salvador R, and Mantilla E. Photooxidant dynamics in the Mediterranean basin in  
7 summer: Results from European research projects. *J Geophys Res* 1997; 102 (D7): 8811-  
8 8823.

9  
10 Millán M, Sanz MJ, Salvador R, Mantilla E. Atmospheric dynamics and ozone cycles related  
11 to nitrogen deposition in the western Mediterranean. *Environ Pollut* 2002; 118 (2): 167–86.

12  
13 Nickovic S, Kallos G, Papadopoulos A, Kakaliagou O. A model for prediction of desert dust  
14 cycle in the atmosphere. *J Geophys. Res* 2001; 106(D16): 18113-18129.  
15 doi:10.1029/2000JD900794.

16  
17 Parra R, Gassó S, Baldasano JM. Estimating the biogenic emissions of non-methane volatile  
18 organic compounds from the North Western Mediterranean vegetation of Catalonia, Spain.  
19 *Sci Tot Env* 2004; 24: 241–259.

20  
21 Parra R, Jiménez P, Baldasano JM. Development of the high spatial resolution EMICAT2000  
22 emission model for air pollutants from the north-eastern Iberian Peninsula (Catalonia, Spain).  
23 *Environ Pollut* 2006; 140: 200–219.

24  
25 Pay M T, Piot M, Jorba O, Gassó S, Gonçalves M, Basart S, et al. A full year evaluation  
26 of the CALIOPE-EU air quality modelling system over Europe for 2004. *Atmos Environ*  
27 2010; 44: 3322-3342. doi: 10.1016/j.atmosenv.2010.05.040.

1 Pérez C, Nickovic S, Baldasano JM, Sicard M, Rocadenbosch F, Cachorro VE. A long  
2 Saharan dust event over the western Mediterranean: Lidar, sun photometer observations, and  
3 regional dust modelling J Geophys Res 2006a; 111(D15214): 1-16,  
4 doi:10.1029/2005JD006579.

5

6 Pérez C, Nickovic S, Pejanovic G, Baldasano JM, Ozsoy E. Interactive dust-radiation  
7 modelling: A step to improve weather forecasts. J Geophys Res 2006b; 111 (D16206): 1-17.  
8 doi:10.1029/2005JD006717.

9

10 Sagers A. Data assimilation in atmospheric chemistry models using Kalman filtering  
11 [dissertation]. Delft University Press: ISBN 90-407-2286-2. 2002.

12

13 Sebald L, Treffeisen R, Reimer E, Hies T. Spectral analysis of air pollutants. Part 2: ozone  
14 time series. Atmos Environ 2000; 34(21): 3503-3509.

15

16 Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, Powers JG. A  
17 Description of the Advanced Research WRF Version 2 NCAR Tech Notes-468. 2005.

18

19 Skamarock WC, Klemp JB. A time-split nonhydrostatic atmospheric model for weather  
20 research and forecasting applications. J Comput Phys 2008; 227(7): 3465-3485.  
21 doi:10.1016/j.jcp.2007.01.037.

22

23 Staudt M, Bertin N, Frenzel B, Seufert G. Seasonal variations in amount and composition of  
24 monoterpenes emitted by young Pinus pinea trees—implications for emission modelling. J  
25 Atmos Chem 2000; 35: 77–99.

26

27 Steinbrecher R, Smiatek G, Köble R, Seufert G, Thelokec J, Hauff K, et al. Intra- and inter-  
28 annual variability of VOC emissions from natural and semi-natural vegetation in Europe and  
29 neighbouring countries. Atmos Environ 2009; 43: 1380-1391.

Taylor KE. Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res* 2001; 106(D7): 7183–7192.

US-EPA. AP-42. 5<sup>th</sup> Edition, Volume VI, Chapter 13, Section 13.2.1. Paved Roads. Technical Report. Guidance on the Use of Models and Other Analyses for Demonstrating Attainment of Air Quality Goals for Ozone, PM<sub>2.5</sub>, and Regional Haze. Technical Report. EPA-454/B-07-002. U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards: Research Triangle Park, NC. 2007.

US-EPA. Guidance on the Development, Evaluation, and Application of Environmental models”. Office of the Science Advisor. Council for Regulatory Environmental Modelling EPA/100/K-09/003 1-99. 2009.

van Loon M, Heemink AW. Kalman filtering for nonlinear atmospheric chemistry models: first experiences, Modelling, Analysis and Simulation (MAS). Technical Report MAS-R9711. ISSN 1386-3703 CWI P.O. Box 94079 1090 GB Amsterdam, The Netherlands. 1997.

van Loon M, Builtjes PJ H, Segers A. Data assimilation of ozone in the atmospheric transport chemistry model LOTOS. *Env Mod and Softw* 2000; 15: 603-609.

van Loon M, Vautard R, Schaap M, Bergstrom R, Bessagnet B, Brandt J, et al. Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble. *Atmos Environ* 2007; 41: 2083–2097. doi:10.1016/j.atmosenv.2006.10.073.

Vautard R, Builtjes P, Thunis P, Cuvelier C, Bedogni M, Bessagnet B, et al. Evaluation and intercomparison of Ozone and PM<sub>10</sub> simulations by several chemistry transport models over

1 four European cities within the CityDelta project. *Atmos Environ* 2007; 41: 173–188.  
2 doi:10.1016/j.atmosenv.2006.07.039.

3  
4 West J, Szopa S, Hauglustaine D. Human mortality effects of future concentrations of  
5 tropospheric Ozone. *C R Geoscience* 2007; 339: 775–783. doi:10.1016/j.crte.2007.08.005.

6  
7 WHO. Health risks of ozone from long-range transboundary air pollution. *World Health*  
8 *Organization* 2008; pp: 111.

9  
10  
11  
12  
13



## 1 Tables

Table 1: Error ratio values calculated in consideration of the bias-adjusted Root Mean Square Error and Correlation for the seasons of the year 2004.

SEASON	ERROR RATIO
Winter	0.40
Spring	0.20
Summer	0.15
Autumn	0.60

2

3

4

Table 2: Statistical comparison between model results and KF-output. The statistics are for the max 1-hr and max 8-hr for the whole domain, for the year 2004. RMSE is expressed in unit of O<sub>3</sub> concentration (µg/m<sup>3</sup>). All the statistics were determined considering all the stations globally. The values in brackets are the min and the max for the stations.

Statistics	Max 1-hr		Max 8-hr	
	Model	KF	Model	KF
COR	0.64 (0.48-0.86)	0.83 (0.53-0.99)	0.65 (0.49-0.84)	0.86 (0.5-0.89)
RMSE	25.50 (17.09-57.44)	17.09 (11.39-21.39)	24.50 (17.24-55.80)	15.80 (13.28-38.23)
MB	-1.3 (-26.15- 51.35)	-2.2 (-11.64-4.85)	4.1 (-23.29-53.06)	-1 (-4.08-5.52)
MNBE (%)	10.60 (-21.34-317.55)	2.50 (-1.85-57.76)	23.79 (-22.74-277.15)	4.80 (-4.09- 55.62)
MNGE (%)	32.48 (15.63-317.55)	19.7 1 (10.05-76.14)	40.87 (15.98- 277.15)	21.60 (10.37-74-16)

**Comment [v2]:** We decided to update this table with the statistics values calculates over the entire cloud of points of figure 5.

5

6

7

Table 3: Categorical Statistics for the model results and the KF-output over the max 1-hr and max 8-hr. All the statistics are calculated for the hourly data of the 82 stations, for all the 2004. The perfect score is reported in parentheses. POD is the Probability of Detection, FAR is the False Alarm Ratio, POFD is the Probability of False Detection, CSI is the Critical Success Index, A is the Accuracy, and B is the Bias. Refer to the text for the interpretation of these parameters.

Statistics (perfect score)	Max 1-hr		Max 8-hr	
	Model results	KF-output	Model results	KF-output
<i>a (hits)</i>	1	21	703	1085
<i>b (false alarm)</i>	21	22	1622	756
<i>c (misses)</i>	126	106	1462	1080
<i>d (correct negative)</i>	27822	27821	23874	24740
POD (1)	0.008	0.165	0.325	0.501
FAR (0)	0.955	0.512	0.698	0.41
POFD (0)	0.001	0.001	0.064	0.030
CSI (1)	0.007	0.141	0.186	0.371
A (1)	0.995	0.996	0.889	0.934
BIAS (1)	0.173	0.339	1.074	0.850

1   **Figures**

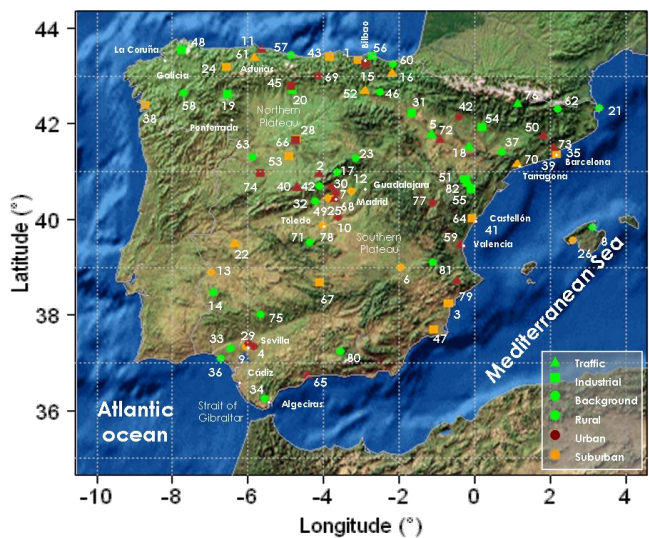


Fig. 1: RedESP stations network measuring ozone concentration in Spain. Different types of stations (U: Urban; S: Suburban; R: Rural; B: Background; I: Industrial; and T: Traffic) are represented by symbols and color codes.

1

Ratio sensitivity. A: RMSE

Ratio sensitivity. B: COR

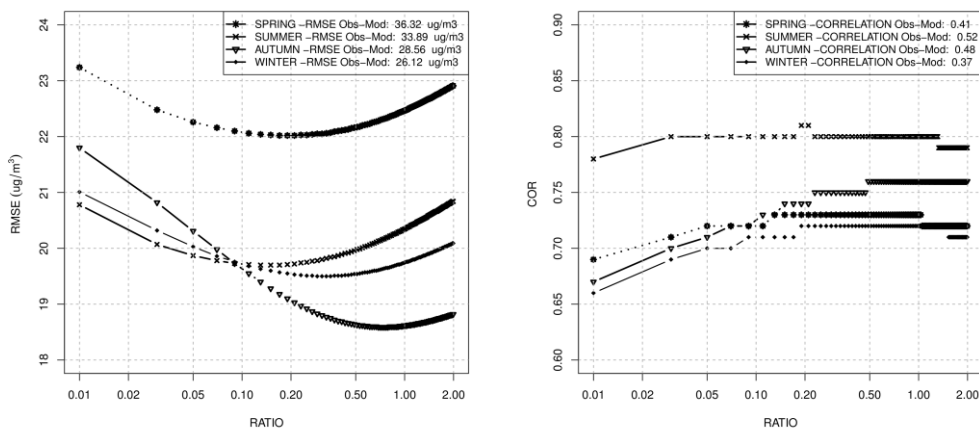
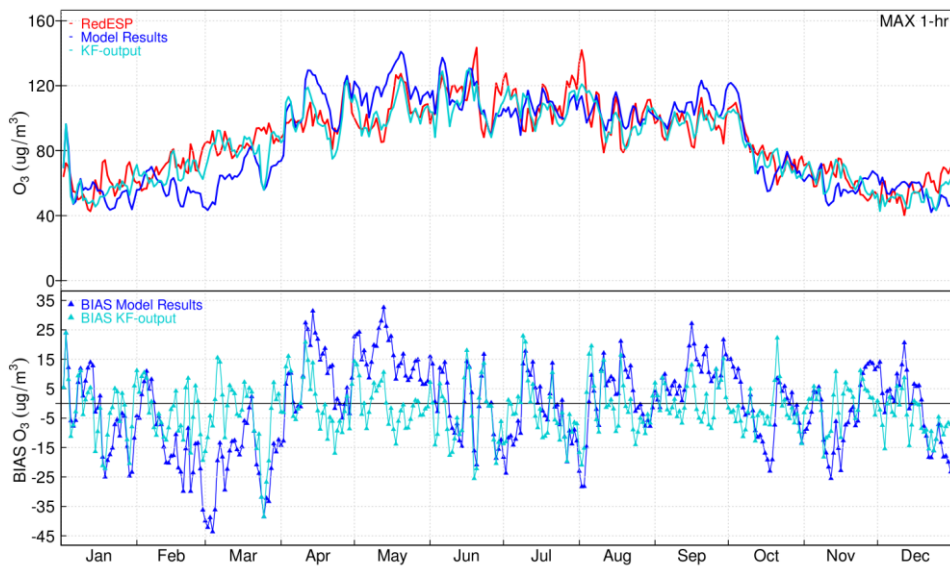


Fig. 2: Seasonal ratio sensitivity for the hourly  $O_3$  concentrations over the 82 RedESP stations. A: Root Mean Square Error and B: Correlation. Values are computed with the ratio raging from 0.01 to 2, plotted on logarithmic scale. Perfect RMSE would be 0, and perfect correlation coefficient 1. On the plots are reported RMSE and COR before the application of the Kalman filter.

2  
3  
4  
5  
6  
7  
8  
9  
10  
11

1



2

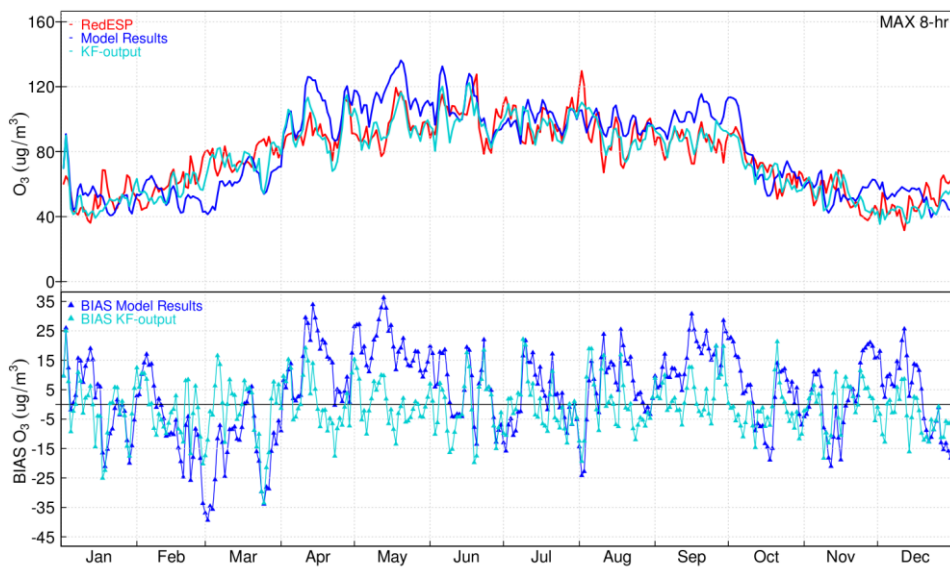
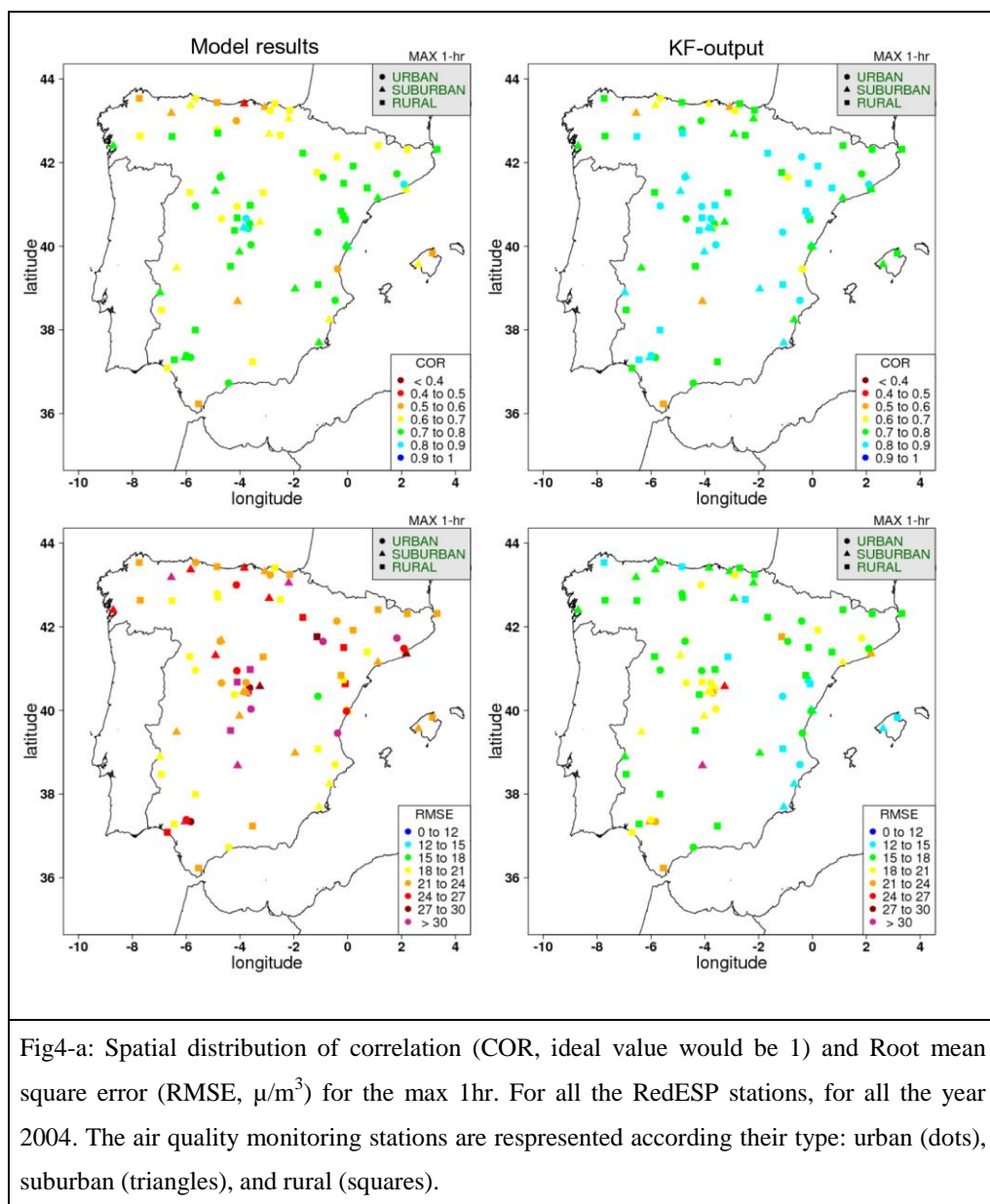


Fig. 3: Time series of the max 1-hr (upper panel) and max 8-hr (lower panel) ground level  $O_3$  concentration ( $\mu g/m^3$ ), averaged for all RedESP stations for the model results and the KF-output, for the year 2004. The plot of the biases ( $\mu g/m^3$ ) is also included.

3

1



2

3

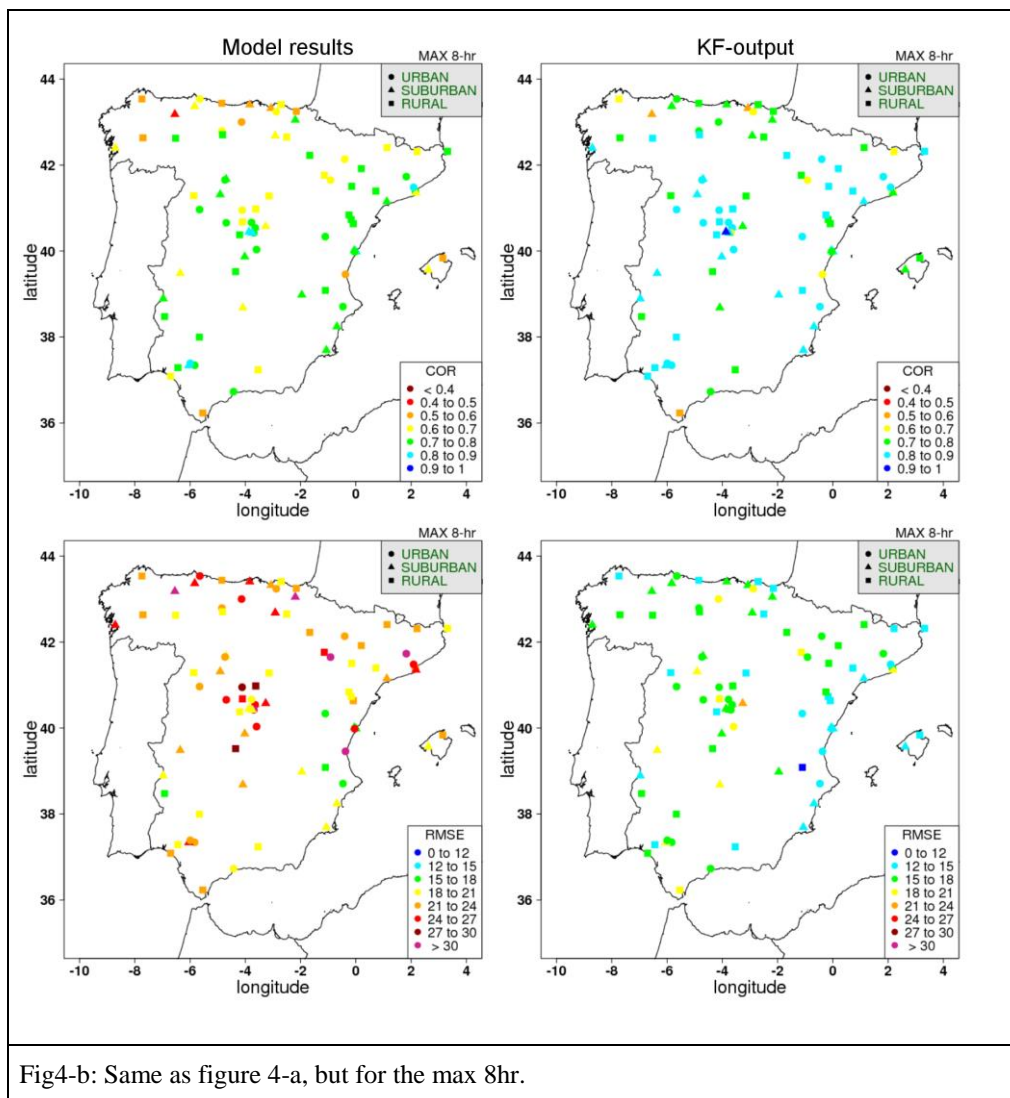


Fig4-b: Same as figure 4-a, but for the max 8hr.

1  
2  
3  
4

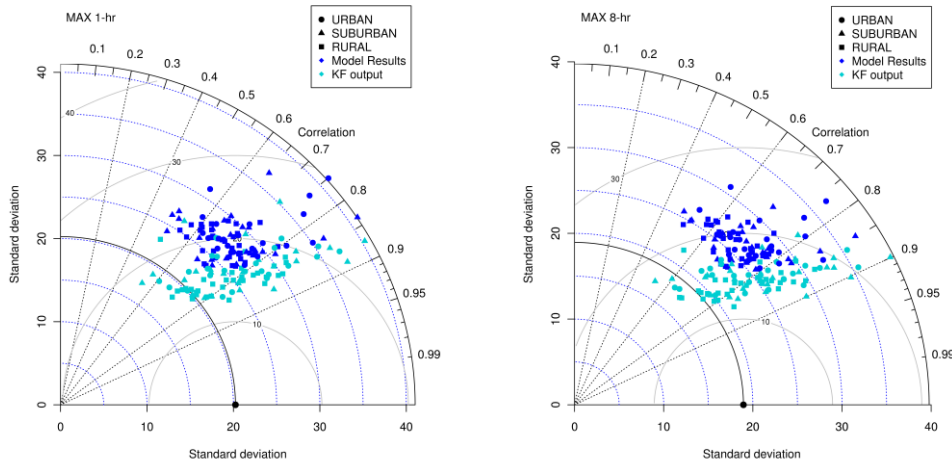


Fig. 5: Taylor Diagrams for the max 1-hr and max 8-hr. All the considered stations are plotted. The different symbols represent the types of stations. On the plots are depicted the Standard Deviation ( $\mu\text{g}/\text{m}^3$ ) of the simulated maximum  $\text{O}_3$  concentration (radius) and the Correlations (cosine of the angle to the horizontal axis). On the horizontal axis is located the standard deviation of the observed values, the closest are the points to this value, the better are simulated the values. The statistics are calculated over the year 2004.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10



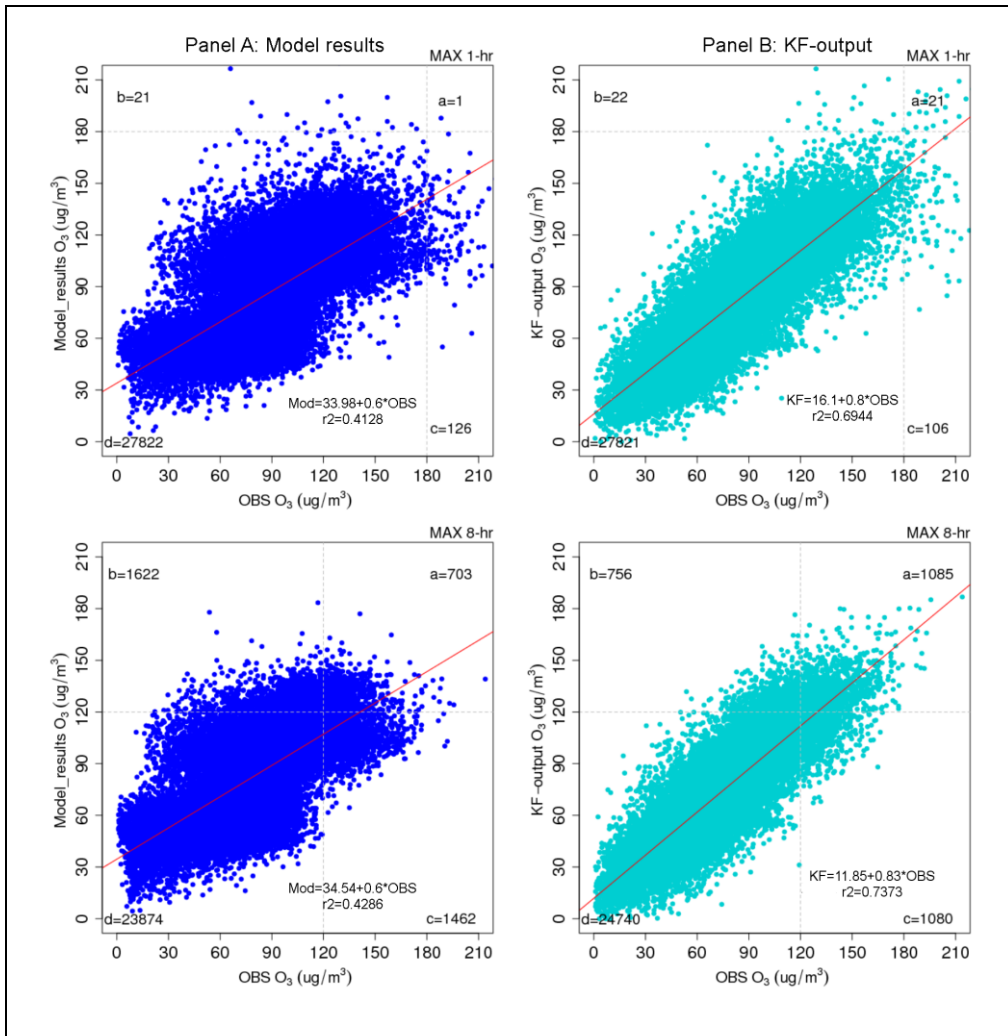


Fig. 6: Predicted versus observed max 1-hr and max 8-hr  $O_3$  concentrations for the year 2004 for all the stations considered. The plotted values are on hourly bases. On the plots are depicted the threshold limit  $120 \mu g/m^3$  and  $180 \mu g/m^3$  as established by the EU for the max 1-hr and max 8-hr respectively. The letters a, b, c, d represent the exceedances that did occur (*hits*), the exceedances that did not occur (*false alarms*), the exceedances that were not predicted but observed (*misses*), and the exceedances that did not occur and were not predicted respectively (*correct negatives*), see section 4.2. The red line is the best-fit line. The linear regression equation and  $r^2$  are reported as well.

1

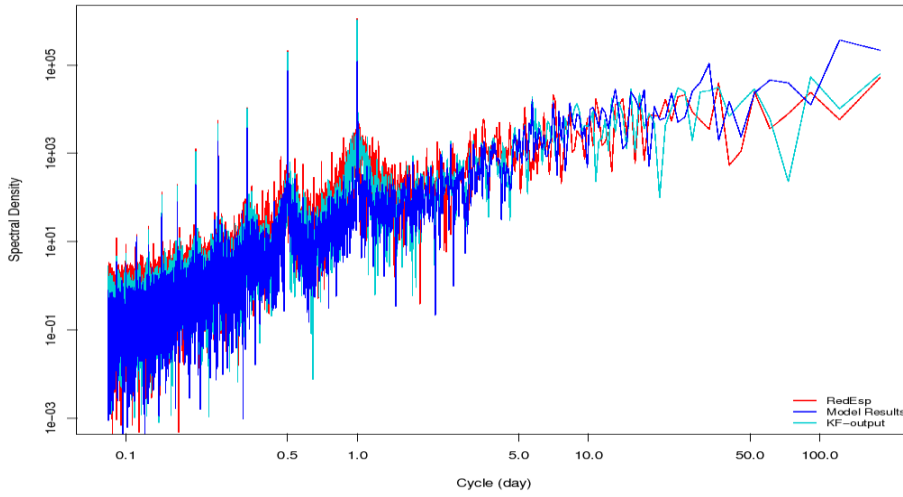


Fig. 7: Spectral analysis of daily O<sub>3</sub> concentrations averaged for the whole year averaged over all the stations. The spectral power of the observations (red), model results (blue) and KF-output (bright blue) are compared. The axes are on logarithmic scale.

2

3

4

5

6

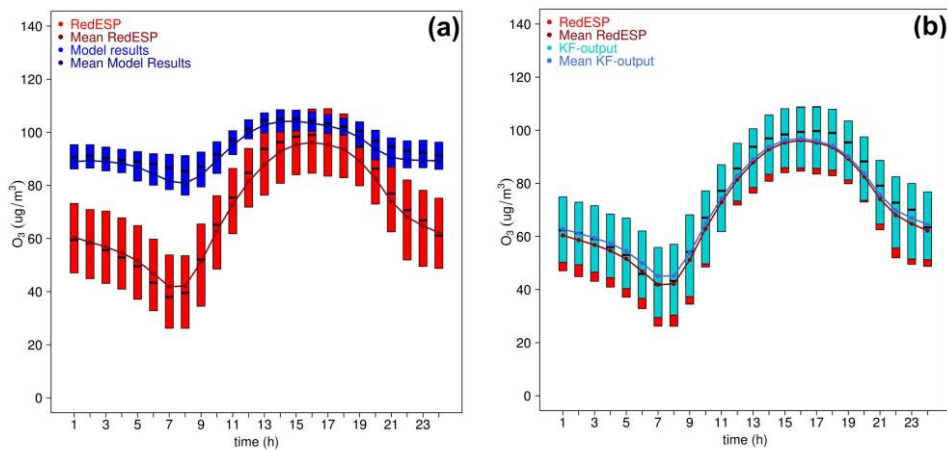


Fig. 8: Ozone daily cycle for hourly averaged concentration in  $\mu\text{g}/\text{m}^3$  for 2004 over all the RedESP stations. Model results are represented in (a) and KF-output in (b). The box plots represent the median, the 75th percentile (top) and the 25th percentile (bottom). The over-plotted lines represent the mean concentrations.

1    **List of Tables:**

2    Table 1: Error Ratio

3    Table 2: Domain wide statistics

4    Table 3: Categorical analysis

5

6    **List of figures:**

7    Figure 1: Map of stations

8    Figure 2: Ratio sensitivity

9    Figure 3: Time series of max O<sub>3</sub>

10   Figure 4: Spatial Map of Statistics (max 1-hr and max 8-hr)

11   Figure 5: Taylor Diagram

12   Figure 6: Scatter Plot

13   Figure 7: Spectral Analysis

14   Figure 8: Diurnal Cycle

15

16

17

18