



Predictability of European air quality : assessment of 3 years of operational forecasts and analyses by the PREV'AIR system

Cécile Honore, Laurence Rouil, Robert Vautard, Matthias Beekmann, Bertrand Bessagnet, Anne Dufour, Christian Elichegaray, Jean-Marie Flaud, Laure Malherbe, Frédérik Meleux, et al.

► To cite this version:

Cécile Honore, Laurence Rouil, Robert Vautard, Matthias Beekmann, Bertrand Bessagnet, et al.. Predictability of European air quality : assessment of 3 years of operational forecasts and analyses by the PREV'AIR system. Journal of Geophysical Research: Atmospheres, 2008, 113 (D4), pp.art D04301. 10.1029/2007JD008761 . ineris-00963111

HAL Id: ineris-00963111

<https://ineris.hal.science/ineris-00963111>

Submitted on 17 Sep 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Predictability of European air quality: Assessment of 3 years of operational forecasts and analyses by the PREV'AIR system

Cécile Honoré,¹ Laurence Rouil,¹ Robert Vautard,² Matthias Beekmann,³ Bertrand Bessagnet,¹ Anne Dufour,⁴ Christian Elichegaray,⁵ Jean-Marie Flaud,³ Laure Malherbe,¹ Frédéric Meleux,¹ Laurent Menut,⁶ Daniel Martin,⁴ Aline Peuch,⁴ Vincent-Henri Peuch,⁴ and Nathalie Poisson⁵

Received 6 April 2007; revised 10 July 2007; accepted 16 October 2007; published 22 February 2008.

[1] For the first time, the long-term evaluation of an operational real-time air quality forecasting and analysis system is presented, using error statistics over 3 consecutive years. This system, called PREV'AIR, is the French air quality forecasting and monitoring system. It became operational in 2003 as a result of a cooperation between several public organizations. The system forecasts and analyzes air quality throughout Europe, with a zoom over France, for regulatory pollutants: ozone (O₃), particulate matter with diameter smaller than 10 μm (PM₁₀), and nitrogen dioxide (NO₂). The ability of PREV'AIR to forecast, up to 3 days ahead, photochemical and particle pollution over the domains considered is demonstrated: daily ozone maxima forecasts correlate with observations with 0.75–0.85 mean coefficients; U.S. Environmental Protection Agency acceptance criteria relative to the forecast accuracy for high concentrations and daily maxima are met for more than 90% of the measurement sites. For NO₂ and PM₁₀, the performance corresponds to the state of the art. The contribution of weather forecast errors to air quality predictability is addressed: ozone daily maxima forecast errors are not dominated by meteorological forecast errors; for rural stations, only 6% (15% and 25%, respectively) of the error variance is due to meteorological forecast errors on the first 24 (48 and 72, respectively) hours. The Model Output Statistics procedure, implemented in PREV'AIR, is proved to improve ozone forecasts, especially when photochemical pollution episodes occur. The PREV'AIR real-time analysis procedure, based on a kriging method, provides an accurate and comprehensive description of surface ozone fields over France.

Citation: Honoré, C., et al. (2008), Predictability of European air quality: Assessment of 3 years of operational forecasts and analyses by the PREV'AIR system, *J. Geophys. Res.*, 113, D04301, doi:10.1029/2007JD008761.

1. Introduction

[2] For several decades, the adverse effects of air pollutants on human health and the environment have been demonstrated. As a consequence, the reduction of atmospheric pollution impacts has become a major challenge for current environmental policies. In Europe, despite more than 20 years devoted to the definition and the implementation of regulations and laws regarding the pollutants of

utmost importance in relation to human health, air pollution is still declared to be a real concern for the public and the authorities. Current health studies demonstrate that atmospheric pollutants are responsible for increasing breathing troubles (asthma and other respiratory diseases) in the population, with a non negligible impact on the morbidity and loss of life expectancy statistics. Ambitious studies such as APHEIS (see <http://www.apheis.net/>) or those carried out by the *World Health Organization (WHO)* [2006] or by *Pope et al.* [2002] clearly show that significant improvements are still needed to manage and control the impacts of air pollution on health.

[3] Facing the challenge of reducing the impacts of air pollution on health, one is left with two options: improving air quality or preventing exposure. In Europe, efforts in controlling emissions are a major pillar of the 6th Environmental Action program and the Thematic Strategy on Air Pollution, adopted in September 2005. The technical project “Clean Air For Europe” [Amann et al., 2005] has provided tools to assess the efficiency of the current legislation and a basis for its revision [Cuvelier et al., 2007]. On the prevention side, the objective is to anticipate pollution

¹Institut National de l'Environnement Industriel et des Risques, Chronic Risks Division, Parc Technologique Alata, Verneuil-en-Halatte, France.

²Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre-Simon Laplace, Gif sur Yvette, France.

³Laboratoire Inter-Universitaire des Systèmes Atmosphériques, Faculté des Sciences et Technologie, Créteil, France.

⁴Météo France, Paris, France.

⁵Agence de l'Environnement et de la Maîtrise de l'Energie, Paris, France.

⁶Laboratoire de Météorologie Dynamique, Institut Pierre-Simon Laplace, Ecole Polytechnique, Palaiseau, France.

events in order to warn the public in advance, particularly its most sensitive fraction (the elderly, young children and asthmatics), as well as to help the authorities take efficient emergency control measures. Air quality forecasting is the tool that can help to deal with such objectives.

[4] Statistical forecasting approaches, based on links between airborne pollutant concentrations and parameters like emissions, wind speed, temperature, etc., have been widely used and generally give satisfactory results, at least for ozone daily maxima [Zeldin and Thomas, 1975; Simpson and Layton, 1983; Robeson and Steyn, 1990]. In the last decade or so, progress in computing technologies has allowed the use of simplified three-dimensional models [Vautard et al., 2001] and then full three-dimensional models for daily forecasts [Tilmes et al., 2002; Cope et al., 2004; McHenry et al., 2004; Vaughan et al., 2004; McKeen et al., 2005]. With respect to the statistical approaches, deterministic three-dimensional models provide a more comprehensive picture of air quality for a given time period and geographical area, indicating for instance city plume directions; moreover, they offer the possibility of assessing the efficiency of a given emission scenario. More recently, ensembles of different models have also been used in order to improve single model forecasts, in experimental settings during limited time periods [Delle Monache and Stull, 2003; Delle Monache et al., 2006]. Operational systems yielding daily forecasts of air quality over Europe from regional three-dimensional models have been developed in several European countries (see, e.g., http://www.eurad.uni-koeln.de/index_e.html), as well as in several smaller regional areas (see, e.g., <http://www.airparif.asso.fr/>). However, an evaluation of air quality predictability, defined here as the quantitative level that a correct forecast of the pollutant concentrations can be made, taking into account several regulated pollutants, at regional or smaller scales and over several years, has not been reported to date.

[5] The purpose of the present paper is to give a quantitative assessment of the French national air quality forecasting and monitoring system (known as PREV'AIR), which became operational in 2003 as a result of a close cooperation between several public organizations. The PREV'AIR consortium includes the French National Institute for Industrial Environment and Risks (INERIS), where it has been implemented; the Agency for Environment and Energy Management (ADEME); the Pierre-Simon Laplace Institute (IPSL), the National Scientific Research Center (CNRS) and the French Meteorological Office (Météo France). It is supported by the French Ministry for Ecology.

[6] The evaluation was carried out over 3 consecutive years (2004, 2005 and 2006) for three pollutants: Ozone (O_3), particulate matter with diameter smaller than $10\text{ }\mu\text{m}$ (PM_{10}), and nitrogen dioxide (NO_2). Such an evaluation is a prerequisite to the development of European air quality forecasting systems such as those foreseen in the Global Monitoring of Environment and Security (GMES) projects related to the atmosphere.

[7] One key issue about air quality predictability is to identify its limiting factors. In meteorology, weather predictability is limited by our ability to accurately describe the current state of atmospheric parameters because the system is intrinsically chaotic. By contrast, the impact of inaccurate initial conditions on air quality predictability is limited,

especially in the long term, due to little sensitivity to initial conditions [Honoré, 2000; Elbern and Schmidt, 2001; Blond and Vautard, 2004]. Instead, predictability is thought to be limited by forcing parameters: meteorology, emissions and boundary conditions. The contribution of weather forecast errors to air quality predictability as lead time increases is addressed here. Beyond this analysis, our findings help in defining potential improvements to air quality forecasting systems.

[8] The representation of the immediate current state of regional air quality is also an issue for operational air quality management. The real-time description of air quality is usually provided by monitoring data. Over the few past decades, air quality monitoring networks have grown up in many regions of the world. In most cases, they consist of a set of ground-based monitoring stations, with well-known characteristics in term of proximity to emission sources (industries, urban areas, roads) and of geographical peculiarities (influence of the local meteorology, for instance). Examples of such networks are EuroAirnet in Europe [EEA, 1999] and the State and Local Air Monitoring Stations (SLAMS) network in the United States [Demerjian, 2000]. However, these networks give a spatially discrete description of the concentration fields. Before real-time regional monitoring can be improved with space-borne measurements, which could take more than a decade, models can be used as smart interpolators of the monitoring observations in order to produce near-real time analyses (in the meteorological sense) of concentrations [Elbern and Schmidt, 2001; Blond et al., 2003], using data assimilation techniques. The PREV'AIR system provides ozone analyses in near-real time. The evaluation of this part of the system is also the purpose of this article.

[9] Section 2 presents a description of the PREV'AIR operational system, its architecture, input data and outputs. In section 3, we present a quantitative evaluation of the daily forecasts. Section 4 is devoted to the evaluation of the analysis system. Last, section 5 contains a discussion and concluding remarks.

2. Air Quality Forecasting System

2.1. General Description

[10] The French national air quality forecasting and monitoring system (PREV'AIR) is designed to provide (Figure 1a) forecasts of air quality, up to three days ahead, in order to inform the public and stakeholders sufficiently in advance about pollution episodes; (Figure 1b) a diagnostic, as accurate as possible, of current concentration fields of air pollutants, hereafter called “analyses”; and (Figure 1c) information about the accuracy of the forecasts. The chain of processes leading to the three “products” is illustrated in Figure 1: The system ingests input data from various origins and uses numerical models to produce a forecast of different pollutants. Near-real time air quality observations are combined with the forecasts in order to produce analyses and are used later on to verify the forecasts and quantify the system's skill in predicting concentrations. The accuracy analysis is necessary in order to understand the limits of the forecasting system. To provide in real time a comprehensive picture of air quality, its patterns and the origin of air pollution episodes, forecasts are produced at several spatial

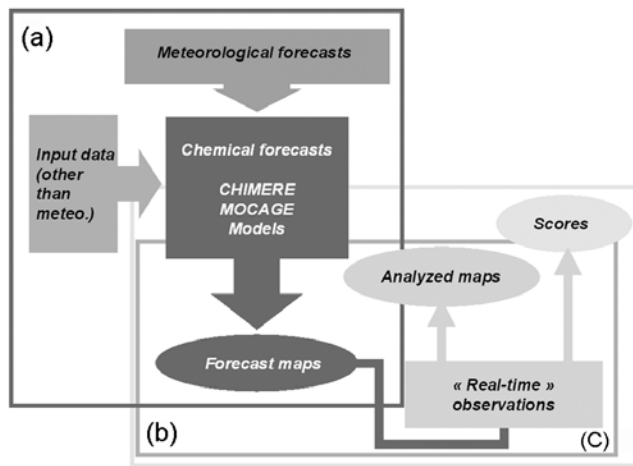


Figure 1. General diagram of the PREV'AIR system: (a) forecasting system, (b) analyses, and (c) assessment of the system's skill.

scales: France, Western Europe (see in Figure 2 the extension of these domains) and the global scale.

2.2. Numerical Air Quality Models

[11] The system is based on three-dimensional deterministic models dedicated to air quality simulation.

[12] The IPSL/INERIS/LISA CHIMERE model [Schmidt *et al.*, 2001; Bessagnet *et al.*, 2004] has been used in several

studies, including the analysis of European air quality extreme episodes such as the 2003 heat wave [Vautard *et al.*, 2005a] and the assessment of past and future trends in air quality over decadal timescales throughout Europe [Vautard *et al.*, 2006; Szopa *et al.*, 2006]. The model participated in intercomparisons [Van Loon *et al.*, 2007] and in the European effort of evaluation of emission control policies [Cuvelier *et al.*, 2007; Thunis *et al.*, 2007]. CHIMERE has also extensively been evaluated at the scale of large urban areas in Europe [Hodzic *et al.*, 2005] and intercompared to other urban air quality models [Vautard *et al.*, 2007]. Model uncertainty has been addressed using Constrained Monte Carlo techniques [Beekmann and Derognat, 2003]. The model can be downloaded from the Web site <http://euler.lmd.polytechnique.fr/chimere/>.

[13] The Météo France MOCAGE model covers scales from the regional to the planetary one, and extending from the surface up to the middle stratosphere. The model comprises several levels of two-way nested domains, the parent global grid providing fully consistent boundary conditions to the inner grids [Peuch *et al.*, 1999]. The settings of this model enable a wide range of scientific applications to be covered, from the study of climate-chemistry interactions [Teyssède *et al.*, 2007], global-scale distributions of species [Cathala *et al.*, 2003; Pradier *et al.*, 2006] to “chemical weather” forecasting down to the regional scale.

[14] These two models are run every day and produce routine forecasts within PREV'AIR. Results for O₃, NO₂ and PM₁₀ are displayed on the Web site <http://www.>

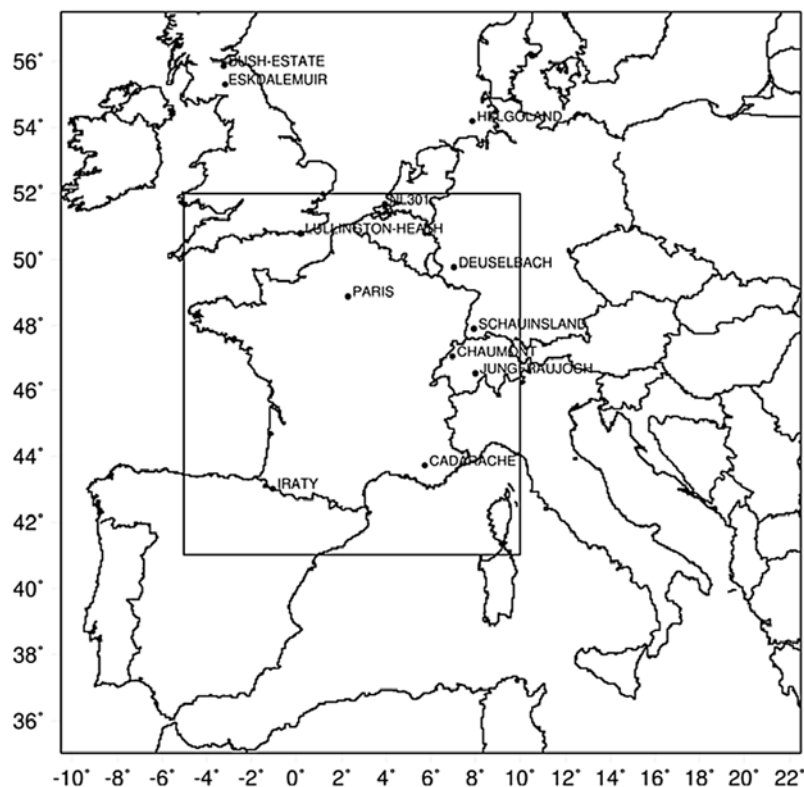


Figure 2. European and French domains in the PREV'AIR system. The locations of some measurement stations are also displayed.

prevair.org/. Forecasts are provided to forecasters in order to formulate messages and recommendations to the public, broadcast by French television public channels. Only one model forecast is displayed over a given domain: MOCAGE ozone forecasts are displayed at a global scale while CHIMERE forecasts are shown at European and French scales.

2.2.1. Geometric Configuration

[15] The horizontal resolution used for the European scale is 0.5° (approximately 50 km) in longitude and latitude; 0.15° in longitude and 0.1° in latitude at the national (French) scale (approximately 10 km). MOCAGE runs at the global scale, with a horizontal resolution of 4° . CHIMERE simulates the atmosphere below 500 hPa only. Below this altitude, 20 levels are taken into account in MOCAGE vs. 8 in CHIMERE. 47 vertical levels are actually considered by MOCAGE to describe the atmosphere up to 5 hPa (approximately 35 km). Thus only MOCAGE can handle the stratospheric contribution to the ozone levels in the troposphere.

2.2.2. Meteorology

[16] CHIMERE and MOCAGE are two offline chemistry transport models: external meteorological models calculate the meteorological data needed by these models. MOCAGE uses meteorological data resulting from forecasts carried out by Météo France, using the operational meteorological models ARPEGE [Courtier *et al.*, 1991]. MOCAGE is forced every three or six hours with pressure, temperature, wind and moisture fields.

[17] Regarding CHIMERE, the 5th Penn State mesoscale meteorological model MM5 [Dudhia, 1993], version 3.6, with modifications on ice cloud microphysics and boundary layer described in the work of Chiriac *et al.* [2005], is run daily over Europe and France to generate the meteorological forecasts with a 36 km and 18 km resolution respectively. MM5 is forced by the analyses and forecasts from the Global Forecast System (GFS, previously known as AVN/MRF) operated daily by the American National Centers for Environmental Prediction (NCEP). Grid nudging to the GFS analyses/forecasts is used to force MM5 in order for the model not to diverge from the initial forecast.

2.2.3. Gas Phase Chemistry

[18] CHIMERE uses the MELCHIOR gas-phase chemical mechanism [Lattuada, 1997]. From the complete mechanism, a reduced mechanism system of equations has been derived [Derognat, 2002] according to the concept of chemical operators [Aumont *et al.*, 1996]: the reduced mechanism includes 44 chemical species and about 120 reactions. As for MOCAGE, it takes into account about 120 gaseous species and more than 350 chemical reactions, based on the RACM and REPROBUS chemical mechanisms [see Dufour *et al.*, 2004, and references therein]. Such a comprehensive mechanism makes it possible to account for the gaseous chemistry in the troposphere and in the stratosphere as well.

2.2.4. Aerosols

[19] In PREV'AIR, CHIMERE has been providing aerosol forecasts since the winter of 2004 over the European domain (not over the refined French domain due to computer time constraints). The internally mixed particles are assumed to be made of anthropic primary particulate matter (PPM), the composition of which is unspecified; sulfates,

nitrates, ammonium, secondary organic species (SOA) and telluric dust or particulate matter resuspended by wind and turbulence. The latter accounts for a large fraction of the coarse particles [Putaud *et al.*, 2004], and is modeled following Vautard *et al.* [2005b]. The main aerosol processes considered in CHIMERE are coagulation, deposition (dry and wet), absorption of the semivolatiles and nucleation of sulfuric acid [Bessagnet *et al.*, 2004].

2.3. Input Data

[20] Besides meteorological fields, the PREV'AIR system requires other types of input.

2.3.1. Routine Observations of Air Quality

[21] Near-real time observations of O_3 , NO_2 and PM_{10} concentrations are used within PREV'AIR: in combination with model prediction, for the production of analyses (see section 2.6); for the assessment of the system's skill in predicting concentrations. Besides, observations are used to train Model Output Statistics (MOS) forecasts (see section 2.5) used to a priori correct the ozone concentration field forecasts from possible biases at individual stations.

[22] In France, observations are carried out by official local bodies in charge of air quality monitoring. They are collected at the national level on an hourly or three-hourly basis (depending on the severity of the pollution event) and stored in the near-real time air quality database Base de données en Temps Réel (BASTER). This database has been developed and is managed by the National Agency for Environment and Energy Management (ADEME). Observations from other European institutions in charge of air quality management (from the Regional Agency for Environmental Prevention (ARPA) in Piemonte and Emilia-Romagna, Italy; from the Federal Environment Agency (UBA) in Germany) are also retrieved in near-real time by the PREV'AIR system. These air quality data are automatically downloaded on a daily basis, or even more frequently in the event of a pollution episode. The total number of rural, suburban and urban stations from which near-real time measurements are available is respectively: 138, 195 and 310 for O_3 ; 85, 144 and 341 for NO_2 ; 62, 96 and 241 for PM_{10} .

2.3.2. Emission Data

[23] At the European and French scales, emissions are derived from the EMEP yearly totals (see for example Vestreng, 2003] and the Web site <http://www.emep.int/>), calculated within the framework of the Convention on Long-range Transboundary Air Pollution (CLRTAP). For CHIMERE, emissions are projected on the model grid, which can be of higher resolution. Urban areas (according to the Global Land Cover Facility land use, see <http://glcf.umd.edu/>) are taken into account: a greater quantity of emissions is projected there. Yearly totals are then disaggregated using monthly, daily and hourly factors and a speciation of nonmethane volatile organic compounds (NMVOC) [Passant, 2002]. At global scale, MOCAGE uses emissions from the GEIA and EDGAR inventories [Michou and Peuch, 2002].

2.3.3. Boundary Concentrations

[24] At the European scale, CHIMERE uses climatological monthly means, computed with the LMDzINCA model [Hauglustaine *et al.*, 2004; Folberth *et al.*, 2005], for a dozen chemical gaseous species (O_3 , NO_2 , and NMVOC)

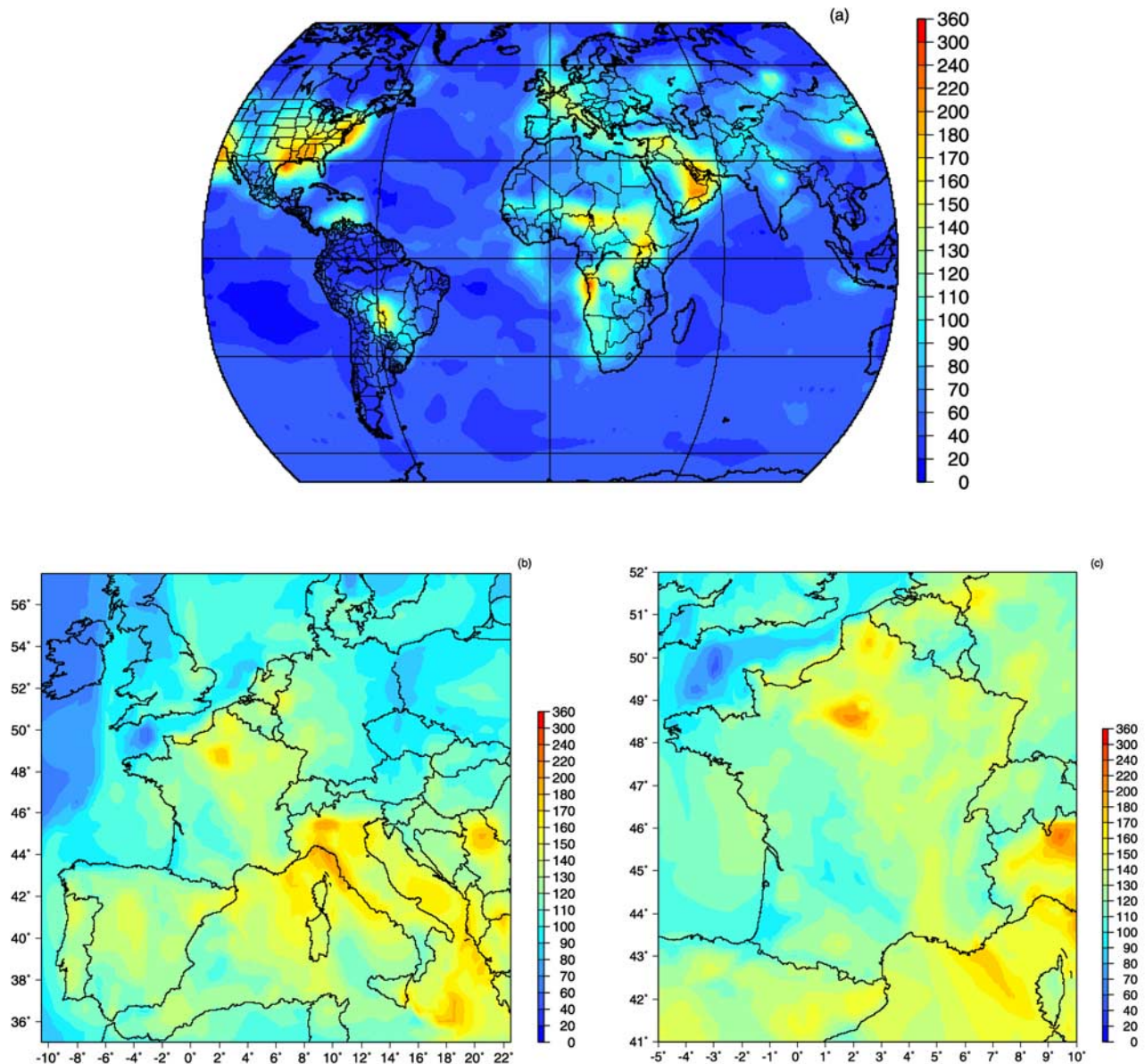


Figure 3. Ozone daily maximum D + 0 forecast, 30 June 2006, over (a) the globe, (b) Europe, and (c) France. Concentrations are in $\mu\text{g}/\text{m}^3$. They are displayed for the first layer of the model.

and the GOCART monthly climatologies for aerosol concentrations [Ginoux *et al.*, 2001]. The zoomed higher-resolution CHIMERE version over France is forced at its boundaries by the results obtained from the model version run over Western Europe. At the global scale, concentrations on top of the MOCAGE domain (5 hPa) are derived from climatologies computed with another version of the model, across the whole stratosphere and mesosphere, run in a climatological mode.

2.3.4. Initial Concentrations and System Cycle

[25] The PREV'AIR system is a cycle of operations, each being performed once a day. The day when the forecast is available is called D + 0 (day + 0); the day after D + 0 (resp. D + 1) is called D + 1 (resp. D + 2).

[26] All CHIMERE forecasts actually begin on the previous day (D − 1) at 00:00 Universal Time (UT), and are

first forced by meteorological analyses (instead of forecasts) until D − 1 at 12:00 UT. Thus the meteorological MM5 forecast is forced by short-term GFS forecasts only from 12:00 UT to midnight on D − 1. The resulting CHIMERE simulation for D − 1 is therefore based on meteorological data close to be free of forecast errors. CHIMERE runs on a 32 bits biprocessor personal computer, with 4 Go of RAM, operated under Linux. The whole forecast starts at 19:00 UT of D − 1 and is completed in 5 hours.

[27] MOCAGE simulations are sliced by day of forecasts and run on the operational chain of Météo France. D − 1, D + 0, D + 1 and D + 2 runs start respectively at 15:10 UT, 18:25 UT, 21:25 UT of D − 1. One day of forecast by MOCAGE is completed in 3 hours. Fields are delivered to the PREV'AIR system as soon as they are available, all along the run.

[28] At initial time, CHIMERE and MOCAGE use the result of the previous day's 24-hour forecast as initial conditions.

2.4. An Example of Ozone Forecast

[29] As an example, Figure 3 shows the ozone daily maximal concentrations forecasted at the global, European and French scales on 30 June 2006, for $D + 0$.

[30] On the global scale, large ozone plumes are simulated, associated with high anthropogenic emission areas (over northeast America) but also originating from biogenic activity (over Africa). These photochemical plumes are potentially advected at an intercontinental scale, mostly in the free troposphere and as reservoir species like PAN. This day was associated with high ozone concentrations in some areas in Europe. There are some differences between the PREV'AIR European and French forecasts: ozone daily maxima forecasted are higher at the fine scale than those obtained with the coarse scale version, over several areas of France. One can distinguish the impact of the higher-resolution emission inventory at the national scale in particular: higher NO_x emissions in the core center of Paris are responsible for ozone concentration decrease in this area, as seen on the higher-resolution map.

2.5. Model Output Statistical Forecast

[31] Model Output Statistics (MOS) is a classical procedure in meteorology used to correct model forecasts at individual stations [Glahn and Lowry, 1972; Klein and Glahn, 1974]. A site-dependent regression model, trained over past data, is applied to forecast parameters. In PREV'AIR, a MOS forecast is issued in addition to the raw forecast for ozone daily maxima predicted by CHIMERE in France.

[32] MOS relies on a training procedure: observations over a past period and simultaneous forecasts are required. Training seasons and implementation of the statistical adaptation must be consistent, deficiencies in the air quality forecasting models being a priori dependent on the season. During the summer of 2006 (from 06/15/2006 to 09/15/2006), MOS was applied in PREV'AIR, using as training period the summers of 2003 to 2005, with the same model version as the one implemented during the summer of 2006. In the current application, the forecast error $O_{3,obs}(s) - O_{3,mod}(s)$ is regressed, at each monitoring site s , from the predicted 2 m temperature $T_{2m,mod}(s)$ issued from the MM5 forecast, and the predicted ozone daily maximum $O_{3,mod}(s)$, thus giving the estimated forecast error (EFE(s)):

$$O_{3,obs}(s) - O_{3,mod}(s) \approx \text{EFE}(s) = \alpha \cdot T_{2m,mod}(s) + \beta \cdot O_{3,mod}(s) + \gamma.$$

This choice is motivated by a tendency of the model to underestimate high ozone concentration values found, especially in hot summer spells.

[33] A different set of multiple regression coefficients is calculated for each site and each forecast lead time. Then MOS daily maxima $O_{3,MOS}(s)$ are calculated for each monitoring station, as the sum of the ozone forecast and the estimated forecast error:

$$O_{3,MOS}(s) = O_{3,mod}(s) + \text{EFE}(s).$$

[34] The MOS forecasts issued for each monitoring station are then interpolated over the whole modeling domain by the same kriging method as the one used for the analyses, presented in section 2.6. The kriging of the "innovations" is applied, taking here as innovations the difference between the MOS forecasts and the raw model forecasts.

[35] The choice of the measurement sites used to correct ozone concentration forecasts is a key stage in the MOS procedure: the stations must be selected in keeping with the type of information delivered by PREV'AIR, which depend on the resolution of the model. In practice, for the implementation of the MOS forecast in PREV'AIR, rural stations are selected in priority. When no rural station is present in a large region, suburban stations are selected and, in the absence of such stations, urban stations are selected provided that the influence of local sources of pollution and local meteorology is minor. In practice, for the implementation of the MOS forecast in PREV'AIR, 101 stations are retained, including 39 (out of about 50) rural, 27 (out of about 110) suburban and 32 (out of about 210) urban.

[36] Figure 4a and 4b show an example of MOS forecast for the surface ozone daily maxima, established for 30 June 2006, together with the difference field (namely, raw – MOS forecast concentrations). Negative (resp. positive) differences are in blue (resp. in red) and correspond to areas where the raw forecast is lower (resp. larger) than the MOS forecast. The strongest negative differences are located in the southeast region near Nice. The strongest positive differences are located in the Alps regions. A systematic evaluation of the skill improvement due to the application of MOS is given at the end of section 3.1.

2.6. Concentration Analyses

[37] The analysis method is designed to assess, as accurately as possible, near-real time surface concentration fields of ozone. Actually, PREV'AIR analysis is carried out over France only, but the approach can easily be generalized to a larger, for example European, domain. The observations retrieved in near-real time from the BASTER database are used in combination with $D - 1$ ozone daily maxima. We use one of the methods proposed by Blond *et al.* [2003], based on the kriging of the differences between simulated and observed values, often called innovations in meteorology. Kriging methods have the advantage of providing spatial interpolations that necessitate few assumptions and give robust results. Few assumptions are needed in kriging methods; a sensitivity analysis on the kriging parameters has been performed, enabling to select the most appropriate parameters.

[38] Like for the MOS procedure, the choice of the measurement sites is a crucial stage in the analysis procedure: the monitoring stations selected must deliver concentrations representative of the gridded concentrations. In practice, the same criteria as in the MOS procedure guide the choice of monitoring stations: rural stations are selected in priority, then suburban stations and urban stations, provided that the influence of local sources of pollution and local meteorology is minor. In France, 49 rural, 49 suburban and 25 urban stations are finally selected.

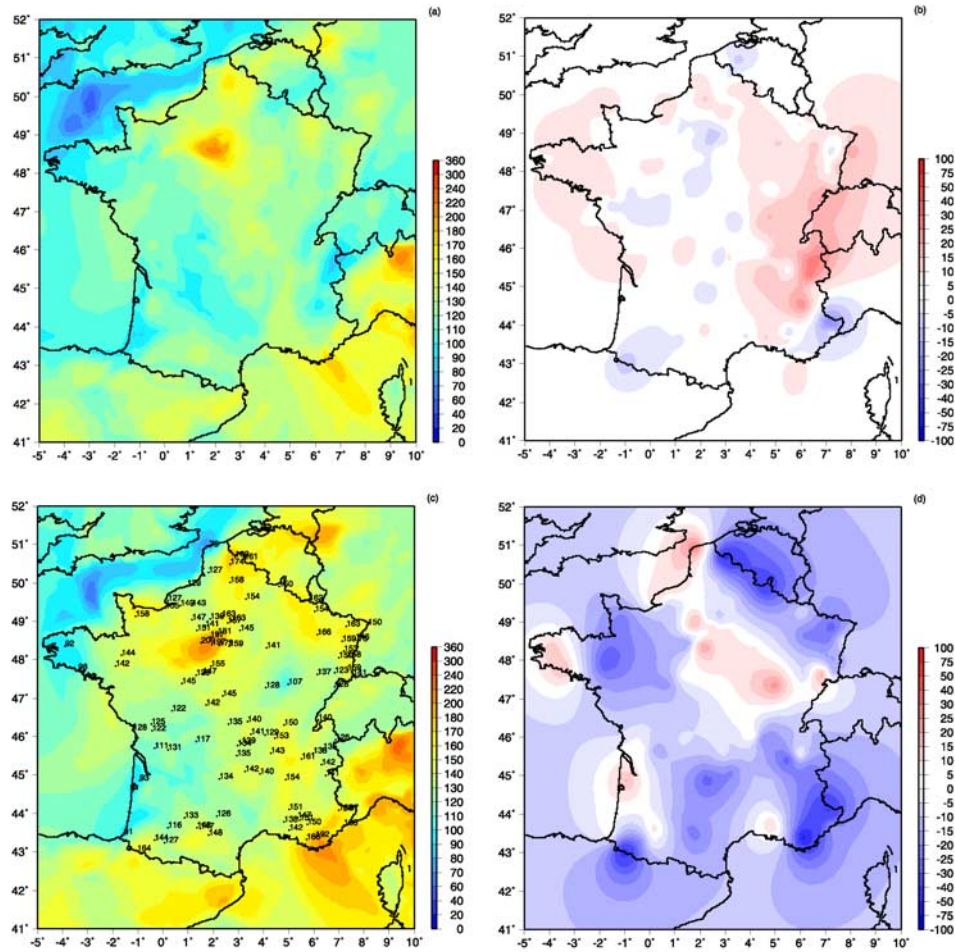


Figure 4. (a) Ozone daily maximum D + 0 MOS forecast, 30 June 2006, France. (b) Difference field (raw – MOS forecast concentrations). (c) Analysis of ozone daily maximum, 30 June 2006, France. (d) Difference field (modeled – analyzed concentrations). Concentrations are in $\mu\text{g}/\text{m}^3$. They are displayed for the first layer of the model.

[39] At a given location s , the analyzed concentration is calculated from the following equation:

$$Z_a(s) = Z_b(s) + \sum_{k=1,p} w_k(s) \cdot (Y_o(s_k) - Z_b(s_k)),$$

where $Z_a(s)$ refers to the analyzed concentration at site s ; $Z_b(s)$ refers to the corresponding simulated value; $Y_o(s_k)$ is the measured concentration at site s_k and $w_k(s_k)$ are the weights derived from the kriging constraints (see *Blond et al.* [2003] for more details about the method). Innovations $Y_o(s_k) - Z_b(s_k)$ are estimated at each monitoring site s_k .

[40] The kriging method used here is “exact”: at the measurement sites, the analyzed concentration is equal to the observed concentration. Analyses can be calculated on an hourly basis or for daily maxima; the second option is operational in the current PREV’AIR system.

[41] Figure 4c and 4d show an example of analysis of surface ozone daily maxima, established for 30 June 2006, together with the difference field (model – analysis); the difference with the ozone field calculated by PREV’AIR varies between -55 and $+24 \mu\text{g}/\text{m}^3$. The strongest negative differences (in blue) are located around Toulon and in

northeast of France. The strongest positive differences (in red) are located southeast of Paris, in Bourgogne. The error in analyses obtained with this method is evaluated in section 4.

3. Evaluation of the PREV’AIR Regional Forecasts

3.1. Methodology

[42] It is not a straightforward task to compare model outputs with in situ measurements. Indeed, these two kinds of data are in nature different [*Riccio et al.*, 2006]: 3D air quality models compute the evolution of pollutant concentrations on a grid; the concentrations can be thought of as volume averages over each grid cell and time averages over the integration time step. Observations are available from fixed measurement sites; they are local data and may be (and usually are) influenced by local processes. A problem one has to face when evaluating a model against measurement data is therefore to guarantee the representativeness of the observation sites used [*McNair et al.*, 1996]. Representativeness depends on the characteristics of the site (topography, proximity to emission sources) and on the chemical

Table 1. Number of Rural (RUR), Suburban (SUB), and Urban (URB) Background Sites Used in This Study for Different Countries and Pollutants

Country	O ₃			NO ₂			PM ₁₀		
	RUR	SUB	URB	RUR	SUB	URB	RUR	SUB	URB
Belgium	5	3	0	0	0	0	1	1	1
Czech Republic	10	6	0	0	0	0	0	0	0
France	43	50	75	20	35	90	3	20	46
Germany	56	22	0	0	0	0	32	15	0
Netherlands	12	0	0	0	0	0	8	0	0
Switzerland	7	1	0	0	0	0	0	0	0
United Kingdom	16	2	0	0	0	0	0	0	0
Total for Europe	149	84	75	0	0	0	44	36	47

species under evaluation (lifetime). The selection of representative observation sites also depends on the model itself (typically, the horizontal and vertical resolution) and on the purpose of the evaluation [Schmidt *et al.*, 2001].

[43] However, PREV'AIR is an operational forecasting system and its performance has been evaluated against as many available observations as possible. We have not evaluated the system against measurements carried out at traffic or industrial sites, because it is beyond the scope of the system; for the long-term evaluation of the PREV'AIR system performed here, we have used observational data from rural, suburban and urban sites. Table 1 summarizes the number of stations available per country for the evaluation of the PREV'AIR forecasts.

[44] Figure 5 (for ozone), Figure 6 (for PM₁₀) and Figure 7 (for NO₂) show the spatial distribution of the monitoring sites throughout Europe. Measurement stations are spread out over different countries. Unfortunately, we have not yet managed to use data from all European countries, in particular from southern countries (Spain and Italy).

[45] Observed concentrations are compared to forecasts in the first model layer, bilinearly interpolated at the measurement point. This choice is debatable for mountain sites when there is a strong altitude difference between the actual and model topographies. However, it is an objective and simple choice. For ozone, the evaluation is made for daily maxima (computed based on hourly modeled concentrations) while for other pollutants (PM₁₀, NO₂) it is made on daily mean concentrations. Only those monitoring stations for which at least 75% of the daily measurements are available are considered. Skill scores are considered for each observation site separately; for all observation sites in Europe sorted by type (rural/suburban/urban). Finally, since ozone episodes occur essentially during spring and summer, ozone daily maxima are evaluated only over this period (April to September). For other pollutants, all yearlong data are used for evaluation; we distinguish between spring/summer and fall/winter seasons.

[46] Classical statistical skill scores are used for the evaluation: bias, root mean square error (RMSE), and correlation. Bias indicates, on average, if the forecasts are under- or overpredicted. In our case, negative values indicate underprediction, whereas positive values indicate overprediction; small values are best. RMSE gives information about the skill in predicting the overall magnitude of the observations. Correlation is a measure of whether forecasts and observations change together (i.e., at the same time and/

or place). The closer the correlation is to one (100% in the figures), the better is the correspondence of extreme values of the two data sets. For ozone daily maxima, the percentages of correct predictions (GF+), false alarms (FA), or missed events (ME) related to the information threshold (180 $\mu\text{g}/\text{m}^3$) are also estimated. Bias and RMSE are in $\mu\text{g}/\text{m}^3$; correlation has no unit, and GF+, FA, and ME are in %.

[47] Error statistics are cumulated over all days of the evaluation periods (3 seasons in 2004, 2005, 2006), for each forecast lead time (D + 0, D + 1, D + 2). Simulated (D − 1) concentrations are also evaluated and compared to forecast concentrations for reference, in order to analyze the impact of meteorological forecast errors. Unless specified, the evaluation is performed on the European domain. The location of some measurement sites that are referred to hereafter is specified in Figure 2.

3.2. Skill of Ozone Daily Maxima Forecasts

[48] Table 2 displays error statistics obtained for the ozone daily maxima computed by the operational system during the springs/summers of 2004 to 2006.

[49] On average, at the European scale, the system overestimates ozone daily maxima observed at rural, suburban and urban sites (the absolute bias ranges from −0.4 to 1.7 $\mu\text{g}/\text{m}^3$, 1.2 to 3.6 $\mu\text{g}/\text{m}^3$ and 3.7 to 5.6 $\mu\text{g}/\text{m}^3$ respectively); the agreement is better for rural and for suburban sites than for urban sites. This can easily be understood since ozone at urban sites undergoes titration with nitrogen monoxide, a process that is underestimated in the model because nitrogen oxides are diluted in large grid cells (see also section below on NO₂).

[50] There is no significant bias in the forecasts of the ozone daily maxima, on average over all stations. However, large positive or negative biases occur at some sites, which compensate on average.

[51] Figure 5b shows the spatial distribution of biases at rural sites for D + 1 forecasts. Biases range from −18.9 $\mu\text{g}/\text{m}^3$ at the high mountain site JUNGFRUJOCH in Switzerland (about 3500 m high) to 14.5 $\mu\text{g}/\text{m}^3$ at ESKDALEMUIR, UK. Median bias is equal to 2.2 $\mu\text{g}/\text{m}^3$. Urban sites are characterized by positive biases (not shown).

[52] A general pattern is that the model overestimates daily ozone maxima over maritime areas and within a belt of several hundred kilometers from the Atlantic coast and underestimates them over continental, central areas. This pattern was also found in simulations of surface ozone [Jonson *et al.*, 2005; Vautard *et al.*, 2006] for the recent parts of the decadal simulations performed in these studies. It has been hypothesized that precursor emissions in central Europe are underestimated, although other factors can explain this, as for instance an overestimation of dry deposition or inaccuracies in the sensitivity of photochemistry to emissions.

[53] RMSE averaged over Europe varies between 16.8 and 19.4 $\mu\text{g}/\text{m}^3$ at rural sites, between 17.6 and 20.1 $\mu\text{g}/\text{m}^3$ at suburban sites and between 17.3 and 19.4 $\mu\text{g}/\text{m}^3$ at urban sites, depending on the lead time (see Table 2). The skill smoothly decreases with lead time. The agreement is best for rural, then better for urban sites than for suburban sites. The spatial distribution of RMSE is shown in Figure 5c at rural sites for D + 1 forecasts: RMSE ranges from 11.7 $\mu\text{g}/\text{m}^3$ at

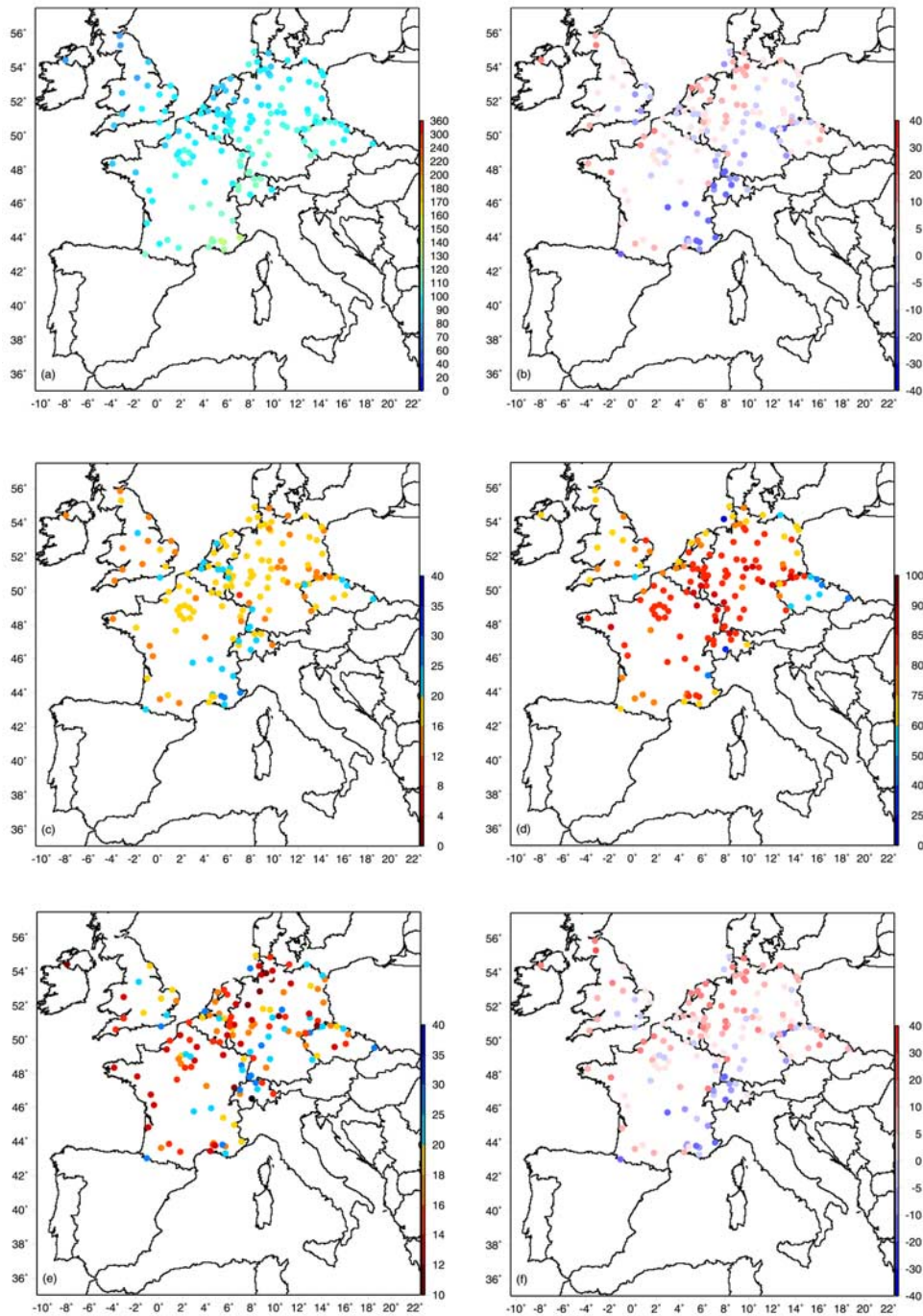


Figure 5. Skill scores for ozone, over the spring/summer of 2004–2006, for the European D + 1 forecast at rural sites. (a) Observed mean (in $\mu\text{g}/\text{m}^3$); (b) bias (in $\mu\text{g}/\text{m}^3$); (c) RMSE (in $\mu\text{g}/\text{m}^3$); (d) correlation (in %) of ozone daily maxima; (e) NGE_120 (in %); and (f) UPPA (in %).

DEUSELBACH, Germany to $29.5 \mu\text{g}/\text{m}^3$ at CADARACHE, France. Median RMSE is equal to $17.8 \mu\text{g}/\text{m}^3$.

[54] RMSE is generally weak in the western and central parts of France and becomes greater in areas with more

complex terrain. Skill is poorer along the Mediterranean coast where land/sea breezes are poorly resolved by the meteorological model at the resolution considered (about 40 km). The development of sharp ozone plumes in the

Figure 6. Skill scores for the PM_{10} daily mean, for the European D + 1 forecast, over (left) the spring/summer of 2004–2006 and (right) the fall/winter of 2004–2006. (a) Observed PM_{10} daily mean concentration (in $\mu\text{g}/\text{m}^3$); (b) bias (in $\mu\text{g}/\text{m}^3$); (c) RMSE (in $\mu\text{g}/\text{m}^3$); and (d) correlation (in %). Scores are displayed at all sites: diamond symbols refer to rural sites, small circles to urban sites, and large circles to suburban sites.

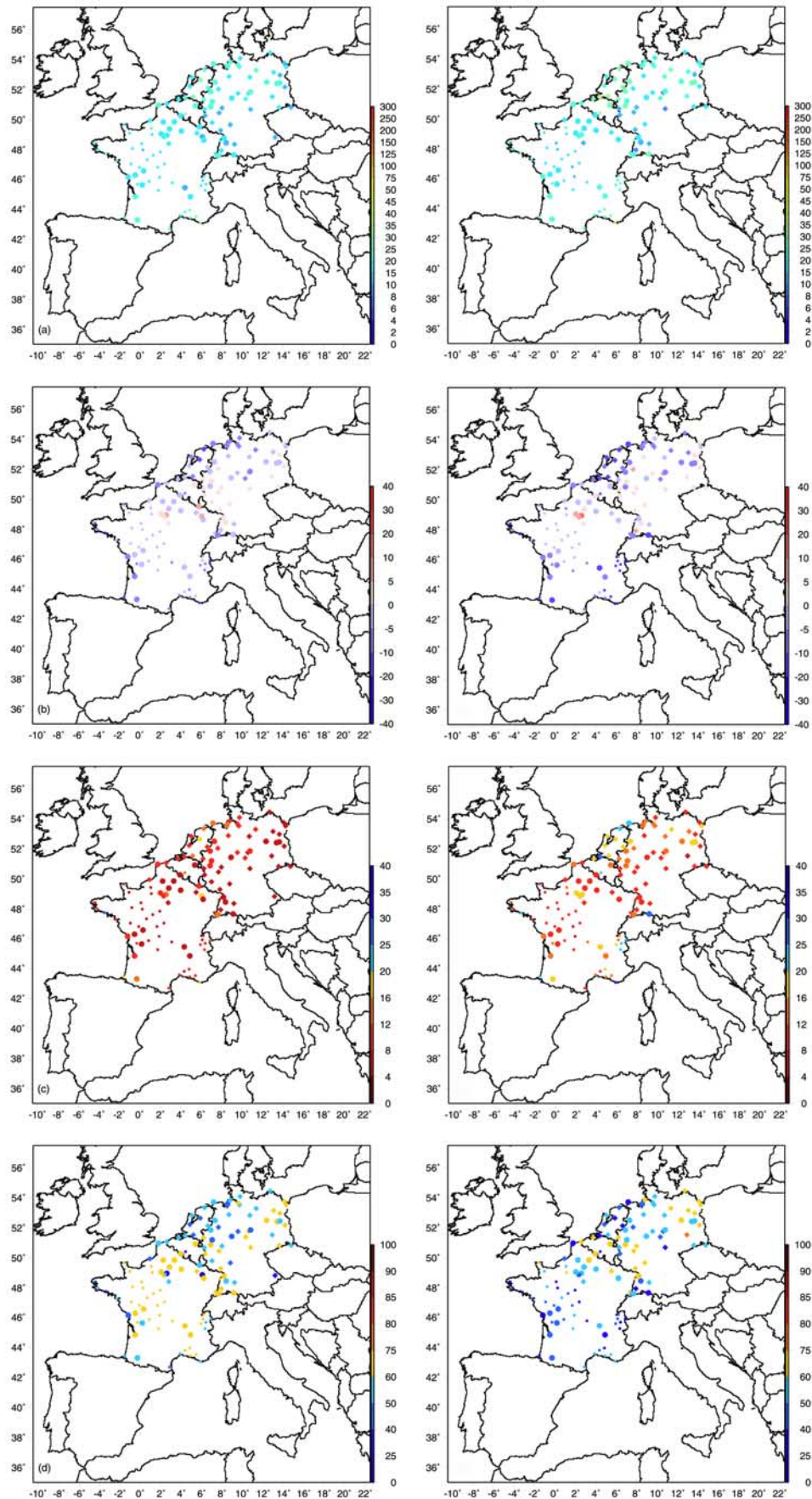


Figure 6

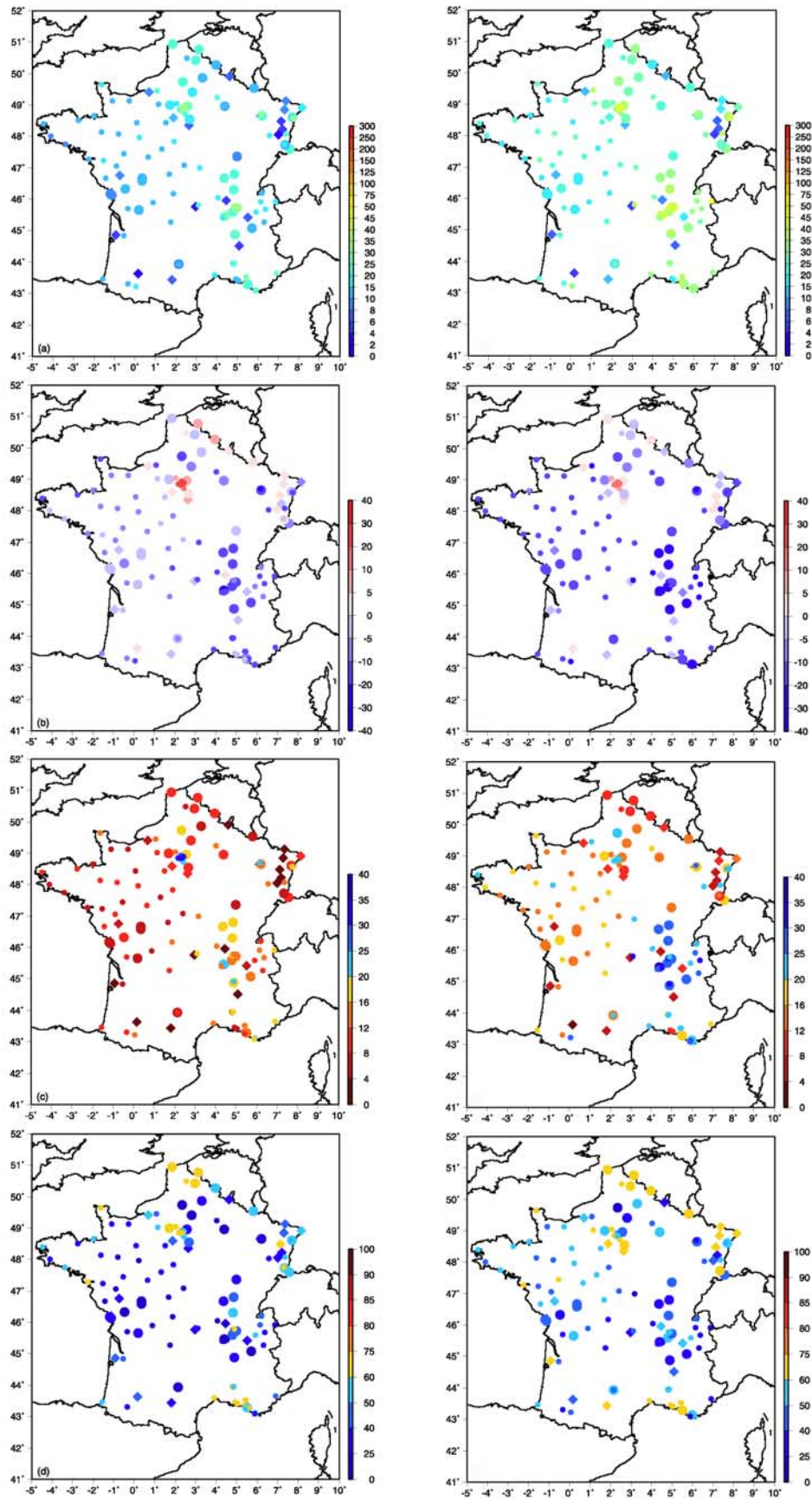


Figure 7. Same as Figure 6, but for the NO₂ daily mean for the France D + 1 forecast.

Table 2. Observed Mean of Ozone Daily Maxima (in $\mu\text{g}/\text{m}^3$) and Skill Scores for the Ozone Daily Maxima Over the Spring/Summer of 2004–2006 for the European Forecast^a

	Lead Time	RUR	SUB	URB
Mean obs.	...	100.1	102.0	99.3
Bias	D – 1	1.7	3.6	5.6
	D + 0	0.9	2.6	4.8
	D + 1	0.0	1.6	4.1
	D + 2	–0.4	1.2	3.7
RMSE	D – 1	16.8	17.6	17.3
	D + 0	17.3	18.2	17.7
	D + 1	18.2	18.9	18.3
	D + 2	19.4	20.1	19.4
Corr.	D – 1	0.83	0.83	0.84
	D + 0	0.81	0.82	0.82
	D + 1	0.79	0.80	0.81
	D + 2	0.76	0.77	0.78

^aBias (modeled – observed daily maxima, in $\mu\text{g}/\text{m}^3$); RMSE (in $\mu\text{g}/\text{m}^3$); correlation (no unit). Scores are computed separately for each station type: rural, suburban, and urban (using 50,000, 33,000, and 30,000 observations, respectively). They are displayed as a function of lead time: D – 1, D + 0, D + 1, and D + 2.

wake of the city of Marseille due to the large industrial area nearby [Cros *et al.*, 2004; Drobinski *et al.*, 2007] makes the comparison between observations and simulations from coarse-resolution models even more difficult. Skill is also poorer in mountainous areas, where the complexity of the terrain combines the difficulty of determining the right model layer to use for comparison with observations.

[55] Mean correlation ranges from 0.76 to 0.84, with decreasing values as the forecast lead time increases. Mean correlation is slightly higher for urban and for suburban sites than for rural sites. The spatial distribution of correlation coefficients is shown in Figure 5d at rural sites for D + 1 forecasts: it ranges from 22 at HELGOLAND, Germany to 88 at DEUSELBACH, Germany. Correlation is low at JUNGFRAUJOCH, Switzerland (36), in the Czech Republic, in the UK, in the southeastern part of France, along the North Sea and the Baltic Sea. Median correlation is equal to 81. Some rural and suburban sites have very poor correlation, whereas correlation at urban sites is almost always higher than 65.

[56] Although the performance of the system is quite good (see next paragraph), it is not able to catch the small increase in concentrations, from urban to suburban sites and from rural to suburban sites.

[57] The day-to-day variability of ozone daily maxima is fairly well forecasted but the skill of the forecast clearly depends on the lead time, with a degradation in performance, except for the bias, due to meteorological forecast errors. However, comparing the above indices between D – 1 and other lead times, one concludes that ozone daily maxima forecast errors are not dominated by meteorological forecast errors. Indeed, the forecast error variance (the square of the RMSE) can be separated into two parts: the first one is an error variance that is close to be free of meteorological errors (the square of the RMSE for D – 1); the second one is an error that relates to the meteorological forecast error. Subtracting the square of the RMSE for D – 1 from that of D + 0 (resp. D + 1 and D + 2), for rural stations, it is found that only 6% (resp. 15% and 25%) of

the forecast error variance is due to meteorological forecast error.

3.2.1. Comparison With Other State-of-the-Art Operational Forecasting Systems

[58] McKeen *et al.* [2005] and Pagowski *et al.* [2006] computed bias, RMSE and correlation of D + 0 hourly concentrations forecast over the Eastern USA and Southern Canada for July and August 2004. Skill scores were computed for seven AQ models over 350 sites and for ensemble forecasts. They are displayed in Table 3 together with the equivalent scores computed for the PREV'AIR system. Note that these scores refer to ozone hourly concentrations; they are not as good as the equivalent scores computed for the ozone daily maxima. Indeed, the evolution of ozone over the course of the day is much harder to capture, especially during nighttime when mixing processes are slowed and titration of ozone by nitrogen monoxide becomes more influent (note the strong positive bias). Moreover, knowledge of emission diurnal profiles is of primary importance in order to derive ozone diurnal cycles consistent with the observations; it is probable that these profiles are not known with the required accuracy throughout Europe. Although the scope of the two studies is different (area, time period, length of the data sets), these figures show that the performance of the PREV'AIR system, with respect to ozone forecasts, is on the high side as compared to that of other air quality forecasting systems: PREV'AIR skill scores always lie in the lower (upper for correlation) range of the values computed for the seven models; RMSE is even 5 $\mu\text{g}/\text{m}^3$ below. Note that the best scores are associated with the ensemble forecast; this ensemble approach is rather promising in order to increase substantially the performance of air quality forecasting systems.

[59] In order to evaluate whether the PREV'AIR ozone forecasts would conform to U.S. Environmental Protection Agency (EPA) recommendations, we computed two performance indicators prescribed by the EPA [U.S. Environmental Protection Agency, 1991]: the Normalized Gross Error for a cutoff value of 120 $\mu\text{g}/\text{m}^3$ (NGE₁₂₀, %) and the Unpaired Peak Prediction Accuracy (UPPA, %). The former index is evaluated on hourly concentrations and gives information on the forecast accuracy for higher concentrations; it is computed by averaging the differences, in absolute value, between observations above 120 $\mu\text{g}/\text{m}^3$ and the corresponding forecasts, normalized by the observations. The latter gives information on the forecast accuracy for daily maxima; it is computed by averaging the differences between observations and the corresponding forecasts of the peak value, normalized by the observations (the term “unpaired” refers to the fact that observed and forecast maxima might occur at different times). The EPA

Table 3. Skill Scores (Bias, RMSE in $\mu\text{g}/\text{m}^3$, and Correlation) for the Ozone D + 0 Hourly Forecasts Compared to Those Obtained by Pagowski *et al.* [2006]^a

	PREV'AIR	Pagowski <i>et al.</i> [2006]
Bias	12.3	10.6, 16.0, 18.6, 23.1, 29.6, 62.2, 17.6 (1.8)
RMSE	28.2	33.0, 36.2, 37.6, 48.2, 42.8, 74.9, 33.0 (21.1)
Corr.	0.67	0.67, 0.62, 0.63, 0.70, 0.65, 0.55, 0.72 (0.81)

^aFigures in parentheses refer to ensemble forecasts.

Table 4. Contingency Tables for the Ozone Daily Maxima Over the Spring/Summer of 2004–2006 at the European Scale and for the 180 $\mu\text{g}/\text{m}^3$ Threshold^a

	Lead Time	RUR	PUB	URB
GF+	D – 1	34.8	35.4	40.6
	D + 0	27.9	27.2	35.0
	D + 1	20.4	20.4	22.5
	D + 2	15.0	15.1	20.2
FA	D – 1	0.7	0.9	0.9
	D + 0	0.6	0.8	0.8
	D + 1	0.4	0.4	0.5
	D + 2	0.3	0.5	0.5
ME	D – 1	65.2	64.6	59.4
	D + 0	72.1	72.8	65.0
	D + 1	79.6	79.6	77.5
	D + 2	85.0	84.9	79.8

^aAbbreviations are as follows: GF+, percentage of correct forecasts above threshold; FA, percentage of false alarms; ME, percentage of missing events. Scores are displayed as a function of lead time and station type. The number of observations above threshold is equal to 750, 700, and 430 at the rural, suburban, and urban sites, respectively.

recommends upper limit values of 35% for NGE₁₂₀, and $\pm 20\%$ for UPPA. The two indicators were computed for the operational system for the springs/summers of 2004 to 2006.

[60] On average, at the European scale, both indicators lie well below the EPA recommendations: NGE₁₂₀ ranges from 13 to 15% at urban sites, 14 to 16% at suburban sites and 19 to 21% at rural sites; UPPA lies in the range of 4 to 5% at rural sites, 6 to 7% at suburban sites and 8 to 9% at urban sites. There is a depredation of these indicators with the lead time.

[61] Figure 5e and 5f display an overview of the spatial distribution of the two indices (rural sites, D + 1 forecast): NGE₁₂₀ ranges from 8% at BUSH-ESTATE, UK to 48% at the high mountain site JUNGFRUJOCH, Switzerland, again somehow out of the distribution; actually, NGE₁₂₀ lies well below EPA criteria for all sites except for JUNGFRUJOCH. Median NGE₁₂₀ is equal to 15.5%. There is a larger dispersion in NGE₁₂₀ computed at rural sites than at suburban and urban sites. Highest NGE₁₂₀ (>25%) are found at elevated, rural sites (CHAUMONT, Switzerland; IRATY, France; SCHAUMSLAND, Germany) or at seaside rural sites (NL-301, Netherlands, LULLINGTON-HEATH, UK; HELGOLAND, Germany). Lowest NGE₁₂₀ (<11%) are computed at sites of all kinds throughout Europe.

[62] UPPA ranges from –19% at JUNGFRUJOCH, Switzerland to 23% at ESKDALEMUIR, UK; actually, UPPA lies well below EPA criteria for almost all sites: only

1 out of 149 rural sites (5 out of 84 suburban sites and 3 out of 75 urban sites) are above the acceptance criteria. Median UPPA is equal to 5.3%. Note the difference in the distributions depending on station type: ozone daily maxima are overestimated at almost all sites, except for about 30% of the rural sites. The distribution of low and high UPPA does not exhibit specific patterns.

3.2.2. Contingency Tables

[63] Table 4 displays the percentages of correct predictions, false alarms and missing events related to the information threshold (180 $\mu\text{g}/\text{m}^3$).

[64] Skill scores associated with threshold criteria are very tough for the model: the percentage of correct forecasts above threshold (GF+) ranges from 41% (urban sites, D – 1) to 15% (rural sites, D + 2); at D + 0, GF+ ranges from 28% (rural sites) to 35% (urban sites). This is in accordance with the fact that, at high ozone concentrations, the system underestimates ozone daily maxima (not shown). Whatever the lead time, GF+ is higher at urban sites; this might be due to a compensation between underestimation and ozone underestimation. There is a large decrease in GF+ with lead time, in accordance with the underestimation increase, whatever the station type (bias ranges from –27 to –43 $\mu\text{g}/\text{m}^3$ for rural sites; from –26 to –41 $\mu\text{g}/\text{m}^3$ for suburban sites; from –22 to –38 $\mu\text{g}/\text{m}^3$ for urban sites). The counterpart of the rather low GF+ rate is a low false alarm rate: in every case, FA is lower than 1%.

3.2.3. MOS Forecasts

[65] The impact of the MOS procedure (see section 2.5) on the ozone daily maxima forecasted in France has been evaluated: Table 5 displays the error statistics obtained for the ozone MOS forecasts compared to those obtained for the raw forecasts in France. For the sake of conciseness, we only display D + 0 results. Note that, reading Table 5, scores for MOS forecasts correspond to the sum of both numbers given in a column (raw forecast + difference). When computing the error statistics, two sets of stations are considered: the first one is made up of stations taken into account in the training stage of the MOS procedure (39 rural sites, 38 suburban sites, 36 urban sites); the second set is made up of stations not used in the elaboration of the MOS procedure (4 rural sites, 11 suburban sites, 35 urban sites). The first set enables the direct impact of the MOS procedure on the ozone forecasts to be assessed, whereas the second set is used to perform an evaluation of the MOS procedure together with the kriging method (see section 2.5).

[66] The MOS procedure generally results in a shifting of biases from negative values to positive values at all station types (except at rural sites of set 2). The difference between

Table 5. Observed Mean of Ozone Daily Maxima (in $\mu\text{g}/\text{m}^3$) and Skill Scores (Bias, RMSE in $\mu\text{g}/\text{m}^3$, and Correlation) for the Ozone Daily Maxima Over the Summer of 2006 for the French Raw Forecast (D + 0)^a

	Set 1			Set 2		
	RUR	SUB	URB	RUR	SUB	URB
Mean obs.	109.5	110.2	106.0	112.0	110.9	107.6
Bias	–2.8 (+5.1)	–1.7 (+4.3)	–0.7 (+3.1)	–5.4 (+2.1)	–2.3 (+3.5)	–3.6 (+5.4)
RMSE	19.4 (–1.7)	18.7 (–0.2)	19.3 (–1.2)	18.0 (+0.4)	20.2 (–0.3)	20.0 (–1.7)
Corr.	0.85 (+0.03)	0.87 (–)	0.85 (+0.02)	0.82 (–0.01)	0.84 (–)	0.85 (+0.02)

^aThe differences in scores between the MOS and raw forecasts are indicated in parentheses; figures in bold indicate better skill scores for the MOS forecast. Scores on the left are computed for set 1 (using 2970, 2970, and 2770 observations over rural, suburban, and urban sites, respectively), and on the right, for set 2 (using 260, 830, and 2760 observations over rural, suburban, and urban sites, respectively).

Table 6. Same as in Table 5, but for an Observation Cutoff Value of $120 \mu\text{g}/\text{m}^3$ ^a

	Set 1			Set 2		
	RUR	SUB	URB	RUR	SUB	URB
Mean obs.	150.2	151.4	148.8	142.4	150.9	149.5
Bias	−13.0 (+7.3)	−11.0 (+6.8)	−10.2 (+5.0)	−10.8 (+3.1)	−13.2 (+4.8)	−10.3 (+5.7)
RMSE	25.2 (−4.9)	24.1 (−2.8)	24.1 (−3.1)	22.3 (−0.5)	27.1 (−2.3)	24.2 (−3.1)
Corr.	0.59 (+0.07)	0.62 (+0.03)	0.57 (+0.05)	0.61 (−0.01)	0.56 (+0.01)	0.61 (+0.03)

^aScores on the left are computed for set 1 (using 1100, 1100, and 920 observations over rural, suburban, and urban sites, respectively), and on the right, for set 2 (using 100, 310, and 970 observations over rural, suburban, and urban sites, respectively).

raw and MOS forecasts is almost independent of lead time. Absolute biases are often higher for the MOS than for the raw forecasts.

[67] RMSE are, as expected, lower for the MOS than for the raw forecasts. On sites of set 1, the difference reaches $1.7 \mu\text{g}/\text{m}^3$ at rural sites ($D + 0$) and $1.2 \mu\text{g}/\text{m}^3$ at urban sites ($D + 0$ and $D + 1$). As for sites of set 2, the impact of the MOS procedure on RMSE is of the same order of magnitude whatever the lead time at suburban and urban sites (between 1.2 and $1.7 \mu\text{g}/\text{m}^3$ at urban sites); at rural sites, kriging slightly worsens the skill of the forecast.

[68] For sites of set 1, correlation is always greater for the MOS forecasts. The impact of the MOS procedure on correlation at sites of set 2 is of the same order of magnitude at suburban and urban sites; as for RMSE, skill improvement is lower at rural sites of set 2, but the difference is almost not significant.

[69] For both sets of stations, the MOS procedure generally improves on average the representation of the ozone daily maxima in terms of RMSE and correlation. Table 6 displays the error statistics computed for the $D + 0$ ozone forecasts, for a cutoff value of $120 \mu\text{g}/\text{m}^3$ on observed ozone daily maxima. From Table 6, it is clear that the impact of the MOS procedure is higher when ozone episodes take place: all skill scores are improved whatever the type and set of stations considered.

[70] Contingency tables were computed for the MOS forecasts and compared to those obtained for the raw forecasts; they are displayed in Table 7. As seen before, skill scores associated with threshold criteria are very tough for the model: the percentage of good raw forecasts above threshold (GF+) reaches 47% (urban sites, set 2, $D - 1$) and can be 0 (rural sites, set 2, $D + 2$); at $D + 0$, GF+ ranges from 35% (urban sites, set 1) to 23% (rural sites, set 1). In all cases, FA is lower than 2%. Since the MOS procedure results in a shifting of biases toward higher values at all station types, the percentage of good forecasts above threshold is higher for the MOS than for the raw forecasts: GF+ reaches 64% at urban sites (set 2, $D - 1$); at $D + 0$, GF+ ranges from 51% (suburban sites, set 1) to 25% (rural

sites, set 2) and the rate of false alarms is always lower than 3%. Thus the MOS procedure appears to be an efficient tool to improve the skill of the system with respect to threshold exceedances.

3.3. Skill of PM₁₀ Daily Averages

[71] Table 8 displays error statistics obtained for the PM₁₀ daily mean forecasted by PREV'AIR during 2004, 2005 and 2006, for the two main seasons. In most countries, PM measurements based on the TEOM device are corrected by a coefficient. This correction is applied because of the evaporation of the volatile fraction of particles (ammonium nitrate) in the device principle. However, in France, such a correction is not applied, leading to an underestimation of actual PM levels in the measurements [Bessagnet *et al.*, 2005]. A limitation to the evaluation that is made hereafter is the fact that we have no information about the chemical composition of PM₁₀.

[72] Whatever the season, on average at the European scale, PREV'AIR forecasts underestimate PM₁₀ mean concentrations on all types of stations (up to 1.8 , 2.9 and $3.5 \mu\text{g}/\text{m}^3$ in summer and up to 2.7 , 5.3 and $5.7 \mu\text{g}/\text{m}^3$ in winter at rural, suburban and urban sites respectively). As expected, biases are lower at rural sites than at suburban and urban sites, due to too coarse model resolution for cities. The bias in summertime might be attributed to the simplified scheme that describes the buildup of secondary organic aerosols: it accounts in a limited manner for the formation of SOA, especially those of biogenic origin which might be a large fraction of the aerosol in the summer months. Also, sea salts are not included in the current version of the model, which can explain the systematic negative bias observed at coastal stations in winter. Moreover, the aqueous and heterogeneous chemistry is hard to accurately simulate in wintertime, as it depends on meteorological parameters difficult to predict, such as cloud water content.

[73] Negative biases are quite general. Large biases occur at some sites, near the coasts or in urban areas, as can be seen in Figure 6b, which shows the spatial distribution of biases at all sites for $D + 1$ forecast. Around the Paris area,

Table 7. Same as in Table 4, but for the Ozone Daily Maxima Over the Summer of 2006 for the French Raw Forecasts ($D + 0$)^a

	Set 1			Set 2		
	RUR	PUB	URB	RUR	PUB	URB
GF+	23.0 (+13.5)	24.7 (+26.6)	35.1 (+5.1)	25.0 (−)	31.4 (+5.7)	32.0 (+11.7)
FA	1.4 (−)	1.2 (+1.2)	0.8 (+0.3)	1.6 (−)	1.4 (+0.2)	1.3 (+0.6)
ME	77.0 (−13.5)	75.3 (−26.6)	64.9 (−5.1)	75.0 (−)	68.6 (−5.7)	68.0 (−11.7)

^aThe differences in scores between the MOS and raw forecasts are indicated in parentheses; bold figures indicate better scores for the MOS forecasts. Scores are evaluated with respect to the $180\text{-}\mu\text{g}/\text{m}^3$ threshold. The number of observations above threshold at rural, suburban, and urban sites is 120, 150, and 95, respectively, for set 1 and 4, 35, and 100, respectively, for set 2.

Table 8. Same as Table 2, but for the PM₁₀ Daily Mean for the European Forecast Over the Spring/Summer of 2004–2006 and the Fall/Winter of 2004–2006^a

		Lead Time	RUR	SUB	URB
<i>Summer</i>					
Mean obs.	...		17.7	19.2	19.0
Bias	D – 1		–1.5	–2.8	–3.5
	D + 0		–1.3	–2.6	–3.3
	D + 1		–1.6	–2.8	–3.5
	D + 2		–1.8	–2.9	–3.5
RMSE	D – 1		8.4	8.3	8.3
	D + 0		8.4	8.4	8.4
	D + 1		8.5	8.7	8.6
	D + 2		8.8	8.9	8.7
Corr.	D – 1		0.53	0.55	0.61
	D + 0		0.53	0.53	0.59
	D + 1		0.51	0.50	0.58
	D + 2		0.46	0.47	0.56
<i>Winter</i>					
Mean obs.	...		18.8	21.7	20.3
Bias	D – 1		–1.5	–4.2	–4.9
	D + 0		–1.5	–4.2	–4.7
	D + 1		–2.3	–4.9	–5.4
	D + 2		–2.7	–5.3	–5.7
RMSE	D – 1		12.5	13.5	13.5
	D + 0		12.9	13.9	14.1
	D + 1		13.1	14.0	14.1
	D + 2		13.2	14.3	14.2
Corr.	D – 1		0.53	0.47	0.41
	D + 0		0.53	0.45	0.40
	D + 1		0.51	0.45	0.35
	D + 2		0.50	0.43	0.33

^aScores are computed separately for each station type: rural, suburban, and urban (using 14,000, 14,500, and 18,000 observations, respectively, in spring/summer and 7900, 8500, and 10,200 observations, respectively, in fall/winter).

positive biases are found, possibly due to an overestimation of emissions. Winter biases are coherent with summer biases.

[74] On average, RMSE lies between 8 and 9 $\mu\text{g}/\text{m}^3$ in summertime and between 12.5 and 14 $\mu\text{g}/\text{m}^3$ in wintertime. In summer, RMSE is approximately of the same order, on average, at all station types, whereas in winter, it is larger at suburban and urban sites than at rural sites.

[75] As the amplitude of PM₁₀ concentration and variability are larger in winter than in summer, it is interesting to compare normalized skill. Table 9 displays normalized biases and normalized root mean square errors (NMSE). These indicators are computed by adding up relative differences between observed and forecast concentrations (i.e., the ratio of the absolute differences to the observed concentrations). Normalized biases and NMSE are slightly higher in winter than in summer at suburban and urban sites. On the contrary, at rural sites, normalized biases and NMSE are more than twice as high in winter than in summer. This indicates a clear loss of skill from summer to winter.

[76] Correlation is higher in summertime than in wintertime. This better general performance in summer suggests that dynamic processes are better simulated in summer than in winter, or that concentrations are less sensitive to errors in the dynamics, which can simply be due to the fact that boundary layers are higher. In winter, low planetary boundary layer heights and weak turbulence make concentrations

very sensitive to many parameters and hence less predictable. Moreover, as discussed previously, observations can be impaired due to nitrate evaporation effects in winter. Then, correlation in summertime is highest at urban sites, then at suburban sites and finally at rural sites (it ranges from 0.56 to 0.61 for urban sites and from 0.46 to 0.55 for rural and suburban sites), whereas, in wintertime, the order is reversed: highest correlation is found at rural sites, then at suburban and finally at urban sites (it ranges from 0.50 to 0.53 at rural sites; from 0.43 to 0.47 at suburban sites and from 0.33 to 0.41 at urban sites). These orders of correlation are suggestive of missing or badly represented aerosol processes in summer in rural areas. These could be explained by forest fires (missing in our emission inventory), or lack of knowledge of secondary organic aerosol formation.

[77] Once again, the system is not able to catch the increase in concentrations from rural to suburban sites and from suburban to urban sites. One observes a slight decrease in the performance from D + 0 to D + 2, for all skill scores except for the bias of the system. The better skill scores are associated with the simulated (D – 1) PM₁₀ mean concentrations (not shown). This is again due to the increase of error in weather forecasting.

3.3.1. Evaluation Against Performance Criteria

[78] *Boylan and Russell* [2006] recommend the following indices to be used for model performance evaluation: the Mean Fractional Bias (MFB, %) and the Mean Fractional Error (MFE, %). These indices are computed based on relative differences, i.e., the ratio of the absolute differences to the average of the model and observation. They are evaluated on daily mean concentrations. For three-dimensional particulate matter modeling, Boylan and Russell propose performance goals: MFB and MFE less than or equal to ± 30 and 50%, and criteria: MFB and MFE less than or equal to ± 60 and 75%. The two indicators were computed for the operational system for the springs/summers and falls/winters of 2004 to 2006.

Table 9. Normalized Bias and NMSE for the PM₁₀ Daily Mean for the European Forecast Over the Spring/Summer of 2004–2006 and the Fall/Winter of 2004–2006^a

		Lead Time	RUR	SUB	URB
<i>Summer</i>					
Nbias	D – 1		3.5	–8.7	–16.3
	D + 0		4.5	–7.5	–15.2
	D + 1		3.6	–8.3	–15.7
	D + 2		3.5	–7.9	–15.4
NMSE	D – 1		50.0	44.6	41.6
	D + 0		50.2	45.6	41.8
	D + 1		51.1	47.2	42.7
	D + 2		55.8	48.2	43.5
<i>Winter</i>					
Nbias	D – 1		20.3	–10.8	–18.7
	D + 0		18.4	–11.4	–18.4
	D + 1		13.6	–14.6	–20.8
	D + 2		11.7	–15.7	–22.0
NMSE	D – 1		111.1	51.8	55.8
	D + 0		105.6	54.6	58.9
	D + 1		106.6	55.9	61.6
	D + 2		106.2	56.5	60.2

^aScores are computed separately for each station type based on the same observations as in Table 8.

Table 10. Same as Table 2, but for the NO₂ Daily Mean for the France Forecast Over the Spring/Summer of 2004–2006 and the Fall/Winter of 2004–2006^a

	Lead Time	RUR	SUB	URB
<i>Summer</i>				
Mean obs.	...	7.2	17.3	20.2
Bias	D – 1	0.5	–5.6	–3.1
	D + 0	0.4	–5.8	–3.5
	D + 1	0.3	–5.8	–3.7
	D + 2	0.3	–5.8	–3.7
RMSE	D – 1	5.3	12.0	17.3
	D + 0	5.3	12.1	16.8
	D + 1	5.4	12.1	16.5
	D + 2	5.5	12.2	16.7
Corr.	D – 1	0.57	0.43	0.62
	D + 0	0.55	0.42	0.61
	D + 1	0.54	0.41	0.61
	D + 2	0.52	0.40	0.60
<i>Winter</i>				
Mean obs.	...	12.5	27.7	31.8
Bias	D – 1	–1.2	–12.3	–11.4
	D + 0	–1.4	–12.6	–11.8
	D + 1	–1.6	–12.8	–12.0
	D + 2	–1.7	–12.9	–12.0
RMSE	D – 1	7.9	18.6	21.8
	D + 0	8.0	18.8	21.4
	D + 1	8.1	19.1	21.7
	D + 2	8.3	19.2	21.9
Corr.	D – 1	0.70	0.49	0.58
	D + 0	0.69	0.48	0.58
	D + 1	0.68	0.46	0.56
	D + 2	0.67	0.45	0.56

^aScores are computed separately for each station type: rural, suburban, and urban (using 8200, 16,600, and 44,400 observations, respectively, in spring/summer and 6000, 12,000, and 33,000 observations, respectively, in fall/winter).

[79] On average at the European scale, both indicators lie well below the performance criteria: in summer, MFB ranges from –16 to –15% at urban sites, –9 to –8% at suburban sites and 3.5 to 4.5% at rural sites; MFE lies in the range of 37 to 39% at rural sites, 36 to 38% at suburban sites and 40 to 41% at urban sites; in winter, MFB ranges from –22 to –19% at urban sites, –16 to –11% at suburban sites and 12 to 20% at rural sites; MFE lies in the range of 50 to 54% at rural sites, 47 to 53% at suburban sites and 55 to 61% at urban sites. In summertime, the indicators are in compliance even with the performance goals. There is a depredation of these indicators with the lead time. As expected, the compliance with the criteria is far better in summertime than in wintertime.

3.4. Skill of NO₂ Forecasts

[80] We only evaluated NO₂ concentration forecasts for the zoom model version because of the higher resolution, making it easier to compare NO₂ in large emissions areas like cities. Table 10 displays skill scores computed for the NO₂ daily mean forecasted by PREV'AIR during 2004, 2005 and 2006. As for PM₁₀, the evaluation is carried out separately for spring/summer and fall/winter seasons.

[81] Figure 7b shows the spatial distribution of biases at all sites for D + 1 forecast. Whatever the season, the system underestimates on average the NO₂ mean concentrations at suburban and urban sites. This is probably due to a lack of resolution: up to 6 and 4 $\mu\text{g}/\text{m}^3$ in summer and to 13 and

12 $\mu\text{g}/\text{m}^3$ in winter at suburban and urban sites respectively. However, with its 10 km resolution, the system exhibits a relative decrease in concentrations from urban to suburban sites and from suburban to rural sites. As for PM₁₀, NO₂ forecasts are overestimated around Paris. At rural sites, the bias is much smaller.

[82] On average, RMSE lies between 5 $\mu\text{g}/\text{m}^3$ (rural stations), 12 $\mu\text{g}/\text{m}^3$ (suburban stations) and 16.5–17 $\mu\text{g}/\text{m}^3$ (urban stations) in summertime compared to 8 $\mu\text{g}/\text{m}^3$ (rural stations), 19 $\mu\text{g}/\text{m}^3$ (suburban stations) and 21–22 $\mu\text{g}/\text{m}^3$ (urban stations) in wintertime; RMSE is lower at rural and suburban sites than at urban sites.

[83] Biases and RMSE are larger in wintertime than in summertime, as are observed concentrations. In order to compare the system's performance throughout the seasons and station types, we have derived normalized biases and NMSE (not shown). From this analysis, we observe that the system performance with respect to RMSE is worse at rural sites than at suburban and urban sites. Contrarily to PM₁₀, there is no clear loss of skill in wintertime for the NO₂ forecasts.

[84] At rural and suburban sites, higher correlation is found in winter than in summer. As for PM₁₀, in summer, higher correlation is found at urban sites (0.60), whereas in wintertime it is found at rural sites (0.70).

[85] Finally, as for ozone and PM₁₀, the skill moderately decreases as lead time increases.

4. Evaluation of Ozone Analyses

[86] We also evaluated the near-real time ozone daily maxima analyses produced over the fine scale domain (France). Table 11 displays the error statistics obtained for the ozone analyses compared to those obtained for the raw simulations (D – 1, zoomed over France). As for the MOS forecast evaluation, two sets of stations are considered: the first one is made up of stations taken into account in the analysis procedure (43 rural sites, 23 suburban sites, 14 urban sites); the second set is made up of independent stations, not used in the analysis procedure (no rural sites, 26 suburban sites, 57 urban sites) and is used to evaluate the resulting analyses applied to independent sites.

[87] For both sets of stations, the analysis procedure improves the representation of the D – 1 ozone daily maxima on average for all indicators and all measurement sites. For sets 1 and 2, averaged bias on the ozone daily maxima is lower than 2.1 $\mu\text{g}/\text{m}^3$ (3.3%); RMSE ranges between 3.8 and 5.5 $\mu\text{g}/\text{m}^3$ for set 1 (respectively 4.4 and 5.7%) and between 12.2 and 13.1 $\mu\text{g}/\text{m}^3$ for set 2 (respectively 12.2 and 14.4%); correlation reaches almost 1.0 on average for stations of set 1; increases from 0.85 to 0.93 at suburban stations of set 2 and from 0.82 to 0.91 at urban stations.

[88] These results, and particularly the ones obtained for set 2, show how the kriging method actually succeeds in propagating efficiently the discrete corrections computed at some specific, well-chosen sites. They also demonstrate the reliability of the analyses for areas not covered by observations.

5. Conclusions and Perspectives

[89] Since spring 2003, the PREV'AIR air quality forecasting and monitoring system (<http://www.prevair.org/>) has

Table 11. Skill Scores for the Ozone Daily Maxima Over the Spring/Summer of 2004–2006 for the French Raw Simulation^a

	Set 1			Set 2		
	RUR	SUB	URB	RUR	SUB	URB
Mean obs.	105.0	103.2	98.9	...	104.9	100.5
Bias	1.5 (−1.4)	3.1 (−2.7)	6.1 (−5.4)	...	4.1 (−2.0)	2.0 (−0.8)
RMSE	17.0 (−11.5)	16.7 (−12.4)	16.7 (−12.4)	...	17.6 (−5.4)	18.3 (−5.2)
Corr.	0.83 (+0.15)	0.84 (+0.15)	0.83 (+0.16)	...	0.85 (+0.08)	0.82 (+0.09)

^aThe differences in scores between the analyses and raw simulations are indicated in parentheses; figures in bold indicate better skill scores for the analyses. Scores are computed for set 1 (using 11,820, 6560, and 3650 observations over rural, suburban, and urban sites, respectively) and for set 2 (using 7200 and 15,380 observations over suburban and urban sites, respectively).

been working on an operational, routine basis. It delivers forecasts of ozone, nitrogen dioxide and particulate matter concentrations, up to three days ahead, over the Globe, Europe and France, based on two three-dimensional deterministic chemistry transport models, CHIMERE (regional and national scale) and MOCAGE (global scale) forced by appropriate emissions and weather forecasts. We have presented a quantitative evaluation of the forecasts obtained with one of these models (CHIMERE), over 3 consecutive a and for three pollutants (O_3 , NO_2 and PM_{10}): the comparison between the pollutant concentrations and a set of observations available over Europe shows the ability of PREV'AIR to predict the evolution of photochemical and particulate pollution over the domains considered. The performance of the system is quite satisfactory for ozone; in particular, the EPA acceptance criteria relative to forecast accuracy for high concentrations and daily maxima are met by the system at more than 90% of the measurement sites. The system however underestimates high daily maxima, which is an important item of information for decision makers. It also fails to predict reliable ozone concentrations in complex terrain areas (mountain tops or valleys) due to the lack of resolution. To a lesser extent, city centre concentrations are also less well captured. The system skill decreases with the lead time of the forecast, due to increasing meteorological error.

[90] PM_{10} forecasts are less accurate and suffer from a general underestimation, except in some areas such as around Paris. Summertime daily means are better forecast than wintertime means. NO_2 forecasts are also poorer than ozone forecasts, especially in urban areas where underestimation is found.

[91] In addition to the purely deterministic simulation/forecast methods, alternative techniques have been implemented in the PREV'AIR system, in order to improve the forecasts: Model Output Statistical (MOS) forecasts, based on the concept of training, are issued. We evaluated these MOS forecasts and showed that their skill is slightly greater than the deterministic forecasts, with relatively large improvements in specific areas and for high ozone levels.

[92] Analyses are also provided by the system. A kriging method, in which model outputs are combined with available near-real time observations, is used to build analyses of ozone over France. The skill of the analyses has been evaluated and we demonstrated that the method propagates the correction of the simulation error in areas where no observations are carried out, thus providing the most comprehensive picture of ozone maxima over the French territory available.

[93] By its design and objectives, the PREV'AIR system offers a framework and some elements to answer questions related to air pollution and its impact on public health. Short-term effects of ozone on health and in terms of mortality in the event of a pollution episode are now acknowledged [WHO, 2000; Stedman, 2004]; recent studies [WHO, 2003] have shown that chronic exposure to moderately high ozone levels also affects human health. PREV'AIR can contribute to estimating the public's exposure to ozone.

[94] Recently, it has been shown [Derwent *et al.*, 2003] that the average values of ozone concentrations (or basic concentrations) have been increasing regularly over the last few decades, whereas the extreme values tend to decrease as abatement measures of primary pollutant emissions become effective. Nevertheless, the photochemical episode of August 2003, exceptional by its duration (from 4 to 12 August 2003) and geographical extent (Western Europe), confirms that acute pollution episodes remain a recurrent problem. In this context, the second interest of PREV'AIR lies in its potential of real-time information, with the possibility of anticipating photochemical or particulate pollution events on a large scale over Europe and of warning the populations at risk. It can then be attempted to reduce the exposure of these populations (by acting either on the causes of the pollution episodes or on the activity of vulnerable subjects).

[95] Among the possible developments of PREV'AIR, two particularly promising avenues may be mentioned. First, the system performance could be improved by taking observations more fully into account: analyzed concentrations should be integrated in the forecasting chain at the initialization stage. In the longer term, the implementation in PREV'AIR of data assimilation procedures should make it possible to integrate three-dimensional observations of pollutant concentrations (resulting from LIDAR or satellite measurements). Second, as mentioned before, ensemble forecasting is a rather promising approach for substantially increasing the performance of air quality forecasting systems. If a pollution episode occurs, air quality forecasts are of primary importance to decision makers. However, the question arises of the reliability of the forecasts. Instead of one "binary" answer, a probabilistic response (e.g., probability of going beyond an air quality threshold in a given area) would be more appropriate. The realization of "ensembles" of forecasts, characterized by different input data, and/or carried out with several chemistry transport models, will be the next step in the development of operational air quality forecasting systems.

References

- Amann, M., I. Bertok, J. Cofala, F. Gyarfas, C. Heyes, Z. Klimont, W. Schöpp, and W. Winiwater (2005), Clean Air for Europe (CAFE) program final report, Int. Inst. for Appl. Syst. Anal., Laxenburg, Austria.
- Aumont, B., A. Jaeger-Voirol, B. Martin, and G. Toupance (1996), Tests of some reduction hypothesis in photochemical mechanisms: Application to air quality modeling in the Paris area, *Atmos. Environ.*, **30**, 2061–2077.
- Beekmann, M., and C. Derognat (2003), Monte Carlo uncertainty analysis of a regional-scale transport chemistry model constrained by measurements from the Atmospheric Pollution Over the Paris Area (ESQUIF) campaign, *J. Geophys. Res.*, **108**(D17), 8559, doi:10.1029/2003JD003391.
- Bessagnet, B., A. Hodzic, R. Vautard, M. Beekmann, S. Cheinet, C. Honoré, C. Liousse, and L. Rouil (2004), Aerosol modeling with CHIMERE—Preliminary evaluation at the continental scale, *Atmos. Environ.*, **38**, 2803–2817.
- Bessagnet, B., A. Hodzic, O. Blanchard, M. Lattuati, O. Le Bihan, H. Marfaing, and L. Rouil (2005), Origin of particulate matter pollution episodes in wintertime over the Paris Basin, *Atmos. Environ.*, **39**, 6159–6174.
- Blond, N., and R. Vautard (2004), Three-dimensional ozone analyses and their use for short-term ozone forecasts, *J. Geophys. Res.*, **109**, D17303, doi:10.1029/2004JD004515.
- Blond, N., L. Bel, and R. Vautard (2003), Three-dimensional ozone data analysis with an air quality model over the Paris area, *J. Geophys. Res.*, **108**(D23), 4744, doi:10.1029/2003JD003679.
- Boylan, J. W., and A. G. Russell (2006), PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models, *Atmos. Environ.*, **40**, 4946–4959.
- Cathala, M.-L., J. Pailleux, and V.-H. Peuch (2003), Improving global simulations of the upper troposphere—lower stratosphere with sequential assimilation of MOZAIC data, *Tellus, Ser. B*, **55**, 1–10.
- Chiriac, M., R. Vautard, H. Chepfer, M. Haeffelin, Y. Wanherdrick, Y. Morille, A. Protat, and J. Dudhia (2005), The ability of MM5 to simulate ice clouds: Systematic comparison between simulated and measured fluxes and lidar/radar profiles at SIRTa atmospheric observatory, *Mon. Weather Rev.*, **134**, 897–918.
- Cope, M. E., et al. (2004), The Australian Air Quality Forecasting System. part I: Project description and early outcomes, *J. Appl. Meteorol.*, **43**, 649–662.
- Courtier, P., C. Freydyer, J. F. Geleyn, F. Rabier, and M. Rochas (1991), The ARPEGE project at METEO-France, paper presented at the ECMWF Seminar, Reading, U. K., 9–13 Sept.
- Cros, B., et al. (2004), The ESCOMPTE program: An overview, *Atmos. Environ.*, **69**, 241–279.
- Cuvelier, C., 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., and R. B. Stull (2003), An ensemble air quality forecast over western Europe during an ozone episode, *Atmos. Environ.*, **37**, 3469–3474.
- Delle Monache, L., T. Nipen, X. Deng, Y. Zhou, and R. Stull (2006), Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction, *J. Geophys. Res.*, **111**, D05308, doi:10.1029/2005JD006311.
- Demerjian, K. L. (2000), A review of national monitoring networks in North America, *Atmos. Environ.*, **34**, 1861–1884.
- Derognat, C. (2002), Pollution photooxydante à l'échelle urbaine et interaction avec l'échelle régionale, Ph.D. thesis, Univ. Paris VI, Paris.
- Derwent, R. G., et al. (2003), Photochemical ozone formation in north west Europe and its control, *Atmos. Environ.*, **37**, 1983–1991.
- Drobinski, P., et al. (2007), Regional transport and dilution during high-pollution episodes in southern France: Summary of findings from the Field Experiment to Constraint Models of Atmospheric Pollution and Emissions Transport (ESCOMPTE), *J. Geophys. Res.*, **112**, D13105, doi:10.1029/2006JD007494.
- Dudhia, J. (1993), A nonhydrostatic version of the Penn State/NCAR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front, *Mon. Weather Rev.*, **121**, 1493–1513.
- Dufour, A., M. Amodei, G. Ancellet, and V.-H. Peuch (2004), Observed and modelled “chemical weather” during ESCOMPTE, *Atmos. Res.*, **74**, 161–189, doi:10.1016/j.atmosres.2004.04.013.
- EEA (1999), Criteria for EUROAIRNET, *Tech. Rep. 12*, EEA Air Qual. Monit. and Inf. Network, Eur. Environ. Agency, Copenhagen.
- Elbern, H., and H. Schmidt (2001), Ozone episode analysis by four-dimensional variational chemistry data assimilation, *J. Geophys. Res.*, **106**, 3569–3590.
- Folberth, G., D. A. Hauglustaine, J. Lathière, and F. Brocheton (2005), Impact of biogenic hydrocarbons on tropospheric chemistry: Results from a global chemistry-climate model, *Atmos. Chem. Phys. Disc.*, **5**, 1680–1735.
- Ginoux, P., M. Chin, I. Tegen, J. M. Prospero, B. Holben, O. Dubovik, and S.-J. Lin (2001), Sources and distributions of dust aerosols simulated with the GOCART model, *J. Geophys. Res.*, **106**, 20,255–20,273.
- Glahn, H. R., and D. A. Lowry (1972), The use of model output statistics (MOS) in objective weather forecasting, *J. Appl. Meteorol.*, **11**, 1203–1211.
- Hauglustaine, D. A., F. Hourdin, L. Jourdain, M. A. Filiberti, S. Walters, J. F. Lamarque, and E. A. Holland (2004), Interactive chemistry in the Laboratoire de Météorologie Dynamique general circulation model: Description and background tropospheric chemistry evaluation, *J. Geophys. Res.*, **109**, D04314, doi:10.1029/2003JD003957.
- Hodzic, A., R. Vautard, B. Bessagnet, M. Lattuati, and F. Moreto (2005), On the quality of long-term urban particulate matter simulation with the CHIMERE model, *Atmos. Environ.*, **39**, 5851–5864.
- Honoré, C. (2000), La photochimie de l'ozone à l'échelle urbaine: Un système dynamique non linéaire, Ph.D. thesis, Univ. Paris VI, Paris.
- Jonson, J. E., D. Simpson, H. Fagerli, and S. Solberg (2005), Can we explain the trends in European ozone levels?, *Atmos. Chem. Phys.*, **6**, 51–66.
- Klein, W. H., and H. R. Glahn (1974), Forecasting local weather by means of model output statistics, *Bull. Am. Meteorol. Soc.*, **55**, 1217–1227.
- Lattuati, M. (1997), Contribution à l'étude du bilan de l'ozone troposphérique à l'interface de l'Europe et de l'Atlantique nord: Modélisation lagrangienne et mesures en altitude, Ph.D. thesis, Univ. Paris VI, Paris.
- McHenry, J. N., W. F. Ryan, N. L. Seaman, C. J. Coats, J. Pudykiewicz, S. Arunachalam, and J. M. Vukovich (2004), A real-time Eulerian photochemical model forecast system: Overview and initial ozone forecast performance in the northeast U.S. corridor, *Bull. Am. Meteorol. Soc.*, **85**, 525–548.
- McKeen, S., et al. (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.
- McNair, L. A., R. A. Harley, and A. G. Russell (1996), Spatial inhomogeneity in pollutant concentrations, and their implications for air quality model evaluation, *Atmos. Environ.*, **30**, 4291–4301.
- Michou, M., and V.-H. Peuch (2002), Surface exchanges in the multiscale chemistry and transport model MOCAGE, *Water Res.*, **15**, 173–203.
- Pagowski, M., et al. (2006), Application of dynamic linear regression to improve the skill of ensemble-based deterministic ozone forecasts, *Atmos. Environ.*, **40**, 3240–3250.
- Passant, N. R. (2002), Speciation of UK emissions of non-methane volatile organic compounds, *Rep. AEAT/ENV/0545*, AEA Technol., London.
- Peuch, V.-H., M. Amodei, T. Barthet, M.-L. Cathala, B. Josse, M. Michou, and P. Simon (1999), MOCAGE, Modèle de Chimie Atmosphérique à Grande Echelle, paper presented at the Workshop on Atmospheric Modelling, Météo France, Toulouse, France, Dec.
- Pope, C. A., R. T. Burnett, M. J. Thun, E. E. Calle, D. Krewski, K. Ito, and G. D. Thurston (2002), Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution, *JAMA*, **287**, 1132–1141.
- Pradier, S., J. L. Attié, M. Chong, J. Escobar, V.-H. Peuch, J.-F. Lamarque, B. Khattatov, and D. Edwards (2006), Evaluation of 2001 springtime CO transport over West Africa using MOPITT CO measurements assimilated in a global chemistry transport model, *Tellus, Ser. B*, **58**, 163–176.
- Putaud, J.-P., et al. (2004), A European aerosol phenomenology—2: Chemical characteristics of particulate matter at kerbside, urban, rural and background sites in Europe, *Atmos. Environ.*, **38**, 2579–2595.
- Riccio, A., G. Barone, E. Chianese, and G. Giunta (2006), A hierarchical Bayesian approach to the spatio-temporal modeling of air quality data, *Atmos. Environ.*, **40**, 554–566.
- Robeson, S. M., and D. G. Steyn (1990), Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations, *Atmos. Environ., Part B*, **24**, 303–312.
- Schmidt, H., C. Derognat, R. Vautard, and M. Beekmann (2001), A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in western Europe, *Atmos. Environ.*, **36**, 6277–6297.
- Simpson, R. W., and A. P. Layton (1983), Forecasting peak ozone levels, *Atmos. Environ.*, **17**, 1649–1654.
- Stedman, J. R. (2004), The predicted number of air pollution related deaths in the UK during the August 2003 heatwave, *Atmos. Environ.*, **38**, 1087–1090.
- Szopa, S., D. A. Hauglustaine, R. Vautard, and L. Menut (2006), Future global tropospheric ozone changes and impact on European air quality, *Geophys. Res. Lett.*, **33**, L14805, doi:10.1029/2006GL025860.
- Teyssède, H., et al. (2007), A new tropospheric and stratospheric Chemistry and Transport Model MOCAGE-Climat for multi-year studies: Evaluation of the present-day climatology and sensitivity to surface processes, *Atmos. Chem. Phys.*, **7**, 5815–5860.
- Thunis, P., et al. (2007), Analysis of model responses to emission-reduction scenarios within the CityDelta project, *Atmos. Environ.*, **41**, 208–220.

- Tilmes, S., et al. (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.
- U.S. Environmental Protection Agency (1991), Guidance for regulatory application of the urban airshed model (UAM), report, Off. of Air Qual. Plan. and Standards, Research Triangle Park, N. C.
- Van Loon, M., et al. (2007), Evaluation of long-term ozone simulations from seven regional air quality models and their ensemble, *Atmos. Environ.*, **41**, 2083–2097.
- Vaughan, J., et al. (2004), A numerical daily air quality forecast system for the Pacific Northwest, *Bull. Am. Meteorol. Soc.*, **85**, 549–561.
- Vautard, R., M. Beekmann, J. Roux, and D. Gombert (2001), Validation of a hybrid forecasting system for the ozone concentrations over the Paris area, *Atmos. Environ.*, **35**, 2449–2461.
- Vautard, R., C. Honoré, M. Beekmann, and L. Rouïl (2005a), Simulation of ozone during the August 2003 heat wave and emission control scenarios, *Atmos. Environ.*, **39**, 2957–2967.
- Vautard, R., B. Bessagnet, M. Chin, and L. Menut (2005b), On the contribution of natural aeolian sources to particulate matter concentrations in Europe: Testing hypotheses with a modelling approach, *Atmos. Environ.*, **39**, 3291–3303.
- Vautard, R., S. Szopa, M. Beekmann, L. Menut, D. A. Hauglustaine, L. Rouïl, and M. Roemer (2006), Are decadal anthropogenic emission changes in Europe consistent with surface ozone observations?, *Geophys. Res. Lett.*, **33**, L13810, doi:10.1029/2006GL026080.
- Vautard, R., et al. (2007), Evaluation and intercomparison of ozone and PM10 simulations by several chemistry transport models over 4 European cities within the CityDelta project, *Atmos. Environ.*, **41**, 173–188.
- Vestreng, V. (2003), Review and revision of emission data reported to CLRTAP, status report, EMEP, Meteorol. Synthesizing Cent.-West, Norwegian Meteorol. Inst., Oslo.
- World Health Organization (2000), *Air Quality Guidelines for Europe*, 2nd ed., *WHO Reg. Publ. Eur. Ser.*, vol. 91, Geneva, Switzerland.
- World Health Organization (2003), Health aspects of air pollution with particulate matter, ozone and nitrogen dioxide: Report on a WHO Working Group, Bonn, Germany.
- World Health Organization (2006), Health risks of particulate matter from long-range transboundary pollution, Regional Off. for Eur., World Health Org., Copenhagen.
- Zeldin, M. D., and M. D. Thomas (1975), Ozone trends in the eastern Los Angeles Basin corrected for meteorological variations, paper presented at the International Conference on Environmental Sensing and Assessment, Las Vegas, Nev., 14–19 Sept.
- M. Beekmann and J.-M. Flaud, Laboratoire Inter-Universitaire des Systèmes Atmosphériques, Faculté des Sciences et Technologie, 61 avenue du Général de Gaulle, F-94010 Créteil Cedex, France.
- B. Bessagnet, C. Honoré, L. Malherbe, F. Meleux, and L. Rouïl, Institut National de l'Environnement Industriel et des Risques, Chronic Risks Division, Parc Technologique Alata, BP 2, F-60550 Verneuil-en-Halatte, France. (frederik.meleux@ineris.fr)
- A. Dufour, D. Martin, A. Peuch, and V.-H. Peuch, Météo France, 1, quai Branly, F-75007 Paris, France.
- C. Eliegaray and N. Poisson, Agence de l'Environnement et de la Maîtrise de l'Energie, 27, rue Louis Vicat, F-75737 Paris Cedex 15, France.
- L. Menut, Laboratoire de Météorologie Dynamique, Institut Pierre-Simon Laplace, Ecole Polytechnique, F-91128 Palaiseau Cedex 1, France.
- R. Vautard, Laboratoire des Sciences du Climat et de l'Environnement, Institut Pierre-Simon Laplace, Orme des merisiers, Centre de Saclay, F-91190 Gif sur Yvette Cedex, France.