



HAL
open science

Three-dimensional ozone data analysis with an air quality model over the Paris area

N. Blond, Liliane Bel, Robert Vautard

► **To cite this version:**

N. Blond, Liliane Bel, Robert Vautard. Three-dimensional ozone data analysis with an air quality model over the Paris area. *Journal of Geophysical Research*, 2003, 108 (D23), 10.1029/2003JD003679 . hal-02116922

HAL Id: hal-02116922

<https://hal.science/hal-02116922>

Submitted on 29 Jan 2021

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.

Three-dimensional ozone data analysis with an air quality model over the Paris area

N. Blond¹

Laboratoire de Météorologie Dynamique, CNRS, École Polytechnique, Palaiseau, France

L. Bel

Probabilités, Statistique et Modélisation, Université Paris-Sud, Orsay, France

R. Vautard

Laboratoire de Météorologie Dynamique, CNRS, École Polytechnique, Palaiseau, France

Received 10 April 2003; revised 9 August 2003; accepted 18 August 2003; published 13 December 2003.

[1] We present a description, an evaluation, and a comparison of four methods designed to produce objective and physically consistent maps of ozone concentration fields. These methods are based on the use of a chemistry-transport model (CTM) and available observations. In most existing analysis systems, the error covariance is modeled using assumptions of homogeneity and isotropy. However, these assumptions may fail in the case of strongly heterogeneous terrain or emission patterns. Therefore we propose a simple method for specifying an anisotropic and heterogeneous background error covariance model in a statistical interpolation. We illustrate the positive impact of the implementation of this model. Since the covariance model is independent of the state of the atmosphere and invariant in time, we simultaneously test kriging techniques that are generally used for spatial interpolations. Kriging is applied to the observed values or to the differences between simulated and observed concentrations, namely the innovations. We perform an objective statistical validation of all methods for data recorded over an entire summer season, in contrast to other evaluations of data assimilation methods made in air quality modeling. We demonstrate that the RMS error of the ozone analyses at the surface is 30–50% smaller than the one from simulations, regardless of the method used. Most of the time, the kriging method applied to the innovations gives results equivalent to that of the anisotropic statistical interpolation, in spite of very different formulations. In addition, we show that this method outperforms the classical kriging method applied to only observations. The information provided by the CTM is therefore essential to a good-quality representation of ozone patterns such as city plumes or urban/suburban gradients. We also use the Atmospheric Pollution Over the Paris Area (ESQUIF) airborne measurements to demonstrate that the anisotropic method efficiently corrects CTM fields in altitude.

INDEX TERMS: 0345 Atmospheric Composition and Structure: Pollution—urban and regional (0305); 0368 Atmospheric Composition and Structure: Troposphere—constituent transport and chemistry; 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; *KEYWORDS:* air quality modeling, chemical data assimilation, statistical interpolation and kriging

Citation: Blond, N., L. Bel, and R. Vautard, Three-dimensional ozone data analysis with an air quality model over the Paris area, *J. Geophys. Res.*, 108(D23), 4744, doi:10.1029/2003JD003679, 2003.

1. Introduction

[2] In the last 20 years, atmospheric chemistry and air quality sciences have undergone increasing progress with the development of sophisticated chemistry-transport models (CTMs) and with the growing extent of operational

monitoring networks. So far, fields of pollutant concentrations have been only represented by either of the three following ways, namely, (1) direct model output maps, (2) maps displaying concentration values at given monitoring stations, or (3) field campaign measurement points. Direct model output concentration maps are useful as a first estimate of the pollutant concentrations anywhere over a given region. However, they are not exact because of model deficiencies. On the other hand, raw measurements suffer from representativeness problems and only provide local information.

¹Now at Laboratoire Interuniversitaire des Systèmes Atmosphériques, Universités Paris 7 et Paris 12, Créteil, France.

[3] The need for observation-consistent three-dimensional quantitative maps of pollutant concentration fields (usually called analyzed maps) is obvious. First, our understanding of the processes participating in photo-oxidant pollution events can only be complete if accurate three-dimensional descriptions of the atmospheric pollutant content are available. Second, inhabitants need to know the quality of the air they breathe regardless of whether or not they live close to monitoring stations. Finally, the routine production of analyzed concentration maps, used as initial fields, may be of particular relevance to air quality forecasting.

[4] In weather forecasting, accurate three-dimensional initial fields are required because forecast errors propagate to later times and eventually dominate. The routine production of analyzed meteorological fields has been a well-known question since forty years. A variety of statistical interpolators have been proposed in order to account for fresh observations when correcting previous forecasts [see, e.g., Daley, 1991]. Presently, the three-dimensional initial fields are constructed using data assimilation methods. In these methods, the observed information is accumulated into a model state by taking advantage of consistency constraints with laws of time evolution and physical properties. The development of elaborate data assimilation methods has led to an increasing interest, especially during the last decade with the pioneering work of Lewis and Derber [1985] and Talagrand and Courtier [1987] on four-dimensional variational (4D-VAR) assimilation. Specific input variables to the model are adjusted to optimize predicted output fields in comparison to observations. Techniques using the Kalman-Levy filter [Kalman, 1960; Kalman and Bucy, 1961; Ghil and Malanotte-Rizzoli, 1991] and variants have also been proposed.

[5] Recently, these advanced data assimilation techniques have been applied in air quality modeling to determine tropospheric pollutant concentration fields [Schmidt, 1999; Elbern and Schmidt, 2001; van Loon and Heemink, 1997]. In addition, they have been demonstrated to be efficient for optimizing model inputs such as emissions [Elbern et al., 2000; Elbern and Schmidt, 2002; van Loon et al., 2000; Mendoza-Dominguez and Russell, 2001; Chang et al., 1996, 1997; Gilliland and Abbitt, 2001]. Their main advantage is to provide best consistency between model input and output fields. However, they are computationally intensive and memory expensive. This is a major difficulty in operational real time applications for air quality diagnosis and forecasting.

[6] The purpose of this work is strictly to produce routine ozone concentration maps, so a full data assimilation scheme is not necessarily required. Therefore we propose and evaluate methods based on statistical interpolation [Daley, 1991] and ordinary kriging [Cressie, 1993]. These methods are based on the use of a CTM ozone concentration outputs and surface ozone measurements. Their advantage is to present a compromise between numerical cost and physical consistency.

[7] In atmospheric chemistry, statistical interpolation has been mainly used for stratospheric data assimilation [Jeuken et al., 1999]. In the troposphere, such an interpolation has been tested by Finkelstein [1984] to produce acid precipitation analyses. The method consists of interpolating differ-

ences between observations and simulations and adding the resulting field to the initial simulation. The weights of the interpolation depend on covariances of the observation errors, of the CTM output errors and of their cross-covariances. Therefore the success of the method relies on accurate knowledge of the error statistics in both observed data and CTM forecasts.

[8] In most existing analysis systems, the CTM error covariance, usually called the background error covariance, is modeled using assumptions about homogeneity and isotropy. However, these assumptions are often violated in the case of pollutants analysis. These violations especially occur where there are strong concentration gradients (near large sources, topographic barriers). Similar isotropy violations have been noticed in the representation of wind or temperature near a front [Desroziers, 1997]. Riishøjgaard [1998] studied the assimilation of nadir total ozone observations in a weather forecast model. He proposed a simple way to construct an anisotropic error covariance model. The method consists of multiplying the standard isotropic covariance, based on a distance criterion, by a function that decreases with the norm of the differences between values of the CTM simulated field. Following Riishøjgaard's work, Hoelzemann et al. [2001] developed a heterogeneous and anisotropic error covariance model to produce ozone analysis over Europe. The error covariance is based on a urban/rural classification of the CTM grid points in addition to the distance criterion. In this paper, we propose a simple way for specifying background error covariance in terms of the background covariance itself. Error covariances between two points are predicted using a function that decreases with the covariances between the respective values of the simulated field.

[9] In the statistical interpolation approach the background error covariance is modeled assuming ergodicity of the forecasting system. Under this assumption the background error covariance is substituted by an average over a time sequence of ozone residuals (observed minus CTM simulated values). This time sequence leads to a climatological description of the CTM error. The main disadvantage of the approach is that the background error covariance is treated as being decoupled from the actual pollution state of the atmosphere. This is why we also test spatial interpolation techniques. The error statistics are no longer assumed to be stationary in time. Time averages are substituted by spatial averages over data available only at the analysis time.

[10] Several studies have been conducted, which were aimed at spatially interpolating observations of environmental parameters. Most of them are based on kriging formalism [Finkelstein, 1984; Fedorov, 1989]. Ionescu et al. [2000] used thin plate splines to interpolate nitrogen dioxide observation data. They obtained fields of the pollutant over Paris area. Fuentes [2001] and Fuentes and Raftery [2001] combined outputs of a numerical model and measurements to produce maps of pollutant concentrations. In the study by Fuentes [2001] the model outputs are used to perform a nonstationary covariance model by a spectral method. In the study by Fuentes and Raftery [2001], model outputs and measurements are prior of a Gibbs sampling used to deduce posterior conditional distribution of the true process in a Bayesian melding approach. Both methods seem to be

computationally intensive and would not be suitable for routine hourly analyses. In the present study, the CHIMERE CTM [Vautard *et al.*, 2001] is used to capture the non-stationarity of the ozone fields in a fast and simple way. Kriging is not only applied to observations but also to ozone residuals. In this way, the model gives a first estimate of the geographical distribution of the pollutant concentrations.

[11] We discuss the advantages and disadvantages of the proposed methods when used in an operational forecasting and diagnosis setup. We focus on accurate representations of ozone fields on the scale of a large populated city, Paris, and its surroundings (Île-de-France area). The problem of forecast improvement is not considered in this work. The 2-D or 3-D ozone maps are produced using surface ozone observations and possibly ozone CHIMERE outputs. Their quality are objectively estimated using the leave-one-out method over the same surface measurements. Additionally, in order to validate 3-D analyzed ozone patterns in altitudes (typically a few hundred meters), we use airborne measurements from the Atmospheric Pollution Over the Paris Area (ESQUIF) field campaign [Menut *et al.*, 2000; Vautard *et al.*, 2003a].

[12] In section 2 we briefly described the CTM. Section 3 presents the observations used in this study. The mathematical formulations of statistical interpolation and kriging techniques are given in section 4. The results are presented in section 5. We conclude with a summary and a discussion in section 6.

2. Multiscale CTM

2.1. Model Brief Description

[13] CHIMERE [Vautard *et al.*, 2001; Schmidt *et al.*, 2001] is a 3-D Eulerian chemistry-transport model. This model is designed to allow fast simulations and forecasts of photo-oxidant concentrations. It has been extended to a multiscale version. A coarse-grid ($0.5^\circ \times 0.5^\circ$) version covers western Europe, and forces a refined-grid small-scale version at the boundaries (one-way nesting). The small-scale version of the model is centered on the city of Paris and covers an area of $150 \text{ km} \times 150 \text{ km}$. Its horizontal resolution is of $6 \text{ km} \times 6 \text{ km}$. The domain and grid are represented in Figure 1a.

[14] The model formulation is based on the mass continuity equation for several chemical species in every grid cell [Seinfeld and Pandis, 1998]. The time numerical solver is the TWOSTEP method, which has been originally described by Verwer [1994] for gas-phase chemistry. The TWOSTEP method is applied here to integrate all processes including advection and diffusion as proposed by Schmidt *et al.* [2001]. Vertical resolution of both the coarse-grid and refined-grid model consists in five layers, going from surface up to 750 hPa. Therefore the model contains the whole boundary layer in anticyclonic conditions over western Europe. For the coarse-grid model, the emissions are derived from the EMEP [Mylona, 1999] annual totals, modulated in time and VOC speciation by *Generation of European Emission Data for Episodes (GENEMIS) Project* [1994] profiles. For the regional domain, anthropic emission data are provided by AIRPARIF, the air quality monitoring network. The spatial scale of these emission data is of 3 km. More information is given by Vautard *et al.* [2003b].

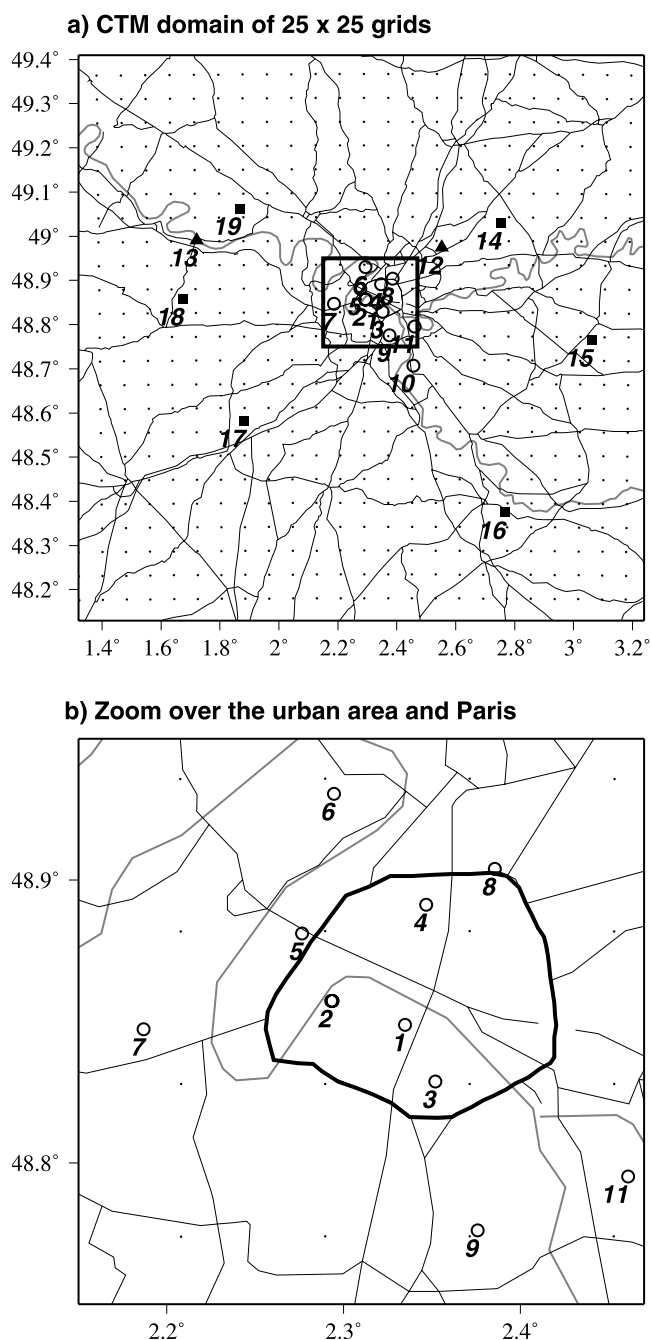


Figure 1. (a) Regional CTM extension and (b) zoom on the city of Paris. The city lies in the center of the maps. The thicker line represents the highway around Paris. Open circles, solid squares, and solid triangles denote urban, rural, and suburban surface monitoring stations, respectively. The names of these stations are: 1. PARIS06, 2. PARIS07, 3. PARIS13, 4. PARIS18, 5. NEUILLY, 6. GENNEVILLIERS, 7. GARCHES, 8. AUBERVILLIERS, 9. VITRY, 10. MONTGERON, 11. CRETEIL, 12. TREMBLAY, 13. MANTES, 14. MONTGE-EN-GOELE, 15. SAINTS, 16. FONTAINEBLEAU, 17. SONCHAMP, 18. PRUNAY, 19. FREMAINVILLE.

Biogenic emissions data base has been developed by *Derognat et al.* [2003] and has a resolution of 6 km. The land-use data are derived from the RIVM data basis [*van de Velde et al.*, 1994]. Following the concept of chemical operators, a reduced chemical mechanism has been derived from the original one [*Lattuati*, 1997]. It describes 116 reactions with 44 gaseous species.

[15] Advection is modeled using the piecewise parabolic method third-order scheme [*Carpenter et al.*, 1990]. Vertical mixing is parameterized by a diffusion profile [*O'Brien*, 1970] which depends only on the height of the boundary layer. This height is calculated from Richardson number profiles. Dry deposition is modeled following the resistance analogy presented by *Wesely and Hicks* [1977] and *Erisman et al.* [1994]. Photolytic rates are modulated by cloudiness. Boundary concentrations for the continental setup are fixed for fourteen species. These species have long lifetime and are relevant to photo-oxidant formation. Their concentrations are taken from a climatology of monthly mean data produced by the global MOZART CTM [*Hauglustaine et al.*, 1998].

[16] The CTM is at all scales forced by the European Centre for Medium-Range Weather Forecasts (ECMWF) meteorological short range forecasts. These data include temperature, pressure, wind, humidity and cloudiness fields. Their horizontal resolution is of about 0.5° . They are described in 10 vertical hybrid sigma-pressure levels for summer 1999 (18 for summer 2001), covering the boundary layer. Given the geographical nature of the city of Paris, which is located far away from any coasts and any significant topography, the meteorological variables can reasonably be represented by the large-scale meteorological model of ECMWF. Meteorological data have a time resolution of six hours and are linearly interpolated to one hour intervals.

[17] The chemistry and physical processes of the model are optimized enough (reduced chemical mechanism, low vertical resolution, use of large-scale meteorology data inputs, etc.) that long simulations or real-time forecasts are possible on a workstation. Typically 5 CPU minutes are required for one simulated day on a PC. However, the model remains realistic and allows for quantitatively reasonable ozone concentrations at various spatial scales. More details about the model are given by *Vautard et al.* [2001] and *Schmidt et al.* [2001].

2.2. CTM Errors

[18] The CHIMERE CTM errors are due to several causes: (1) modeling conceptual limitations (approximate description of processes, lack of an upper to mid-troposphere in the CTM, use of coarse meteorology), (2) model setup (horizontal and vertical resolutions), and (3) input data errors (imperfect emissions, meteorological errors, approximate land use data and surface resistance, boundary conditions, etc.). The combination of these errors leads to a total model error, which is usually called the background error. *Schmidt et al.* [2001] showed with CHIMERE model that the RMS error of simulated ozone daily maxima is of the order of 10 ppb on continental scale. *Vautard et al.* [2001] also found this order of magnitude on regional scale for the Paris area. They also use CHIMERE and show that one of the major difficulty of the CTM is to capture the correct city plume direction under very light wind conditions. However,

in most cases (light to moderate wind speed), the structure and amplitude of the ozone field was correctly simulated. Therefore we expect the CTM to add valuable information to the interpolation methods, especially when there is spatial variability in the ozone field.

3. Observations

3.1. Surface Hourly Ozone Concentrations

[19] The surface ozone observations are provided by AIRPARIF. Three types of ozone monitoring sites are used: urban stations, suburban stations and rural stations. The typology and discrimination of these sites are defined by European standards.

[20] Urban stations are located where the minimum population density is 4000 inhabitants/km² within a radius of 1 km. Suburban stations are situated under the direct influence of the city. Both type of stations provide a representative estimate of pollutant concentration in populated areas. They are located at some distance from all direct sources of pollution. Most of them are in the city of Paris. In contrast, rural stations are located about 50 km away from the city of Paris and monitor pollution advected by the wind.

[21] The distribution of monitoring sites is shown in Figure 1. There are 19 monitoring sites including six rural sites, 2 suburban sites and 11 urban sites.

3.2. Airborne Measurements

[22] The ESQUIF project (Étude et Simulation de la Qualité de l'air en Île-de-France) [*Vautard et al.*, 2003a] was designed to improve the understanding of photochemical pollution events in the region of Paris. Within the framework of this project, a large number of observation flights have been performed around the agglomeration [*Menut et al.*, 2000]. The airborne measurements have been carried out for several purposes, which include the study of ozone production in the Paris pollution plume and the improvement in emission inventories [*Vautard et al.*, 2003b].

[23] These flights were performed during high-ozone-concentration episodes during Intensive Observation Periods (IOPs) which lasted several days. The ozone airborne measurements used in the present article are taken from two aircraft, the ARAT and the DIMONA aircrafts. DIMONA aircraft was used to study the chemical composition of the air masses entering and leaving Paris area. The ARAT flight was designed to study the chemical regimes throughout the region. Details of the measurements are given by *Vautard et al.* [2003a].

[24] The above measurements are not used directly in the analysis methods. However, they allow for an evaluation of the methods' skills in representing ozone in the middle of the boundary layer. Table 1 lists the flight times and altitudes and CTM level for comparison.

3.3. Observation Error

[25] The knowledge of instrumental characteristics allows us to estimate observation uncertainties. On the other hand, representativeness error are difficult to evaluate.

[26] In the present study, we consider no instrumental biases. Even though this statement is hard to verify in practice, the observations are assumed on average to be

Table 1. Information About Flights of Aircrafts DIMONA and ARAT During the ESQUIF Campaign

Flight	Day	Time, UTC	MAF ^a	Level ^b
DIMONA	25 June 1999	1232–1638	270	2
DIMONA	2 July 1999	0752–0990	295	2
DIMONA	2 July 1999	1235–1660	292	2
DIMONA	16 July 1999	1157–1563	299	2
DIMONA	17 July 1999	0787–1010	262	2
DIMONA	17 July 1999	1498–1840	322	2
DIMONA	18 July 1999	0798–1038	270	2
DIMONA	18 July 1999	1223–1627	330	2
ARAT	17 July 1999	1124–1397	847	3
ARAT	17 July 1999	1518–1782	906	3
ARAT	29 July 1999	1110–1363	830	3
ARAT	29 July 1999	1486–1715	781	3
ARAT	30 July 1999	1212–1472	849	3

^aMean values of the altitude of the flight (in meters) computed over the time period indicated in the third column.

^bCTM level to compare with observations.

representative of an environment whose dimensions are at least comparable to the model grid size. Indeed, since an observation is a hourly mean, it is representative of the mean concentration advected into the monitoring station during that time. For wind speeds, at least of $1-2 \text{ m s}^{-1}$, the space spanned by the flow is about 5 km. Hence one hourly-mean observation should be comparable to the instantaneous concentration in one CTM grid cell. In other words, we assume no systematic representativeness biases. If there is one, it is assumed to be due to model deficiencies. Thus the observation error only consists in a random error, including instrumental and representativeness random errors.

[27] The representativeness of observations is controlled by small-scale features like local emissions or surface characteristics. The representativeness error depends on the same features and also on the resolution of the analysis grid. To estimate properly the random part of this error, it is necessary to have several monitoring stations in the same model box. As shown in Figure 1, this is a quite unusual situation, except in the Paris center, where two or three monitor stations can be in the same model box. The standard deviation of the observations within one box is calculated every day at each hour. It is an estimate of the representativeness error. For instance, this standard deviation is about 5 ppb at 1500 UTC.

4. Analysis Methods

[28] The analysis schemes developed in this paper follow the general framework of statistical interpolation [Daley, 1991] and kriging techniques [Cressie, 1993]. The aim is to provide an estimate $Z_{t,h}^a(\mathbf{s})$ of a field concentration value $Z_{t,h}(\mathbf{s})$ at any location \mathbf{s} , at day t ($t = 1, \dots, T$), and hour h ($h = 1, \dots, 24$). For this purpose, we have a set of K spatially distributed measurement values $Y_{t,h}^o(\mathbf{s}_k)$ where $\mathbf{k} = 1, \dots, K$ and \mathbf{s}_k is the location of the k -th monitoring station. We also use a prior estimate of the concentration field $Z_{t,h}^b(\mathbf{s})$, which is often called the first guess or the background field. In this setting, $Z_{t,h}^a(\mathbf{s})$, which is called the analysis, is given by

$$Z_{t,h}^a(\mathbf{s}) = Z_{t,h}^b(\mathbf{s}) + \sum_{k=1}^K \lambda_{t,h}^k(\mathbf{s}) \left(Y_{t,h}^o(\mathbf{s}_k) - Z_{t,h}^b(\mathbf{s}_k) \right). \quad (1)$$

The $\lambda_{t,h}^k(\mathbf{s})$ are weighting functions which have to be determined.

[29] Equation (1) shows that the analyzed field is a correction of the first-guess field by a linear combination of the innovations, $Y_{t,h}^o(\mathbf{s}_k) - Z_{t,h}^b(\mathbf{s}_k)$. The methods proposed in the present study differ in the way the weights $\lambda_{t,h}^k(\mathbf{s})$ are calculated and in the background fields. These background fields can be zero or fields simulated by CTM CHIMERE.

4.1. Isotropic Statistical Interpolation (ISI) Method

[30] This is by far the most common method in meteorology for producing daily analyses. The reader is referred to Daley [1991] for a complete mathematical description and meteorological applications of the method. We now recall only briefly the methodology, and especially the way to construct the background error covariance.

[31] Assume temporarily that the background and the observations are unbiased. Here, background fields $Z_{t,h}^b$ are the hourly ozone fields simulated by CTM CHIMERE at the first level. The optimized weighting functions $\hat{\lambda}_{t,h}^k(\mathbf{s})$ are obtained by minimizing the mean square analysis error, e.g., $E(Z_{t,h}^a(\mathbf{s}) - Z_{t,h}(\mathbf{s}))^2$. $E()$ denotes the statistical mean. The minimization problem leads to a linear system involving: (1) covariances of the background error, $c_{t,h}^b(\mathbf{s}_k, \mathbf{s}_l)$, (2) covariances of the observation error $c_{t,h}^o(\mathbf{s}_k, \mathbf{s}_l)$, and (3) their cross-covariance function. Since the causes of errors in the first guess and in the observations are presumably independent, the cross-covariance is assumed to be zero. Moreover, we assume that observation errors of two distinct monitoring stations are uncorrelated, e.g., $c_{t,h}^o(\mathbf{s}_k, \mathbf{s}_l) = \sigma_{t,h}^{2o}(\mathbf{s}_k) \delta_{kl}$. δ_{kl} is the Kronicker delta function ($\delta_{kl} = 0$ if $k \neq l$ and $\delta_{kl} = 1$ if $k = l$). In contrast, the background errors are *a priori* spatially dependent. In this setting, optimized weighting functions verify the following system: $\forall k = 1, \dots, K$

$$\sum_{l=1}^K \hat{\lambda}_{t,h}^l(\mathbf{s}) \left[c_{t,h}^b(\mathbf{s}_k, \mathbf{s}_l) + \sigma_{t,h}^{2o}(\mathbf{s}_k) \delta_{kl} \right] = c_{t,h}^b(\mathbf{s}_k, \mathbf{s}). \quad (2)$$

[32] The computation of the weighting functions requires the preliminary knowledge of the covariances of the background errors and the variances of the observation errors, $\sigma_{t,h}^{2o}(\mathbf{s}_k)$. Because these quantities are not available directly, they need to be estimated in a statistical sense.

4.1.1. Error Covariance Modeling

[33] The best source of information about errors is the innovations $Y_{t,h}^o(\mathbf{s}_k) - Z_{t,h}^b(\mathbf{s}_k)$. Assuming that observation errors are uncorrelated and independent on CTM errors, it can easily be shown that the innovation covariance $c_{t,h}^i(\mathbf{s}_k, \mathbf{s}_l)$ equals to $c_{t,h}^b(\mathbf{s}_k, \mathbf{s}_l) + \sigma_{t,h}^{2o}(\mathbf{s}_k) \delta_{kl}$. Therefore covariance $c_{t,h}^b(\mathbf{s}_k, \mathbf{s}_l)$ between two distinct monitoring station locations \mathbf{s}_k and \mathbf{s}_l ($k \neq l$, $\delta_{kl} = 0$) can be directly estimated from $c_{t,h}^i(\mathbf{s}_k, \mathbf{s}_l)$. In contrast, covariance $c_{t,h}^b(\mathbf{s}_k, \mathbf{s})$ between a monitoring station location \mathbf{s}_k and any location \mathbf{s} , where there is generally no available measurements, needs to be computed through a covariance model. To construct this model, we use the observational method [Hollingsworth and Lönnberg, 1986]. It consists in assuming that (1) the forecasting system is ergodic, and then the error statistics can be obtained from an average over a long time

sequence of innovations (2) $c_{t,h}^b(\mathbf{s}_k, \mathbf{s}_l)$ is homogeneous and isotropic.

[34] The first assumption implies that the error covariances do not depend on day t . However they depend on hour h , e.g., $c_{t,h}^b(\mathbf{s}_k, \mathbf{s}_l) = c_h^b(\mathbf{s}_k, \mathbf{s}_l)$ and $\sigma_{t,h}^{2o}(\mathbf{s}_k) = \sigma_h^{2o}(\mathbf{s}_k)$. The weighting functions do not depend on t , e.g., $\hat{\lambda}_{t,h}^k(\mathbf{s}) = \hat{\lambda}_h^k(\mathbf{s})$. They are optimized for each hour h of the day. In this way, the pronounced diurnal cycle of the ozone concentrations is reproduced. The second assumption implies that $c_h^b(\mathbf{s}_k, \mathbf{s}_l)$ between two locations \mathbf{s}_k and \mathbf{s}_l depends on their distance $r = \|\mathbf{s}_k - \mathbf{s}_l\|$, e.g., $c_h^b(\mathbf{s}_k, \mathbf{s}_l) = c_h^b(r)$.

[35] In order to determine function $c_h^b(r)$, innovation covariance coefficients are usually computed for innovation time series at pairs of monitoring station locations. Then they are plotted versus their corresponding distance. At zero distance, the plot provides average information about the background and observation errors. At nonzero distance it gives the averaged background error covariance function.

[36] The methodology to construct the background error covariance model is here achieved in two steps. First, a model of the background error correlation $\rho_h^b(r)$ and second a model of the innovation variance $\sigma_h^{2i}(\mathbf{s})$ are built. The error covariance model is then the product of the two separate models.

[37] The background error correlation function $\rho_h^b(r)$ is estimated by a fit of available innovation correlation coefficients, $\rho_h^i(\mathbf{s}_k, \mathbf{s}_l)$. These coefficients are plotted versus distance $r = \|\mathbf{s}_k - \mathbf{s}_l\|$ for each hour h . In order to ensure the symmetric positive definiteness of background error covariance matrix [Gaspari and Cohn, 1999], we use the scaled Balgovind *et al.* [1983] correlation function:

$$\rho_h^b(r) = \rho_h^b(0)(1 + r/a_h) \exp(-r/a_h), \quad (3)$$

where $\rho_h^b(0)$ and a_h are the regression coefficients.

[38] Under homogeneous conditions, the innovation variance does not depend on location \mathbf{s} , e.g., $\sigma_h^{2i}(\mathbf{s}) = \sigma_h^{2i}$. It is estimated by taking the mean value of the innovation variances.

[39] Finally, the background error covariance is written

$$c_h^b(r) = \sigma_h^{2i} \rho_h^b(r). \quad (4)$$

We notice that it is not site-specific (e.g., it does not depend on location).

[40] σ_h^{2i} is actually the sum of the background error variance, $\sigma_h^{2b} = c_h^b(0)$, and the observation error variance, σ_h^{2o} . Both variances are assumed to be homogeneous. $\rho_h^b(0)$ is the ratio of σ_h^{2b} to $\sigma_h^{2b} + \sigma_h^{2o}$. Therefore term $\sigma_{t,h}^{2o}(\mathbf{s}_k)$ in equation (2), which is equaled to the constant value σ_h^{2o} , can easily be estimated from $\rho_h^b(0)$ and σ_h^{2i} .

[41] Figure 2 shows the innovation correlation coefficients for 0700 UTC plotted versus their corresponding distances. The innovation correlation coefficients are calculated for innovation time series at pairs of stations. The time series cover summer 1999, from 1 May through 30 September, corresponding to $T = 153$ days. We remark that the innovation correlation coefficients vary from 0.4 to 0.8 for distances smaller than 10 kilometers. They also vary over a very wide range of values for larger distances. Moreover,

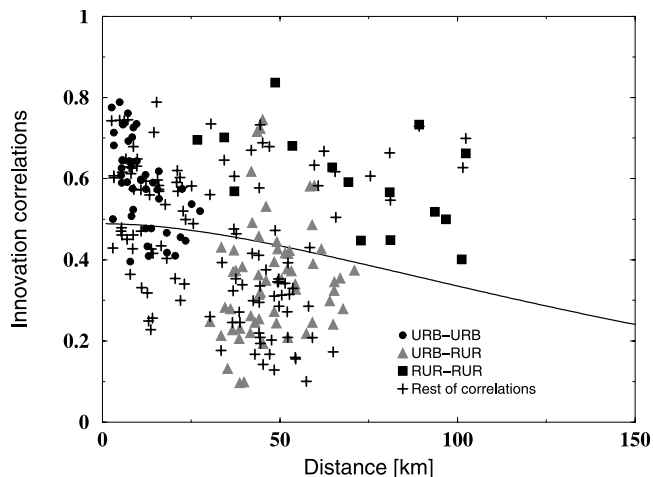


Figure 2. Innovation correlation coefficients, $\rho_h^i(\mathbf{s}_k, \mathbf{s}_l)$, calculated for innovation time series (over summer 1999) at pairs of stations (k, l) for $h = 0700$ UTC, versus the distance $r = \|\mathbf{s}_k - \mathbf{s}_l\|$. The squares stand for pairs of rural stations, the circles stand for pairs of urban stations, the triangles stand for the pairs of the type [rural;urban], and the plus symbols stand for the rest of correlations.

fairly high correlation coefficients (0.7–0.8) are found even if distance between monitoring stations is larger than 50 km. These coefficients are of the same magnitude as the ones calculated at some pairs of close monitoring stations. These long-distance correlations occur for pairs of rural stations while pairs of the type [rural;urban] display low correlation coefficients. The explanation of such a behavior is simple. Ozone is titrated by nitrogen monoxide emissions during peak driving hours in shallow mixed layers over the urban area. Therefore its concentration is very weak and quite heterogeneous. In contrast, The concentrations advected over rural stations from other areas are homogeneous. Both observations and CTM outputs exhibit such a behavior. Therefore the innovations at rural locations are strongly correlated with one another, but are weakly correlated with those in the urban areas. As shown in Figure 2, the isotropy assumption is clearly violated.

[42] During the daytime, other isotropy violations occur (not shown), especially because of the city plume development. In other regions with significant topography or coastal areas, the isotropy can even fail to a much larger extent. In general, we expect that the error correlation between two locations depends more on the fact that the locations belong to the same spatial ozone patterns than on their distance. This is why we propose an anisotropic approach based on this principle in section 4.2.

[43] Finally, this isotropic covariance model can only be applied for 2-D fields. Indeed, the isotropy is obviously violated in the vertical direction because of the stratification of the atmosphere. The distance criterion is not adapted. Therefore, in the following, this method will only be used for 2-D analysis of surface ozone concentrations.

4.1.2. Bias Model

[44] We have temporally assumed that neither observations nor first guesses are biased. That is not true in practice, since model uncertainties lead to systematic errors and observations can suffer from instrumental or representative-

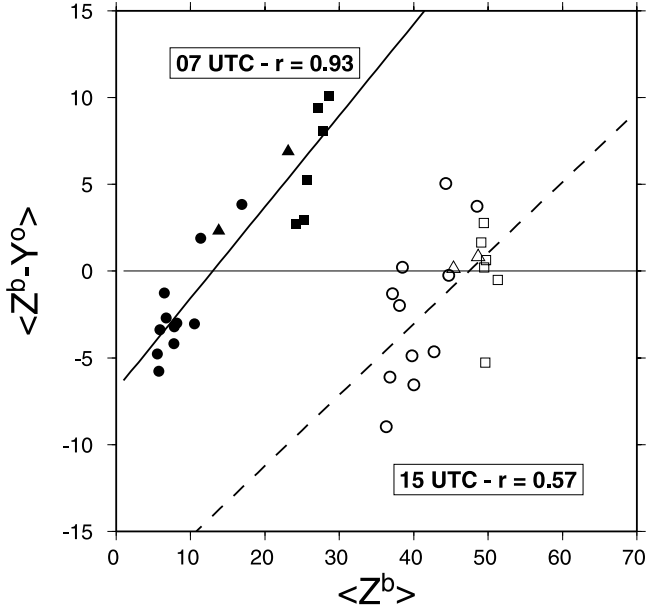


Figure 3. Averages of innovations $\langle Z^b(\mathbf{s}_k) - Y^o(\mathbf{s}_k) \rangle$ (in ppb), versus averages of ozone concentration values $\langle Z^b(\mathbf{s}_k) \rangle$ (in ppb) simulated and interpolated on station locations \mathbf{s}_k . Solid symbols stand for 0700 UTC, and open symbols stand for 1500 UTC. Linear regression lines are also presented: solid line for 0700 UTC and dashed line for 1500 UTC. r denotes the correlation regression coefficient. Circles, squares, and triangles denote urban, rural, and suburban surface observation sites, respectively.

ness biases. Instrumental and representativeness biases are assumed to be zero (see section 3.3). These assumptions mean that the innovation biases are assumed to be entirely due to CTM first-guess biases. In order to obtain an unbiased estimated field, we need to remove potential biases from the first guess, $Z_{t,h}^b(\mathbf{s})$ for each \mathbf{s} before applying equation (1). Therefore we also need to model these biases.

[45] Figure 3 displays averages of innovations versus averages of ozone concentrations simulated at each monitoring stations for two specific hours of the day, 0700 UTC and 1500 UTC. The averages are computed over summer 1999. We can see that the model tends to overestimate ozone concentrations in rural areas where ozone concentrations are highest. It is especially the case for 0700 UTC. The model also tends to underestimate concentrations over the urban area, displaying low ozone concentration means. This especially happens at 1500 UTC.

[46] Finally at any location within the studied area, the first guess bias is modeled using a linear regression between bias and CTM ozone concentration mean. It is removed from the background field $Z_{t,h}^b$ before applying the analysis algorithm each day t at each hour h . This yields 24 different sets of linear coefficients per CTM box.

4.2. Anisotropic Statistical Interpolation (ASI) Method

[47] This new method is based on the assumption that innovation correlations are higher when the two locations frequently belong to the same ozone pattern and lower when they do not. The methodology is nearly the same as in section 4.1. However, instead of modeling the background

error correlation $\rho_h^b(\mathbf{s}_k, \mathbf{s}_l)$ as a function of the distance $r = \|\mathbf{s}_k - \mathbf{s}_l\|$, we model it as a function of the ozone concentration correlation, $\rho_h^{[O_3]}(\mathbf{s}_k, \mathbf{s}_l)$. An exponential fit is again carried out to model these correlations (see equation (6)). We also assume that innovation variance $\sigma_h^{2i}(\mathbf{s})$ depends linearly on ozone concentration variance $\sigma_h^{2[O_3]}(\mathbf{s})$ (see equation (7)). In contrast with ISI method, error variances, $\sigma_h^{2b}(\mathbf{s})$ and $\sigma_h^{2o}(\mathbf{s})$, are no longer assumed to be homogeneous. They depend on location \mathbf{s} . Finally, the background error covariance model is written:

$$c_h^b(\mathbf{s}_k, \mathbf{s}_l) = \sigma_h^i(\mathbf{s}_k) \sigma_h^i(\mathbf{s}_l) \rho_h^b(\mathbf{s}_k, \mathbf{s}_l), \quad (5)$$

with

$$\rho_h^b(\mathbf{s}_k, \mathbf{s}_l) = e_h \left(1 + \left(1 - \rho_h^{[O_3]}(\mathbf{s}_k, \mathbf{s}_l) \right) / b_h \right) \times \exp \left(- \left(1 - \rho_h^{[O_3]}(\mathbf{s}_k, \mathbf{s}_l) \right) / b_h \right), \quad (6)$$

and

$$\sigma_h^{2i}(\mathbf{s}) = \alpha_h \sigma_h^{2[O_3]}(\mathbf{s}) + \beta_h. \quad (7)$$

$-1 \leq \rho_h^{[O_3]}(\mathbf{s}_k, \mathbf{s}_l) \leq 1$. α_h , β_h , e_h and b_h are the regression coefficients. These coefficients are computed at each hour h . As $c_h^b(\mathbf{s}_k, \mathbf{s}_l)$ depends on ozone concentration correlation pattern, it is site-specific.

[48] Figure 4 shows the innovation correlation coefficients for 0700 UTC, plotted versus corresponding coefficients of correlation of observed ozone concentrations. Coefficients are calculated for time series at pairs of stations. The spread of points is clearly smaller than when distance is used (see Figure 2). The innovation correlations computed at pairs of urban stations or pairs of the type [rural; urban]

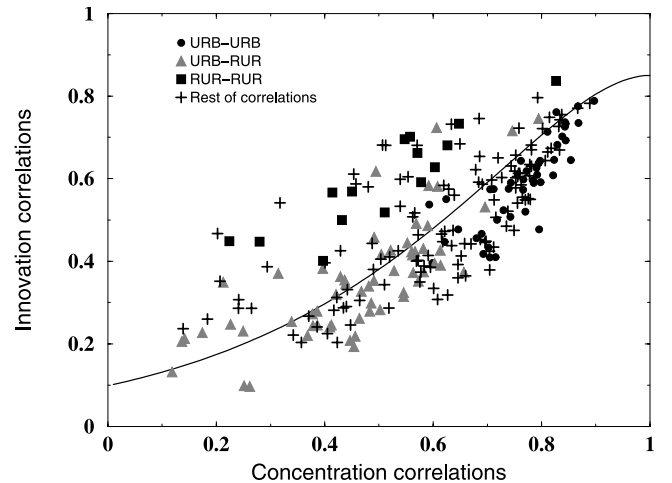


Figure 4. Innovation correlation coefficients $\rho_h^i(\mathbf{s}_k, \mathbf{s}_l)$ computed for innovation time series (over summer 1999) at pairs of stations (k, l) for $h = 0700$ UTC versus the corresponding correlation coefficients of observed ozone values $\rho_h^{[O_3]}(\mathbf{s}_k, \mathbf{s}_l)$. The squares stand for pairs of rural stations, the circles stand for pairs of urban stations, the triangles stand for the pairs of the type [rural; urban], and the plus symbols stand for the rest of correlations.

are better modeled. Nevertheless, we notice that innovation correlations computed at pairs of rural stations still remain underestimated.

[49] As the ozone concentration correlations cannot be computed for sites where no measurements are available, we use the correlation computed on simulated ozone concentrations. Since the ozone concentration correlations given by the CTM may contain biases relative to observed ones, they are corrected using a linear regression.

[50] In the same way, $\sigma_h^2(\mathbf{s})$ model is estimated from the variances of observed values. It is computed for any \mathbf{s} using variances of simulated ozone concentrations, which are not particularly biased relative to variances of observed values.

[51] Finally, we remark that the anisotropic covariance model proposed here can be a priori used to produce 3-D analyses using only surface measurements. Indeed, the method takes as an error covariance predictor, the covariance of the simulated ozone concentration field itself regardless of its 3-D structure. However, the skill of the method above the surface is difficult to estimate since we do not have a large series of upper air ozone measurements. The ESQUIF airborne measurements is precisely used to make an a posteriori verification of the quality of upper air analyses in section 5.2.

[52] Both statistical interpolation methods proposed here strongly rely on the assumption of time stationarity of the background error covariance. They lead to weighting functions $\widehat{\lambda}_h^k(\mathbf{s})$, which do not depend on the state of the atmosphere of the day t . The weighting functions do not account for daily changes in meteorology and emissions, which could modify the information content of a given measurement. This is why we now investigate the use of kriging methods. In these methods, weighting functions are computed at each analysis, and are time dependent.

4.3. Observation Kriging Method (OBK)

[53] In geostatistics, estimating a random field at location \mathbf{s} from observations at locations \mathbf{s}_k is classically achieved by the kriging method. The analyzed value at a given point \mathbf{s} reads:

$$Z_{t,h}^a(\mathbf{s}) = \sum_{k=1}^K \lambda_{t,h}^k(\mathbf{s}) Y_{t,h}^o(\mathbf{s}_k). \quad (8)$$

[54] The method follows the general formalism of equation (1), but now the CTM outputs are not used. Therefore the first-guess field $\mathbf{Z}_{t,h}^b$ is zero. Ordinary kriging method [Cressie, 1993] is performed under the assumptions of spatial intrinsic stationarity and isotropy. These assumptions read

$$E(Z(\mathbf{s}_k) - Z(\mathbf{s}_l)) = 0, \quad (9)$$

$$\frac{1}{2} E(Z(\mathbf{s}_k) - Z(\mathbf{s}_l))^2 = \gamma_{t,h}(\|\mathbf{s}_k - \mathbf{s}_l\|). \quad (10)$$

[55] The weighting functions $\widehat{\lambda}_{t,h}^k(\mathbf{s})$ are optimized at each hour h and day t . We minimize the mean square analysis error with an additional constraint to force the sum of weights to be equal to one. This constraint is required to

obtain an unbiased estimate (μ is a Lagrangian multiplier). The criterion to minimize is written:

$$E\left(Z_{t,h}^a(\mathbf{s}) - Z_{t,h}(\mathbf{s})\right)^2 + 2\mu \left(\sum_{k=1}^K \lambda_{t,h}^k(\mathbf{s}) - 1\right).$$

[56] The minimization problem leads to the following system:

$$\begin{cases} \forall k = 1, \dots, K, \\ \sum_{l=1}^K \widehat{\lambda}_{t,h}^l(\mathbf{s}) \gamma_{t,h}(\mathbf{s}_k, \mathbf{s}_l) - \mu = \gamma_{t,h}(\mathbf{s}_k, \mathbf{s}), \\ \sum_{l=1}^K \widehat{\lambda}_{t,h}^l(\mathbf{s}) = 1. \end{cases} \quad (11)$$

[57] System (11) involves the semivariogram function $\gamma_{t,h}(\|\mathbf{s}_k - \mathbf{s}_l\|) = \gamma_{t,h}(r)$ with $r = \|\mathbf{s}_k - \mathbf{s}_l\|$. The observation error variance $\sigma_{t,h}^{2,o}(\mathbf{s}_k)$ is assumed to be homogeneous and constant in time, e.g., $\sigma_{t,h}^{2,o}(\mathbf{s}_k) = \sigma_h^{2,o}$. In practice, hourly values $\sigma_h^{2,o}$ are the same values computed when using ISI method.

[58] To model the semivariance $\gamma_{t,h}(r)$, an experimental variogram $\gamma_{t,h}^{\text{exp}}$ is first computed using observation data available at hour h , day t . We construct some distance classes N_j ($j = 1, \dots, J$) with a length interval of $L = 15$ km, and a tolerance of $\tau = 10$ km. Each class, N_j , is defined by

$$N_j = \left\{ (\mathbf{s}_k; \mathbf{s}_l); \left| \|\mathbf{s}_k - \mathbf{s}_l\| - (2j-1)\frac{L}{2} \right| < \frac{L}{2} + \tau \right\}. \quad (12)$$

[59] Over each distance class, we compute

$$2\gamma_{t,h}^{\text{exp}}(r_j) = \frac{1}{|N_j|} \sum_{N_j} \left(Y_{t,h}^o(\mathbf{s}_k) - Y_{t,h}^o(\mathbf{s}_l) \right)^2, \quad (13)$$

where $|N_j|$ is the number of elements into the class N_j . r_j is the mean distance of the class. It is defined by

$$r_j = \frac{1}{|N_j|} \sum_{N_j} \|\mathbf{s}_k - \mathbf{s}_l\|. \quad (14)$$

[60] Second, an exponential model with nugget effect (noted NUG_h) is used to fit the experimental variogram. This model is written by

$$\begin{cases} \gamma_{t,h}(r) = c_{t,h} \left(1 - \exp\left(-\frac{r}{d_{t,h}}\right) \right) + NUG_h & r > 0, \\ \gamma_{t,h}(0) = 0, \end{cases} \quad (15)$$

where $d_{t,h}$ and $c_{t,h}$ are the regression coefficients.

[61] This function is conditionally negative definite. It produces a permissible model [Cressie, 1993]. Moreover, it has only two parameters to be fitted and it allows a quite wide range of situations. In our case, the nugget effect characterizes the observation error variance. For consistency with ISI method, we set the nugget to a value of $NUG_h = \sigma_h^{2,o} = 25 \text{ ppb}^2$ when $h = 1500$ UTC. The variability of the semivariogram is reflected in the statistics on parameters $d_{t,h}$ and $c_{t,h}$ given in Table 2 for $h = 1500$ UTC.

Table 2. Statistics Computed on Coefficients $d_{t,h}$ and $c_{t,h}$ of the Semivariogram Model^a

	Mean	Min.	Q25	Med.	Q75	Max.
$d_{t,15}$	1393	0.0056	4	15	45	17, 353
$c_{t,15}$	6437	0	59	141	445	233, 860

^aThese statistics are given for $h = 1500$ UTC over the period from the 1 May to 1 September 1999 ($t = 1, \dots, 153$). Min., minimum; Q25, quantile 25%, Med., medium; Q75, quantile 75%; Max., Maximum.

[62] When $d_{t,h}$ and $c_{t,h}$ are large, $\gamma_{t,h}(r)$ varies almost linearly with r , and goes to infinity when r is large (high variability). This high variability occurs when a strong urban plume is developed. In such cases, at same large distance r , differences $Y^o(\mathbf{s} + \mathbf{r}) - Y^o(\mathbf{s})$ between observations located in and out of the urban plume are higher than differences between observations located both out of the urban plume. The assumption of spatial constant mean of the true ozone concentration field made when using an ordinary kriging (see equation (9)) fails. In order to prescribe the geographical location of the plume and limit such nonstationarities, we propose in section 4.4 to use CTM outputs.

[63] Three experimental variograms are shown in Figure 5a for three typical days at 1500 UTC. For 17 July 1999, the variogram displays an increase of the spatial variation of $Z_{t,h}(\mathbf{s} + \mathbf{r}) - Z_{t,h}(\mathbf{r})$ when considering point pairs at large distance. As we will show later, three rural monitoring stations (FREMAINVILLE, PRUNAY, MANTES) are located in the urban plume.

4.4. Innovation Kriging Method (INK)

[64] The present method uses the same ordinary kriging method as in section 4.3. However, we interpolate the innovations instead of the observations. The idea is actually to use the CTM outputs as a predictor of the spatial trend of the field in order to reduce its variability. If the CTM predicts plume direction relatively well, the innovations vary less than the observations and equation (9) is more justified.

[65] Such method consists actually in estimating the CTM error field, $\mathbf{Z}_{t,h}^b - \mathbf{Z}_{t,h}^o$, using the K innovations, $Z_{t,h}^b(\mathbf{s}_k) - Y_{t,h}^o(\mathbf{s}_k)$, as “observations.” The analysis formula has the exact same form as in equation (1). However, weighting functions $\hat{\chi}_{t,h}^k(\mathbf{s})$ significantly differ from methods ASI or ISI. The variogram model is built here from the innovations instead of the ozone concentrations. The nugget is the same as in section 4.3.

[66] Some experimental and fitted variograms are displayed in Figure 5b for the same three days as in Figure 5a. For 17 July 1999, the variogram does no longer go to infinity at large distances. For 18 and 29 July 1999, variograms are now almost a positive constant (except at the origin). In these cases, the error field almost behaves as a white noise.

[67] As ISI method, INK method could only be used to correct the first level of the CTM.

5. Method Validation

[68] The four methods described in section 4 are now tested over the region of Paris during summer 2001.

5.1. Leave-One-Out Validation Using Ground-Based Observations

[69] In order to objectively evaluate the efficiency of the analysis methods, a leave-one-out method is used. Sequentially, observations collected during summer 2001 at one single observation site are omitted in the analysis process. These omitted observations are afterward statistically compared with the resulting analyzed field. Error statistics are calculated at each hour h by the root mean square analysis residuals RMS_h :

$$RMS_h(\mathbf{s}_k) = \sqrt{\frac{1}{T} \sum_{t=1}^T \left(Z_{t,h}^a(\mathbf{s}_k) - Y_{t,h}^{o*}(\mathbf{s}_k) \right)^2}, \quad (16)$$

where $Y_{t,h}^{o*}(s_k)$ ($t = 1, \dots, T$) denotes the omitted observations.

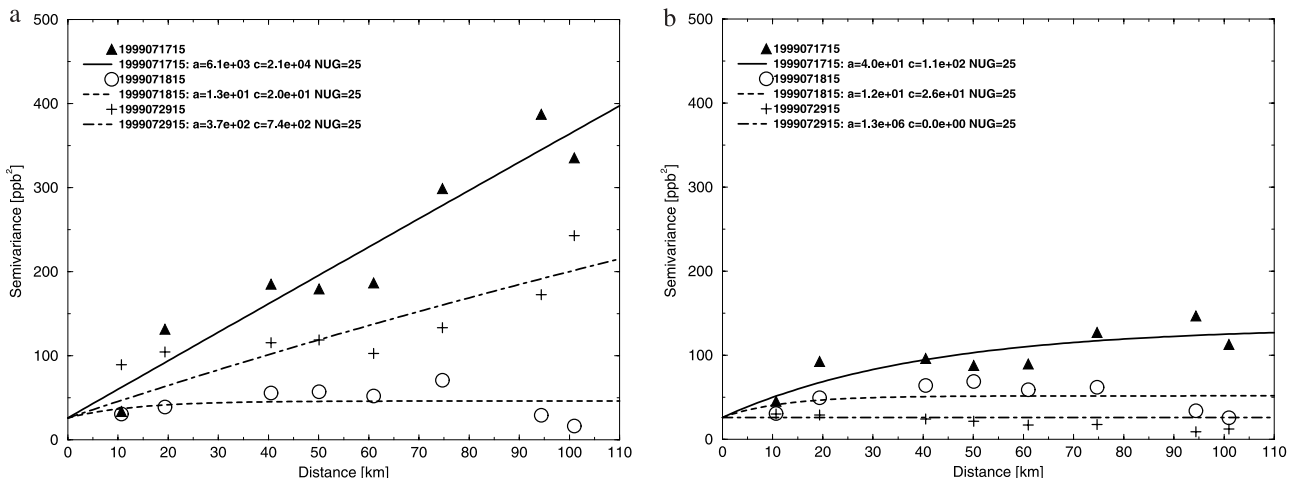


Figure 5. Experimental (symbols) and fitted (lines) variograms used in (a) OBK method and (b) INK method for three days at 1500 UTC of summer 1999: triangles and solid line for 17 July 1999, circles and dashed line for 18 July 1999, crosses and dot-dashed line for 29 July 1999. The semivariance is in ppb^2 , and the distance is in kilometers.

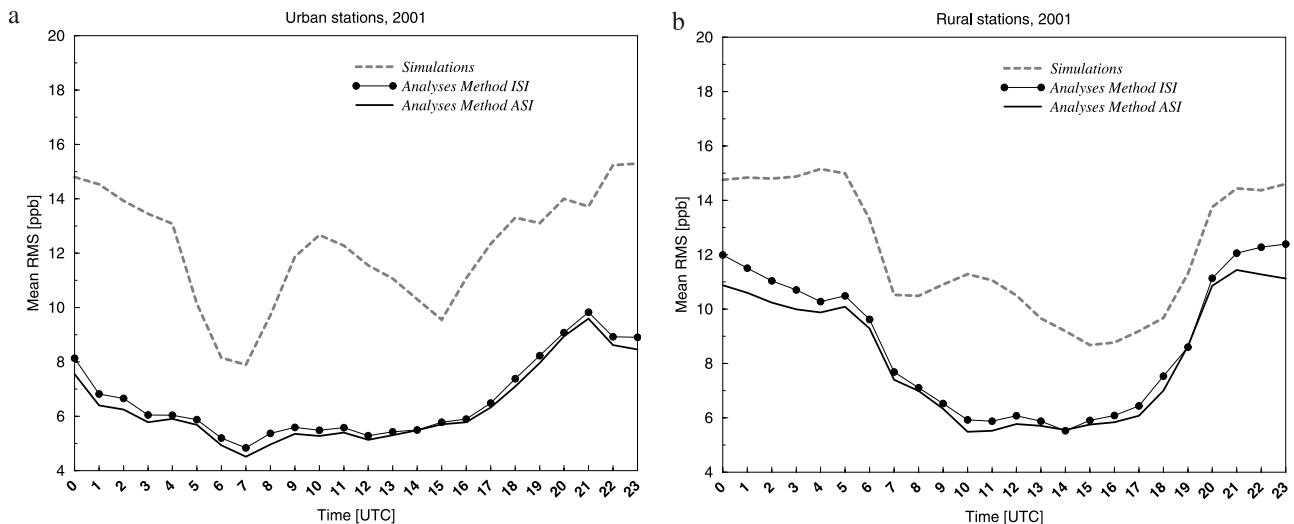


Figure 6. Mean diurnal cycles of the RMS analyses residuals computed over observations sites: (a) urban sites and (b) rural sites. The analyses are produced using ISI method (black solid line and circles) or method ASI (black solid line). Statistics are computed over summer 2001 using the leave-one-out method. The dashed line corresponds to the mean diurnal cycle of the RMS innovations. The RMS is in ppb, and time is in UTC.

[70] The statistical analysis methods (ISI and ASI) are trained over observed and simulated data available during summer 1999 and verified over data from summer 2001. Therefore all statistics, including biases and covariances are calculated from a set of data independent on the testing data. Summers are defined as running from 1 May to 30 September.

5.1.1. Isotropic and Anisotropic Methods

[71] We first compare ISI and ASI methods in order to investigate their relative skill. Figure 6 shows the mean diurnal cycles of the RMS analyses residuals and RMS innovations. These RMS error are computed over urban or rural observations sites. The analyses are produced using methods ISI or ASI.

[72] We note that the RMS analyses residuals are significantly smaller than the RMS innovations, regardless of the method used. Ozone analyses are clearly of better quality than the initial simulations. By combining observations and simulations, the RMS error is reduced of about 30 to 50%. We also note that RMS reduction is larger at urban sites than at rural sites. This is mainly caused by the high density of urban sites. It is also due to the larger adjustments needed at urban locations since model performance is often worse than at rural location.

[73] Moreover all curves display a marked diurnal cycle with values larger during nighttime than daytime. Several factors could explain this behavior. First the CTM itself simulates better ozone concentrations during daytime than nighttime, as shown by the RMS innovations. Indeed, nighttime ozone concentrations are very sensitive to the thickness of the mixing layer. However, during nighttime, the variability of this thickness is not well described, especially because of the low vertical resolution of the CTM. The first CTM layer is fixed at 50-m thick, while the mixing layer can be smaller at night. Second, because of titration with local nitrogen monoxide emissions, ozone structures at ground level could be of smaller scale than during daytime. Large horizontal gradients could cause representativeness problems. Since all other calculations

depend on innovation statistics, they could simply be modulated by the CTM skill variation.

[74] Finally the comparison between the two methods displays small differences (about 1 ppb) during nighttime and emission hours (mainly at peak traffic hours) in favor of the anisotropic method. These differences are often higher over rural stations because in data-sparse areas, the first guess correction is completely controlled by the background error correlation structure. While in data-dense area, like the city of Paris, the first guess corrections are less sensitive to the background error covariance shape and tend to average the measurements.

[75] Figure 7a details the RMS errors for 0700 UTC. Statistics display large fluctuations from one station to another. By comparison with ISI method, ASI allows almost systematically more reduction of RMS error. The maximum reduction is -17% . We note that the RMS goes up for station MONTGE-EN-GOELE for both ASI and ISI methods. Several factors could explain this behavior. First, on that monitoring station, the modeled bias (5 ppb) overestimates by a factor 2 the average of innovations (2.5 ppb). The standard deviations which allow to compare methods besides the bias problem are reduced of about 12% for both methods. Second, correlations of observed ozone concentrations between that monitoring station and the other ones are also specially overestimated by the correlations calculated with the CTM. The corresponding background error covariances are also probably overestimated. The monitoring station may be influenced by small-scale features like local emission sources or dynamical processes not resolved by the CTM. It is located on a house side in a town of about 8000 inhabitants. Figure 7b details the RMS errors for 1500 UTC. At this time, no significant differences are observed between ISI and ASI methods.

[76] The present results confirms that the use of a heterogeneous and anisotropic background error covariance has a positive impact on the analyses. During emission hours and nighttime, errors are not correlated between rural and urban

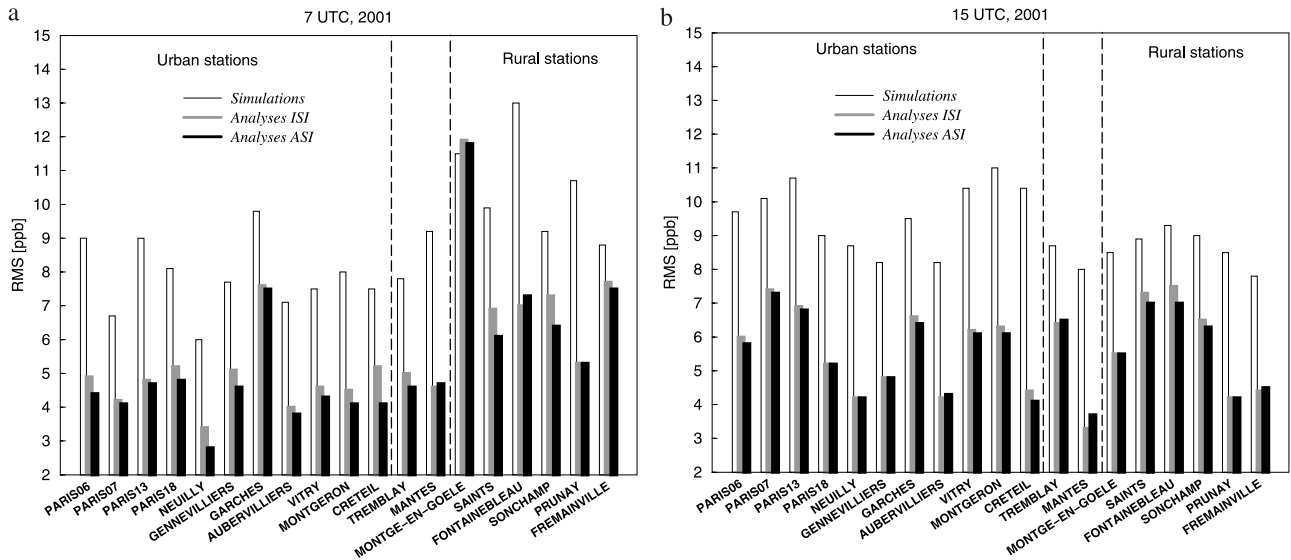


Figure 7. RMS innovations (open bars) and RMS analyses residuals on each observations sites (a) at 0700 UTC and (b) at 1500 UTC. The analyses are produced using ISI method (gray bars) or ASI method (solid black bars). Statistics are computed over summer 2001 using the leave-one-out method. Dashed vertical lines separate urban, suburban, and rural monitoring stations.

areas, while they are strongly correlated over rural area. In these cases, assumption about isotropy of the background error covariance fails. In contrast, during the afternoon, in strongly mixed air, statistics are more homogeneous between rural and urban stations and the two methods are equivalent.

5.1.2. Statistical Interpolation and Kriging

[77] In this section, we compare ASI method with the two kriging methods. Kriging is also applied every day of summer 2001 at each hour. In the present study, we only show results for 1500 UTC for comparison. In this way, we especially investigate the use of time varying weighting functions or the use of the CTM, regardless of the isotropic

or anisotropic assumption. The comparison ISI/ASI showed that the isotropic assumption does not generate major failures during afternoon peak ozone hours (see section 5.1.1).

[78] Figure 8a presents RMS analyses residuals over observations sites for 1500 UTC. It shows that ASI method is often more efficient than INK method. Indeed, in case of ASI method, a specific bias model is built while in INK method, bias is accounted for by the constraint on the sum of weights and is finally assumed to be uniform over the regional domain. In order to compare the methods besides the bias problem, we present in Figure 8b the standard deviations of analyses residuals over observations sites for

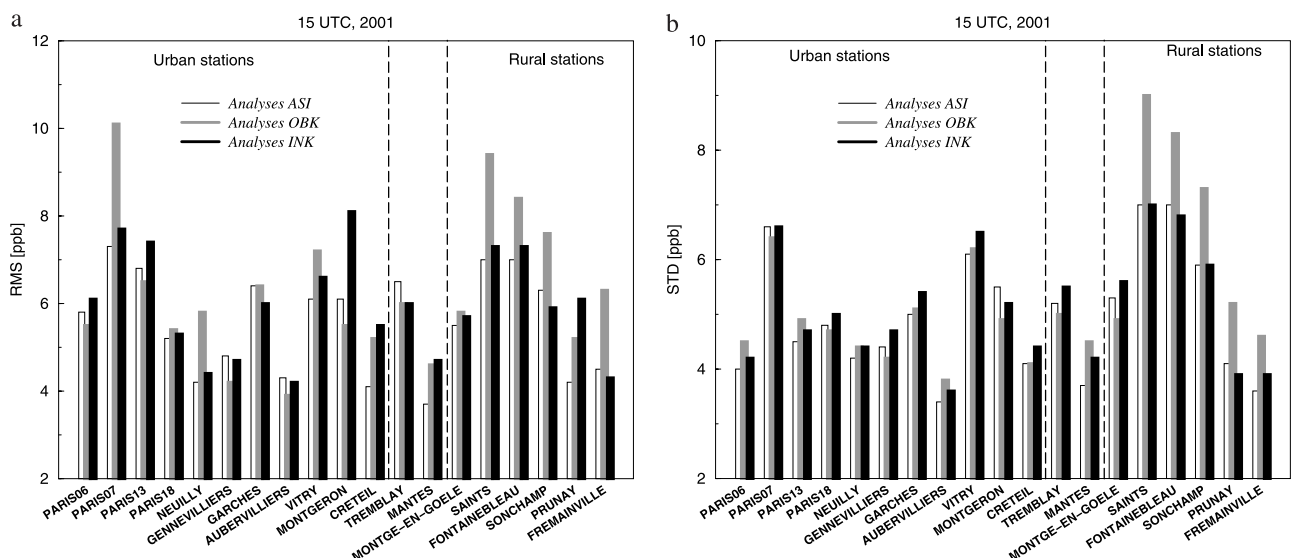


Figure 8. (a) RMS analyses residuals on each observations sites at 1500 UTC. (b) Same as Figure 8a for standard deviations. The analyses are produced using ASI method (open bars), OBK method (gray bars), and INK method (solid black bars). Statistics are computed over summer 2001 using the leave-one-out method. Dashed vertical lines separate urban, suburban, and rural monitoring stations.

the same hour, 1500 UTC. Contrary to RMS analyses residuals, the standard deviations are nearly the same for methods ASI and INK. These results show that besides the bias problem, methods ASI and INK are almost strictly equivalent at 1500 UTC.

[79] RMS analyses residuals (like standard deviations) results also show that OBK method is less accurate than the other methods during the afternoon in rural areas where the network is sparse. At 1500 UTC the RMS analyses residuals can be 2 ppb larger for OBK method in comparison with the other methods. We especially noticed that OBK method is less efficient during high-pollution events, when a strong ozone plume is developed. In these cases mean concentration values are often higher in rural area (outside the Paris city) than in urban area (inside Paris city). However, OBK method consists in looking for a true field $Z_{t,h}$ with a spatial constant mean. There are more many urban stations than rural stations. Therefore the kriging is mainly controlled by urban observations and the quality of analyses is better over the urban area. When the CTM is relatively correct as concerns plume direction, the innovations are more homogeneous than the observations. Finally, applying INK method to interpolate innovations is more efficient.

[80] When computed over the training year 1999, the statistics display similar behavior, demonstrating that the results are robust to inter-annual variability.

5.1.3. Case Studies

[81] In order to further illustrate the differences between the methods, we now show individual cases. Figure 9 shows the CTM simulation, three analyses resulting from methods ASI, OBK and INK, and the available observations for 17 July 1999 at 1500 UTC.

[82] The pattern predicted by the CTM is a strong ozone plume in the northwestern area with a “bubble” shape. The patterns analyzed by methods ASI and INK add small corrections to the model prediction. In contrast, OBK method gives a very different pattern with high ozone values spread all over the northwestern side of the domain. This latter structure is probably unrealistic since high values on the sides do not correspond to any significant precursor sources upstream. Nevertheless, the pattern matches quite closely the observations. On this day, another measurement was available in the city of Chartres, located in the domain’s southwest corner. The monitoring station belongs to another operational monitoring network, LIGAIR. The ozone value reached 77 ppb at 1500 UTC. This value matches better the ASI pattern than the OBK and the INK patterns. Indeed, simulated ozone concentrations at station SONCHAMP and station CHARTRES-FULBERT location are highly correlated (correlation coefficient: 0.99), which agree with the correlation coefficient of observed ozone concentrations (0.96). Thus increased ozone level on one station has the same effect onto the other one. ASI method is able to extend the plume more than the INK. However, for the same reason, ASI keeps the overestimation at MANTES, because ozone concentration on this station is correlated with stations GARCHES and PRUNAY ones.

[83] INK method has the advantage of time-varying weights. It is suited to atypical days when the CTM fails to represent the ozone concentration even in a qualitative manner and when errors, which are usually very correlated, become independent. Such days are rare but it happens

especially when the cloudiness is not well forecasted. It was the case of 24 May 2001. The meteorological conditions on that day were characterized by very light northeast winds, and stationary boundary-layer clouds developed in the afternoon only in the southern part of the domain. While the wind prediction was fairly correct, the ECMWF meteorological data strongly underestimated clouds. As a result, the CTM develops a strong city plume in the southwestern part of the domain, which is not observed (see Figure 10a). The observed ozone pattern seems even reversed with higher values in the North of the domain. Moreover, the CTM does not respect the ozone gradient observed over the center of Paris.

[84] For this case, the ASI and INK analyses (see Figures 10c and 10d) at 1500 UTC display large differences. The ASI has more difficulties in decreasing ozone level in the south of the domain and increasing ozone level over the eastern SAINTS station than the INK technique. These differences highlight the advantage of using INK for special days when large gradients occur between stations that are usually correlated. For instance, this is the case between SAINTS and FONTAINEBLEAU (the two east/southeast rural stations). Their average ozone concentration correlation coefficient is quite high (0.91).

[85] Finally, a comparison between Figure 10b and Figure 10d shows that analysis INK is very close to analysis OBK. In this case, INK method gives almost all weight to the observations and few to the first guess.

5.2. Validation of Upper Air ASI Analyses

[86] Long range transport of pollutants often implies vertical exchange between the boundary layer and the free troposphere. Therefore the evaluation of the vertical pollutant distribution is an important issue. However, the lack of routine data above ground does not permit to evaluate the CTM 3-D error field (bias, background error covariance). In the present section, we investigate the extension of ASI method to altitudes higher than the surface.

[87] Two assumptions are made. First, the background error covariance model, estimated using surface observations and simulations in the CTM first level, is used to model background error covariances between any locations above the surface up to the CTM top. As ozone concentrations in and out the boundary layer are less correlated than ozone concentrations in the boundary layer (almost homogeneous), the CTM error is assumed to follow the same behavior.

[88] Second, the CTM bias is assumed to be constant in the vertical direction. However, as we have no information about the bias value out of the boundary layer, its correction is only performed in that layer. In other words, information contained in surface observations is assumed to be relevant to higher altitude when the air is sufficiently mixed.

[89] These two assumptions can not be directly verified since there are no continuous monitoring except at the ground. However, the ESQUIF field campaign [Menut *et al.*, 2000] during summer 1999 provides a basis for a comparison between 3-D analyses and a number of airborne measurements.

[90] In these settings, 3-D ozone analyses are produced by ASI method using only the ground-based measurements (e.g., the airborne measurements are not used) but are compared in altitude to the ESQUIF airborne measurements.

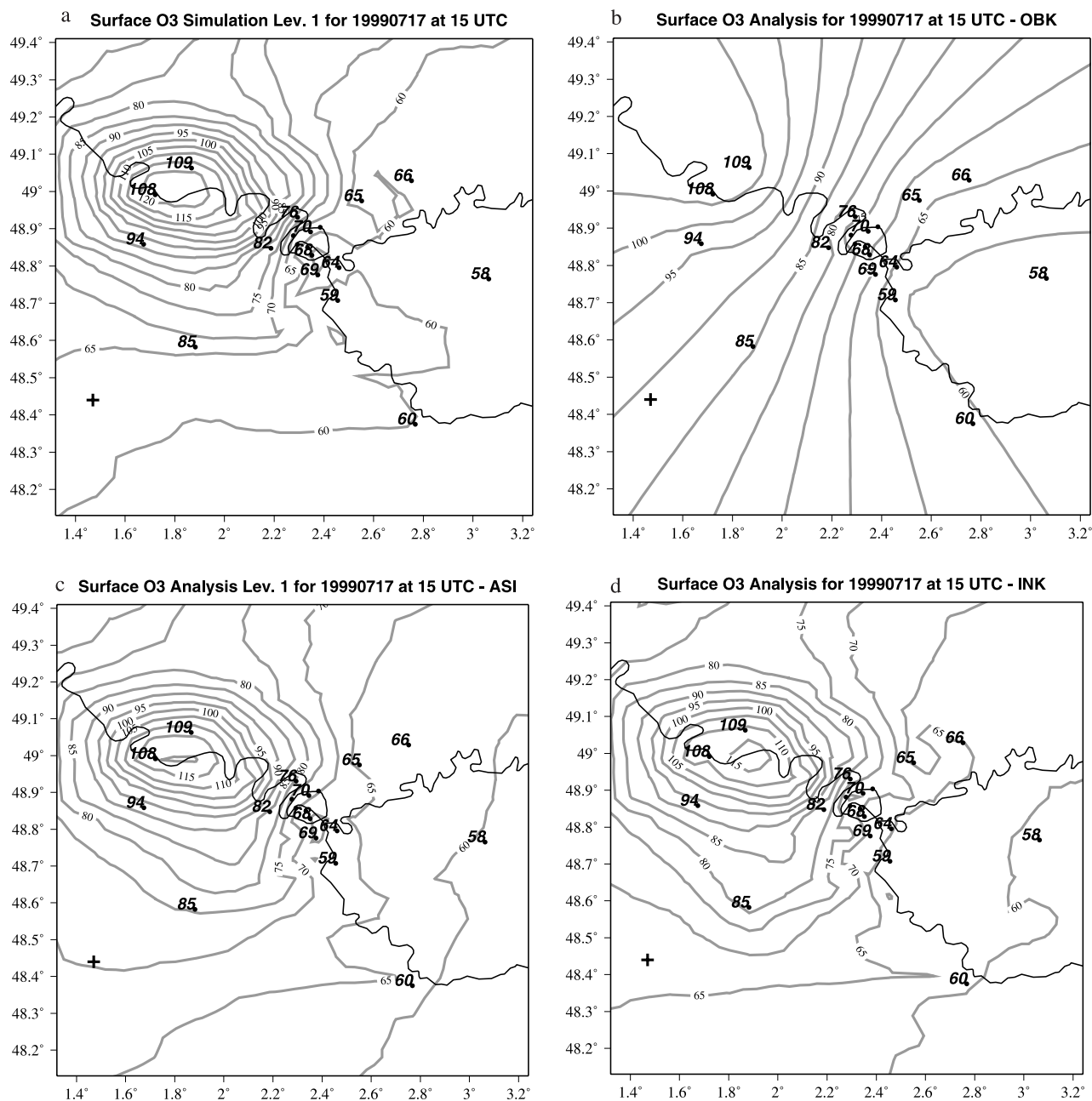


Figure 9. Comparison of some grounded observations (black numbers) with simulated or analyzed ozone mixing ratios (gray contours) in the first CTM level for 17 July 1999 at 1500 UTC. (a) CTM raw simulations, (b) analysis OBK, (c) analysis ASI, and (d) analysis INK. Ozone mixing ratio are in ppb. The plus sign denotes the monitoring station CHARTRES-FULBERT.

[91] Figure 11 shows results of such a comparison. The airborne measurements of ozone (12 flights over the region, 7 DIMONA flights and 5 ARAT flight covering 7 distinct IOP days) simulated, and analyzed values are displayed. Compared to the raw CTM simulations (Table 1 gives the CTM level for comparison), the RMS reduction is almost systematically observed. Only for a flight during the 17 July afternoon are the analyses slightly poorer than the simulation. This day was already well simulated by the CTM (see section 5.1.3) and surface measurements do not bring any significant additional information.

[92] As an example, Figure 12 shows a comparison of ARAT measurements with CTM ozone simulation and analyses for 30 July 1999 around 1500 UTC (IOP 8 of the ESQUIF campaign). The model strongly underestimates the whole ozone field, probably because of an underestimate in the large-scale simulation forcing the boundaries. The analysis using only surface measurements drastically reduces the underestimate and improves the CTM behavior in the urban plume (the simulations are not degraded if they are already closed to the observations). However, some small-scale structures still remain unresolved by the analyses involving only ground-based observations.

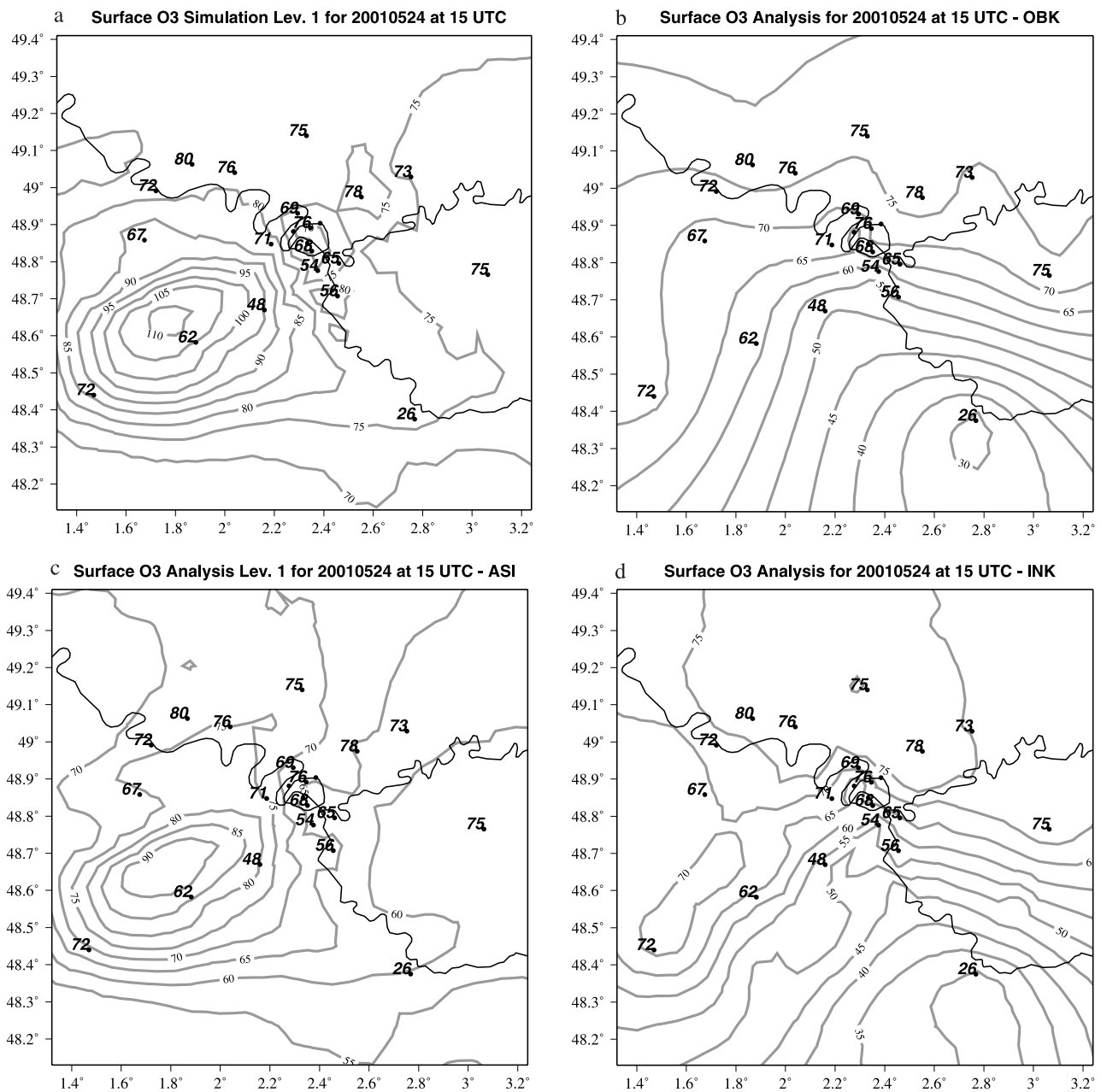


Figure 10. Comparison of some grounded observations (black numbers) with simulated or analyzed ozone mixing ratios (gray contours) in the first CTM level for 24 May 2001 at 1500 UTC. (a) CTM raw simulations, (b) analysis OBK, (c) analysis ASI, and (d) analysis INK. Ozone mixing ratio are in ppb.

[93] These results are important for future forecast experiments. By such a correction in the vertical direction at 1500 UTC (when the boundary layer is thickest), we expect to correct a part of the upper level ozone, retained in the residual layer during the night and introduced again in the convective boundary layer during its development the day after. We expect that an initialization of the model forecast with analyses will have a positive impact on short-time ozone forecasts.

6. Summary and Discussion

[94] In this study, we presented a discussion and comparison of four analysis methods which are designed to

produce 2-D or 3-D ozone fields as close as possible to real state. Two of the methods are adapted from meteorology where routine analyses are necessary for daily numerical weather forecasting. A classical isotropic statistical interpolation (ISI) is proposed, which uses the assumption of isotropy of error statistics. Since this assumption may fail in the case of strongly heterogeneous terrain or emission patterns, we propose a new anisotropic method (ASI). These two methods actually aim at correcting the ozone concentration fields simulated by a chemistry-transport model (CTM) with the routine observations of the regional air quality monitoring network AIRPARIF.

[95] The two other methods are kriging techniques generally used for spatial interpolation. They directly interpo-

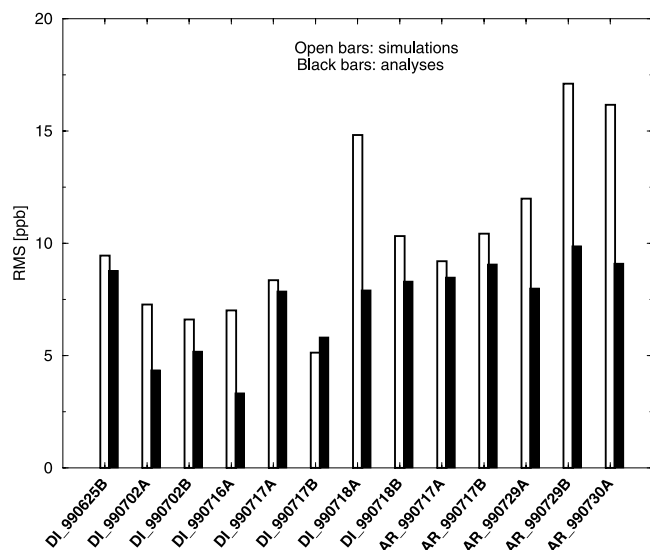


Figure 11. RMS of differences between aircraft measurements of ozone and simulated (open bars) or analyzed (solid bars) values. Results given for 12 flights around Paris. Analyses have been produced using ASI method. DI corresponds to the DIMONA aircraft, and AR corresponds to the ARAT aircraft. A stands for morning flights, and B stands for flights in the afternoon. DI_990716A corresponds to DIMONA aircraft measurements of 16 July 1999 in the morning.

late surface observations (OBK method) or the differences between the CTM concentrations and the observations (INK method). In the first case, the CTM information is not used while in the second case, kriging is applied as a correction method.

[96] Each method relies on strong assumptions about error statistics and about the concentration field itself. ISI, OBK, and INK methods assume isotropy of statistical properties. ISI and ASI are calculated from statistics over a sufficient number of CTM simulation days and corresponding observations. They provide an interpolator whose parameters do not vary with time but depend only on station distribution. In contrast, OBK and INK methods are based on spatial statistics. They presumably are more suitable to represent day to day variations in error statistics. OBK is the only method which does not assume a first-guess field given by the CTM. Finally, ASI is the only method which can provide corrections in three dimensions rather than in two dimensions. The efficiency of these 3-D corrections is assessed using the airborne measurements during the ESQUIF field campaign.

[97] The intercomparison of the methods is performed with a leave-one-out verification technique. It shows that on average the most accurate methods are ASI and INK. Therefore the additional information provided by the CTM simulation is of particular value, especially in the sparsest parts of the monitoring network. Comparison between ISI and ASI shows that the isotropic assumption does not seem to generate major failures during peak ozone hours in the afternoon. However, this assumption provides systematically poorer interpolations than those

calculated using ASI, for data taken during night and emissions times.

[98] Above all, the mean reduction of the RMS error relative to the CTM simulations is large for all methods. The leave-one-out method actually shows a reduction of the order of 30–50% at the surface. These results are obtained not only over the training period (summer 1999) but also over the validation period (summer 2001). The improvement of upper air ASI analyses relative to the raw simulations is shown to be of the same order of magnitude. It corresponds not only to the reduction of biases but also to a reduction of random errors. This is an important result for a model designed to forecast pollution episodes. We also demonstrate that the INK method can give more accurate ozone maps than ASI in exceptional cases. This happens when the CTM fails to represent the ozone concentration even in a qualitative manner, and when the resulting errors which are usually correlated, become independent. In these cases, the advantage of INK is due to use of spatial statistics rather than time statistics.

[99] There are various types of applications of these analysis techniques. The first one is the routine production of maps at the surface level. This is a strong requirement for pollution monitoring especially away from monitoring station locations. Since the INK method seems more efficient in extreme cases and has otherwise skills equivalent to that of ASI, its use is recommended for routine analyses production. However, the relative skills of the different methods are likely to change depending on the region. Indeed, our comparisons could yield different results in the presence of strong terrain/emission heterogeneities such as coasts or mountains. The assumptions of homogeneity and isotropy made in INK method (not in ASI method) may fail.

[100] For the purpose of initializing CTMs for air-quality forecasts, the three-dimensional character of ASI seems more appropriate. However, analyses performed for a large scale

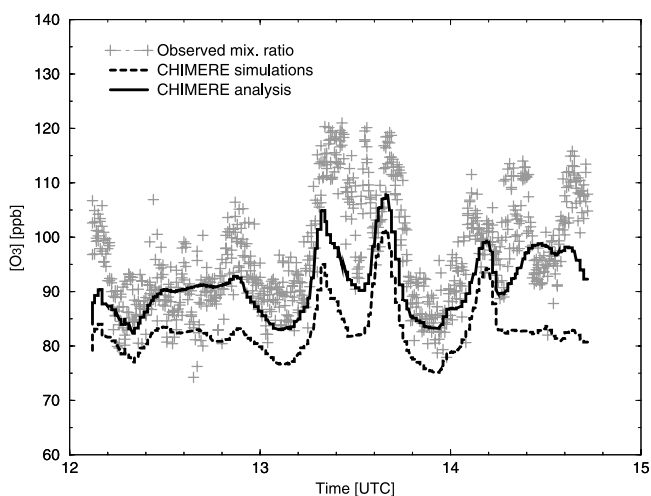


Figure 12. Comparison of ARAT aircraft measurements of ozone with simulated (dashed black line) and analyzed values (solid black line; produced using ASI method) of ozone concentrations at 1400 UTC calculated by CHIMERE for 30 July 1999 in the third level.

(e.g., European scale) should be more efficient for 1–2 day forecasts. Since, the wind may transport air masses over several tens or hundreds of kilometers, the relevant observational information does not necessarily lie within the regional domain. A series of experimental forecasts have been performed in real-time using the continental version of the CTM during summer 2001. Some ozone analyses has been also produced. The results will be presented in a future article.

[101] Future investigation will extend this study to include analyses of other pollutants such as nitrogen oxides or particle matter. Preliminary experiments performed on nitrogen dioxide, routinely monitored in the AIRPARIF network, show promising results.

[102] **Acknowledgments.** The authors would like to thank H. Schmidt, LMD, Palaiseau, France, for many helpful discussions. Thanks are given to Olivier Talagrand, LMD-ENS, Paris, for fruitful advice and his interest in this work. We thank the following institutions gratefully: Météo-France and the European Centre for Medium-Range Weather Forecasts (ECMWF) for giving us free access to the meteorological database and French Air quality network AIRPARIF for the free access to surface ozone data. The aircraft teams from the Institut National des Sciences de l'Univers (INSU) and Swiss agency METAIR are acknowledged for efforts concerning the observations with ARAT and DIMONA aircrafts, respectively. Financial support for the study was provided by the Agence de l'Environnement et De Maîtrise de l'Énergie (ADEME) and the TOTAL-FINA-ELF company. Finally, thanks to the reviewers of this paper, who made several useful suggestions to clarify an earlier version.

References

- Balgovind, R., A. Dalcher, M. Ghil, and E. Kalnay, A stochastic-dynamic model for the spatial structure of the forecast error statistics, *Mon. Weather Rev.*, **111**, 701–722, 1983.
- Carpenter, R. L., Jr., K. K. Droegemeier, P. R. Woodward, and C. E. Hane, Application of the piecewise parabolic method (PPM) to meteorological modeling, *Mon. Weather Rev.*, **118**, 586–612, 1990.
- Chang, M. E., D. E. Hartley, C. Cardelino, and W. L. Chang, Inverse modeling of biogenic isoprene emissions, *Geophys. Res. Lett.*, **23**, 3007–3010, 1996.
- Chang, M. E., D. E. Hartley, C. Cardelino, D. Haas-Laursen, and W. L. Chang, On using inverse methods for resolving emissions with large spatial inhomogeneities, *J. Geophys. Res.*, **102**, 16,023–16,036, 1997.
- Cressie, N. A. C., *Statistics for Spatial Data*, revised ed., John Wiley, Hoboken, N. J., 1993.
- Daley, R., *Atmospheric Data Analysis*, Cambridge Univ. Press, New York, 1991.
- Derognat, C., M. Beekmann, M. Baeumle, D. Martin, and H. Schmidt, Effect of biogenic volatile organic compound emissions on tropospheric chemistry during the Atmospheric Pollution Over the Paris Area (ESQUIF) campaign in the Ile-de-France region, *J. Geophys. Res.*, **108**(D17), 8560, doi:10.1029/2001JD001421, 2003.
- Desroziers, G., A coordinate change for data assimilation in spherical geometry of frontal structures, *Mon. Weather Rev.*, **125**, 582–600, 1997.
- Elbern, H., and H. Schmidt, Ozone episode analysis by four-dimensional variational chemistry data assimilation, *J. Geophys. Res.*, **106**, 3569–3590, 2001.
- Elbern, H., and H. Schmidt, The development of a 4D-variational chemistry data assimilation scheme for initial value and emission rate estimates, paper presented at GLOREAM Symposium, Fundação para a Ciênc. e a Tecnol., Aveiro, Portugal, Sept. 2002.
- Elbern, H., H. Schmidt, O. Talagrand, and A. Ebel, 4D-variational data assimilation with an adjoint air quality model for emission analysis, *Environ. Modell. Software*, **15**, 539–548, 2000.
- Erismann, J. W., W. van Pul, and G. Wyers, Parameterization of surface resistance for the quantification of atmospheric deposition of acidifying pollutants and ozone, *Atmos. Environ.*, **28**, 2595–2607, 1994.
- Fedorov, V., Kriging and other estimators of spatial field characteristics (with special reference to environmental studies), *Atmos. Environ.*, **23**, 174–184, 1989.
- Finkelstein, P., The spatial analysis of acid precipitation data, *J. Clim. Appl. Meteorol.*, **23**, 52–62, 1984.
- Fuentes, M., A high frequency kriging approach for non-stationary environmental processes, *Environmetrics*, **12**, 469–483, 2001.
- Fuentes, M., and A. Raftery, Model validation and spatial interpolation by combining observations with outputs from numerical models via Bayesian melding, *Tech. Rep. 403*, Dep. of Stat., Univ. of Wash., Seattle, 2001.
- Gaspari, G., and S. E. Cohn, Construction of correlation functions in two and three dimensions, *Q. J. R. Meteorol. Soc.*, **125**, 723–757, 1999.
- Generation of European Emission Data for Episodes (GENEMIS) Project, EUROTRAC annual report 1993, part 5, EUROTRAC Int. Sci. Secr., Garmisch-Partenkirchen, Germany, 1994.
- Ghil, M., and P. Malanotte-Rizzoli, Data assimilation in meteorology and oceanography, *Adv. Geophys.*, **33**, 141–266, 1991.
- Gilliland, A., and P. J. Abbitt, A sensitivity study of the discrete Kalman Filter (DKF) to initial condition discrepancies, *J. Geophys. Res.*, **106**, 17,939–17,952, 2001.
- Hauglustaine, D. A., G. P. Brasseur, S. Walters, P. J. Rasch, J.-F. Müller, L. K. Emmons, and M. A. Carroll, MOZART: A global chemical transport model for ozone and related chemical tracers: 2. Model results and evaluation, *J. Geophys. Res.*, **103**, 28,291–28,336, 1998.
- Hoelzemann, J., H. Elbern, and A. Ebel, PSAS and 4D-var data assimilation for chemical state analysis by urban and rural observation sites, *Phys. Chem. Earth, Part B: Hydrol. Oceans Atmos.*, **26**, 807–812, 2001.
- Hollingsworth, A., and Lönnberg, The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, *Tellus, Part A*, **38**, 111–136, 1986.
- Ionescu, A., Y. Candau, E. Mayer, and I. Colda, Analytical determination and classification of pollutant concentration fields using air pollution monitoring network data: Methodology and application in the Paris area, during episodes with peak nitrogen dioxide levels, *Environ. Modell. Software*, **15**, 565–573, 2000.
- Jeuken, A. B. M., H. J. Eskes, P. F. J. van Velthoven, H. M. Kelder, and E. V. Holm, Assimilation of total ozone satellite measurements in a three-dimensional tracer transport model, *J. Geophys. Res.*, **104**, 5551–5563, 1999.
- Kalman, R. E., A new approach to linear filtering and prediction problems, *Trans. ASME, Ser. D, J. Basic Eng.*, **83**, 35–45, 1960.
- Kalman, R. E., and R. S. Bucy, New results in linear filtering and prediction theory, *J. Basic Eng.*, **83**, 95–108, 1961.
- Lattuat, M., 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 6, Paris, 1997.
- Lewis, J. M., and J. C. Derber, The use of adjoint equations to solve a variational adjustment problem with advective constraints, *Tellus, Part A*, **37**, 309–322, 1985.
- Mendoza-Dominguez, A., and A. G. Russell, Estimation of emission adjustments from the application of four-dimensional data assimilation to photochemical air quality modeling, *Atmos. Environ.*, **35**, 2879–2894, 2001.
- Menut, L., et al., Measurements and modeling of atmospheric pollution over the Paris area: An overview of the ESQUIF project, *Ann. Geophys.*, **18**, 1467–1481, 2000.
- Mylona, S., EMEP emission data: Status report 1999, *EMEP/MSC-W Note 1/99*, Res. Note 26, Norw. Meteorol. Inst., Oslo, 1999.
- O'Brien, J. J., A note on the vertical structure of the eddy exchange coefficient in the planetary boundary layer, *J. Atmos. Sci.*, **27**, 1213–1215, 1970.
- Riishøjgaard, L. P., A direct way of specifying flow-dependent background error correlations for meteorological analysis systems, *Tellus, Ser. A*, **50**, 42–57, 1998.
- Schmidt, H., Vierdimensionale Datenassimilation nach der Variationsmethode für ein mesoskaliges Chemie transport modell, *Mitt. Inst. Geophys. Meteorol. Univ. Köln 132*, Cologne, Germany, 1999.
- Schmidt, H., C. Derognat, R. Vautard, and M. Beekmann, A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in western Europe, *Atmos. Environ.*, **35**, 6277–6297, 2001.
- Seinfeld, J. H., and S. N. Pandis, *Atmospheric Chemistry and Physics*, John Wiley, Hoboken, N. J., 1998.
- Talagrand, O., and P. Courtier, Variational assimilation of meteorological observations with the adjoint vorticity equation, I: Theory, *Q. J. R. Meteorol. Soc.*, **113**, 1311–1328, 1987.
- van de Velde, R., W. Faber, V. Katwijk, H. J. Scholten, T. J. M. Thewessen, M. Verspuj, and M. Zevenbergen, The preparation of a European land-use data base, *RIVM Rep. 712401001*, Natl. Inst. of Public Health and the Environ., Bilthoven, Netherlands, 1994.
- van Loon, M., and A. W. Heemink, Kalman filtering for non-linear atmospheric chemistry models: First experiences, *Tech. Rep. MAS-R9711*, Cent. voor Wisk. en Inf., Amsterdam, March 1997.

- van Loon, M., P. J. H. Builtjes, and A. J. Segers, Data assimilation of ozone in the atmospheric transport chemistry model LOTOS, *Environ. Modell. Software*, 15, 603–609, 2000.
- Vautard, R., M. Beekmann, J. Roux, and D. Gombert, Validation of a hybrid forecasting system for the ozone concentrations over the Paris area, *Atmos. Environ.*, 35, 2449–2461, 2001.
- Vautard, R., et al., A synthesis of the Air Pollution Over the Paris Region (ESQUIF) field campaign, *J. Geophys. Res.*, 108(D17), 8558, doi:10.1029/2003JD003380, 2003a.
- Vautard, R., et al., Paris emission inventory diagnostics from ESQUIF airborne measurements and a chemistry transport model, *J. Geophys. Res.*, 108(D17), 8564, doi:10.1029/2002JD002797, 2003b.
- Verwer, J. G., Gauss-Seidel iteration for stiff ODEs from chemical kinetics, *SIAM J. Sci. Comput.*, 15, 1243–1250, 1994.
- Wesely, M. L., and B. B. Hicks, Some factors that affect the deposition rates of sulfur dioxide and similar gases on vegetation, *J. Air Pollut. Control Assoc.*, 27, 1110–1116, 1977.

L. Bel, Probabilités, Statistique et Modélisation, Université Paris-Sud, Batelle 425, F-91405 Orsay Cedex, France. (liliane.bel@math.u-psud.fr)

N. Blond, Laboratoire Interuniversitaire des Systèmes Atmosphériques, 61 Avenue du Général de Gaulle, UMR 7583-CNRS, F-94010 Créteil Cedex, France. (blond@lisa.univ-paris12.fr)

R. Vautard, Laboratoire de Météorologie Dynamique, CNRS, École Polytechnique, F-91128 Palaiseau Cedex, France. (vautard@lmd.polytechnique.fr)