HAL
open science

Three-dimensional ozone analyses and their use for short-term ozone forecasts

N. Blond, Robert Vautard

To cite this version:

N. Blond, Robert Vautard. Three-dimensional ozone analyses and their use for short-term ozone forecasts. Journal of Geophysical Research, 2004, 109 (D17), 10.1029/2004JD004515 . hal-02414785

HAL Id: hal-02414785

https://hal.science/hal-02414785v1

Submitted on 1 Feb 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 analyses and their use for short-term ozone forecasts

N. Blond[1] and R. Vautard

Laboratoire de Météorologie Dynamique, École Polytechnique, Palaiseau, France

Received 6 January 2004; revised 9 March 2004; accepted 17 May 2004; published 3 September 2004.

[1] A statistical interpolation method is evaluated for routine production of ozone three-dimensional fields over western Europe. These fields are used for initializing short-term ozone forecasts issued from a chemistry-transport model. We mainly address two questions: (1) To what extent can the use of surface ozone observation data improve the description of ozone fields relative to raw simulations? (2) Does the use of ozone analysis improve short-term forecasts of the troposphere's chemical composition? The method consists of combining ozone simulations with surface ozone measurements. The resulting analyses are compared with independent observations in a statistical way over a long period of time (four consecutive summers). The improvement of the root-mean-square (RMS) error of the analyses relative to the raw simulations is ~30%. The short-term (1–2 days in advance) ozone forecasts are improved on average (by ~1 ppb of RMS error) if ozone analyses are used for initialization. The improvement is almost lost after a lead time of 36 hours. However, in cases where a model error propagates throughout the model domain, the improvement can be much larger (~10 ppb). We analyze one such case. *INDEX TERMS:* 0345 Atmospheric Composition and Structure: Pollution—urban and regional (0305); 0365 Atmospheric Composition and Structure: Troposphere—composition and chemistry; 0368 Atmospheric Composition and Structure: Troposphere—constituent transport and chemistry; 3337 Meteorology and Atmospheric Dynamics: Numerical modeling and data assimilation; *KEYWORDS:* air pollution, ozone forecast, data assimilation

Citation: 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.

1. Introduction

[2] Air quality forecasting is a challenging scientific problem that has recently, in the last 20 years, received a high priority in many industrialized countries because of the increasing consciousness of the effect of pollutant emissions on health and the environment. Several chemistry-transport models (CTMs) have been developed in order to better understand the laws driving the chemistry and transport of air pollutants and to forecast exceedances of their information and alert thresholds. Such models cover spatial scales from regional (e.g., for Los Angeles, *Lu et al.* [1997a, 1997b]; for Atlanta, *Cardelino et al.* [2001]; for Milan, *Silibello et al.* [1998]; and for Paris, *Vautard et al.* [2001]) to continental (for Europe, *Schmidt et al.* [2001], *Hass et al.* [1997], and *Mosca et al.* [1998]). They require an array of variables as input data (emissions, initial and boundary conditions, meteorology...) that are in most cases quite uncertain and, in some, difficult to collect in real time. As a consequence, comparisons between forecasts and observations (routine surface and a few airborne measurements) still show important discrepancies for ozone and for its precursors. Recently, five Eulerian air pollution forecasting systems running over Europe have been compared by *Tilmes et al.* [2002].

[3] Since the work of *Lorenz* [1963], we know that the meteorological system is chaotic. This means that errors resulting from initial field uncertainties grow with time and eventually dominate the forecast error after a few days. Therefore the implementation of numerical weather prediction (NWP) requires the specification of accurate initial conditions. The problem of determining best initial fields has become of great practical importance in meteorology and has been the subject of many studies for ~40 years [*Daley*, 1991, 1997]. It has been shown that these weather forecast errors can be greatly reduced using data assimilation algorithms. These algorithms consist of combining outputs of a numerical model with some observation data by taking advantage of consistency constraints with laws of time evolution and physical properties. The results of data assimilation are called "analyses."

[4] The routine production of analyzed meteorological fields was first performed using a series of statistical linear interpolations (references are given by *Talagrand* [1997]). During the last decade, advanced data assimilation schemes have been proposed. For instance, the four-dimensional

[1]Now at Laboratoire Interuniversitaire des Systèmes Atmosphériques, Universités Paris 7–Denis Diderot et Paris 12–Val de Marne, Créteil, France.

Copyright 2004 by the American Geophysical Union.
0148-0227/04/2004JD004515

variational (4D-VAR) assimilation scheme is described by *Lewis and Derber* [1985] and *Talagrand and Courtier* [1987]. In this scheme, initial conditions of the model forecast are adjusted to optimize predicted output fields relative to observations over a period of time. Other advanced techniques based on the Kalman-Levy filter [*Kalman*, 1960; *Kalman and Bucy*, 1961] have also been developed [*Ghil and Malanotte-Rizzoli*, 1991; *Evensen*, 1994].

[5] As the photochemical system highly depends on meteorological conditions, any forecast model is limited by the quality of the meteorological inputs. However, until now, no study has shown that the photochemical system itself is also chaotic. The chaotic behavior is discussed by *Kleinman et al.* [1997], *Stewart* [1993, 1995], *White and Dietz* [1984], *Krol* [1995], *Honoré et al.* [2000], and *Honoré* [2000]. In practice, as far as photochemistry is concerned, CTMs are not as sensitive to the initial conditions as weather forecast models. On the basis of our experience with the model that will be used in this article, over a continental area like western Europe, two simulations initialized differently eventually converge after 3–5 days. The continental photochemical system seems more sensitive to errors in surface emissions [*Menut*, 2003; *Schmidt and Martin*, 2003]. Nevertheless, errors in the initial state can propagate from one day to another. The use of analyses for initializing air quality forecasts is therefore expected to improve these forecasts. The present article intends to quantify this improvement.

[6] The advanced data assimilation techniques referenced above have been recently applied in air quality modeling. They have been demonstrated in a few cases to be efficient for optimizing model inputs such as initial fields [*Elbern and Schmidt*, 2001] and 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 the 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 as well as for validation over a long period of time.

[7] The analysis technique used in this study is based on statistical interpolation [*Daley*, 1991]. Surface observations are combined with the outputs of the CHIMERE chemistry-transport model [*Schmidt et al.*, 2001]. The method, called anisotropic statistical interpolation (ASI), does not produce fields as physically consistent as four-dimensional variational assimilation. However, it is much lighter in terms of computational cost and implementation for operational use. It also allows statistical validation. The ASI method has been initially used for representing as accurately as possible the ozone fields around the Paris area [*Blond et al.*, 2003]. We proposed a new modeling of an anisotropic error covariance model that allows for better interpolations than those produced using a classical isotropic error covariance model (especially for data taken during nighttime and emission times). Moreover, we showed that the root-mean-square (RMS) error of the ozone analyses at the surface was 30–50% smaller than that of CHIMERE simulations. The Etude et Simulation de la Qualité de l'Air en Île de France (ESQUIF) [*Menut et al.*, 2000] airborne measurements have also been used to prove that the ASI method efficiently corrects CTM fields well above the ground within the boundary layer using only surface measurements. The approach of the present article is actually twofold: First, we quantify the quality of analyses obtained by the same method but at a much larger scale, over Europe, and second, we quantify the potential improvement of ozone forecasts over Europe using analyses as initial conditions.

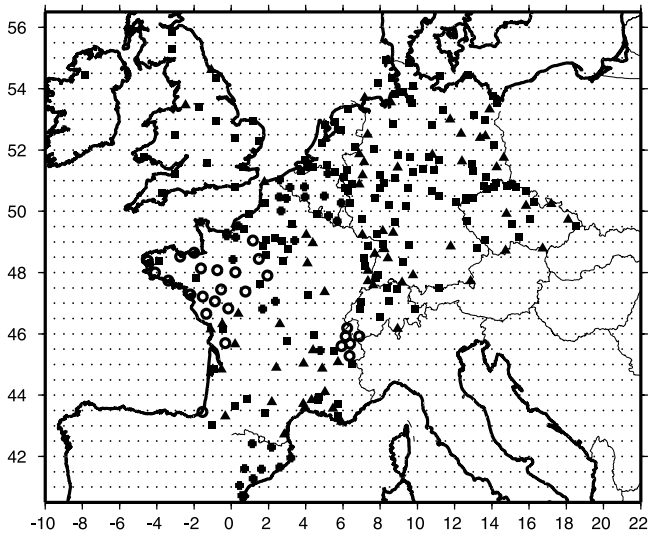

Figure 1. Continental CHIMERE CTM extension and locations of 232 ozone monitoring stations where measurements were available in summer 2002. Open circles, solid squares, and solid triangles denote urban, rural, and suburban surface monitoring stations, respectively. Crosses show monitoring stations whose characteristics are unknown.

[8] The paper is organized into six sections. In section 2, the CHIMERE model is briefly described. Section 3 describes the surface ozone observations. The mathematical formulation of the ASI method is presented in section 4. The evaluation of the ozone analyses and of the simulation improvements is discussed in section 5. We summarize and conclude the work in section 6.

2. Chemistry-Transport Model

[9] CHIMERE is a three-dimensional (3-D) Eulerian chemistry-transport model. It has been used over several scales, from the urban scale [*Vautard et al.*, 2001; *Menut*, 2003; *Beekmann and Derognat*, 2003; *Vautard et al.*, 2003a, 2003b] to the regional and continental scales [*Schmidt et al.*, 2001; *Schmidt and Martin*, 2003]. In this article, we use a continental configuration over western Europe with a horizontal resolution of $0.5° \times 0.5°$. The domain and grid are represented in Figure 1.

[10] The model formulation is based on the mass continuity equation for several chemical species in every grid cell. The time numerical solver is the TWOSTEP method [*Verwer*, 1994], which is applied to integrate all processes, including transport, chemistry, emission, and diffusion as proposed by *Schmidt et al.* [2001]. The time step is 10 min.

Vertical discretization consists of six layers going from the surface up to 700 hPa with a first layer fixed at 50-m high. Therefore the model contains the whole boundary layer in anticyclonic conditions over western Europe. The emissions are derived from the European Monitoring and Evaluation Programme (EMEP) [*Mylona*, 1999] annual totals, modulated in time and volatile organic compound (VOC) speciation by *Generation of European Emission Data for Episodes Project* (*GENEMIS*) [1994] profiles. The biogenic emissions database is described by *Derognat et al.* [2003]. The land use data are derived from the National Institute of Public Health and the Environment (RIVM) database [*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. Photolytic rates are modulated by cloudiness [*Vautard*, 2001]. Boundary concentrations for the continental setup are fixed for 14 species. These species have a long lifetime and are relevant to photo-oxidant formation. Their concentrations are taken from a climatology of monthly mean data produced by the global Model for Ozone and Related Chemical Tracers (MOZART) CTM [*Horowitz et al.*, 2003]. More details about the CTM are given by *Schmidt et al.* [2001] and *Vautard et al.* [2003a, 2003b] and by the Laboratoire de Météorologie Dynamique on the Web site http://euler.lmd.polytechnique.fr/chimere.

[11] The CTM is forced by the European Centre for Medium-Range Weather Forecasts (ECMWF) meteorological short-range forecasts. These data have a time resolution of 3 hours and are linearly interpolated to 1-hour intervals. They include temperature, pressure, wind, humidity, and cloudiness fields. Their horizontal resolution is 0.25° × 0.25°. Vertically, the model uses 10 layers following hybrid sigma-pressure levels for summer 1999 and 18 levels for summers 2000, 2001, and 2002.

[12] For the purpose of this article, we performed long-term simulations and ozone analyses over all of the summers from 1999 to 2002 from May to September. In the process of validation of analyzed ozone fields, the statistical error covariance model is built from testing data over 1999 and used for verification over the other years. The series of initial-value experiments have been performed over the year 2002. In this case, we evaluate the improvement of forecasts initialized with analyses relative to raw forecasts.

3. Observations

[13] Measurements of ground-level ozone are collected from several European air quality monitoring agencies. These agencies are listed in Table 1. The distribution of monitoring sites is shown in Figure 1. 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.

[14] From all available sites, 232 European ones were selected for the analysis experiment. In order to have stations that are representative for ozone of an area of the size of a grid cell, our choice is as follows. First, all available rural sites are retained. Second, where possible, we fill large areas with no observations with suburban stations, and if not, urban stations located in small or medium-size towns. This work has been done with the assistance of experienced agency staff. In this way, we avoid the largest representativeness biases. More information about these biases is given by *Schmidt et al.* [2001]. We assume in the following that observations are unbiased from the instrumental and representativeness points of view. However, we also assume that the used observations may contain random errors or noise.

4. Anisotropic Statistical Interpolation (ASI Method)

[15] We now briefly recall the anisotropic statistical interpolation method that is used here to produce three-dimensional ozone analyses over western Europe. This method has been previously used to produce ozone analyses over the Paris area using a regional version of CHIMERE and some surface ozone observations. It has been compared to the classical isotropic statistical interpolation and to the kriging method applied on observations alone. Therefore the reader is referred to *Blond et al.* [2003] and *Blond* [2002] for further details and discussion about the method and for references to other existing skilled analysis algorithms (e.g., kriging [see also *Wackernagel et al.*, 2004], thin plate splines [*Ionescu et al.*, 2000], and the Bayesian melding approach [*Fuentes and Raftery*, 2001]).

[16] The aim is to provide an estimate $Z^a_{t,h}(\mathbf{s})$ of a field concentration value $Z_{t,h}(\mathbf{s})$ at any location $\mathbf{s}$, on day t ($t = 1, 2, \ldots, T$) and at hour h ($h = 1, 2, \ldots, 24$). For this purpose, we have a set of K spatially distributed measurement values $Y^o_{t,h}(\mathbf{s_k})$, where $\mathbf{k} = 1, 2, \ldots, K$ and $\mathbf{s_k}$ is the location of the kth monitoring station. We also use a prior estimate of the concentration field $Z^b_{t,h}(\mathbf{s})$, which is often called the "first-guess" or the "background" field. In this setting, $Z^a_{t,h}(\mathbf{s})$, which is called the "analysis," is given by

$$Z^a_{t,h}(\mathbf{s}) = Z^b_{t,h}(\mathbf{s}) + \sum_{k=1}^{K} \lambda^k_{t,h}(\mathbf{s}) \left[Y^o_{t,h}(\mathbf{s_k}) - Z^b_{t,h}(\mathbf{s_k}) \right]. \quad (1)$$

[17] The $\lambda^k_{t,h}(\mathbf{s})$ are weighting functions that have to be determined. The background fields $\mathbf{Z^b_{t,h}}$ are the hourly ozone fields simulated by the CTM over western Europe in the lowest layer.

[18] In order to obtain an unbiased estimated field, we need first to remove biases from the first guess and the observations before applying the analysis algorithm. The observation biases are assumed to be zero (see section 3). The first-guess bias is modeled as given by *Blond et al.* [2003] using a linear regression between bias and CTM ozone concentration average, separately for each hour h of the day. The bias is also estimated for each $\mathbf{s}$.

[19] The optimized weighting functions $\widehat{\lambda}^k_{t,h}(\mathbf{s})$ are then obtained by minimizing the mean square analysis error, e.g., $E[Z^a_{t,h}(\mathbf{s}) - Z_{t,h}(\mathbf{s})]^2$. $E[\ ]$ denotes the expectation calculated here as the time average. The minimization problem leads to a linear system involving (1) covariances of the background error $c^b_{t,h}(\mathbf{s_k}, \mathbf{s_l})$, (2) covariances of the observation error $c^o_{t,h}(\mathbf{s_k}, \mathbf{s_l})$, 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

Table 1. List of European Organizations Involved in the PIONEER Project

Country	Organization[a]	Location
France	AASQA des Bouches-du-Rhône, du Var et du Vaucluse (AIRMARAIX)	Provence-Alpes-Côte d'azur
	AASQA en Île-de-France (AIRPARIF)	Île-de-France
	AASQA en Languedoc-Roussillon (AIR LANGUEDOC-ROUSSILLON)	Languedoc Roussillon
	AASQA en Poitou-Charentes (ATMO-POITOU-CHARENTES)	Poitou-Charentes
	AASQA en Bourgogne (ATMOSFAIR)	Bourgogne
	AASQA de l'Etang de Berre et l'Ouest Bouches-du-Rhône (AIRFOBEP)	Provence-Alpes-Côte d'azur
	AASQA en Auvergne (ATMOAUVERGNE)	Auvergne
	AASQA de la région Centre (LIGAIR)	Région Centre
	AASQA des Pays de la Loire (AIR PAYS DE LA LOIRE)	Pays de La Loire
	AASQA en Picardie (ATMO-PICARDIE)	Picardie
	AASQA en Champagne-Ardennes (ATMO-CHAMPAGNE-ARDENNES)	Champagne-Ardenne
	Comité pour le contrôle de la pollution atmosphérique dans le Rhône et la région lyonnaise (COPARLY)	Rhône-Alpes
	AASQA sur les départements de Savoie, de Haute-Savoie et de l'Ain (AirDes2Savoies)	Rhône-Alpes
	Association pour le contrôle et la préservation de l'air dans la région grenobloise (ASCOPARG)	Rhône-Alpes
	Association de mesure de la pollution atmosphérique sur Saint Etienne et le Département de la Loire (AMPASEL)	Rhône-Alpes
	AASQA dans la Drôme et l'Ardèche (ASQUADRA)	Rhône-Alpes
	Surveillance de la pollution de l'air de Roussillon et ses environs (SUPAIRE)	Rhône-Alpes
France	AASQA en Bretagne (AIRBREIZH)	Bretagne
	AASQA en Aquitaine (AIRAQ)	Aquitaine
	Observatoire de la qualité de l'air en Haute Normandie (AIR NORMAND)	Haute-Normandie
	Association pour la mise en œuvre du réseau d'étude, de mesure et d'alerte pour la prévention de la pollution atmosphérique dans la zone de Lille-Roubaix-Tourcoing (AREMA)	Nord-Pas-de-Calais
	AASQA Flandre–Côte d'Opale (OPALAIR)	Nord-Pas-de-Calais
	Observatoire régional de l'air en Midi-Pyrénées (ORAMIP)	Midi-Pyrenées
	AASQA en Alsace (ASPA)	Alsace
Germany	Umweltbundesamt (UBA)	
Switzerland	Swiss Agency for the Environment, Forests and Landscape Nationales Beobachtungsnetz für Luftfremdstoffe (NABEL/SAEFL)	
Netherlands	Landelijk Meetnet Luchtkwaliteit (LML)	
Italy	Agenzia Regionale Prevenzione e Ambiente (ARPA)	Emilia-Romagna
United Kingdom	Department of the Environment, Transport and the Regions (DETR)	
Czech Republic	Czech Hydrometeorological Institute	
Spain	Generalitat de Catalunya (GENCAT)	

[a]AASQA, Association agréée pour la surveillance de la qualité de l'air.

two distinct monitoring stations are uncorrelated, i.e., $c^o_{t,h}(\mathbf{s_k}, \mathbf{s_l}) = \sigma^{2o}_{t,h}(\mathbf{s_k})\delta_{kl}$. Here, δ_{kl} is the delta function ($\delta_{kl} = 0$ if $k \neq l$ and $\delta_{kl} = 1$ if $k = l$). Finally, the optimized weighting functions are the solutions of the following system ($\forall k = 1, 2, \ldots, K$):

$$\sum_{l=1}^{K} \widehat{\lambda}^l_{t,h}(\mathbf{s}) \left[c^b_{t,h}(\mathbf{s_k}, \mathbf{s_l}) + \sigma^{2o}_{t,h}(\mathbf{s_k})\delta_{kl} \right] = c^b_{t,h}(\mathbf{s_k}, \mathbf{s}). \tag{2}$$

[20] The computation of the weighting functions requires the preliminary knowledge of the covariances of the background errors and the variances of the observation errors, $\sigma^{2o}_{t,h}(\mathbf{s_k})$. The error covariance model consists of distinct models for correlations and for variances. The background error correlation $\rho^b_h(\mathbf{s_k}, \mathbf{s_l})$ is modeled as a function of the ozone concentration correlation, $\rho^{[O_3]}_h(\mathbf{s_k}, \mathbf{s_l})$. This model assumes that CTM error correlations are higher when the two locations frequently belong to the same ozone pattern and lower when they do not. In practice, the background error correlation is estimated by a least squares fit of available innovation correlation coefficients, $\rho^i_h(\mathbf{s_k}, \mathbf{s_l})$, plotted versus the corresponding correlation coefficients of observed ozone values $\rho^{[O_3]}_h(\mathbf{s_k}, \mathbf{s_l})$. An exponential fit is carried out (see equation (4)). However, as the ozone concentration correlations cannot be computed for sites where no measurements are available, we use the correlation computed on simulated ozone concentrations. In the same way, the innovation variance $\sigma^{2i}_h(\mathbf{s})$ is modeled as a linear function of the ozone concentration variance $\sigma^{2[O_3]}_h(\mathbf{s})$ (see equation (5)). Here, σ^{2i}_h is the sum of the background error variance, $\sigma^{2b}_h = c^b_h(0)$, and the observation error variance, σ^{2o}_h, and $\rho^b_h(0)$ is the ratio of σ^{2b}_h to $\sigma^{2b}_h + \sigma^{2o}_h$. Therefore term $\sigma^{2o}_{t,h}(\mathbf{s_k})$ in equation (2) can easily be estimated from $\rho^b_h(0)$ and σ^{2i}_h.

[21] Finally, the background error covariance model is written

$$c^b_h(\mathbf{s_k}, \mathbf{s_l}) = \sigma^i_h(\mathbf{s_k})\sigma^i_h(\mathbf{s_l})\rho^b_h(\mathbf{s_k}, \mathbf{s_l}), \tag{3}$$

with

$$\rho^b_h(\mathbf{s_k}, \mathbf{s_l}) = e_h \left\{ 1 + \left[1 - \rho^{[O_3]}_h(\mathbf{s_k}, \mathbf{s_l}) \right] / b_h \right\} \times \exp\left\{ -\left[1 - \rho^{[O_3]}_h(\mathbf{s_k}, \mathbf{s_l}) \right] / b_h \right\}, \tag{4}$$

$$\sigma^{2i}_h(\mathbf{s}) = \alpha_h \sigma^{2[O_3]}_h(\mathbf{s}) + \beta_h, \tag{5}$$

where $-1 \leq \rho^{[O_3]}_h(\mathbf{s_k}, \mathbf{s_l}) \leq 1$. Here, α_h, β_h, e_h, and b_h are the regression coefficients. These coefficients are computed at each hour h.

[22] This background error covariance is used to correct ozone concentrations throughout the model layers although only surface observations are used to build it. During the afternoon, in a well-mixed boundary layer, the correlation between low-altitude and high-altitude ozone is elevated, both in the model and in reality. Thus we expect ozone model errors to be highly correlated as well. More details about the use of the same error covariance model for the whole of the 3-D model domain are given by *Blond et al.* [2003].

5. Results

[23] In this section, we first evaluate the ozone analyses by comparing them with independent observations. Then, we use the ozone analyses to initialize the forecast model and discuss the resulting improvement of the short-term forecasts.

5.1. "Leave-One-Out" Method

[24] A leave-one-out method (also called "data withholding") is used in order to evaluate objectively the quality of ozone analyses. Sequentially, observations at a single observation site are omitted in the ASI analysis process. These omitted observations are then statistically compared with the resulting analyzed field. Error statistics are given at an hour h by the root-mean-square (RMS_h) of the analysis-minus-observation residuals and the corresponding bias (BIAS_h), the correlation, etc. These error statistics are then compared to those of the simulation-minus-observation residuals. For the analyses, RMS_h is given by

$$\mathrm{RMS}_h(\mathbf{s_k}) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left[Z^a_{t,h}(\mathbf{s_k}) - Y^{o*}_{t,h}(\mathbf{s_k})\right]^2}, \qquad (6)$$

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

[25] The bias is computed by

$$\mathrm{BIAS}_h(\mathbf{s_k}) = \frac{1}{T}\sum_{t=1}^{T}\left[Z^a_{t,h}(\mathbf{s_k}) - Y^{o*}_{t,h}(\mathbf{s_k})\right]. \qquad (7)$$

5.2. Error Statistics for Ozone Simulations and Analyses

[26] Three-dimensional ozone analyses are produced while applying the ASI method to the raw simulations of CHIMERE and some surface ozone observations. These analyses are performed for each day of summers 1999–2002 for h = 1500 UTC. The bias and error covariance models are first computed for observed and simulated data from summer 1999. They are afterward used to produce the analyses for data from summers 2000, 2001, and 2002 running from 1 May to 31 August. Therefore all of the following error statistics are calculated from a set of data independent of the testing data. Figure 2 summarizes the error statistics obtained for summer 2001 for simulation-minus-observation and analysis-minus-observation residuals.

[27] Figure 2a shows the bias distribution. In the case of raw simulations the distribution is centered around 4.4 ppb. The model tends to overestimate ozone concentrations. This especially happens near the main source areas of primary emissions. The nitrogen monoxide is diluted in large grid cells, implying therefore an overestimation of ozone. This dilution also changes the chemical regime. The atmosphere, which is usually VOC sensitive, becomes NO_x sensitive. This change also implies an ozone overestimation. In contrast, the distribution of analysis-minus-observation bias is centered at 0.2 ppb. Ninety-five percent of these biases are between −5 ppb and 5 ppb.

[28] Figure 2b shows that most raw model correlations lie in the 0.8–0.9 range. A detailed stationwise inspection shows that only a few monitoring stations display low correlation (below 0.7). These stations are generally located near coastal or mountainous areas where higher resolution is required to correctly simulate local transport of pollutants because of complex wind flows. Low correlations are also found for monitoring stations located at high altitude. In these cases, the quite low variability of the ozone concentrations observed there is not well reproduced. This variability depends also on phenomena that are not accounted for in the model. For instance, intermittent intrusions of ozone-rich air masses from the high troposphere or the stratosphere cannot be simulated. The analysis-observation correlations mostly lie between 0.9 and 1. Thus the ozone variability is much better reproduced by the analyses than by the raw simulations. High correlations are increased by ~10%. Correlations over coastal areas and low-level mountains are often increased by ~15–20%. However, correlations over mountains like the Pyrénées and the Alps remain poor.

[29] Figure 2c presents the percentage of "good simulations" (errors smaller than 15 ppb). We deduce that the number of "large errors" is often smaller than 15% and is 5% on average. Largest errors are found over monitoring stations often influenced by city plumes. Two situations can occur. First, the width of the urban plume is smaller than the model grid size and is therefore highly diluted in a coarse-resolution simulation. Second, the direction of the urban plume can be wrong, which is a typical error in very weak wind conditions [*Vautard et al.*, 2001]. In this case, high ozone concentrations are simulated at the wrong place.

[30] Finally, Figure 2d presents the RMS error distribution. The RMS error mean value is 10 ppb for the simulations and 6.6 ppb for the analyses. These values are consistent with those found in previous studies [*Schmidt et al.*, 2001; *Vautard et al.*, 2001]. Error statistics are also comparable with those obtained from other chemistry-transport models [*Tilmes et al.*, 2002; *Elbern and Schmidt*, 2001; *Cardelino et al.*, 2001]. RMS is mostly reduced in areas where the density of observations is high. However, this is not always the case, as is shown in section 5.3.

[31] In order to evaluate the interannual variability of the simulation/analysis skills, we show error statistics obtained from four consecutive years, from 1999 to 2002, in Table 2. In each case the analyses are produced using the error covariance model deduced from the summer of 1999 and the leave-one-out method. All simulations were performed with the same emissions database. Therefore they do not take into account the weak emission reductions, in particular, in the road transport sector, during this period. The emission reduction is estimated to be less than 10% in France over the period 1990–2002 [*Centre Interprofessionnel Technique d'Etudes de la Pollution Atmosphérique* (*CITEPA*), 2003].

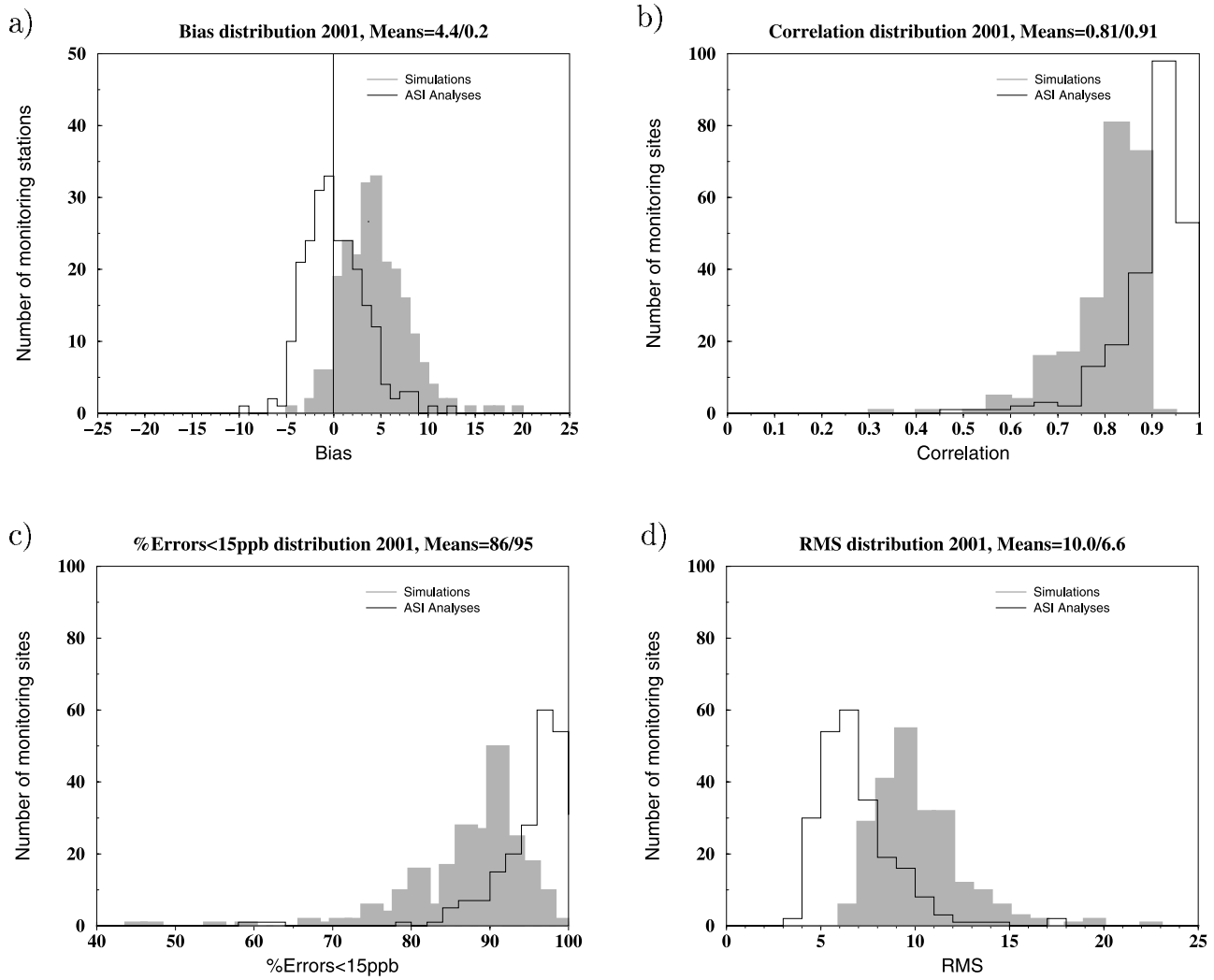

Figure 2. Error statistics over simulation-minus-observation (grey solid bars) and analysis-minus-observation (black open bars) residuals for h = 1500 UTC. These statistics have been computed over 232 monitoring stations throughout western Europe. (a) Bias in ppb, (b) RMS errors in ppb, (c) correlations, and (d) number of "good simulations" in percent. Mean values are written on the top of each panel's simulations/analyses.

This should not significantly affect our results. In Table 2, simulated and analyzed ozone concentrations are compared to observations at 1500 UTC. We show the bias, the RMS error, and the correlation. We also show the fraction of good simulations (E15) and the same fraction applied only to cases when either the forecast or the observation exceeds 50 ppb (E15/50ppb). The latter statistic is of particular relevance because it allows for quantifying the CTM skills during high-ozone-pollution events, i.e., where the population is concerned and needs to be informed. Finally, we show the persistence, a no-skill forecast procedure that simply assumes that the ozone will not change in 24 hours. This latter forecast has been performed for summer 1999.

Table 2. Mean Error Statistics Over Simulation-Minus-Observation and Analysis-Minus-Observation Residuals for 1500 UTC[a]

	BIAS		RMS		COR		E15		E15/50	
	S.	A.	S.	A.	S.	A.	S.	A.	S.	A.
1999	3.2	0.1	9.8	6.5	0.75	0.88	87	95	75	88
2000	4.1	0.1	9.8	6.5	0.78	0.89	87	95	74	85
2001	4.4	0.2	10.0	6.6	0.81	0.91	86	95	75	85
2002	4.4	0.2	10.8	6.9	0.72	0.87	84	95	66	86
Persistence for 1999	6.1		13.1		0.56		77		57	

[a]S., simulation-minus-observation; A., analysis-minus-observation. Bias and RMS errors are in ppb. COR, correlation; E15, number of "good simulations" in percent. E15/50, number of good simulations in percent when the observation or the simulation data is over 50 ppb.

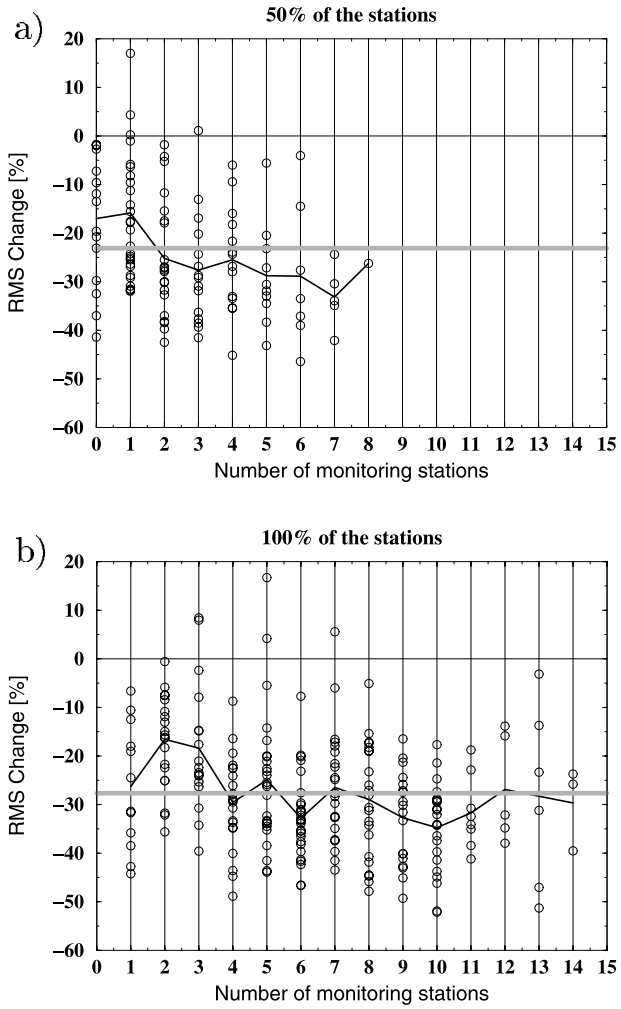

Figure 3. RMS error change due to the use of the ASI algorithm versus the number of monitoring stations surrounding the analyzed grid point within a radius of 100 km. Each circle corresponds to a different observation site. The black line shows the mean value of the change computed over sites surrounded with the same number of monitoring stations. The grey line denotes the mean value of the change computed over all monitoring stations.

[32] We notice that persistence has the worst values for bias, RMS error, correlation, and E15. This indicates that the model CHIMERE is of particular relevance to forecast ozone. The statistics for the raw simulations of CHIMERE and for the analyses are very similar from one year to another. However, the CHIMERE forecasts are not as good for the year 2002 as they are for the other years. In particular, the ozone variability is not well reproduced by the CTM. Indeed, the mean correlation is 0.72 as opposed to 0.81 for 2001. Moreover, only 66% of the ozone values over 50 ppb are well forecasted. Several reasons may explain these larger CTM errors. The meteorological conditions during summer 2002 were not favorable to the development of high-ozone-pollution events. Indeed, a series of fronts crossed western Europe, regularly perturbing ozone diurnal formation and sometimes generating intense convection episodes difficult to simulate. In a cloudy or convective atmosphere, mixing and photolysis rates may be misrepresented, leading to enhanced errors in ozone simulation. In sections 5.3–5.6, we focus on summer 2002 and analyze one particular episode.

5.3. Sensitivity of the Ozone Analyses to the Density of Observation Data

[33] A series of experiments have been performed in order to analyze the influence of the density of observation data on the analyses. We computed the change of the RMS error when only 25%, 50%, 75%, or 100% of the available observation data were used in the ozone analysis. We used a leave-one-out method and averaged the results over all observation sites. These experiments reveal that relative to raw simulations, the RMS error is globally reduced by −17%, −23%, −26%, and −28% when we increase the number of observations.

[34] The change in the RMS error has also been studied versus the number of observation data surrounding the analysis point within a radius of 100 km. Figures 3a and 3b illustrate the results obtained when 50% and 100% of the observation data are assimilated, respectively. Figure 3a shows that the RMS error can be quite reduced in data-sparse areas even where no observation is available within a radius of 100 km (the RMS improvement is between 2 and 40%). Moreover, statistics display large fluctuations from one station to another. Both figures show that in general, the higher the number of observation data around the analysis point, the higher the RMS error reduction. This reduction can reach 50%, as shown in Figure 3b. For a few sites the analyses are worse than the simulations because of a lack of representativeness.

5.4. Time Series Comparisons

[35] In order to further show how well the analysis procedure captures the time variability of observed ozone concentrations, we show in Figure 4 time series of daily ozone mixing ratios in ppb from 1 May to 31 August 2002 for five European locations. These series correspond to values taken at 1500 UTC. Surface ozone observations are compared with both simulations and analyses. The analyses are again obtained using the leave-one-out method. The monitoring stations have been chosen because they are representative of several types of environments. Monitoring sites SONCHAMP and BERLIN-MARIENFELDE are near Paris and Berlin, respectively. They monitor pollution advected by the wind from the cities, while VOSGES-HAUTES is relatively far from any emission source and located in the French Vosges mountains at 1100-m elevation. Finally, NL-107 (middle eastern Netherlands) and LADYBOWER (central England) are much more influenced by nitrogen monoxide emissions. For both stations the hourly mean of the observations is of the order of 30–40 ppb, instead of ∼50 ppb for the other sites.

[36] Though meteorological conditions during summer 2002 were not so often conducive to ozone pollution, some episodes still developed over western Europe. These episodes occurred on 15–16 May, 2–5 June, 17–27 June, 8–9 July, 28 July to 2 August, and 17–23 August. It is often noticed that the ozone episodes move slightly eastward, in this case, from stations SONCHAMP and VOSGES-HAUTES (or NL-107) to BERLIN-MARIENFELDE. A

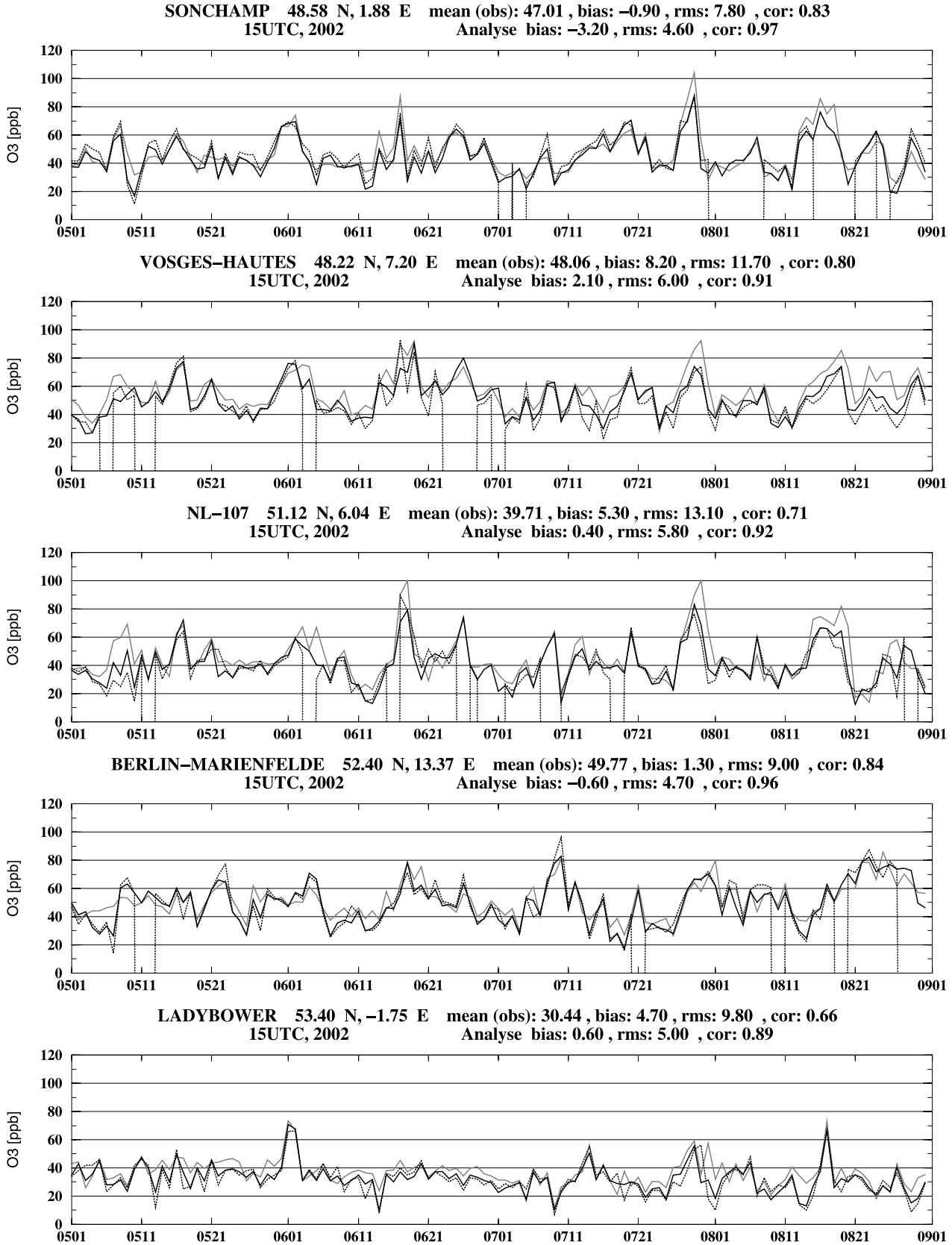

Figure 4. Time series of daily ozone mixing ratios in ppb from 1 May to 31 August 2002 taken every day at 1500 UTC. Dashed line, observations; solid grey lines, raw simulations; solid black lines, analyses. Above each panel are indicated the name and geographical location of the respective observation site, mean of the observed data, biases and RMS error of the simulated (top row) and analyzed (bottom row) mixing ratios relative to the observed mixing ratios, and the corresponding correlation coefficients.

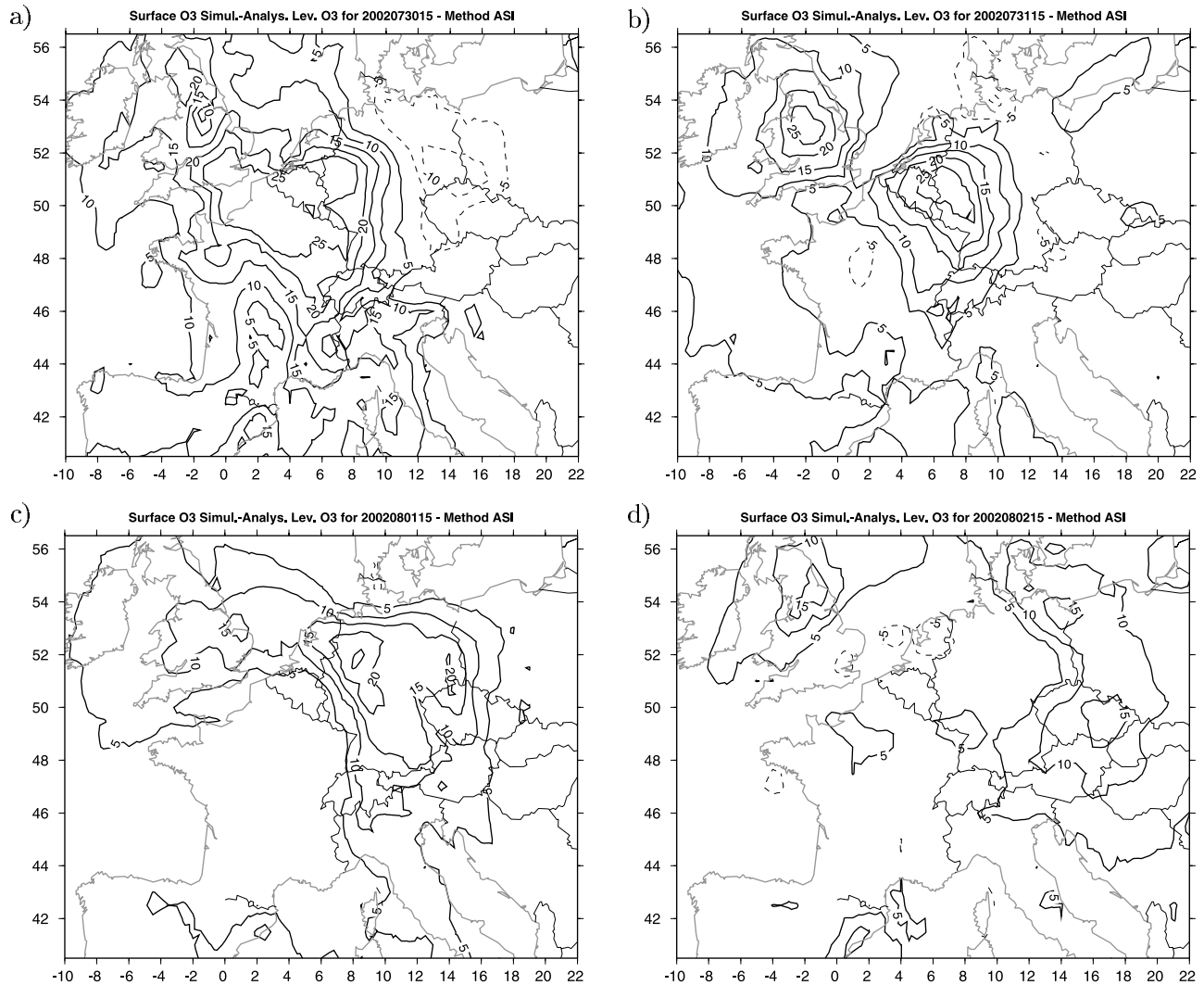

Figure 5. Maps of the differences between ozone simulations and analyses for four consecutive days of summer 2002. These maps are computed for 1500 UTC: (a) 30 July 2002, (b) 31 July 2002, (c) 1 August 2002, and (d) 2 August 2002. The differences are in ppb. Solid contours show positive values, and dashed contours show negative values.

time lag of 2–3 days is often noted (see, for example, the episode of 2–5 June). The episode of 28 July to 2 August is particularly interesting and is studied in section 5.5.

[37] We note that the CTM correctly represents the ozone variability over all types of observation stations. It especially locates all of the high-ozone episodes. However, it often overestimates the ozone concentrations and especially the ozone peaks. Ozone analyses are much closer to the observations than to the simulations. The overestimations are mostly corrected. Biases are drastically reduced except over SONCHAMP. This monitoring station is the only observation site often influenced by strong and thin ozone plumes in the Paris region. It is located downstream from the city of Paris during the anticyclonic situations. At this rural station the ASI correction is therefore mainly controlled by the surrounding ozone observations, which are always less elevated.

[38] We also note that the ASI method corrects ozone concentrations at station VOSGES-HAUTES well although the station is located at altitude. In this case, observations are compared to the simulations and the analyses of the second level of the CTM.

5.5. Illustrative Cases of Ozone Analyses

[39] In order to illustrate the ozone analysis results, we now focus on an individual case, the episode from 28 July to 2 August 2002, which is a very interesting situation. With the exception of western parts of France, the meteorological conditions on 28 July were characterized by sunny hot weather with high temperatures (above 30°C) and light winds until the beginning of the afternoon. Late in the day, a talweg in altitude with a cold pseudofront at the surface penetrated western Europe from the Atlantic, stimulating cloud development and the growth of thunderstorms in the front. During the next few days the convective episode moved eastward with a very similar situation. The CTM does not account for deep convection and therefore had difficulties in representing the strong dispersion of

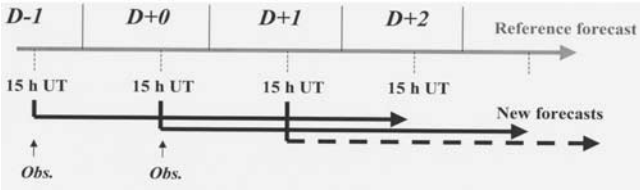

Figure 6. Scheme showing the implementation of the forecast experiments with and without assimilation of surface ozone observations.

pollutants during this episode. During the following 6 days it actually develops a high-ozone region over large parts of the western Europe with few exceedances of the ozone information threshold (90 ppb). However, these high concentrations were in most cases not observed or much lower (see Figure 4). The model error reaches 15–20 ppb on 28 July, 20–25 ppb on 29 July, and more than 25 ppb on the next 2 days. From 30 July the error seems to propagate eastward with the low-pressure system (from SONCHAMP and VOSGES-HAUTES to NL-107 and BERLIN-MARIENFELDE in Figure 4).

[40] The maps of the differences between ozone simulations and analyses for 30 July to 2 August are displayed in Figure 5. Again, they are calculated for 1500 UTC. These differences give an estimate of the CTM error although the analysis itself also contains errors. They illustrate the spatial distribution of the forecast error.

[41] On 30 July 2002 (Figure 5a), the CTM overestimates by more than 10 ppb (up to 25 ppb) the ozone mixing ratios over a large part of France and the Netherlands. On 31 July 2002 (Figure 5b) the overestimations are found in western Germany (but also in central England), and they are found over central to eastern Germany on the next 2 days (Figures 5c and 5d). During the next days the largest errors clearly seem to propagate eastward.

[42] In section 5.6, this question of error propagation is addressed by performing initial-value experiments. If, at least in some cases, model error is due to error propagation and not to "instantaneous error" in meteorology, emission, or chemistry, simulations initialized with the analyses should be improved relative to noninitialized simulations. The objective of section 5.6 is therefore to quantify the importance and beneficial impact of properly defined initial values to improve the CTM predictive skills.

5.6. Forecast Error Statistics

[43] In this section, our objective is to assess the potential of using the ASI method for improving short-term ozone forecasts. We present results of several forecast experiments performed with the assimilation of surface observations or without any assimilation.

[44] These experiments have been performed over summer 2002 following the scheme described in Figure 6. We first perform a forecast without assimilation over the whole summer. This forecast is referred to as the "reference forecast" in the following. Second, we produce 3-D ozone analyses for each day at 1500 UTC using the ASI method. Then, we use the resulting ozone analyses to initialize a series of 3-day forecasts. These forecasts are referred to as the "new forecasts." No more observations are introduced after the ozone analyses initialization. Finally, we compare all of the new 3-day forecasts with the reference forecast.

[45] An afternoon hour (1500 UTC) is chosen for initializing the forecast because it is the time when the boundary layer is the most developed and mixed. Therefore we expect the information from ozone surface observations to be most representative for the whole boundary layer. The error covariance model derived from the 1999 summer ozone data is used, so that the experiment simulates a real-time forecast setup. The only difference from a real forecast is due to the fact that we do not use meteorological forecasts

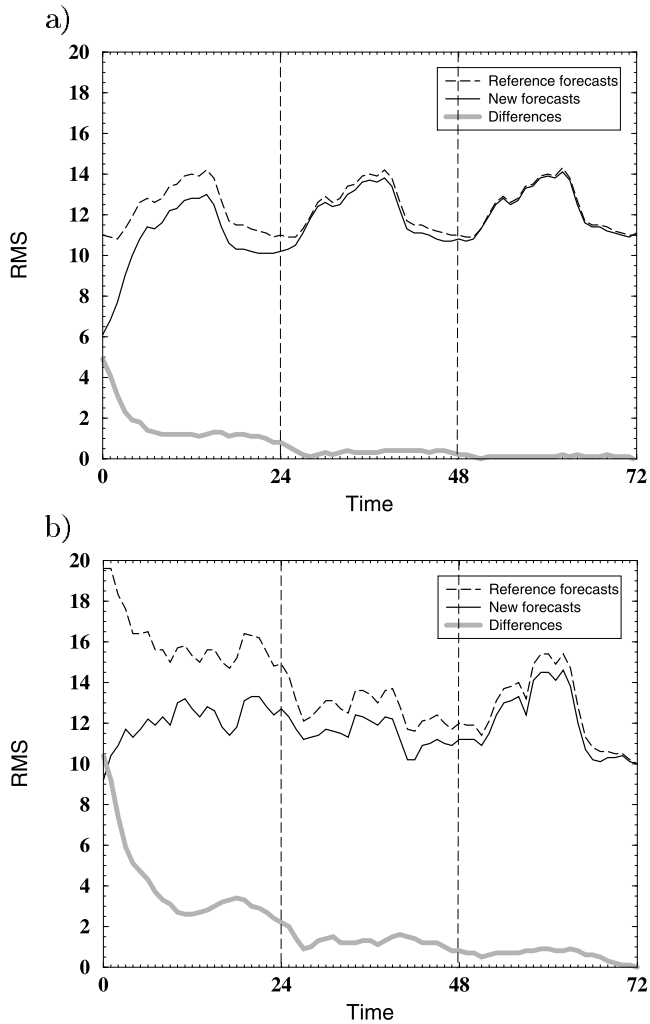

Figure 7. Hourly mean RMS error of the simulated ozone mixing ratios relative to the observed mixing ratios. The RMS error is averaged over the available European observations and over a given time period. (a) Average over the whole summer 2002 (from 1 May to 31 August 2002). (b) Average over the short time period of 30 and 31 July and 1 August 2002. This RMS is plotted as a function of the time, in hours after the initialization. The initialization of the 3-day forecasts is done at 1500 UTC (time zero). Black dashed line, reference forecasts (not initialized with ozone analyses); black solid line, forecasts initialized at time zero with 3-D ozone analyses; grey solid line, difference between the two forecasts. RMS error is in ppb.

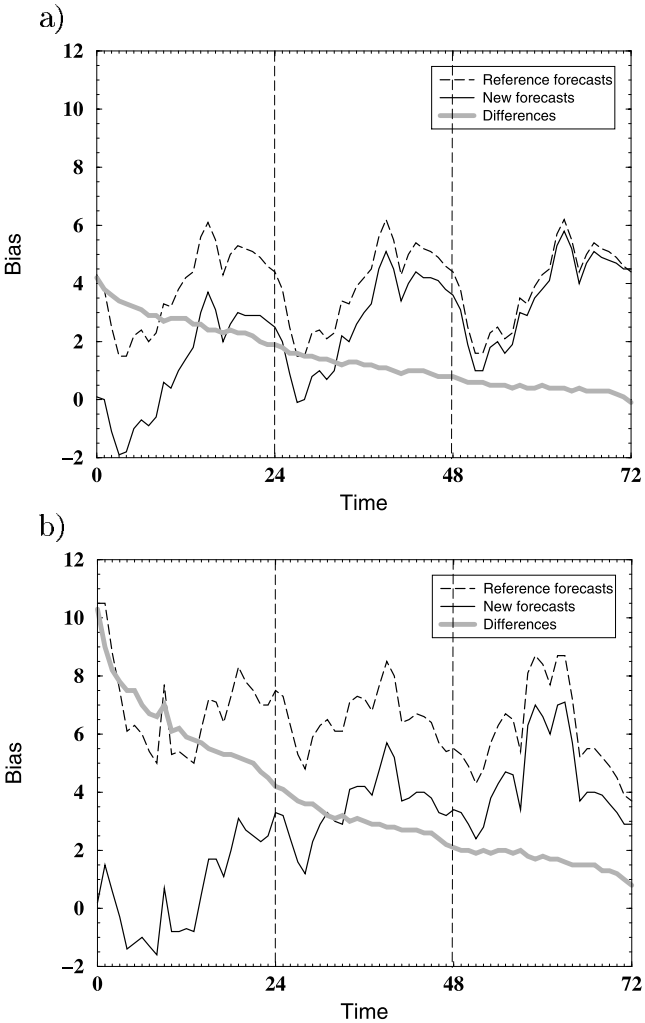

Figure 8. Hourly mean bias of the simulated ozone mixing ratios relative to the observed mixing ratios. The bias is averaged over the available European observations and over a given time period. (a) Average over the whole summer 2002 (from 1 May to 31 August 2002). (b) Average over the short time period of 30 July, 31 July, and 1 August 2002. This bias is plotted as a function of the time, in hours after the initialization. The initialization of the 3-day forecasts is done at 1500 UTC (time zero). Black dashed line, reference forecasts (not initialized with ozone analyses); black solid line, forecasts initialized at time zero with 3-D ozone analyses; grey solid line, difference between the two forecasts. Bias is in ppb.

with the corresponding lead times, but short-term (0–12-hour) forecasts.

[46] Ozone forecast performance is presented in Figures 7 and 8by RMS error and bias graphs, respectively. These statistical measures are computed over two periods of time: the whole summer of 2002 (Figures 7a and 8a) and a 3-day episode on 30 and 31 July and 1 August 2002 (Figures 7b and 8b). They are also computed over all observation sites. Finally, these hourly average RMS error and bias values are plotted versus the time passed after the ozone analysis initializations.

[47] The RMS error of the reference forecast has a clear diurnal cycle that is identical from one simulation day to the next. If real meteorological forecasts were used, the RMS would increase with time. RMS error is larger during nighttime (2000–0500 UTC, e.g., 5–15 hours after the initialization time) than during daytime. Several factors could explain this behavior. First, nighttime ozone concentrations are very sensitive to the thickness of the shallow nocturnal boundary layer, which may be too coarsely represented in the model. Second, because of the titration with local nitrogen monoxide emissions, ozone structures at ground level could be of smaller scale than during the daytime. Large horizontal gradients could cause representativeness problems.

[48] On average (Figure 7a), the new forecasts have a significantly smaller RMS than the reference ones during the first 24 hours of simulation, but beyond that time, the initialized simulation quality becomes identical to the reference one.

[49] Over the 3-day episode (Figure 7b) the reference simulation shows a large RMS error of 20 ppb on 30 July at the initialization time. This is consistent with the maps presented in Figure 5. The RMS error rapidly decreases during the next few hours. In this case, the impact of the ozone analyses is significant up to 48 hours after the initialization time. During the 11 hours following the initialization time, the differences between the forecasts continuously decrease. Afterward, they slightly increase. This increase is probably due to the late morning downward mixing of air masses from the upper residual layer, where previous-day observations have corrected the ozone concentrations. Indeed, the correction of the ozone concentrations in the whole boundary layer at 1500 UTC has a positive impact on the day after.

[50] Figure 8 presents the ozone forecast performance in terms of biases. As shown in Figure 8a, the bias of the reference forecast is always positive (see discussion of the bias in section 5.2). It also shows a clear diurnal cycle. Between 1500 and 1800 UTC the bias decreases because the urban ozone plumes progressively diffuse. The ozone gradients are therefore less strong, and the observations are more representative. Afterward, the bias increases because of the increase of NO_x emissions and the decrease of the atmospheric boundary layer (ABL) (the low vertical resolution of the model induces some systematic errors). It decreases again when the NO_x emissions decrease and the ABL increases. The bias of the new forecasts has a similar evolution but is much smaller. This figure shows a bias reduction of ~2 ppb 24 hours after the initialization time for the whole summer and 4 ppb for the 3-day episode (Figure 8b).

[51] The spatial distribution of the differences between new initialized simulations and analyses for 30 and 31 July and 1 and 2 August is displayed in Figure 9, in the same format as in Figure 5, to which it should be compared. For the four cases the simulations are issued with an initialization at 1500 UTC on the previous day. The large errors are not reduced much for 30 July, showing that these errors are not due to a propagation of previous errors. In contrast, the errors are partially (halfway) corrected for 31 July. The ozone data assimilation allows an error reduction of ~10 ppb over central England, western Germany, and the Benelux

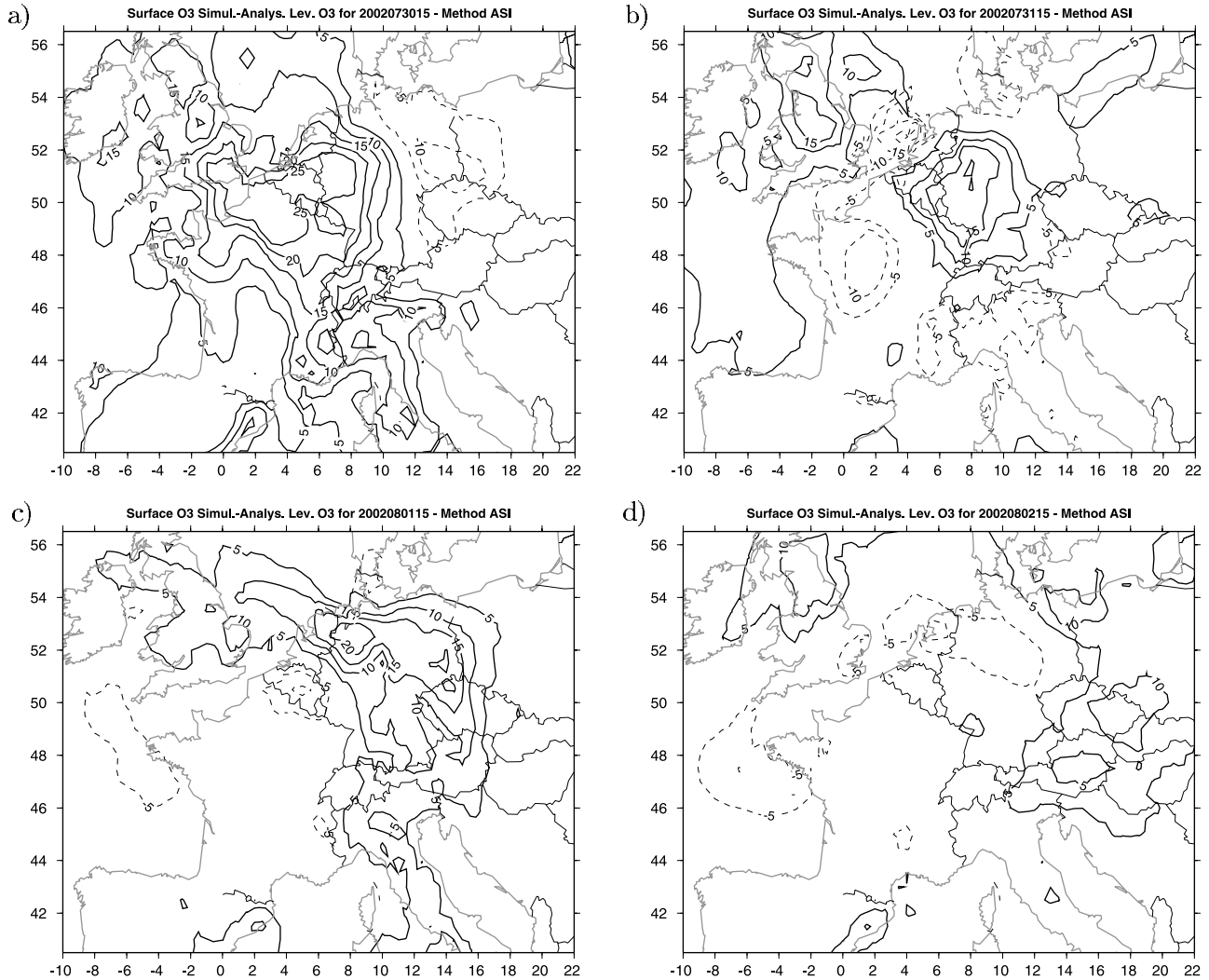

Figure 9. Maps of the differences between ozone initialized simulations and analyses for four days of summer 2002. The forecasts have been initialized using ozone analyses on the previous day at 1500 UTC. The maps are shown for a lead time of 24 hours after initialization and therefore also at 1500 UTC: (a) 30 July 2002, (b) 31 July 2002, (c) 1 August 2002, and (d) 2 August 2002. The differences are in ppb. Solid contours show positive values, and dashed contours show negative values.

countries. The errors made on 30 July have been advected eastward and can be corrected by a proper ozone initialization. Similar conclusions hold for the following 2 days.

6. Summary and Conclusion

[52] In this study, the potential of data assimilation of surface ozone concentrations in a chemistry-transport model at the scale of the European continent is investigated. Ozone data analyses have two major applications: (1) the representation of accurate ozone maps for the information of populations and (2) the initialization of ozone forecasts. The present article evaluates the accuracy of a previously developed analysis method in fulfilling these two applications.

[53] The outputs of the continental-scale version of the CHIMERE chemistry-transport model [*Schmidt and Martin*, 2003] are combined with surface ozone measurements using a statistical interpolation method. This method was previously used to produce ozone analyses over the Paris region [*Blond et al.*, 2003]. Its advantage is to provide an anisotropic error covariance model that is extendable in the vertical direction. As for the regional ozone analyses, the coarse grid ozone analyses are objectively evaluated in a statistical way using a leave-one-out method. We show that the RMS error of analyses is reduced by ~30% relative to the raw simulations.

[54] Next we investigate the potential improvement of using the analyses to initialize ozone forecasts using the same chemistry-transport model. The initialization time is chosen as 1500 UTC, for two reasons. First, this hour often corresponds to the time of ozone daily maxima, and second, it is also the time when the boundary layer is most developed and therefore surface information is better spread in the vertical direction. The evaluation of the skill is achieved by a series of initial-value experiments starting each day of summer 2002. Depending on the case, ozone analysis initialization improves the simulation for 24–

48 hours later. In an operational framework, ozone initialization should therefore improve forecasts for the next day.

[55] However, inspection of individual cases shows that surface ozone analysis initialization only allows a partial correction of the total error, because initial conditions are not the only source of uncertainty. Emissions and meteorology, as well as boundary concentrations, also contribute significantly to the global model uncertainty. All these parameters should also be optimized, as has been proposed in recent studies [*Elbern and Schmidt*, 2002].

[56] The conclusions we reach in this paper are in agreement with those obtained by *Elbern and Schmidt* [2001], who used a much more computationally intensive data assimilation algorithm (4D-Var). Several research questions remain to be addressed in the future. In this study, we only use surface ozone observation data. Accurate observations of precursors (NO_x or volatile organic compounds) may also help in defining proper initial conditions. The use of measurements from spaceborne sensors, which have the ability to view large areas of the atmosphere simultaneously, could also bring in the future valuable information to force chemistry-transport models (CTMs). The method proposed here, which is very easy to implement in practice, can be applied in all these cases.

[57] **Acknowledgments.** Financial support of the study was provided by the Agence de l'Environnement et de Maîtrise de l'Énergie (ADEME) and the TOTAL-FINA-ELF company. The authors would like to thank H. Schmidt from the Max Planck Institute for Meteorology in Hamburg for many helpful discussions and fruitful advice. The following institutions are gratefully thanked: Météo-France and the European Center for Medium-Range Weather Forecasts (ECMWF), for allowing free access to the meteorological database, and European organizations cited in Table 1, for giving us free access to surface ozone data. Most of the work presented here has been tested in real time in the framework of Ozone Forecast and Uncertainty Over Europe and Regions (PIONEER), which has been supported by the French National Research Programme on Atmospheric Chemistry.

References

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.

Blond, N. (2002), Assimilation de données photochimiques et prévision de la pollution troposphérique, Ph.D. report, Lab. de Météorol. Dyn., École Polytech, Palaiseau, France.

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.

Cardelino, C., M. Chang, J. St. John, B. Murphey, J. Cordle, R. Ballagas, L. Patterson, K. Powell, J. Stogner, and S. Zimmer-Dauphinee (2001), Ozone predictions in Atlanta, Georgia: Analysis of the 1999 ozone season, *J. Air Waste Manage. Assoc.*, *51*, 1227–1236.

Centre Interprofessionnel Technique d'Etudes de la Pollution Atmosphérique (CITEPA) (2003), Inventaire des émissions de polluants atmosphériques en France: Series sectorielles et analyses étendues, report, Paris.

Chang, M. E., D. E. Hartley, C. Cardelino, and W. L. Chang (1996), Inverse modeling of biogenic isoprene emissions, *Geophys. Res. Lett.*, *23*, 3007–3010.

Chang, M. E., D. E. Hartley, C. Cardelino, D. Haas-Laursen, and W. L. Chang (1997), On using inverse methods for resolving emissions with large spatial inhomogeneities, *J. Geophys. Res.*, *102*, 16,023–16,036.

Daley, R. (1991), *Atmospheric Data Analysis*, Cambridge Univ. Press, New York.

Daley, R. (1997), Atmospheric data assimilation, *J. Meteorol. Soc. Jpn.*, *75*(1B), 319–329.

Derognat, C., M. Beekmann, M. Baeumle, D. Martin, and H. Schmidt (2003), 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.

Elbern, H., and H. Schmidt (2001), Ozone episode analysis by four-dimensional variational chemistry data assimilation, *J. Geophys. Res.*, *106*, 3569–3590.

Elbern, H., and H. Schmidt (2002), The development of a 4D-variational chemistry data assimilation scheme for initial value and emission rate estimates, paper presented at GLOREAM Symposium, Dep. of Environ. and Plann., Univ. of Aveiro, Aveiro, Portugal, Sept.

Elbern, H., H. Schmidt, O. Talagrand, and A. Ebel (2000), 4D-variational data assimilation with an adjoint air quality model for emission analysis, *Environ. Modell. Software*, *15*, 539–548.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*, 10,143–10,162.

Fuentes, M., and A. Raftery (2001), 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.

Generation of European Emission Data for Episodes Project (GENEMIS) (1994), EUROTRAC annual report 1993, part 5, EUROTRAC Int. Sci. Secr., Garmisch-Partenkirchen, Germany.

Ghil, M., and P. Malanotte-Rizzoli (1991), Data assimilation in meteorology and oceanography, *Adv. Geophys.*, *33*, 141–266.

Gilliland, A., and P. J. Abbitt (2001), A sensitivity study of the discrete Kalman filter (DKF) to initial condition discrepancies, *J. Geophys. Res.*, *106*, 17,939–17,952.

Hass, H., P. J. H. Builtjes, D. Simpson, and R. Stern (1997), Comparison of model results obtained with several European regional air quality models, *Atmos. Environ.*, *31*(19), 3259–3279.

Honoré, C. (2000), La photochimie de l'ozone à l'échelle urbaine, un système dynamique non linéaire, mémoire de thèse, Univ. Pierre et Marie Curie–Paris 6, Paris.

Honoré, C., R. Vautard, and M. Beeckmann (2000), Photochemical regimes in urban atmospheres: The influence of dispersion, *Geophys. Res. Lett.*, *27*, 1895–1898.

Horowitz, L. W., et al. (2003), A global simulation of tropospheric ozone and related tracers: Description and evaluation of MOZART, version 2, *J. Geophys. Res.*, *108*(D24), 4784, doi:10.1029/2002JD002853.

Ionescu, A., Y. Candau, E. Mayer, and I. Colda (2000), 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.

Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *J. Basic Eng.*, *83*, 35–45.

Kalman, R. E., and R. S. Bucy (1961), New results in linear filtering and prediction theory, *Trans. ASME, Ser. D, J. Basic Eng.*, *83*, 95–108.

Kleinman, L. I., P. H. Daum, J. H. Lee, Y.-N. Lee, L. J. Nunnermacker, S. R. Springston, L. Newman, J. Weinstein-Lloyd, and S. Sillman (1997), Dependence of ozone production on NO and hydrocarbons in the troposphere, *Geophys. Res. Lett.*, *24*, 2299–2302.

Krol, M. C. (1995), Comment on "Multiple steady states in atmospheric chemistry," by R. W. Stewart, *J. Geophys. Res.*, *100*, 11,699–11,702.

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. Pierre et Marie Curie–Paris 6, Paris.

Lewis, J. M., and J. C. Derber (1985), The use of adjoint equations to solve a variational adjustment problem with advective constraints, *Tellus, Ser. A*, *37*, 309–322.

Lorenz, E. N. (1963), Deterministic nonperiodic flow, *J. Atmos. Sci.*, *20*, 130–141.

Lu, R., R. P. Turco, and M. Z. Jacobson (1997a), An integrated air pollution modeling aystem for urban and regional scale: 1. Model formulation and numerics, *J. Geophys. Res.*, *102*, 6063–6079.

Lu, R., R. P. Turco, and M. Z. Jacobson (1997b), An integrated air pollution modeling system for urban and regional acale: 2. Evaluation of model performance using SCAQS data, *J. Geophys. Res.*, *102*, 6081–6098.

Mendoza-Dominguez, A., and A. G. Russell (2001), Estimation of emission adjustements from the application of four-dimensional data assimilation to photochemical air quality modeling, *Atmos. Environ.*, *35*, 2879–2894.

Menut, L. (2003), Adjoint modeling for atmospheric pollution process sensitivity at regional scale, *J. Geophys. Res.*, *108*(D17), 8562, doi:10.1029/2002JD002549.

Menut, L., et al. (2000), Measurements and modeling of atmospheric pollution over the Paris area: An overview of the ESQUIF project, *Ann. Geophys.*, *18*, 1467–1481.

Mosca, S., R. Bianconi, R. Bellasio, and G. K. W. Graziani (1998), ATMES II: Evaluation of long-range dispersion models using data of the 1st ETEX release, *EUR 17756 EN*, p. 252, Eur. Commun., Brussels.

Mylona, S. (1999), EMEP emission data: Status report 1999, *EMEP/MSC-W Note 1/99*, Oslo.

Schmidt, H., and D. Martin (2003), Adjoint sensitivity of episodic ozone in the Paris area to emissions on the continental scale, *J. Geophys. Res.*, *108*(D17), 8561, doi:10.1029/2001JD001583.

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.*, *35*, 6277–6297.

Silibello, C., G. Calori, G. Brusasca, G. Catenacci, and G. Finzi (1998), Application of a photochemical grid model to Milan metropolitan area, *Atmos. Environ.*, *32*, 2025–2038.

Stewart, R. W. (1993), Multiple steady states in atmospheric chemistry, *J. Geophys. Res.*, *98*, 20,601–20,611.

Stewart, R. W. (1995), Dynmics of the low to high NOx transition in a simplified tropospheric photochemical model, *J. Geophys. Res.*, *100*, 8929–8943.

Talagrand, O. (1997), Assimilation of observations: An introduction, *J. Meteorol. Soc. Jpn.*, *75*(1B), 191–209.

Talagrand, O., and P. Courtier (1987), Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory, *Q. J. R. Meteorol. Soc.*, *113*, 1311–1328.

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.

van de Velde, R., W. Faber, V. Katwijk, H. J. Scholten, T. J. M. Thewessen, M. Verspuy, and M. Zevenbergen (1994), The preparation of a European land-use database, *RIVM Rep. 712401001*, Bilthoven, Netherlands.

van Loon, M., P. J. H. Builtjes, and A. J. Segers (2000), Data assimilation of ozone in the atmospheric transport chemistry model LOTOS, *Environ. Modell. Software*, *15*, 603–609.

Vautard, R. (Ed.) (2001), Etude et Simulation de la qualité de l'air en Île-de-France, rapport final, Inst. Pierre Simon Laplace, Paris.

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., et al. (2003a), A synthesis of the Air Pollution Over the Paris Region (ESQUIF) field campaign, *J. Geophys. Res.*, *108*(D17), 8558, doi:10.1029/2003JD003380.

Vautard, R., et al. (2003b), Paris emission inventory diagnostics from ESQUIF airborne measurements and a chemistry transport model, *J. Geophys. Res.*, *108*(D17), 8564, doi:10.1029/2002JD002797.

Verwer, J. G. (1994), Gauss-Seidel iteration for stiff ODEs from chemical kinetics, *SIAM J. Sci. Comput.*, *15*, 1243–1250.

Wackernagel, H., C. Lajaunie, N. Blond, C. Roth, and R. Vautard (2004), Geostatistical risk mapping with chemical transport model output and ozone station data, *J. Ecol. Model.*, in press.

White, W. H., and D. Dietz (1984), Does the photochemistry of the atmosphere admit more than one steady state?, *Nature*, *309*, 242–244.

N. Blond, Laboratoire Interuniversitaire des Systèmes Atmosphériques, 61, Av. 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, France. (vautard@lmd.polytechnique.fr)