Accounting for model error in air-quality forecasts: an application of 4DEnVar to the assimilation of atmospheric composition using QG-Chem 1.0

Emanuele Emili[1], Selime Gürol[1], and Daniel Cariolle[1]

[1]CECI UMR5318 CNRS/CERFACS, Toulouse, France

Correspondence to: Emili (emili@cerfacs.fr)

Abstract. Model errors play a significant role in air-quality forecasts. Accounting for them in the data assimilation (DA) procedures is decisive to obtain improved forecasts. We address this issue using a reduced-order coupled chemistry-meteorology model based on quasi-geostrophic dynamics and a detailed tropospheric chemistry mechanism, which we name QG-Chem. This model has been coupled to the software library for data assimilation OOPS (Object Oriented Prediction System) and used to assess the potential of the 4DEnVar algorithm for air-quality analyses and forecasts. The assets of 4DEnVar include the possibility to deal with multivariate aspects of atmospheric chemistry and to account for model errors of generic type. A simple diagnostic procedure for detecting model errors is proposed, based on the 4DEnVar analysis and one additional model forecast. A large number of idealized data assimilation experiments are shown for several chemical species of relevance for air-quality forecasts (O_3, NO_x, CO and CO_2), with very different atmospheric lifetimes and chemical couplings. Experiments are done both under a perfect model hypothesis and including model error through perturbation of surface chemical emissions. Some key elements of the 4DEnVar algorithm such as the ensemble size and localization are also discussed. A comparison with results of 3D-Var, widely used in operational centers, shows that, for some species, analysis and next day forecast errors can be halved when model error is taken into account. This result was obtained using a small ensemble size, which remain affordable for most operational centers. We conclude that 4DEnVar has a promising potential for operational air-quality models. We finally highlight areas that deserve further research for applying 4DEnVar to large scale chemistry models, i.e., localization techniques, propagation of analysis covariance between DA cycles and treatment for chemical non-linearities. QG-Chem can provide a useful tool in this regard.

1 Introduction

In recent years, Data Assimilation (DA) of atmospheric constituents has become a key tool for providing more accurate forecasts and reanalyses of the atmospheric composition. The increasing availability of chemical observations from both satellites and ground-based instruments allowed to reduce the uncertainty of atmospheric chemistry models in a large number of applications. Utilization of DA can be found in the modeling of volcanic ash (Lu et al., 2016), in operational air-quality forecasts at continental scale (Marécal et al., 2015) or in the reanalysis of the global atmospheric composition at decennial scale (van der A et al., 2010; Inness et al., 2013). Data assimilation can also be used to infer surface fluxes of long-lived chemical compounds

(Thompson and Stohl, 2014; Chevallier et al., 2005). A review of the utilization of data assimilation for the atmospheric composition can be found in Zhang et al. (2012); Bocquet et al. (2015).

The main goal of DA is to reduce the uncertainties of a model through a timely combination of model results and observations. This is generally done by means of correcting the so-called *control variables* of the given model. The choice of the control variables should reflect the largest source of uncertainty of the considered model. In atmospheric chemistry control variables are typically associated with the model initial state (Elbern et al., 1997) or chemical emissions (Chevallier et al., 2005). However, inaccurate identification of the model uncertainty can lead to a wrong adjustment of the model through DA, even though the spread between model predictions and assimilated observations is reduced (Tang et al., 2016). The choice of the control variable, and the approximate knowledge of its uncertainty (also named background error covariance), is therefore critical for the design of an appropriate assimilation algorithm and to ensure correct results of DA.

Principal sources of uncertainty of atmospheric chemistry models include the model initial condition, model ancillary data or parameters, model physical parametrization, chemical mechanism etc. (Beekmann and Derognat, 2003; Mallet and Sportisse, 2006). Since different chemical species can be sensitive to different physical and chemical processes, the main sources of uncertainties can also differ from species to species. For example, long-lived species like CO_2 or CO are mostly sensitive to uncertainties in surface fluxes, which can be corrected using variational algorithms in combination with long assimilation windows of several weeks (Chevallier et al., 2005; Koohkan and Bocquet, 2012). This is possible since the chemical reactivity and the sensitivity to the initial condition are negligible for long integration of the model. Uncertainty in the transport processes is also generally neglected (Babenhauserheide et al., 2015). Data assimilation of short-lived gases like tropospheric O_3 or NO_2, which is encountered in air-quality applications, is instead trickier. O_3 and NO_2 are involved in rapid chemical reactions and sensitive to several model parameters ranging from reaction rates, emissions of primary species, clouds and radiation, boundary layer mixing etc. It is much more difficult in this case to identify a single and predominant source of uncertainty in the model predictions.

In most air-quality operational models a pragmatic choice is currently done by setting the control variable to the initial state of the measured species and using short forecast/assimilation cycles (e.g., one hour). Since ground-based measurements are generally available at hourly frequency, most used sequential DA algorithms such as Optimal Interpolation (OI), 3D-Var or Ensemble Kalman Filters (EnKF) (Marécal et al., 2015) all provide a strong constraint on model trajectories (Wu et al., 2008). This strategy gives robust results for operational analyses because chemical fields are corrected every hour. Since the control variable correspond to the assimilated observations, this also permits to estimate the background error covariance from previous validation of the model against observations (Hollingsworth and Loennberg, 1986), and keeps from the difficult diagnosis of the true model uncertainties.

However, the model dynamics are neglected in OI or 3D-Var DA schemes. Attempts of using ensembles of model analyses to specify dynamically the background error covariance within 3D-Var did not show clear improvements over the static case (Jaumouillé et al., 2012). Specification of cross-correlations between interacting chemical species in the 3D-Var background error covariance matrix is also particularly difficult, because chemical interactions depend on the local concentrations and on meteorological conditions. As a consequence, multivariate chemical DA with 3D-Var schemes has not been yet documented in

the literature. In EnKF systems, the forecast model is used to propagate and estimate the background error covariance, but ad-hoc adjustments are necessary to avoid the collapse of the ensemble variance and obtain realistic covariance matrices for 1 hour forecasts (Gaubert et al., 2014; Constantinescu et al., 2007a). As a result, costly algorithms such as EnKF or 3D-Var hardly give better results that more simple OI for chemical reanalyses (Rouil and the MACC team, 2014). More importantly, very little

5   improvement is obtained, whatever the employed DA algorithm, for the next day model forecast (Wu et al., 2008). Forecasts of reactive gases such as O_3 or NO_2, but also other other pollutants such as aerosols mixtures (PM10 and PM2.5), depend weakly on the initial condition and are more sensitive to model settings such as surface emissions or physical parametrizations. Current operational systems can achieve accurate reanalyses of observed chemical species through DA but parameter estimation and, more in general, model errors must be taken into account in DA to improve chemical forecasts of reactive gases and particles.

10   Some studies evaluated more advanced DA algorithms to jointly correct surface emissions of precursor species and initial condition of observed species. For example, Elbern et al. (2007) employed a 4D-Var scheme in combination with assimilation windows of 24 hours to assimilate O_3, NO_x and SO_2 measurements. A similar study has been done also in the context of a toy-model experiment by Hamer et al. (2015), where only emissions of precursor species are adjusted to improve O_3 forecasts. Results seems promising but still rely on the assumption that the model is *almost* perfect, i.e., that there are no additional

15   sources of uncertainties in the model forecast other than the controlled variables (the initial state and the selected emissions). This can lead to the over-correction of control variables when other non negligible model errors exist, for example due to the meteorological forcing, photo-chemistry coefficients, dry or wet deposition. Concerning EnKF implementations, some authors also tested joint optimization of the chemical state and precursor emissions (Miyazaki et al., 2012; Tang et al., 2011; Constantinescu et al., 2007b). EnKF naturally includes model uncertainties in its formulation, which can be added through

20   stochastic perturbation of model parameters during the ensemble forecast (Evensen, 2003). However, EnKF corrects the model trajectories sequentially. In a typical air-quality context, the emissions of O_3 precursor species (e.g., NOx and VOCs) in the early morning or night can affect the concentration of observed species (e.g., O_3) in the early afternoon, when the photo-chemistry takes place. When using EnKF, the information made available by afternoon measurements cannot be used to correct the model at previous hours.

25   Based on above-mentioned facts, we determine that the following capabilities are needed to further improve DA in air-quality applications:

   – permit simultaneous assimilation and optimization of multiple chemical species (multivariate DA), with possible chemical interactions and very different lifetimes;

   – include model error and account for disparate sources of model uncertainty;

30   – allow long assimilation windows to make best use of the information content of frequent air-quality observations.

This can be accomplished by using the so-called *weak constraint* 4D-Var (Tremolet, 2006), which is an extension of the 4D-Var algorithm that accounts for the model error. On top of the operators already needed to perform the strong constraint 4D-Var (e.g., the tangent linear and adjoint codes of the forecast model), the formulation of the weak constraint 4D-Var requires the

definition of the model error covariances $\mathbf{Q}$, which can be difficult to estimate in real applications (Trémolet, 2007). Recent studies have shown that the linear operators needed in the weak constraint 4D-Var formulation can be approximated through an ensemble of model forecasts. For example, the 4D-Var-EnKS (Mandel et al., 2016) uses an ensemble approach to mimic the tangent linear and adjoint model for the minimization of the weak constraint 4D-Var cost function. The iterative ensemble

5   Kalman smoother (IEnKS, Bocquet and Sakov (2014)) is a nonlinear 4D-EnVar formulated under perfect model assumptions, which can also be used to estimate erroneous model parameters through an augmented state formalism (Bocquet and Sakov, 2013). A major asset of IEnKS for chemistry applications is that it can also account for strong non-linearities of the forecast model. The 4DEnVar method (Desroziers et al., 2014) uses an ensemble of non-linear model trajectories to estimate both the error covariances for the initial condition and the model error, as well as to approximate the tangent linear and adjoint model.

10   These approaches are generally referred in the literature as ensemble-variational EnVar (Lorenc, 2013), as opposed to 'hybrid' methods, which make use of ensembles only to specify error covariances matrices in variational algorithms (Belo Pereira and Berre, 2006). EnVar methods have a major advantage for atmospheric chemistry applications: they avoid the construction of tangent linear and adjoint codes of the forecast model, which are still lacking for most of the operational CTMs, or are becoming very difficult to be maintained due to the rapid evolution of models and computer architectures.

15   The main advantage of the 4DEnVar method is that it permits to account for a generic model error through the addition of stochastic perturbations during the model integration step (like in EnKF). Moreover, it focuses exclusively on the estimation of the model state, which is the only variable that is directly constrained by observations. This avoids the difficult specification of $\mathbf{Q}$ still needed in the 4D-Var-EnKS. As all ensemble based methods, 4DEnVar also naturally supports multivariate chemical DA, with the cross-covariances terms between chemical species being automatically obtained from the ensemble of the non-

20   linear model forecasts.

Variants of the 4DEnVar have been already tested in real numerical weather prediction (NWP) applications (Lorenc et al., 2015). The method has proven to be affordable for large scale operational NWP models, even though the skills of the operational hybrid 4D-Var are not yet matched. To the knowledge of the authors, EnVar type methods have not yet been implemented in air-quality or atmospheric chemistry models and only one study has already examined the potential of EnVar methods for

25   chemical DA (Haussaire and Bocquet, 2016). Note that, operational NWPs are already based on well mature 4D-Var DA systems, whereas very few atmospheric chemistry models do. Therefore, there is a larger room for improvements from EnVar type of algorithms in air-quality models than in NWP. Hence, the main objectives of this study are:

- to present a new atmospheric chemistry toy-model built for assessing and comparing performances of several DA algorithms;

30   - examine the potential and limits of the 4DEnVar algorithm for air-quality analyses, compared to the generally used 3D-Var;

- present a new procedure based on 4DEnVar to improve chemical forecasts on the next day.

The purpose is to examine state-of-the-art DA algorithms in the reactive-gases/air-quality context and, therefore, to guide future developments for the operational DA systems. Four gaseous species with very different lifetimes and chemical mechanisms,

currently well observed either from satellites or from ground-based instruments, are considered for this study (CO, O_3, NO_2 and CO_2). Using a simplified model in this context permits: faster implementation of complex DA algorithms, cheaper numerical experiments and more straightforward interpretation of the DA results (Fairbairn et al., 2013). The latter is particularly true compared to DA experiments done using real observations, with generally unknown error statistics. Compared to already

5   mentioned simplified models (Hamer et al., 2015; Haussaire and Bocquet, 2016), which are, respectively, 0-D and 1-D, the newly proposed model is 3-D and uses the same tropospheric chemistry scheme of operational air-quality models. This allows to reproduce more features of real models, for example the complex interactions of reactive chemistry and large scale advection or the effect of boundary conditions. This also permits to better examine typical issues of DA within large systems, like the emergence of sampling errors due to the finite size of the ensemble and the consequences of localization techniques.

10   Additionally, the use of 3-D fields and operators eases the estimation of numerical costs and possible bottlenecks of the DA algorithm in terms of operational implementation. We remind readers that the objective of this study is to demonstrate the applicability of a DA algorithm that could outperform currently implemented methods in operational centers, but with an acceptable compromise between computational costs and precision. Finally, the toy system has been implemented using the library for data assimilation OOPS (Yannick Trémolet, personal communication) to ease the exchange of assimilation

15   algorithms or toy-models between scientists.

    The paper is outlined as follows. The developed atmospheric chemistry model will be presented in Section 2. A summary of the data assimilation algorithms employed in this study is given in Sec. 3. The first section of the numerical results (Sec. 4.1) presents a number of DA experiments done under the hypothesis of perfect model. A detailed comparison between 4DEnVar and 3D-Var is presented, as well as a sensitivity study on the principal parameters of the 4DEnVar algorithm, i.e., the ensemble

20   size and the localization choices. In the following Sec. 4.2 the effects of a model error are investigated using the 4DEnVar algorithm. A statistical comparison of the 4DEnVar and 3D-Var performances on multiple cycles of analyses and forecasts is presented in Sec. 4.3. Finally, conclusions are given in Sec. 5.

2   Model description

A new atmospheric chemistry low-order model has been developed for this study and named QG-Chem. The objective was

25   to reproduce typical features of chemical fields from large scale Chemical Transport Models (CTM), but maintaining the computational cost low enough to allow a comfortable usage on a personal computer. The meteorological forcing is computed using a two-layer Quasi-Geostrophic (QG) model, representative of mid-latitudes mesoscale dynamics (Pedlosky, 1992). The QG wind field is used to advect the chemical species, which makes QG-Chem a coupled meteorological-chemistry model. This choice permits to examine the behavior of DA in the presence of complex gradients of wind fields and vorticity. Since

30   all DA algorithms make strong assumptions on the model dynamics (Sec. 3), it is important to test them in the presence of advection patterns that can be found in real applications. Nevertheless, the focus of this study remains atmospheric chemistry. Therefore, a detailed tropospheric chemical mechanism has been considered. Aspects of DA concerning the coupling between

meteorology and chemistry are also left for future work, with the present using QG-Chem in a CTM-like mode. The details of the meteorological and chemical models are given next.

2.1 Quasi-geostrophic meteorology

The two-layer QG model is a geophysical fluid model composed of two atmospheric layers of fixed depth and potential temperature. It is a simple model of the atmosphere at mid-latitudes, whose main forcings are represented by the Coriolis force and the orography or surface heating. The governing equation is the conservation

$$\frac{\mathrm{D}q_i}{\mathrm{D}t} = 0 \tag{1}$$

of the potential vorticity $q = (q_1, q_2)$ expressed in non-dimensional variables (Fandry and Leslie, 1984):

$$q_1 = \nabla^2 \psi_1 - F_1(\psi_1 - \psi_2) + \beta y \tag{2}$$

$$q_2 = \nabla^2 \psi_2 - F_2(\psi_2 - \psi_1) + \beta y + R_s \tag{3}$$

where the subscripts 1 and 2 stand for the top and bottom layer, respectively. ∇^2 is the two-dimensional Laplacian, R_s represents orography or heating, β is the (non-dimensionalised) northward variation of the Coriolis parameter at a fixed latitude, F_1 and F_2 couple the layers together being a function of Coriolis force, layer depths, gravity and typical length scale. The stream-function $\psi = (\psi_1, \psi_2)$, whose horizontal derivatives give the horizontal wind field (u_i, v_i), can be considered as the model state vector.

The code of the QG model that is distributed with the OOPS DA library have been used for this study (Y. Trémolet, personal communication). The depth of the two layers, the resolution of the horizontal grid and the integration time step Δt are the main model parameters that can be set at run-time. The dimensional scaling and model orography are fixed, as well as the extension of the domain, which is 12000 km in the zonal direction and 6300 km in the meridional direction. The domain is cyclic in the East-West direction, i.e., the model fields are periodic in this direction. The stream function is set to climatological values at meridional walls (Dirichlet boundary conditions). For all the experiments presented in this study a coarse resolution of approximately 750 km (16x8 grid points, respectively, for the East-West and North-South directions) has been used. We remind readers that the focus of this study is to test chemical DA algorithms in a toy-model framework. Therefore, there is no stringent requirement on the realism of the meteorological fields and no need to reproduce a real atmospheric situation. The only desired property is to obtain wind fields that exhibit typical patterns of the complex atmospheric circulation. A summary of the QG model parameters used in this study is detailed in Table 1.

2.2 Tropospheric chemistry

The state vector of the QG model has been extended to include chemical species. The Regional Atmospheric Chemical Mechanism RACM (Stockwell et al., 1997), which describes 96 chemical species with about 300 reactions, has been implemented. This chemical scheme has been developed for air-quality modeling and is currently used by a number of operational models in Europe (Marécal et al., 2015). Photo-chemistry and its diurnal cycle are included via look-up tables, assuming global

clear-sky conditions. Surface fluxes of chemical species (emissions and dry deposition velocities) are assigned at run-time and are kept constant during the temporal integration of the model. Chemical species are advected by the QG wind field using the semi-lagrangian scheme used for solving the QG governing equation (Eq. 1).

After the advection of the species, their concentrations are updated by addition of the chemical tendencies. They are computed by solving the stiff ODE system that describes the adopted chemical mechanism. The ODE system is of the non-linear form:

$$\partial C/\partial t = f(C) = P(C) - L(C) \cdot C \tag{4}$$

where C represents the local species concentrations, $P(C)$ and $L(C)$ the production and loss terms. The stiffness of the systems comes from the wide range of values that can take the loss terms leading to large range of chemical lifetimes.

The above system is integrated using the Adaptative Semi-Implicit Scheme (ASIS). ASIS is a one-step semi-implicit scheme with prognostic time step. To solve the system of linear equations associated to the semi-implicit scheme, ASIS uses the generalized minimal residual method (Saad and Schultz, 1986) which appears to be very competitive in terms of computation time with good convergence properties. The ASIS solver is mass conservative and adapts its sub-timestep to the adopted tolerance errors as described by Verwer (1994). In our application the ASIS solver uses a 7 s minimum sub-timestep, an absolute error tolerance of 10^4 molecules cm^{-3} and a relative tolerance error of 0.01. Therefore, a common integration time-step (dt) for both the dynamical and chemical solvers is used, which is set to dt=10 minutes to ensure reliable chemistry solutions. The dry deposition is computed for each species before the application of the ASIS solver, i.e., using the main model time step dt. Concentrations are updated according to:

$$C_i(t+1) = C_i(t) - \lambda_i C_i(t) \cdot dt \tag{5}$$

where i denotes the chemical species and λ_i is the corresponding deposition time scale in s^{-1}, proportional to the deposition velocity.

The chemical field is set equal to climatological values on the N-S boundaries, which corresponds to Dirichlet type conditions. The values of surface pressure and temperature used by the chemical mechanism are fixed globally and do not depend on the QG fields. These modeling choices let this study focus on the following main processes of air-quality models: emissions, chemistry and transport. Therefore, only the bottom layer of the QG-Chem model will be analyzed throughout this study. A summary of the configuration of QG-Chem is given in Table 1.

2.3 Description of the case study

A model run of 20 days is performed prior to DA experiments, starting from the initial condition given in Table 2, which corresponds to a homogeneous, relatively clean atmosphere and zonal circulation. Emissions (Table 3) are taken from the study of Crassier et al. (2000), FLUX case, representative of the urban environment of Paris. Spatial heterogeneity of model concentrations is obtained by scaling the reference emissions in Table 3 by 0.01, 1, 0.25, respectively, on the western, central, eastern parts of the domain (Fig. 1). Deposition velocities from the French air-quality model MOCAGE (Marécal et al., 2015),

Table 1. QG-Chem model parameters and non-dimensional scaling factors. The parameters marked by * are fixed globally and only relevant for the chemical mechanism.

Characteristic	Description
Geographical domain	12000 km (E-W) x
	6300 km (N-S)
Zonal resolution	750 km (16 grid points)
Merid. resolution	790 km (8 grid points)
Top layer depth	6 km
Bottom layer depth	4 km
Typical horizontal scale	1000 km
Typical velocity	10 m s^{-1}
Coriolis parameter F	10^{-4}
Merid. gradient of F (β)	1.5 10^{-11}
Orography	Gaussian hill (2 km alt.)
Chemical mechanism	RACM (Stockwell et al., 1997)
Surface pressure *	1000 hPa
Temperature *	24.9 °C
Boundary layer thickness *	1.2 km

averaged over the Paris region during the month of July 2010, are used and set constant over the QG-Chem domain. Values are reported in Table 4. The chemical concentrations at the meridional boundaries are set to the same values as in Table 2. Hence, the presence of N-S boundaries can eventually counterbalance the growth of long-lived species by advection of clean air-masses from outside the domain. This configuration relates to regional air pollution modeling mainly concerning the type

5 and amplitude of chemical emissions, spatial hetereogenity of sources and presence of boundaries.

Results for four key species are considered through the study: nitrogen dioxide (NO_2), ozone (O_3), carbon monoxide (CO) and carbon dioxide (CO_2). The first three are of concern for air-quality, since they have an elevated toxicity and their concentration is strongly related to anthropogenic emissions. The chemical reactivity and typical tropospheric lifetime of NO_2, O_3 and CO is, however, very different. NO_2 typically arises from the oxidation of nitric oxide (NO) in combustion processes. It is

10 highly reactive, lasting in the atmosphere from few hours in summer to some days in winter and it is an ozone precursor. O_3 is a secondary gas, mainly formed by the reaction of nitrogen oxides and hydrocarbons under sunlight. In the tropopshere, it has a typical atmospheric lifetime of 2-3 weeks. CO is produced by partial oxidation of carbon compounds, which can occur in industrial or natural combustion processes. It also participates in O_3 chemistry and has an atmospheric lifetime of about 1-2 months. Finally, CO_2 is of major concern for its effect on climate and has the longest lifetime among all the considered species

Table 2. Initial condition used to initialize the truth simulation. The initial fields are constant over the QG-Chem domain. Same values are also assigned to the meridional boundaries of QG-Chem during all simulations. Values are equal to zero for chemical species that are not listed below. Chemical concentrations are expressed in volume mixing ratio units (vmr).

Variable	Value
Meteorology	$(m \cdot s^{-1})$
(u_1, v_1)	$(40 , 0)$
(u_2, v_2)	$(10 , 0)$
Chemistry	(vmr)
O_2	0.2095
O_3	$30 \cdot 10^{-9}$
CO_2	$310 \cdot 10^{-6}$
OH	$1.5 \cdot 10^{-12}$
HO_2	$1 \cdot 10^{-13}$
H_2O_2	$1 \cdot 10^{-12}$
N_2O	$310 \cdot 10^{-9}$
NO	$0.2 \cdot 10^{-9}$
NO_2	$0.1 \cdot 10^{-9}$
HNO_3	$0.5 \cdot 10^{-9}$
HNO_4	$0.1 \cdot 10^{-9}$
CH_4	$1.6 \cdot 10^{-6}$
CO	$150 \cdot 10^{-9}$
Cl	$1 \cdot 10^{-12}$
$HOCl$	$1 \cdot 10^{-13}$
HCl	$1 \cdot 10^{-12}$
Br	$1 \cdot 10^{-13}$
BrO	$1 \cdot 10^{-13}$
HBr	$1 \cdot 10^{-13}$

(>30 years). However, CO_2 fluxes are not activated in the experiments, which makes this gas behave like a passive tracer in our study.

Averaged model fields (24 hours) on day 20 are shown in Figure 2. We note the presence of zonal gradients of chemical concentrations for species that are strongly related to surface emissions (e.g., NO_2, O_3 and CO), which are higher in the central part of the domain (Fig. 1). The developed cyclonic circulation also allows the accumulation of longer living pollutants (O_3, CO) in correspondence of the central low-pressure system, whereas patterns of short living gases (NO_2) maintain a stronger

Table 3. Surface chemical emissions used for all QG-Chem experiments. Values are equal to zero for chemical species that are not listed below. Geographical scaling is further applied to these values as shown in Fig. 1.

Species	Extended Name	Value (10^9 molec cm^{-2} s^{-1})
NO	Nitric oxide	121.29
CO	Carbon monoxide	2500
CH$_4$	Methane	802
ETH	Ethane	6.25
HC3	Alkanes, alcohols, esters and alkynes	37.67
HC5	Alkanes, alcohols, esters and alkynes	44.43
HC8	Alkanes, alcohols, esters and alkynes	19.14
ETE	Ethene	22.33
OLT	Terminal alkenes	39.67
OLI	Internal alkenes	6.37
TOL	Toluene	9.02
HCHO	Formaldehyde	5.77
ALD	Acetaldehyde and higher aldehydes	14.45
KET	Ketones	5.7
XYL	Xylene	14.55
CSL	Cresol and other aromatics	3.68

similarity with the geographical distribution of the sources. The maps also display the influence of meridional boundary conditions, which produce local minima in O$_3$ and CO fields in correspondence of the advection of clean air masses from outside the domain. The average model trajectory during 24 hours show significant differences among all considered species. Since all species are advected and surface fluxes are constant in time, these features arise from the complex chemical interactions and photo-chemistry. Note for example the daylight increase of O$_3$ as a consequence of NO$_2$ production from NO emissions during night-time and daytime photolysis. In contrast, CO shows an almost linear increase in time, due to a constant surface emission and longer lifetime. Most of the numerical experiments that are shown later in this study start on the above-discussed day.

3   Data assimilation algorithm

We considered two data assimilation algorithms in this study: 3D-Var and 4DEnVar. The first is the simplest type in the family of variational DA algorithms and currently the most used in operational chemical assimilation systems (Marécal et al., 2015). It is taken as a reference against which the benefits of more complex (and costly) algorithms can be assessed. 4DEnVar is an hybrid algorithm that combines benefits of variational and ensemble methods. It is already used in a number of NWP models

Table 4. Surface dry deposition velocity used for all QG-Chem experiments. Values are equal to a fixed and low deposition velocity of 10^{-5} m s^{-1} for chemical species that are not listed below.

Species	Extended name	Value (10^{-3} m s^{-1})
HNO_4	Perinitric acid	1.167
MACR	Metacrolein and other unsaturated monoaldehydes	0.421
HO_2	Hydroperoxy radical	8.665
TOL	Toluene	0.035
H_2O_2	Hydrogen peroxide	16.32
DCB	Unsaturated dicarbonyls	2.735
XYL	Xylene	0.035
GLY	Glyoxal	6.686
HNO_3	Nitric acid	15.31
MO2	Methyl peroxy radical	0.931
N_2O_5	Dinitrogen pentoxide	17.69
OP1	Methyl hydrogen peroxide	2.152
HONO	Nitrous acid	1.430
NO_2	Nitrogen dioxide	1.435
ALD	Acetaldehyde and higher aldehydes	0.638
OP2	Higher organic peroxides	1.348
KET	Ketones	0.622
HCHO	Formaldehyde	1.500
PAA	Peroxyacetic acid and higher analogs	1.317
ONIT	Organic nitrate	0.106
TPAN	Unsaturated PANs	1.047
PAN	Peroxyacetal nitrate and higher saturated PANs	1.099
O_3	Ozone	3.557
NO_3	Nitrogen trioxide	1.316
MGLY	Methyglyoxal ans other alpha-carbonyl aldehydes	2.804
SO_2	Sulfur dioxide	4.463
HKET	Hydroxy ketone	2.954
UDD	Unsaturated dihydroxy dicarbonyl	12.31
CSL	Cresol and other aromatics	1.323
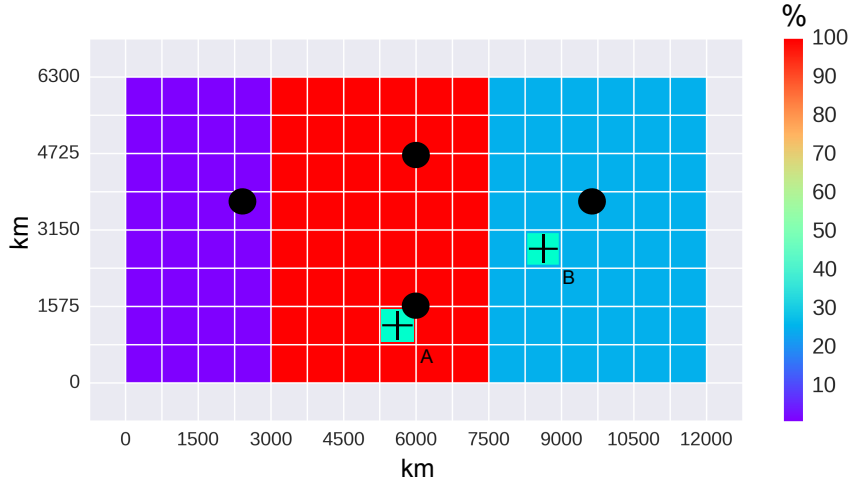
Figure 1. QG-Chem horizontal domain: scaling factor for the chemical surface emissions (in colors), location of the synthetic observations used in the assimilation experiments (black circles) and locations for which time series of DA experiments are displayed (crossed boxes A and B). The numerical grid is displayed using white lines.

(Buehner et al., 2010; Lorenc et al., 2015) and was tested in the framework of meteorological toy-models (Desroziers et al., 2014; Fairbairn et al., 2013). A summary description of the two algorithms is given below, as well as some specific aspects relative to the atmospheric chemistry implementation presented in this study. In the third section, a method based on the post-processing of 4DEnVar output is proposed to correct model biases.

5   The Object Oriented Prediction System (OOPS), a generic software framework to develop data assimilation systems (Y. Trémolet, personal communication), was used to implement and run all the DA experiments described in this study.

3.1 3D-Var

The 3D-Var analysis can be computed after a model forecast time-step by means of minimizing the quadratic cost-function J (Kalnay, 2003):

$$10 \quad J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\delta\mathbf{y}^o - \mathbf{H}\delta\mathbf{x})^T\mathbf{R}^{-1}(\delta\mathbf{y}^o - \mathbf{H}\delta\mathbf{x}) \quad\quad\quad (6)$$

for the increment $\delta\mathbf{x}$ used to correct the previous forecast $\mathbf{x}_b$ (also named background), i.e., $\mathbf{x} = \mathbf{x}_b + \delta\mathbf{x}$ where $\mathbf{x}$ is the control variable (e.g., the 3D chemical state). Here, $\mathbf{B}$ and $\mathbf{R}$ are the background and observation error covariance matrices, $\mathbf{H}$ the linearized observation operator that transforms an increment of the control variable into an increment in the observation space, $\delta\mathbf{y}^o$ the difference between the observations vector $\mathbf{y}^o$ and the previous forecast $\delta\mathbf{y}^o = \mathbf{y}^o - \boldsymbol{H}(\mathbf{x}_b)$, using the non-linear
15 observation operator. The minimization of J can be achieved with standard techniques for the solution of large linear systems: the $\mathbf{B}$-preconditioned Conjugate Gradient algorithm has been used in this study (Derber and Rosati, 1989). The result of the minimization is the analysis increment $\delta\mathbf{x}_a$ (3D). To advance in time, the analysis $\mathbf{x}_b + \delta\mathbf{x}_a$ is used as the new initial condition
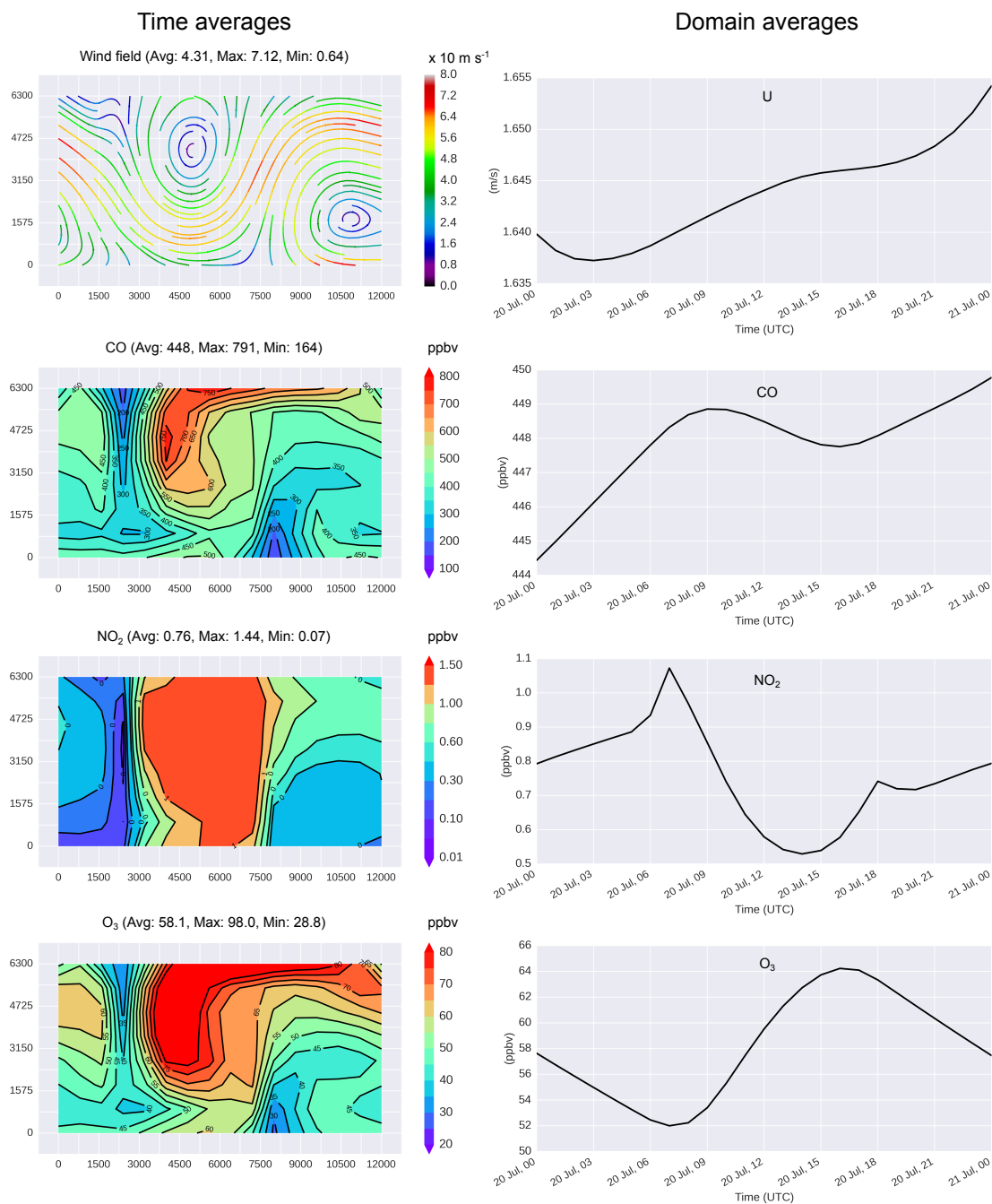
Figure 2. Meteorological and chemical fields of QG-Chem on day 20. Time-averaged fields for a 24 hours period on the left, time series of domain-averaged values for the same 24 hours period on the right. The wind field and the concentration of the four chemical species of interest (CO, NO_2, O_3, CO_2 not shown since constant and equal to 310 ppmv) are shown from top to bottom.

for the following forecast step and so forth. The relative simplicity, efficiency and robustness of the 3D-Var algorithm make it very suitable for operational models. Its practical implementation requires mainly the development of a covariance model for $\mathbf{B}$ (Weaver and Courtier, 2001). Multivariate chemical assimilation can be performed with 3D-Var by extending the 3D control variable $\mathbf{x}$ to contain multiple 3D model variables (chemical species).

5   In this study the control variable $\mathbf{x}$ is set to represent the complete model state, i.e., the stream function ψ plus the 96 chemical species. The covariance matrix $\mathbf{B}$ is modeled through the sequential application of 1D square-root correlation operators and a diagonal matrix, representing the background error standard deviation:

$$\mathbf{B}\delta\mathbf{x} = \mathbf{B}^{T/2}\mathbf{B}^{1/2}\delta\mathbf{x} \tag{7}$$

$$\mathbf{B}^{1/2} = \mathbf{\Sigma}^{1/2}\mathbf{C}_z^{1/2}\mathbf{C}_x^{1/2}\mathbf{C}_y^{1/2}\mathbf{C}_v^{1/2}, \tag{8}$$

10   where $\mathbf{C}_z$, $\mathbf{C}_x$, $\mathbf{C}_y$, $\mathbf{C}_v$ are, respectively, the vertical, zonal, meridional, multivariate correlation operators and $\mathbf{\Sigma}$ is the variance (diagonal matrix). $\mathbf{C}_x$ and $\mathbf{C}_y$ are isotropic homogeneous correlation operators providing Gaussian spatial structures. The other correlation operators are represented by symmetric positive-definite matrices. The following parameters are used to set $\mathbf{B}$: one horizontal length scale that defines the decorrelation scale for the zonal and meridional coordinates, one value for the vertical correlation and one value for the chemical correlation between each couple of model variables. The variance is specified using

15   one global value for each variable. The resulting $\mathbf{B}$ is uniform and homogeneous on the horizontal plane. More complex $\mathbf{B}$ models could be introduced, to account for example for spatial variability and heterogeneity of the background error covariance. However, this is out of the scope of the present study, which is intended to reproduce typical operational chemical DA settings, where $\mathbf{B}$ is usually specified using a single variance and correlation length for each chemical species.

     Only surface observations are considered in this study. Therefore, the observation operator $\mathbf{H}$ is represented by the bi-linear

20   interpolation of model values to the observation location. A diagonal observation error covariance $\mathbf{R}$ is used, as it is the case in most real DA systems.

     One drawback of 3D-Var is that DA results rely strongly on the background error covariance $\mathbf{B}$, which should depend on the previous assimilation cycles and forecast errors (flow-dependence). In practical applications, $\mathbf{B}$ is usually set constant, estimated from verification of previous forecasts (climatological) and/or tuned to provide the best fit of the analyses against

25   independent observations. In the case of multivariate chemical assimilation, the estimation and validity of climatological error covariances between chemical species has not yet been demonstrated. As a consequence, multivariate corrections are normally neglected ($\mathbf{C}_v = \mathbf{I}$). This simplification is also used in this study. Finally, 3D-Var provides a correction to the control variable each time the system is observed (every hour in air-quality applications). This does not allow to exploit the dynamical information contained in observations time series and prevent all estimation of model error terms.

30   3.2   4DEnVar

The 4DEnVar algorithm can solve the main drawbacks of 3D-Var by introducing the temporal dimension in the quadratic cost function (6), similarly to what the classic 4D-Var algorithm does (Dimet and Talagrand, 1986). However, compared to the latter, it avoids the introduction of the tangent linear and adjoint codes of the forecast model. Following the notation in Desroziers

et al. (2014), the 4DEnVar cost function can be written as:

$$J(\underline{\delta \mathbf{x}}) = \frac{1}{2} \underline{\delta \mathbf{x}}^T \underline{\mathbf{B}}_e^{-1} \underline{\delta \mathbf{x}} + \frac{1}{2} (\underline{\delta \mathbf{y}}^o - \underline{\mathbf{H}} \underline{\delta \mathbf{x}})^T \underline{\mathbf{R}}^{-1} (\underline{\delta \mathbf{y}}^o - \underline{\mathbf{H}} \underline{\delta \mathbf{x}}) \tag{9}$$

where all underlined terms are now time-dependent (4D). The control variable $\underline{\mathbf{x}}$ becomes the temporal trajectory of the model state (ψ plus the 96 chemical species in this study). The cost function is computed for an assimilation window that can span several hours or days. The assimilation window is discretized with an arbitrary number of sub-windows, which defines the temporal dimension of the 4D vectors and matrices in (9). The minimization of J returns a 4D vector $\underline{\delta \mathbf{x}}_a$, which provides the analysis trajectory for the entire assimilation window.

The forecast error covariance $\underline{\mathbf{B}}_e$ is estimated from an ensemble of perturbed model trajectories:

$$\underline{\mathbf{B}}_e = \frac{1}{L-1} (\underline{\mathbf{x}}_1', ..., \underline{\mathbf{x}}_L')(\underline{\mathbf{x}}_1', ..., \underline{\mathbf{x}}_L')^T \tag{10}$$

where $\underline{\mathbf{x}}'$ are the four-dimensional perturbation and L denotes the size of the ensemble. It results that $\underline{\mathbf{B}}_e$ describes spatial (3D), multivariate and temporal covariances at once. An analogy with the 4D-Var weak constraint cost function shows that $\underline{\mathbf{B}}_e$ represents a numerical approximation of the linearized forecast model and model error covariance (Desroziers et al., 2014). As for the 3D-Var case, the minimization of J is achieved using a $\mathbf{B}$-preconditioned Conjugate Gradient algorithm (Algorithm n. 3 in Desroziers et al. (2014)). $\underline{\mathbf{H}}$ and $\underline{\mathbf{R}}$ are block diagonal matrices whose blocks are the operators $\mathbf{H}$ and $\mathbf{R}$ defined in Sec. 3.1 for each sub-window.

A nice property of the 4DEnVar algorithm is that the effect of the model error covariance, generally denoted by $\mathbf{Q}$ (Tremolet, 2006), can be introduced easily by adding physical perturbations during the computation of the ensemble of forecasts. This lets the model dynamics develop the complex covariances that derive from physical perturbations, without the need of specify and implement a covariance model for $\mathbf{Q}$. Whenever the sole initial condition of the ensemble is perturbed at the beginning of the assimilation window, the 4DEnVar approximates the 4D-Var algorithm, in the strong constraint formulation. The flexibility of the model error specification in 4DEnVar is a strong asset for atmospheric chemistry DA, where the sources of uncertainty can be highly variable and a function of the chemical species.

The main drawback of 4DEnVar in practical applications to large scale models is related to the finite size of the ensemble. As a consequence the 4D error covariance $\underline{\mathbf{B}}_e$ is not full-ranked and covariance terms may contain statistical noise. The effect of a noisy covariance is the appearance of spurious analysis increments far away from assimilated observations, or between physically uncorrelated model variables (in the multivariate case). Since with 4DEnVar temporal covariances are also estimated through the ensemble, this effect concerns also the time dimension. This is a typical issue with all ensemble-based methods and large systems, and demands the introduction of a *localization* operator, which attenuate non-local increments. The localization is applied to $\mathbf{B}_e$ by:

$$\underline{\mathbf{B}} = \underline{\mathbf{B}}_e \circ \underline{\mathbf{C}} \tag{11}$$

where $\circ$ denotes the Schur product (entrywise product) and $\underline{\mathbf{C}}$ is a 4D correlation operator that damps non-local covariances. The numerical implementation of (11) is made under the following approximations: the same 3D (and multivariate) correlation

operator $\mathbf{C}$ is used for all 4DEnVar sub-windows. It follows that $\underline{\mathbf{C}}$ is a block matrix with all elements set equal to $\mathbf{C}$. Hence, in order to specify $\underline{\mathbf{C}}$, we could use the covariance operator described in (7) by setting the variance terms to one. The choice of parameters of the correlation operators is discussed in the results section. This simplification, also called static localization, reduces significantly the numerical cost of the algorithm (Desroziers et al., 2014) at the price of degraded precision when

5 increasing the length of the DA windows (Bocquet, 2016). The development of localization procedures that are more consistent with the dynamics of the forecast model is an ongoing research topic (Bocquet, 2016; Desroziers et al., 2016) and possible applications to QG-Chem will be considered in a future study.

3.3 Diagnosis of model error and forecast correction with 4DEnVar

One of the objectives of this study is to retrieve model error information from the 4DEnVar solution, which can be potentially

10 used to improve the chemical forecasts for the next day. The 4DEnVar analysis increment $\underline{\delta\mathbf{x}}_a$ accounts for the correction of both the initial condition and the model forecast (model error). However, the control variable $\underline{\delta\mathbf{x}}_a$ only contains the model state. It follows that the correction of the initial condition is simply given by $\delta\mathbf{x}_a^{(t=0)}$, i.e., the first element of the 4D vector $\underline{\delta\mathbf{x}}_a$. The correction of the model error contributes to the values of $\delta\mathbf{x}_a^{(t>0)}$ but a diagnostic procedure has to be applied to retrieve it.

We can compute a forecast trajectory $\underline{\mathbf{x}}_f$ using the non-linear model starting from the updated initial condition $\mathbf{x}_b^{(t=0)} +$

15 $\delta\mathbf{x}_a^{(t=0)}$. Subtracting it from the 4DEnVar analysis at $t > 0$, we obtain

$$\Delta\underline{\mathbf{x}}^{(t>0)} = \underline{\mathbf{x}}_a^{(t>0)} - \underline{\mathbf{x}}_f^{(t>0)}. \tag{12}$$

This difference contains information about the contribution of the model error on the 4D state, as explained below. Let us first define the model error $\boldsymbol{\eta}$ and analysis error $\underline{\mathbf{e}}_a$ as

$$\boldsymbol{\eta}^{(t)} = \mathbf{x}_*^{(t)} - \mathcal{M}_{(t-1)\to(t)}(\mathbf{x}_*^{(t-1)}) \tag{13}$$

20 $$\mathbf{e}_a^{(t)} = \mathbf{x}_a^{(t)} - \mathbf{x}_*^{(t)} \tag{14}$$

where $\mathbf{x}_*^{(t)}$ is the truth at time t, $\mathcal{M}_{(t-1)\to(t)}$ is the integration of the model from time $(t-1)$ to time t, with the temporal discretization matching the length of 4DEnVar sub-windows (Sec. 3.2). From (13), we have

$$
\begin{aligned}
\boldsymbol{\eta}^{(t+1)} &= \mathbf{x}_*^{(t+1)} - \mathcal{M}_{(t)\to(t+1)}(\mathbf{x}_*^{(t)}) \\
&= \mathbf{x}_*^{(t+1)} - \mathcal{M}_{(t)\to(t+1)}(\boldsymbol{\eta}^{(t)} + \mathcal{M}_{(t-1)\to(t)}(\mathbf{x}_*^{(t-1)})) \\
&= \mathbf{x}_*^{(t+1)} - \mathcal{M}_{(t)\to(t+1)}(\boldsymbol{\eta}^{(t)} + \mathcal{M}_{(t-1)\to(t)}(\boldsymbol{\eta}^{(t-1)} + \mathcal{M}_{(t-2)\to(t-1)}(\mathbf{x}_*^{(t-2)}))) \\
&= ... \\
&\simeq \mathbf{x}_*^{(t+1)} - \mathcal{M}_{(0)\to(t)}(\mathbf{x}_*^{(0)}) - \sum_{j=1}^{t}\prod_{l=j+1}^{t+1}\mathbf{M}_l\boldsymbol{\eta}^{(j)}
\end{aligned}
\tag{15}
$$

where we have used the Taylor expansion and assumed that the forecast model can be linearized inside each sub-window: $\mathcal{M}_{(t-1)\to(t)} \sim \mathbf{M}_t$. This equality can be rewritten as

$$\mathbf{x}_*^{(t+1)} - \mathcal{M}_{(0)\to(t)}(\mathbf{x}_*^{(0)}) \simeq \boldsymbol{\eta}^{(t+1)} + \sum_{j=1}^{t} \prod_{l=j+1}^{t+1} \mathbf{M}_l \boldsymbol{\eta}^{(j)} \tag{16}$$

Using the definition (14) and the linear assumption on the model, it yields from (12) that

$$
\begin{aligned}
5 \quad \Delta\mathbf{x}^{(t)} &= \mathbf{x}_a^{(t)} - \mathcal{M}_{(0)\to(t)}(\mathbf{x}_a^{(0)}) \\
&= \mathbf{x}_a^{(t)} - \mathbf{x}_*^{(t)} + \mathbf{x}_*^{(t)} - \mathcal{M}_{(0)\to(t)}(\mathbf{x}_a^{(0)} - \mathbf{x}_*^{(0)} + \mathbf{x}_*^{(0)}) \\
&= \mathbf{e}_a^{(t)} + \mathbf{x}_*^{(t)} - \mathcal{M}_{(0)\to(t)}(\mathbf{e}_a^{(0)} + \mathbf{x}_*^{(0)}) \\
&\simeq \mathbf{e}_a^{(t)} + \mathbf{x}_*^{(t)} - \mathcal{M}_{(0)\to(t)}(\mathbf{x}_*^{(0)}) - \prod_{l=0}^{t} \mathbf{M}_l \mathbf{e}_a^{(0)}
\end{aligned}
\tag{17}
$$

Substituting the equality (16) into the latter, we get

$$10 \quad \Delta\mathbf{x}^{(t)} \simeq \mathbf{e}_a^{(t)} - \prod_{l=0}^{t} \mathbf{M}_l \mathbf{e}_a^{(0)} + \boldsymbol{\eta}^{(t)} + \sum_{j=1}^{t} \prod_{l=j+1}^{t} \mathbf{M}_l \boldsymbol{\eta}^{(j)}. \tag{18}$$

Therefore, if the error $(\mathbf{e}_a^{(t)} - \prod_{l=0}^{t} \mathbf{M}_l \mathbf{e}_a^{(0)})$ is small, the vector $\Delta\mathbf{x}^{(t)}$ represents the contribution of the model error $\underline{\boldsymbol{\eta}}$ on the 4D state, accumulated through the preceding 4DEnVar sub-windows. Hereafter, we will name $\Delta\mathbf{x}^{(t)}$ *effective* model error, to be distinguished from the model error, which is generally associated with $\underline{\boldsymbol{\eta}}$.

In a first instance, $\Delta\underline{\mathbf{x}}$ can be used to diagnose the underlying presence of model error in the forecast. Furthermore, if the
15 model error is stationary along multiple DA windows (bias), $\Delta\underline{\mathbf{x}}$ could be used to correct the forecast for the next window:

$$\tilde{\underline{\mathbf{x}}}_f^{i+1} = \underline{\mathbf{x}}_f^{i+1} - \Delta\underline{\mathbf{x}}^i \tag{19}$$

where the superscript denotes the assimilation window index and the $\tilde{\underline{\mathbf{x}}}_f$ is the corrected forecast. Validation of operational air-quality models shows that biases contribute for a large part of the model uncertainties (Marécal et al., 2015; Zyryanov et al., 2012; Huijnen et al., 2010), which justifies the implementation of the proposed bias correction procedure. Note that this
20 procedure is compatible with model biases that show an hourly variability but are stationary on successive days, as it is found in most of air-quality models (Marécal et al., 2015; Gaubert et al., 2014). Note also that the computation of the effective model error is blind to the type of the underlying model error (e.g., chemical emissions or physical parametrizations). The two main requirements needed to make a useful estimation of $\Delta\underline{\mathbf{x}}$ are: i) analysis errors smaller than model errors and ii) an approximate knowledge of the sources of model error, necessary to generate informative ensembles.

25 Alternatively from the proposed procedure, model error terms $\boldsymbol{\eta}^{(t)}$ could be diagnosed for each sub-window and then applied to the next day forecast on an hourly basis. However, this correction method is more intrusive, because it must be applied during the non-linear model forecast and was not considered for this study.

4 Results and discussion

Numerical experiments are described and discussed in this section. The objective is to assess the performances of the 4DEn-Var algorithm for the assimilation of the four key species: NO_2, O_3, CO and CO_2. In all experiments a model simulation with unperturbed parameters (same as the one in Sec. 2.3) represents the truth. The experiments are performed during the meteorological-chemical situation described in Sec. 2.3. Synthetic observations are generated from the truth applying $\mathbf{H}$ (Sec. 3.1) and by adding a normally distributed error (Table 5). Four observation locations are considered for the experiments (Fig. 1), where chemical species are observed hourly. The relatively low density of the observation network allows to assess DA skills at unobserved locations. Model forecasts are produced by perturbing only the initial condition (perfect model), the surface emissions (model error) or both. DA is performed with either 3D-Var or 4DEnVar, and the obtained analyses are compared to the truth. The meteorology is never observed nor perturbed, which corresponds to use QG-Chem in a CTM-like mode.

Operational air-quality centers collect hourly observations and perform DA typically once per day (Marécal et al., 2015). Therefore, a 24 hours assimilation window is adopted when using the 4DEnVar algorithm, with 1 hour sub-windows matching the observations frequency. For the same reason, 24 sequential cycles of 1 hour are adopted with 3D-Var. The main processes affecting air-quality forecasts (daily emissions, evolution of the mixing layer, photo-chemistry) have a period of approximately 24 hours. Therefore, a 24 hour window permits to account for errors on main model processes. The utilization of longer windows is theoretically possible with 4DEnVar, assuming that the linearization of the model perturbations remains valid. However, the numerical cost of the minimization increases with the windows length when keeping fixed the duration of the sub-windows. The costs-benefit ratio of longer windows will deserve further investigations in real applications.

Section 4.1 presents results based on a perfect model hypothesis and compares 3D-Var and 4DEnVar performances for chemical reanalyses. Section 4.2 introduces model errors and is focused on the estimation of the effective model error using 4DEnVar. Section 4.3 summarizes the results on a larger number of DA windows: 4DEnVar results using the model bias correction procedure are compared to 3D-Var ones for both reanalyses and forecasts.

4.1 Perfect model experiments

Experiments considering a perfect model are presented in this section for one assimilation cycle of 24 hours. Emphasis is placed on the reanalysis capabilities of DA. A 24 hours long forecast is produced by perturbing only the initial condition. Initial perturbations are computed applying $\mathbf{B}^{1/2}$ (Eq. 8) to a Gaussian uncorrelated noise field $N(0, \mathbf{I})$ at $t = 0$. Since chemical concentrations can span different orders of magnitudes in the atmosphere, the standard deviation values set in $\mathbf{B}^{1/2}$ depend on the chemical species (Table 5). The background error standard deviations have been chosen to be about 10-20% of the average field values (Fig. 2). The same horizontal correlation length has been set for all species (750 km). Multivariate correlations in $\mathbf{B}^{1/2}$ are switched off so that perturbations of chemical species are not correlated at $t = 0$. Chemical correlations can, however, arise at $t > 0$ due to chemical couplings.

DA is applied to correct the 24 hours long perturbed forecast. The same covariance matrix $\mathbf{B}$ that was used to produce the initial perturbation is applied at each cycle of the 3D-Var analysis (1 hour). Hence, the background error covariance used in

Table 5. Background (σ_B) and observation (σ_O) error standard deviation used in DA experiments expressed in volume mixing ratio units and, within brakets, in percentage of the average field value in Figure. 2.

Species	σ_B	σ_O
NO_2	$1.2 \cdot 10^{-10}$ (16%)	$0.9 \cdot 10^{-10}$ (11%)
O_3	$12.2 \cdot 10^{-9}$ (21%)	$3.2 \cdot 10^{-9}$ (5%)
CO	$64.8 \cdot 10^{-9}$ (14%)	$8.1 \cdot 10^{-9}$ (2%)
CO_2	$40.5 \cdot 10^{-6}$ (13%)	$20.2 \cdot 10^{-6}$ (6%)

DA is perfectly known at $t = 0$, which represents the best possible case in DA. At $t > 0$ the true $\mathbf{B}$ depends on the observations assimilated at previous steps and on the model dynamics, which makes the use of a fixed $\mathbf{B}$ within 3D-Var a raw approximation. However, this is the typical setting of operational air-quality models.

With 4DEnVar, an ensemble of 16 forecasts is generated perturbing the initial condition with the same $\mathbf{B}$ as above. Therefore,

5   the ensemble size is small compared to the dimensions of the system (16·8·97=12416 variables). When using small size ensembles, the localization of the sample covariance is necessary. In this study, horizontal localization is applied to the 4D ensemble covariance using Eq. (11), by setting a horizontal length scale equal to the double of values used for $\mathbf{B}$ (1500 km). Multivariate localization is performed using a correlation coefficient of 0.5, which was chosen empirically. A sensitivity analysis concerning the ensemble size and the localization choices is presented later in this section. Since we use QG-Chem

10 in a CTM-like configuration and the two layers are chemically not coupled, the vertical terms of the covariance or localization matrix are always set to zero in this study, without any impact on the presented results.

4.1.1   Univariate assimilation

This section compares results of 3D-Var and 4DEnVar DA in univariate settings, i.e., one independent assimilation experiment is performed for each of the four chemical species. The controlled species corresponds to the species that is measured and that

15 is perturbed at the initial time. With 3D-Var, this is obtained setting all terms of $\mathbf{B}$ to zero except for the 3D covariance of the selected species. With 4DEnVar the same is obtained setting to zero all multivariate localization coefficients.

Figure 3 compares the temporal trajectories of the analysis of each species obtained from 3D-Var and 4DEnVar. Each figure provides the temporal trajectories at the two grid points shown in Fig. 1: one located in the polluted region and in correspondence of assimilated observations (grid point A) and one in the cleaner region and slightly displaced from the measurements

20 location (grid point B).

In addition to the comparison of the two DA algorithms, these experiments also permit to assess the impact of the initial perturbation on chemical forecasts. First of all, we remark that O_3 forecasts are very close to the truth values after 24 hours, which is a consequence of the fact that O_3 is strongly controlled by precursor emissions and photo-chemistry. The memory of the initial condition is rapidly lost for O_3, as it was also demonstrated within regional air-quality models (Jaumouillé et al.,

2012; Wu et al., 2008). This is not the case for CO_2 and CO, which have a longer lifetime. Therefore, the initial perturbation is advected by the wind field and the spread between the forecast and the truth lasts longer in time. NO_2, which lasts a few hours in a summer-time atmosphere is practically not sensitive to the perturbation of the initial condition (Fig. 3). Note also that the chemical concentrations are always lower at the clean location (grid point B) than at the polluted one (grid point A), except for CO_2 that is neither emitted nor chemically produced. Again, this is a consequence of the geographical variability of the emission factors (Fig. 1).

We remark that the 4DEnVar provides in general better analyses than 3D-Var for all species. First, it can be observed from Fig. 3 that the analysis time-series obtained with the 4DEnVar are smoother than those resulting from 3D-Var because daily trajectories are optimized at once with 4DEnVar. The sequential aspect of 3D-Var, instead, makes the analysis more sensitive to the random observation errors. This introduces the observed jumps in the analyses.

Figure 4 provides the Root Mean Square Error (RMSE) gain (in %) for every grid point (i,j) of the model domain,

$$\text{RMSE}_{(i,j)}^{gain} = \frac{1}{\bar{X}_{(i,j)}} \left(\text{RMSE}_{(i,j)}^{anl} - \text{RMSE}_{(i,j)}^{fct} \right). \tag{20}$$

Here, $\bar{X}_{(i,j)}$ is the average concentration of the truth values (Fig. 2), RMSE^{fct} and RMSE^{anl} are the absolute values of RMSE (in chemical concentration units) for the forecast and analysis, respectively. The RMSE at the location (i,j) is defined as

$$\text{RMSE}_{(i,j)} = \sqrt{\frac{1}{N} \left(\underline{\mathbf{x}}_{(i,j)}^* - \underline{\mathbf{x}}_{(i,j)} \right)^T \left(\underline{\mathbf{x}}_{(i,j)}^* - \underline{\mathbf{x}}_{(i,j)} \right)}, \tag{21}$$

where $N = 24$ is the total number of sub-windows, $\underline{\mathbf{x}}_{(i,j)}^*$ and $\underline{\mathbf{x}}_{(i,j)}$ denotes the temporal trajectories of the truth and the analysis (or forecast) at the grid point (i,j), respectively. In Fig. 4 blue color means that RMSE values after DA has reduced (DA improved the forecast) and red color means that RMSE values after DA has increased (DA degraded the forecast). Therefore, we can see from these figures that with 4DEnVar, improvements of the RMSE at unobserved locations are more pronounced than those of 3D-Var (more widespread blue regions and less red regions in RMSE gain). This is likely due to a better description of the background error covariance, which is flow-dependent and closer to the true forecast error covariance within 4DEnVar. For example, the background error covariance used to assimilate NO_2 with 3D-Var is highly overestimated for $t > 0$. This happens because, when the model is perfect, the NO_2 field rapidly converges to the truth after a few hours (Fig. 3). In this specific case, the RMSE is degraded even at observed locations (Fig. 4, bottom left plot). This is an instructive example of the effects of incorrectly specified $\mathbf{B}$ within 3D-Var. Finally, note that for other species also 3D-Var is capable of decreasing RMSE at unobserved locations. This effect is a result of the advection of the analysis increments, and, as expected, is more pronounced with long-lived species (CO, CO_2) than with more reactive or emitted gases (O_3, NO_2).

Table 6 reports the minimum, maximum and average of the field values in Fig. 4, with the sign inversed to display positive values for positive gain of DA, negative otherwise. It can be seen that with 4DEnVar, the maximum degradation of RMSE (i.e. the minimum gain in absolute values) is always smaller by a factor 2 to 5 than the maximum gain, and the average RMSE gain is always positive, i.e., DA improves the forecast. The appearance of local but relatively small RMSE degradation can be tolerated in atmospheric chemistry, because the chemical system has a dissipative behavior and errors on the model state cannot grow during the forecast step. Since the error covariance of the initial condition is perfectly known in the experiments setup,

20
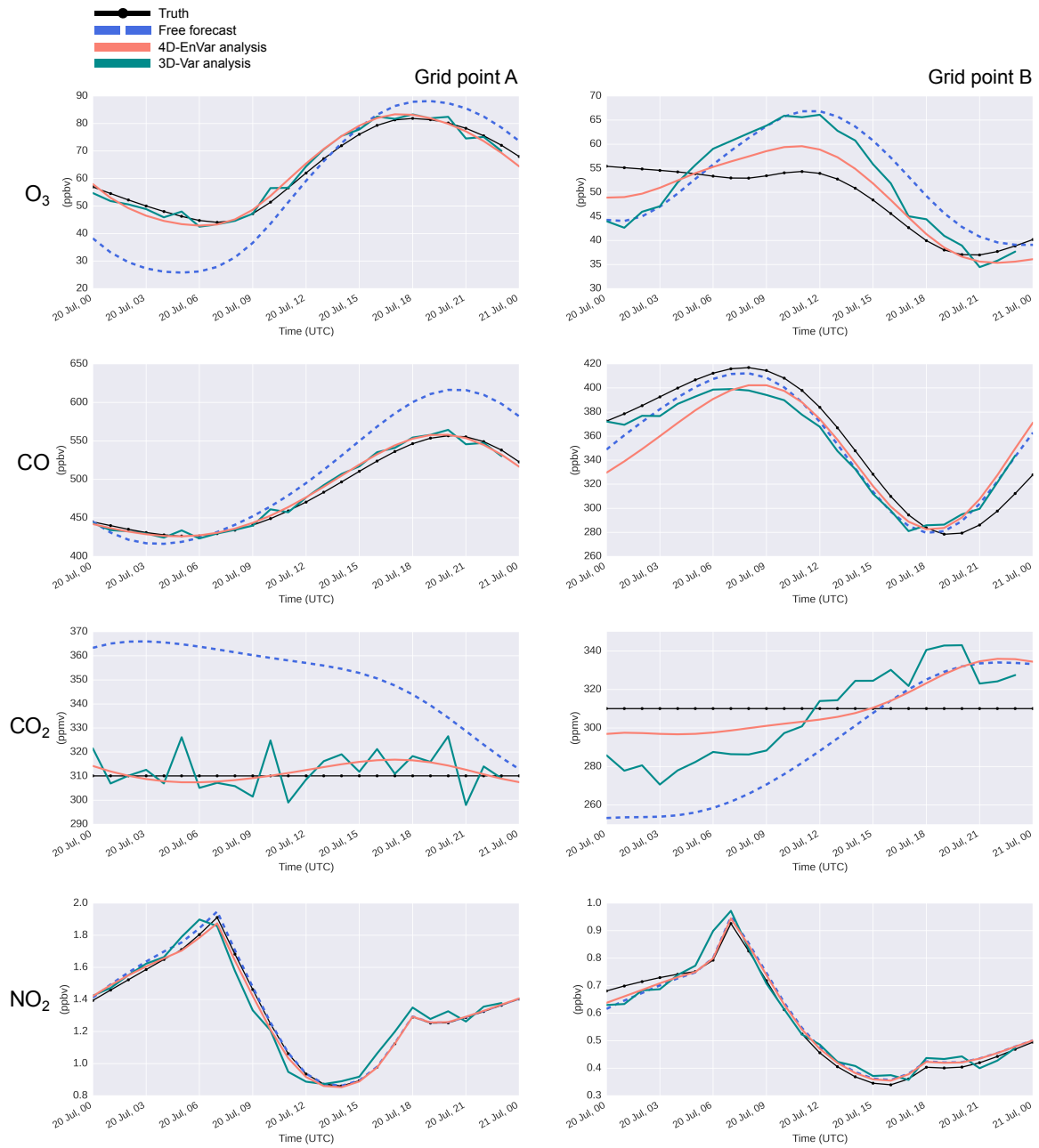
Figure 3. DA results with a perfect model hypothesis. Temporal trajectories for the three assimilated species are shown from top to bottom (O_3, CO and CO_2). Forecast, 3D-Var/4DEnVar analyses and truth are shown in each plot for the grid points A (left plots) and B (right plots) depicted in Fig. 1.
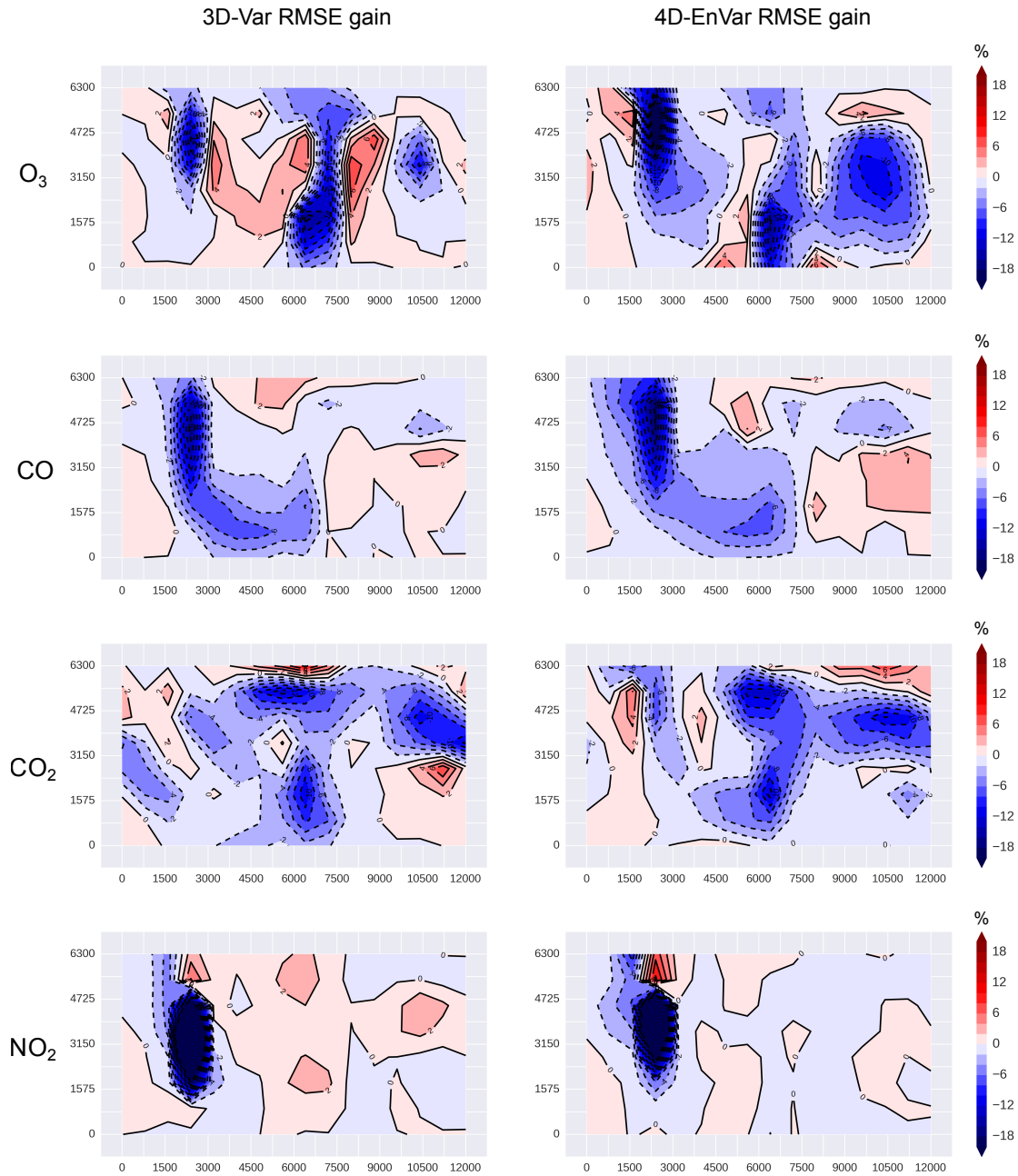
Figure 4. DA results with a perfect model hypothesis. From top to bottom: the RMSE gain (Eq. 20) is displayed, respectively, for O_3, CO, CO_2 and NO_2 assimilation experiments. Blue color means that DA lowered the RMSE, red color means that DA increased the RMSE. Plots on the left are obtained using 3D-Var, on the right using 4DEnVar.

Table 6. Summary statistics of RMSE gain for perfect model DA experiments in Fig. 4. Signs have been inverted compared to Figures, to show positive values when DA reduces the RMSE.

	3D-Var			4D-EnVar		
Species	Max. Gain (%)	Min. Gain (%)	Avg. Gain (%)	Max. Gain (%)	Min. Gain (%)	Avg. Gain (%)
O_3	17.11	-7.67	0.92	27.03	-6.68	2.55
CO	17.15	-3.33	1.22	18.82	-4.11	1.62
CO_2	13.08	-10.60	1.76	13.62	-7.51	1.93
NO_2	57.39	-6.15	0.95	48.07	-10.52	1.21

the degradation of the 4DEnVar analysis is a consequence of the algorithm hypotheses (e.g., the linearization of the forecast model) or of the numerical implementation (e.g., the finite size of the ensemble and the localization approximations).

4.1.2 Multivariate assimilation

A second set of DA experiments has been performed, but perturbing and assimilating the four chemical species at the same time. Therefore, one 24 hours long forecast and the corresponding analysis has been computed for all species, instead of four independent analyses as before. In the case of 3D-Var, elements of **B** related to the four assimilated species are set using the same parameters as in Sec. 4.1.1. Elements related to unobserved species are kept at zero as well as for cross-variable correlations. This leads to a multi-species assimilation, which is not yet multivariate. Effects on unobserved variables or between species are still permitted by chemical couplings in the forecast model.

The gain on RMSE obtained with 3D-Var for the four species (not shown) is very similar to those obtained in Sec. 4.1.1 with independent experiments.

With 4DEnVar, the corresponding multi-species assimilation has been tested by setting to zero the cross-variable correlation coefficients in the localization operator. In addition, a multivariate case has also been examined, by setting the correlation coefficients to 0.5. In both cases results (not shown) were found to be again very similar to those in Fig. 4.

These results indicate that, when the initial condition is solely taken as a source of uncertainty, the chemical coupling between species does not influence much DA. This is confirmed by the ensemble standard deviation of the 4DEnVar experiments in Sec. 4.1.1. For each of the four assimilation experiments the average ensemble standard deviation has been computed for all species (perturbed and not perturbed at $t = 0$). The ensemble standard deviation of unperturbed species stays below 1% of the local concentration (not shown), compared to typical values of about 10-15% for the species that are perturbed initially. Therefore, the perturbation of the sole initial state does not affect significantly the chemical balance. This also justifies neglecting the cross-correlations between chemical species in operational systems that assimilate hourly observations sequentially. This result was expected for weakly reacting species like CO_2 or CO but was not evident for reacting gases such as O_3 and NO_2. A possible reason is that the amount of NO_2 produced hourly by the oxidation of emitted NO is much larger than the applied initial

23

perturbation. Therefore, the O_3 photochemical production, which happens later during the day, is not much influenced by the perturbation of NO_2 at midnight. It would be interesting to verify if similar results also hold when larger perturbations are applied during day-time. A wider exploration of different chemical regimes is left for a future study.

4DEnVar was capable to provide similarly good results as in Sec. 4.1.1 when enabling the cross-variable covariances (and the
5   respective localization). This means that the noise of chemical cross-covariances due to the small ensemble size (16 members) did not degrade the results. This leaves hope for an effective multivariate chemical assimilation, when the role of chemical couplings becomes larger (Sec. 4.2).

4.1.3   Ensemble size and localization

This section examines the impact of the principal parameters of the 4DEnVar algorithm, i.e., the ensemble size (16 members in
10 previous experiments) and the localization length scale, on the analysis RMSE. These two aspects are closely linked with the dimension of the system and the eigenspectrum of the error covariance matrices (Furrer and Bengtsson, 2007). For instance, for a strongly decaying spectrum relatively few ensemble members are enough to provide a good approximation of the error covariance matrix. If this is not the case, a larger ensemble size allows in principle less severe localization. However, no theoretical formulation exists already for the 4DEnVar that links these parameters, e.g., the localization scale with the ensemble
15 size. Moreover, the number of parameters of the localization operator (e.g., the horizontal length scale) might depend on the chemical species. A full exploration of the parameter space becomes rapidly unpractical even in a simplified model framework. Finally, the approximations made in the implementation of the localization operator (Sec. 3.2) can potentially have a larger effect than the choice of the parameters themselves. Therefore an empirical approach has been used in this paper, and we postpone a detailed theoretical analysis to future work.

20   In the empirical approach, first the ensemble size is examined using fixed but reasonable values for the localization operator (Sec. 4.1). The ensemble size can be one of the main limiting factors in operational forecast centers and the smallest ensemble providing better results than 3D-Var for all species was taken. Second, the impact of horizontal and chemical localization are examined keeping the selected ensemble size fixed.

Results for the case study are shown in Fig. 5 (varying ensemble size), Fig. 6 (varying horizontal localization) and Fig. 7
25 (varying cross-variable localization). The minimum, maximum and mean RMSE gain (Table 6) of the analysis are considered to compare 4DEnVar and 3D-Var experiments. Hence, the plots display the differences between the RMSE gain of 4DEnVar minus the 3D-Var one, with the sign opportunely adjusted to display positive values when 4DEnVar beats 3D-Var, negative values otherwise. In general, the desired case is represented by all bars (mean, maximum and minimum gain) being positive. Nevertheless, a positive average gain with the sporadic occurrence of negative maximum/minimum gains can be tolerated, if
30 the values for the minimum gain bar do not become too negative. Since the minimum RMSE gain is already negative (Table 6), negative values for the corresponding bars in Figs. 5, 6, and 7 mean that the degradation of the 4DEnVar analysis is larger than the 3D-Var one, which is not desired.

Results are very similar using 64 or 128 members, for all species, suggesting that some convergence of the assimilation scores is achieved with more than 64 members (Fig. 5). As expected, the accuracy starts to decrease using less than 32 members, with
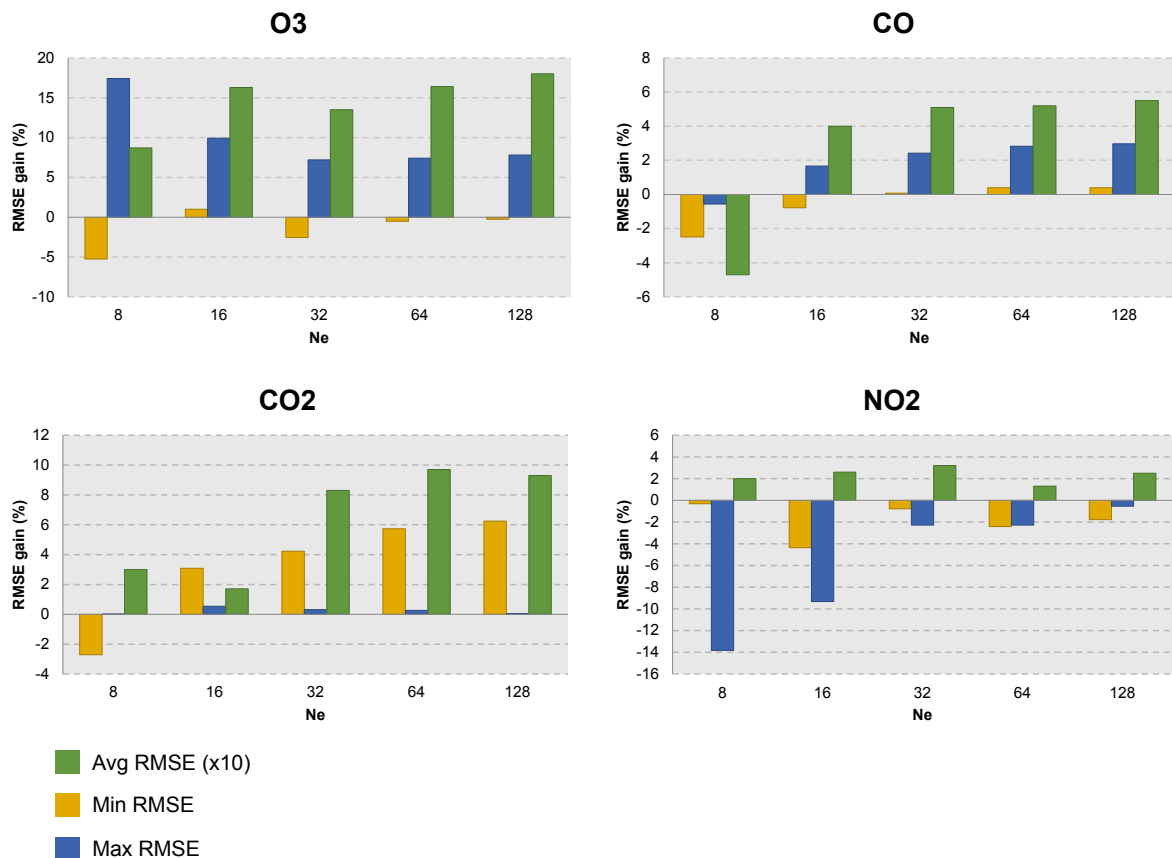
Figure 5. Impact of ensemble size (Ne) on the RMSE gain of the 4DEnVar analyses. Results are shown for the four chemical species O_3, CO, CO_2, NO_2. For each plot values of maximum, minimum and average RMSE gain (Table 6) are shown in different color, compared to reference results obtained with 3D-Var. Positive values mean better RMSE gain with 4DEnVar than with the reference 3D-Var. Average RMSE gain has been multiplied by ten to better highlight the differences among the experiments.

most of the gain of 4DEnVar over 3D-Var being lost using only 8 members. In a few cases, the RMSE gain occasionally decreases when increasing the size of the ensemble. This is due to statistical fluctuations when the ensemble sizes are small. To avoid misinterpretation of statistical noise, the comparison of 3D-Var and 4DEnVar is repeated in Sec. 4.3 for a larger number of DA windows. We remark finally, that results with 16 members satisfy the requirements expressed above: better average

5 RMSE gain than 3D-Var for all species and a limited number of cases of RMSE degradation. Hence, an ensemble size of 16 members is retained. We remind readers that the objective of this study is to demonstrate the applicability of a DA algorithm that could outperform currently implemented methods in operational centers, but with an acceptable compromise between computational costs and precision. Therefore, even if an ensemble size of 32 or 64 would have represented a more accurate option for this study, we found it more valuable to assess the potential of 4DEnVar when computational resources might be

10 limited.
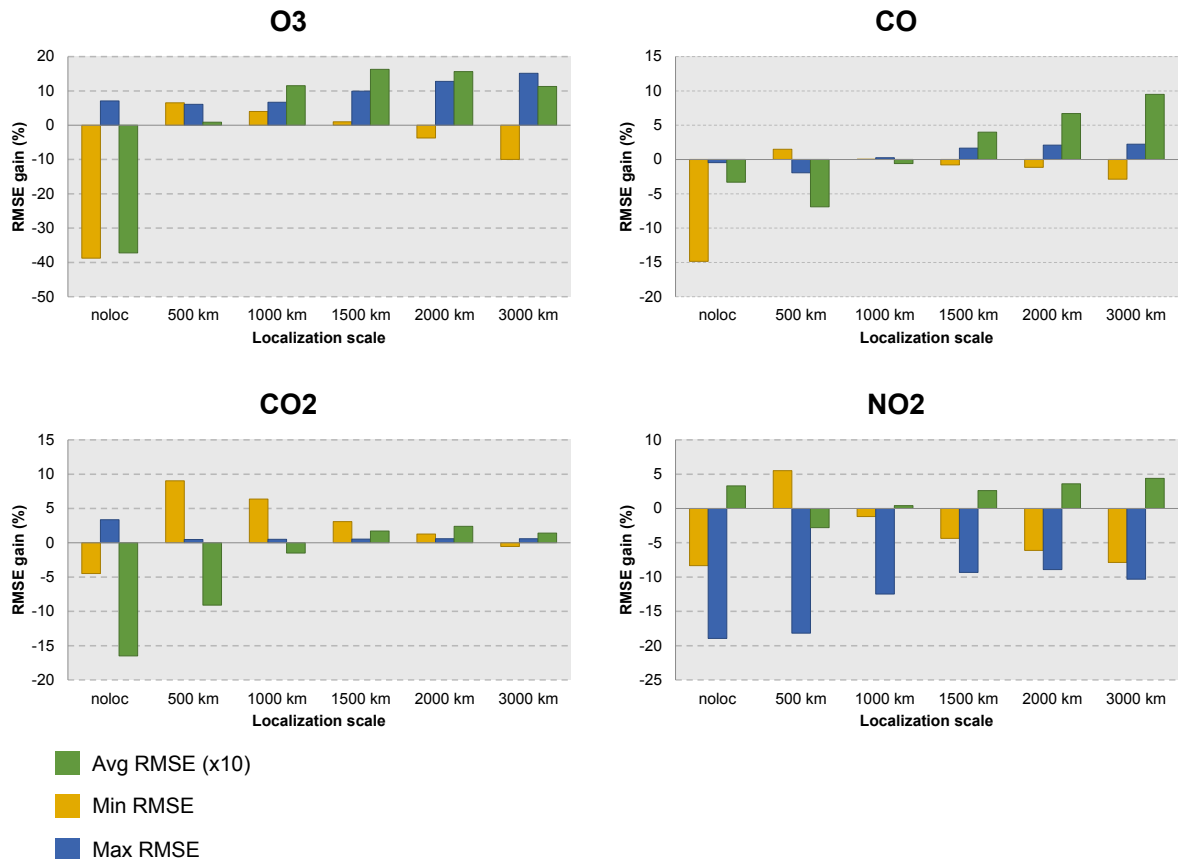
Figure 6. Same as Fig. 5 but for the horizontal length scale of the localization operator ('noloc' when no localization is applied).

The choice of the horizontal localization scale is more delicate, because it is intimately linked to the model dynamics, depends on the ensemble size and on the assumptions made to construct and apply $\underline{C}$ (Sec. 3.2). Fig. 6 shows that horizontal localization is necessary to obtain meaningful results with 4DEnVar. Second, increasing the localization scale to values as high as 3000 km, compared to the 750 km of the initial perturbation scale, has the effect of improving the maximum RMSE gain but also degrading significantly the minimum one. This is not desired, as explained above. The best results, considering employing the same homogeneous and global localization scale for all chemical species, can be found for the value of 1500 km.

The configuration of a multivariate localization, or chemical localization in our study, presents the same issues as the spatial one. A similar approach as in the above paragraph has been taken. The numerical experiments described in Sec. 4.1.2, which considered multivariate cases, are used to compute error differences in Fig. 7. Again, we remark that without any localization the statistical noise of the ensemble covariance degrades significantly the results compared to 3D-Var. A localization coefficient of 0.5 reduces efficiently the effect of noisy correlations in the 4D ensemble covariance, leaving the possibility of describing
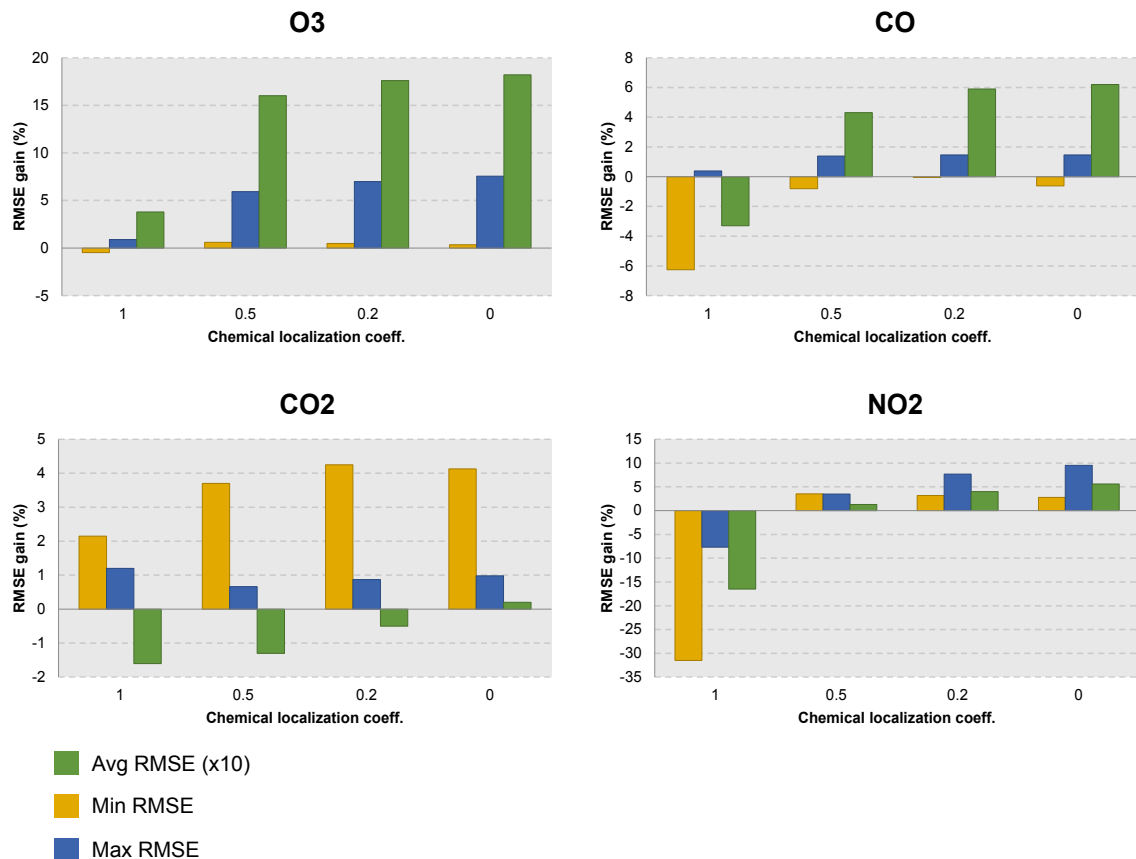
Figure 7. Same as Fig. 5 but for the multivariate correlation coefficient of the localization operator (0 corresponding to univariate DA, 1 for a full multi-variate unlocalized case).

multivariate effects. We remind readers that multivariate effects were found to be very small in perfect model experiments presented until now, but can be much larger when a model error is introduced (Sec. 4.2).

We conclude that a simple localization scheme, based on global and empirically tuned parameters, provides already encouraging results for the application of 4DEnVar to large scale chemistry models. The development of more sophisticated local-
5   ization operators (Bocquet, 2016; Desroziers et al., 2016) and more rigorous methods to estimate their parameters (Ménétrier et al., 2015), represents a current subject of research, and will be the topic of a future study.

4.2   Model error experiments

In the previous section we have shown how 4DEnVar is able to match or even outperform 3D-Var results when the model is perfect. However, the main interest of 4DEnVar for atmospheric chemistry arises when the model is not perfect, i.e., a case that
10   is more easily addressed using ensemble-based methods. In this section 4DEnVar experiments are conducted in the presence

27

of a model error term. A typical source of uncertainty in CTMs is represented by anthropogenic or biogenic emissions (Kok, 2011; Zhao et al., 2011; Ma and van Aardenne, 2004). In the case of reactive species like NO_x, erroneous emissions can impact strongly the formation of secondary species such as O_3 (Lei and Wang, 2014; Sillman, 1999). Errors on surface emissions produce already a complex and rich dynamics and will be used as a test-bed to investigate the effects of model error in DA.
5   Other sources of uncertainties, e.g., chemistry parameters or meteorology, will be addressed in a future study, considering that the same methodology as used here can be applied.

Similar single cycle DA experiments are conducted as in Sec. 4.1. The main difference with the previous section is that experiments are done here during the spin-up period of the model (day 2 to 4). This allows to examine the model error estimation during the pollution build-up period, when the chemical system is in a transient phase and daily cycles of reactive gases are
10   not yet stationary. This represents a more challenging and realistic situation for testing the bias correction procedure (Eq.19), which is fully consistent only in stationary conditions. The true NO emissions (Table 3) are perturbed by a multiplicative factor, which is sampled from a log-normal distribution with mean and sigma equal to 0.5 and 0.8, respectively. Forecast emissions are increased by a multiplicative factor of 2.35, whereas the log-normal distribution has been used to generate emission perturbations for the ensemble of forecasts. The emissions perturbation is constant in time but not in space, due to the geographical
15   variability of emission factors (Fig. 1).

The main objective of this section is to illustrate the application of the bias estimation procedure (Sec. 3.3) on chemical fields. Chemical interactions alone can already give rise to complex temporal dynamics, which can produce unattended behavior within typical hypotheses of most DA schemes (Tang et al., 2016). Therefore, a simplified model-set up is used in this section deactivating the advection of chemical species. This reduces QG-Chem to a collection of 0-D chemistry models and allows to
20   focus on the effects of model errors on chemistry. Except from this, the same exact configuration as before is used for 4DEnVar (initial condition error, ensemble size, localization). DA results in a more general case when both chemistry and advection are activated will follow in Sec. 4.3. Univariate O_3 DA is presented in Sec. 4.2.1. Results of a multivariate DA experiment are presented in Sec. 4.2.2, to examine the combined effects of model error and chemical couplings. Finally, the impact of the model bias correction procedure is evaluated on 48 hours forecasts of several species in Sec. 4.2.3.

25   **4.2.1   Univariate assimilation**

Three DA experiments are shown in Fig. 8: i) activating only the initial perturbation, ii) activating only the model error, iii) with both perturbations activated. Only O_3 observations have been assimilated and the chemical localization coefficients are set to zero to compute univariate analyses.

With the NO emissions increased by a factor of 2.35, the forecast produces higher concentrations of O_3 (Fig. 8, middle
30   plot). We remark also that the effects of model error on O_3 appear later in the day, due to the photo-chemistry, whereas the perturbation of the initial condition has a larger effect in the first part of the day. In this specific case it is also interesting to note that, when both perturbation are applied, a compensation of the two errors cancels out the differences between the forecast and the truth later in the day. This is an example of compensating errors in atmospheric chemistry, which might be hard to detect when comparing model results to observations.
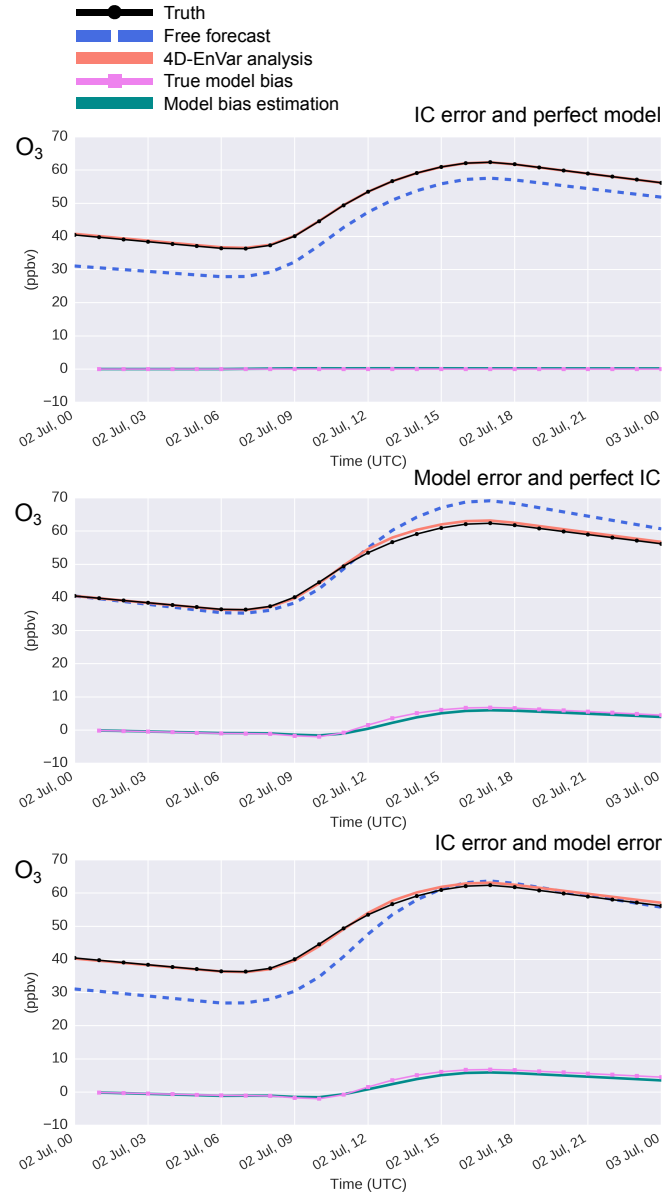
Figure 8. O_3 effective model error estimation in the univariate case (only O_3 assimilated). Temporal trajectories during day 2 for the forecast, the analysis, the truth, the true effective model error and the effective model error estimated using Eq. (12), for the pixel A located close to an observation. From top to bottom the following experiments are shown with the uncertainties being introduced: i) only in the O_3 initial condition (as in Sec. 4.1) ii) only in the forecast model (through surface emissions perturbation) iii) both in the O_3 initial condition and in the forecast model.

The 4DEnVar analyses agree well with the truth in all 3 cases, similarly to what was obtained in Sec. 4.1.1. This satisfies the main requirement for a meaningful computation of the effective model error (Sec. 3.3). The estimation of the effective model
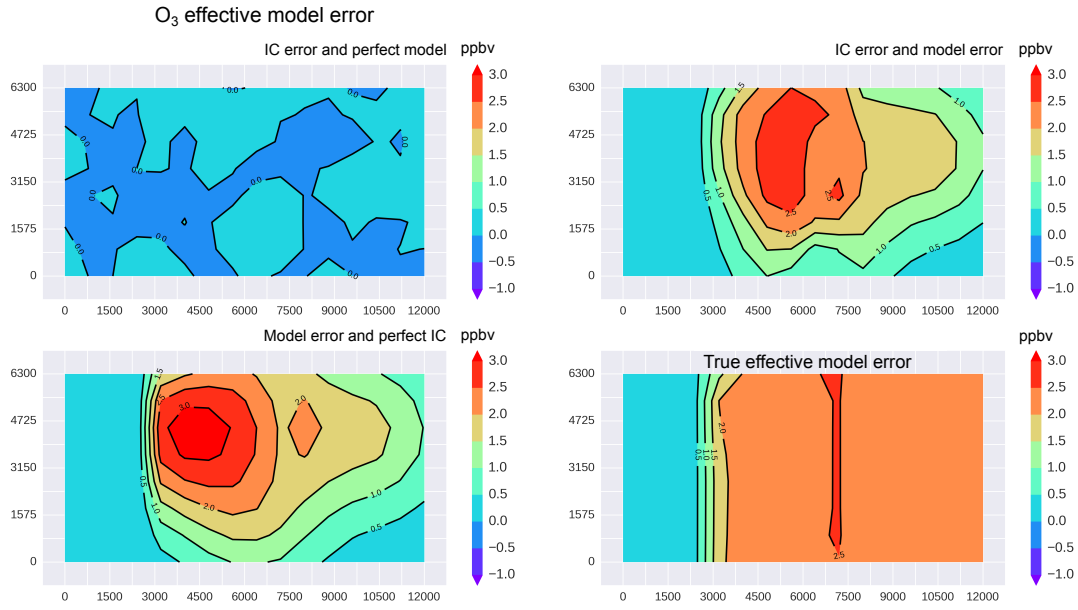
Figure 9. Temporal averages (24 hours) of the estimated effective model error for the three experiments as in Fig. 8. From left to right and from top to bottom, experiments (i), (ii) and (iii) are shown. On the bottom right plot the true effective model error in the experiments (ii) and (iii) is shown (no model error was activated for the first experiment).

error with Eq. (12) is also displayed in Fig. 8. We remind readers that the effective model error is expressed in the same physical units of the model state, i.e., chemical concentration units (ppbv). The *true* effective model error is computed by subtracting the truth from a forecast initialized from the truth, and it is also added to the figures.

The effective model error is zero in the first experiment, which is correctly diagnosed by the procedure. In the second experiment, the estimated effective model error is approximately zero until 10 am and grows to positive values of about 8 ppbv later in the afternoon, which is coherent with the underlying perturbations and chemical mechanism.

The temporal average of the effective model error (Fig. 9) shows that the larger errors are diagnosed in the center of the domain, coherently with the characteristics of the emissions errors (global scaling of NO emissions). The geographical patterns and the differences from the true effective error (bottom right plot in Fig. 9) are a result of the observations location and localization scale. With the proposed procedure, the effective model error estimation relies on the localization scale that has been used for the 4DEnVar analysis. Therefore, the effective model error approaches zero moving far from assimilated observations, no matter what the spatial patterns of the true model error are. In the third experiment, the temporal behavior of the effective model error is well captured at the observed location. However, significant differences with the second experiment are visible in the spatial distribution of the error. Ideally, the estimation of the effective model error should provide exactly the same results in the second and third experiment. Differences arise because DA is not perfect due to: small ensemble size, linearization hypotheses, observation number and observation errors. Also, when adding degrees of freedom to the sources of

30

uncertainty and keeping constant the observation network, DA becomes more challenging: if, for example, observations of O_3 were available only in the late afternoon, no discrimination between the two sources of error could have been achieved in the third experiment. However, thanks to the hourly frequency of O_3 observations, the temporal and large scale features of the model error have been retrieved also in the third case.

4.2.2 Multivariate assimilation

The effects of O_3 assimilation on other chemical species are shown in Fig. 10. These were obtained by repeating the third experiment of Sec. 4.2.1 but setting the chemical localization coefficient to 0.5 instead of zero. Compared to the results discussed in Sec. 4.1.2, multivariate corrections are now very significant. For example, analyzed NO and NO_2 concentrations are almost halved and the initial forecast errors greatly reduced. On the other hand, CO, which was not perturbed initially nor is it strongly coupled to $NO/NO_2/O_3$ concentrations, is not modified at all by the DA. This shows that multivariate aspects of chemical DA can be well captured by the 4DEnVar algorithm, even with a small ensemble.

The effective model error can be computed for all the variables of the state vector, and is plotted in Fig. 10 and 11. The error fields of NO_2 and NO reproduce well the temporal and spatial features of the perturbation on NO emissions that was implemented in the forecast. Note that since NO is rapidly converted into NO_2 during the night, the initial linear increase of the effective model error is only observable on NO_2. On the other hand, during daytime, a strong presence of the effective model error is found for both NO and NO_2. However, the NO_2 and NO errors trajectories are not linear during daytime, due to the complex photo-chemistry.

Strong non-linearities of the chemical system cannot be taken into account by the current implementation of the 4DEnVar algorithm. When the NO/O_3 relationship was strongly non-linear, inaccurate analyses and, therefore, estimations of the effective model error have been found (not shown). This difficulty could be overcome by introducing external loops within 4DEnVar, similarly to how it was already done with the IEnKS algorithm (Haussaire and Bocquet, 2016). This represents the objective of a future study. Alternatively, the combined assimilation of O_3 and NO_2 or a more severe chemical localization can reduce the occurrence of analysis errors in non-linear regimes.

4.2.3 Impact on chemical forecasts

The analysis computed in the previous section has been used to initialize chemical forecasts for the following 48 hours. This was done to evaluate the potential of the forecast bias correction procedure (Eq. 19). The corrected forecasts (CF) of O_3, NO_2 and NO are compared to the initial 3 days long forecast (OF), the 48 hours forecast initialized from the latest available analysis without any correction (AF) and the truth (Fig. 12).

First, the AF converges very rapidly (in about 12 hours) to the OF for all species, confirming the limited effects of the state correction on the chemical forecasts for the next days (Wu et al., 2008). On the other hand, the CF is very close to the truth during the first 24 hours, indicating that the hypothesis of a stationary effective model error on successive days (but hourly variable) seems appropriate in this case. During the third day, a positive forecast correction is still achieved. However,

31
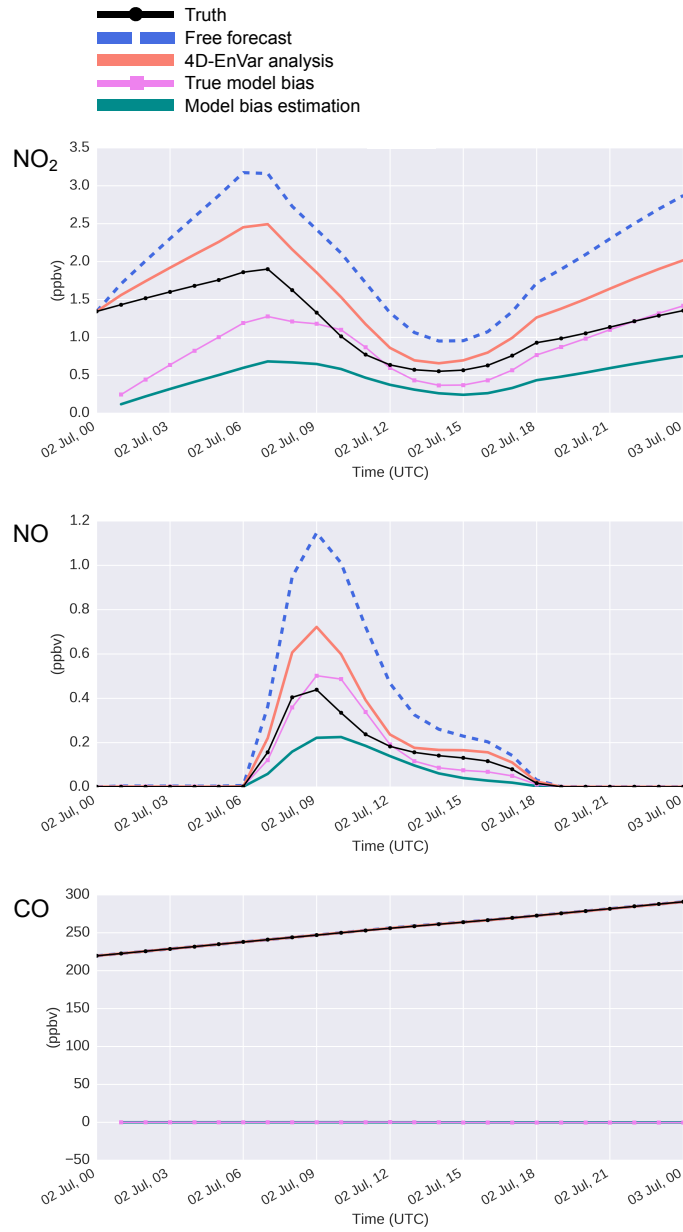
Figure 10. Multivariate effective model error estimation. Temporal trajectories for day 2, as in Fig. 8 but for NO_2, NO, CO species (not assimilated), obtained assimilating only O_3 . Uncertainties were both in the O_3 initial condition and in the forecast model (NO emissions), corresponding to the third experiment in Fig. 8. Only difference with the previous experiment in is that a non-zero multivariate localization coefficient is used here.

chemical concentrations have evolved too much to be efficiently corrected using the bias estimated on the first day. Estimating and correcting the forecast tendencies, instead of the state, could provide a better result for the third day.
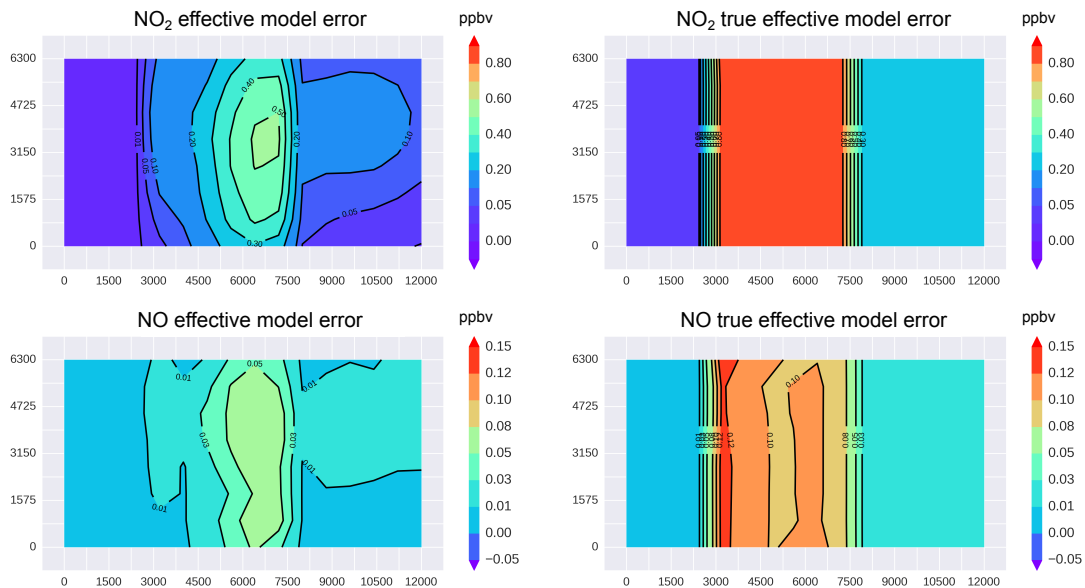
Figure 11. Temporal averages (24 hours) of the estimated effective model error (in ppbv) for the multivariate DA experiment in Fig. 10. Estimated effective model error for NO_2 on top and NO on bottom. On the right the true effective model error is displayed for both species.

Accounting for model errors, either bias or tendencies, for the next day forecast requires the model uncertainties to be stationary. A stationary error has been used in this study since surface emissions were taken constant in time. Most of regional air-quality models seem affected in large part by stationary errors (Marécal et al., 2015) and assuming the persistence of the bias during 24 hours looks reasonable. However, the main sources of uncertainties of real CTMs need first to be identified to allow a meaningful effective model error estimation. Therefore, the implementation of 4DEnVar to a real CTM is necessary to further demonstrate its potential for air-quality forecasts.

4.3 Statistical comparison between 4DEnVar and 3D-Var

Single windows DA experiments have been examined so far, to better analyze new aspects of 4DEnVar DA for atmospheric chemistry. In this section, DA experiments are conducted for multiple consecutive days, to provide a statistically robust comparison between 3D-Var and 4DEnVar. A general case including initial condition and model errors for all four species is considered, and both reanalyses and 24 hours forecasts for the next day are evaluated. The same values as before (multivariate experiments in Sec. 4.1.2) are used for the initial perturbation and for the algorithm settings. Compared to Sec. 4.2, all emissions in Table 3 are now perturbed, using a log-normally distributed scaling factor.

Synthetic observations are generated from a seven days long truth simulation. The four species are assimilated for a period of seven days, starting on day 20. However, at the end of each cycle of 24 hours (forecast, analysis and next day forecast), the initial condition is reinitialized with the truth concentrations. Initial condition perturbations are recomputed advancing the seed of the pseudo-random generator. Therefore, the seven daily cycles are statistically independent. This is done to increase the
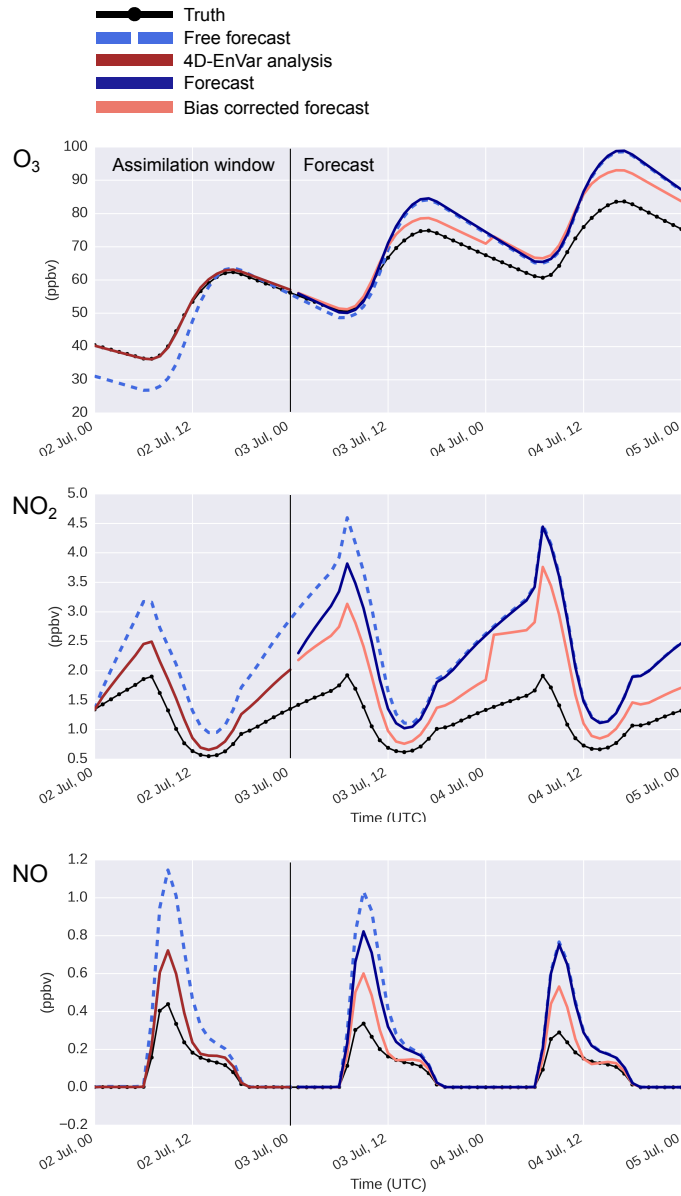
Figure 12. Impact of model bias correction on 48 hours forecasts. Temporal trajectory of free forecast (3 days), analysis (24 hours), truth and 48 hours long forecasts initialized from the latest available analysis, for the pixel A in Fig. 1. DA of O_3 is performed only during the first 24 hours (as in Fig. 10). Curves displayed in dark blue are when no bias correction is applied to the final forecast, in salmon when the effective model error estimation in Fig. 10 is used to correct the final forecast (Eq. 19).

experiment statistics, without having to deal with the propagation of the analysis covariance through consecutive cycles. This aspect will be investigated in a future study.

Table 7. DA RMSEs for seven cycles of analyses/next day forecasts where all considered species are perturbed and assimilated and with model error enabled. ϵ_{min}, ϵ_{max}, ϵ_{avg} (in %) are the minimum, maximum and average values of the relative RMSE (Eq. 21 dived by the truth average) on the QG-Chem domain.

| | Reanalysis | | | | | | Forecast | | | | | |
| | 3D-Var | | | 4D-EnVar | | | 3D-Var | | | 4D-EnVar | | |
Species	ϵ_{min}	ϵ_{max}	ϵ_{avg}	ϵ_{min}	ϵ_{max}	ϵ_{avg}	ϵ_{min}	ϵ_{max}	ϵ_{avg}	ϵ_{min}	ϵ_{max}	ϵ_{avg}
O_3	2.8	25.7	14.3	2.0	26.4	13.3	5.1	23.9	14.4	4.9	21.7	11.3
CO	1.3	42.8	17.1	1.1	27.7	13.1	6.7	90.8	37.3	7.2	50.1	22.1
NO_2	7.7	88.9	41.2	4.1	69.4	23.3	23.6	117.8	74.0	11.3	90.3	37.2
CO_2	3.1	15.3	8.6	1.4	17.5	8.6	1.7	13.7	5.7	1.9	13.6	6.9

The relative RMSE between the experiments and the truth is computed at each grid point for the seven days long period as in Eq. (21). Results are summarized in Table 7: minimum, maximum and average values of the relative RMSEs computed over the QG-Chem domain are reported for 3D-Var and 4DEnVar experiments.

We confirm preceding findings concerning the reanalyses capabilities of 4DEnVar, which provides superior results to 3D-Var for all species. Results are particularly good for species strongly related to emissions (CO and NO_2), which show reduced average RMSE values compared to 3D-Var results. This is a consequence of precisely accounting for emissions uncertainties within 4DEnVar. A similar result was obtained by (Haussaire and Bocquet, 2016). They showed that by using an ensemble forecast of the meteorology, thus partially accounting for model error, the root-mean-square error of IEnKS (a nonlinear 4DEn-Var method) on the low-order online tracer model L95-T was improved by 25% to 50%. Moreover, the 4DEnVar maximum RMSE is also similar or lower than in 3D-Var reanalyses, which suggests that the occurrence of degraded results compared to 3D-Var (negative bars in Fig. 5) is not systematic.

Similar conclusions can be drawn for the RMSE of the next day forecast. We remind readers that the bias correction procedure has been used with 4DEnVar, whereas no correction is applied with 3D-Var. CO and NO_2 forecasts, show significantly lower RMSE with 4DEnVar, due to the forecast bias correction. O_3 forecast improvements are also observed, even if smaller than those for the two precursors. With 4DEnVar, forecasts of CO_2 are slightly worse than with 3D-Var. Since CO_2 is not affected by the considered model error and it is not chemically coupled to other species, the relative bias correction term should be strictly equal to zero. However, the small ensemble size and use of localization introduce statistical noise in the effective model error estimation, which can impact the forecast correction. This issue can be mitigated by selectively setting to zero the chemical localization coefficient between species that are chemically uncoupled. However, results remained on pair with 3D-Var in this study.

We can conclude that the forecast of species related to surface emissions, either directly or through chemical couplings, can be significantly improved when the model error is considered within DA. The forecasts of CO_2, which is a passive tracer

in our study, seem instead dominated by the number of assimilated observation, no matter which DA algorithm is employed. However, this is not the case in real applications, where CO_2 concentration is modulated by uncertain anthropogenic emissions and natural sinks.

We finally remark that EnKF, which also takes advantage of the information available from the ensembles, could have represented an alternative and possibly more accurate reference than 3D-Var for this study. However, it is significantly more costly than 3D-Var, it introduces some difficulties like the definition of an optimal inflation procedure (Constantinescu et al., 2007a; Gaubert et al., 2014) and was not yet available in the OOPS DA library at the time of the study. A more comprehensive comparisons of DA schemes of increasing complexity and cost within QG-Chem is left for a future study, the present being focused especially on the bias correction procedure using 4DEnVar.

5   Conclusions

The objectives of this study were i) to develop a new toy-model framework to test advanced data assimilation (DA) algorithms for atmospheric chemistry ii) demonstrate the potential of the 4DEnVar method for air-quality or more in general tropospheric chemistry applications. In particular, we addressed the questions of: how to jointly assimilate chemical species with very different lifetimes and possible chemical couplings, and how to account for model error to improve forecasts for the next day.

An atmospheric chemistry reduced-order model (QG-Chem) has been developed, based on quasi-geostrophic meteorology and a detailed tropospheric chemistry scheme. It has been used to simulate the complex spatio-temporal patterns of reactive gases (NO_x,O_3) and long-lived species (CO,CO_2) under the effects of emissions, chemistry and transport. QG-Chem has been coupled to a generic library for data assimilation (OOPS) and has been proven to be well suited to perform a large number of DA experiments. Concerning temporal aspects and assimilated observations, the experiments have been designed based on the implementation of chemical DA in operational air-quality forecast centers.

A number of DA experiments have been conducted to compare 4DEnVar analyses with 3D-Var ones in a perfect model hypothesis, in an univariate and a multivariate setting. The sensitivity of 4DEnVar results to the ensemble size and localization method was also assessed. Results with 4DEnVar are generally better for all chemical species even using a small ensemble size of 16 members, provided that ensemble localization, even if basic, is applied. This suggests that considering the linearized model dynamics to derive a flow dependent background error covariance can be beneficial for chemical reanalyses. Thanks to 4DEnVar, this can be obtained without need of tangent linear and adjoint codes of the complex CTM. Multivariate effects were found to be not significant when a perfect model is used, suggesting that multivariate DA goes together with model error for atmospheric chemistry applications.

Clear advantages of using an ensemble method, which is significantly more costly than 3DVar, have been found when model errors were introduced. It has been shown that 4DEnVar is able to take into account heterogeneous errors on chemical emissions and complex chemical couplings to cross-correct precursor species (e.g., NO and NO_2), based only on hourly observations of secondary species (O_3). Using long enough DA windows (24 hours in this study), 4DEnVar allowed to account for the different time-scales of the chemical mechanism and the corresponding effects in DA. For example, the delayed impact of NO

36

emissions errors on afternoon O_3 concentrations was correctly accounted by 4DEnVar. The contribution of model errors on the 4D chemical state can be estimated at the cost of an additional forecast, providing quantitative information on possible forecast biases, and, in the case of stationary errors, a method to correct next day forecasts. This has been tested with success on several independent DA windows, showing that 24 hours forecasts of NO_2 and CO can be twice as accurate with 4DEnVar than with
5   3D-Var. Better results have been also found for O_3 forecasts.

We conclude that 4DEnVar is potentially of high interest for chemical DA. We determine the main benefits to be: the implicit specification of a flow-dependent and multivariate error covariance matrix, the possibility of accounting for model errors through stochastic perturbation, and the opportunity of using DA windows long enough to catch typical features of model errors in air-quality models. The computational cost of 4DEnVar is higher than 3D-Var, but a small ensemble of at least
10   15 to 20 members remain affordable within most operational centers.

The application to a real CTM remains necessary to evaluate if the advantages of 4DEnVar shown in this study hold in real applications, where model uncertainties are not perfectly known, as well as observational ones. Aspects related to the vertical propagation of the information, which have been neglected with QG-Chem, could also represent an additional challenge in real systems. Finally, research on algorithmic aspects of 4DEnVar is needed to implement more accurate localization operators, to
15   account for non-linear chemical regimes and to correctly propagate the analysis covariance through successive DA windows. QG-Chem could represent a useful tool for these type of studies.

Code availability

The QG-Chem code is copyright of the CERFACS laboratory. The sources and the data used in this study are available upon request to E. Emili (emili@cerfacs.fr) or D. Cariolle (cariolle@cerfacs.fr).

References

Babenhauserheide, A., Basu, S., Houweling, S., Peters, W., and Butz, A.: Comparing the CarbonTracker and TM5-4DVar data assimilation systems for CO2 surface flux inversions, Atmospheric Chemistry and Physics, 15, 9747–9763, doi:10.5194/acp-15-9747-2015, http://www.atmos-chem-phys.net/15/9747/2015/, 2015.

5   Beekmann, M. and Derognat, C.: Monte Carlo uncertainty analysis of a regional-scale transport chemistry model constrained by measurements from the Atmospheric Pollution Over the Paris Area (ESQUIF) campaign, Journal of Geophysical Research, 108, 8559, doi:10.1029/2003JD003391, http://www.agu.org/pubs/crossref/2003/2003JD003391.shtml, 2003.

Belo Pereira, M. and Berre, L.: The Use of an Ensemble Approach to Study the Background Error Covariances in a Global NWP Model, Monthly Weather Review, pp. 2466–2489, 2006.

10   Bocquet, M.: Localization and the iterative ensemble Kalman smoother, Quarterly Journal of the Royal Meteorological Society, 142, 1075–1089, doi:10.1002/qj.2711, http://dx.doi.org/10.1002/qj.2711, 2016.

Bocquet, M. and Sakov, P.: Joint state and parameter estimation with an iterative ensemble Kalman smoother, Nonlinear Processes in Geophysics, 20, 803–818, doi:10.5194/npg-20-803-2013, 2013.

Bocquet, M. and Sakov, P.: An iterative ensemble Kalman smoother, Quarterly Journal of the Royal Meteorological Society, 140, 1521–1535,

15   doi:10.1002/qj.2236, 2014.

Bocquet, M., Elbern, H., Eskes, H., Hirtl, M., Žabkar, R., Carmichael, G. R., Flemming, J., Inness, A., Pagowski, M., Pérez Camaño, J. L., Saide, P. E., San Jose, R., Sofiev, M., Vira, J., Baklanov, A., Carnevale, C., Grell, G., and Seigneur, C.: Data assimilation in atmospheric chemistry models: current status and future prospects for coupled chemistry meteorology models, Atmospheric Chemistry and Physics, 15, 5325–5358, doi:10.5194/acp-15-5325-2015, http://www.atmos-chem-phys.net/15/5325/2015/, 2015.

20   Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., and He, B.: Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part II: One-Month Experiments with Real Observations, Monthly Weather Review, 138, 1567–1586, doi:10.1175/2009MWR3158.1, 2010.

Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F. M., Chédin, a., and Ciais, P.: Inferring CO2 sources and sinks from satellite observations: Method and application to TOVS data, Journal of Geophysical Research: Atmospheres, 110, 1–13,

25   doi:10.1029/2005JD006390, 2005.

Constantinescu, E. M., Sandu, A., Chai, T., and Carmichael, G. R.: Ensemble-based chemical data assimilation. I: General approach, Quarterly Journal of the Royal Meteorological Society, 133, 1229–1243, doi:10.1002/qj.76, http://dx.doi.org/10.1002/qj.76, 2007a.

Constantinescu, E. M., Sandu, A., Chai, T., and Carmichael, G. R.: Ensemble-based chemical data assimilation. II: Covariance localization, Quarterly Journal of the Royal Meteorological Society, 133, 1245–1256, doi:10.1002/qj.77, http://dx.doi.org/10.1002/qj.77, 2007b.

30   Crassier, V., Suhre, K., Tulet, P., and Rosset, R.: Development of a reduced chemical scheme for use in mesoscale meteorological models, Atmospheric Environment, 34, 2633–2644, doi:10.1016/S1352-2310(99)00480-X, 2000.

Derber, J. and Rosati, A.: A Global Oceanic Data Assimilation System, Journal of Physical Oceanography, 19, 1333–1347, doi:10.1175/1520-0485(1989)019<1333:AGODAS>2.0.CO;2, http://dx.doi.org/10.1175/1520-0485(1989)019{\T1\textless}1333: AGODAS{\T1\textgreater}2.0.CO2, 1989.

35   Desroziers, G., Camino, J.-T., and Berre, L.: 4D-En-Var: link with weak-constraint 4D-Var and different possible implementations, Quarterly Journal of the Royal Meteorological Society, pp. 2097–2110, doi:10.1002/qj.2325, 2014.

Desroziers, G., Arbogast, E., and Berre, L.: Improving spatial localisation in 4DEnVar, Quarterly Journal of the Royal Meteorological Society, pp. n/a—-n/a, doi:10.1002/qj.2898, http://dx.doi.org/10.1002/qj.2898, 2016.

Dimet, F.-X. L. E. and Talagrand, O.: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, Tellus A, 38A, 97–110, doi:10.1111/j.1600-0870.1986.tb00459.x, http://dx.doi.org/10.1111/j.1600-0870.1986.tb00459.x, 1986.

5   Elbern, H., Schmidt, H., and Ebel, A.: Variational data assimilation for tropospheric chemistry modeling, Journal of Geophysical Research: Atmospheres, 102, 15 967–15 985, doi:10.1029/97JD01213, http://dx.doi.org/10.1029/97JD01213, 1997.

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmospheric Chemistry and Physics Discussions, 7, 1725–1783, doi:10.5194/acpd-7-1725-2007, http://www.atmos-chem-phys.net/7/3749/2007/acp-7-3749-2007.html, 2007.

10   Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practical implementation, Ocean Dynamics, 53, 343–367, doi:10.1007/s10236-003-0036-9, http://dx.doi.org/10.1007/s10236-003-0036-9, 2003.

Fairbairn, D., Pring, S. R., Lorenc, a. C., and Roulstone, I.: A comparison of 4DVar with ensemble data assimilation methods, Quarterly Journal of the Royal Meteorological Society, pp. 281–294, doi:10.1002/qj.2135, 2013.

Fandry, C. B. and Leslie, L. M.: A Two-Layer Quasi-Geostrophic Model of Summer Trough Formation in the Australian Subtropi-
15   cal Easterlies, Journal of the Atmospheric Sciences, 41, 807–818, doi:10.1175/1520-0469(1984)041<0807:ATLQGM>2.0.CO;2, http://dx.doi.org/10.1175/1520-0469(1984)041{\T1\textless}0807:ATLQGM{\T1\textgreater}2.0.CO2, 1984.

Furrer, R. and Bengtsson, T.: Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants, Journal of Multivariate Analysis, 98, 227–255, doi:10.1016/j.jmva.2006.08.003, 2007.

Gaubert, B., Coman, A., Foret, G., Meleux, F., Ung, A., Rouil, L., Ionescu, A., Candau, Y., and Beekmann, M.: Regional scale ozone
20   data assimilation using an ensemble Kalman filter and the CHIMERE chemical transport model, Geoscientific Model Development, 7, 283–302, doi:10.5194/gmd-7-283-2014, http://www.geosci-model-dev.net/7/283/2014/, 2014.

Hamer, P. D., Bowman, K. W., Henze, D. K., Attié, J.-L., and Marécal, V.: The impact of observing characteristics on the ability to predict ozone under varying polluted photochemical regimes, Atmospheric Chemistry and Physics, 15, 10 645–10 667, doi:10.5194/acp-15-10645-2015, http://www.atmos-chem-phys.net/15/10645/2015/, 2015.

25   Haussaire, J.-M. and Bocquet, M.: A low-order coupled chemistry meteorology model for testing online and offline data assimilation schemes: L95-GRS (v1.0), Geoscientific Model Development, 9, 393–412, doi:10.5194/gmd-9-393-2016, http://www.geosci-model-dev.net/9/393/2016/, 2016.

Hollingsworth, A. and Loennberg, P.: The statistical structure of short-range forecast errors as determined from radiosonde data . Part I : The wind field, Tellus, 38, 111–136, 1986.

30   Huijnen, V., Eskes, H. J., Poupkou, A., Elbern, H., Boersma, K. F., Foret, G., Sofiev, M., Valdebenito, A., Flemming, J., Stein, O., Gross, A., Robertson, L., D'Isidoro, M., Kioutsioukis, I., Friese, E., Amstrup, B., Bergstrom, R., Strunk, A., Vira, J., Zyryanov, D., Maurizi, A., Melas, D., Peuch, V.-H., and Zerefos, C.: Comparison of OMI NO2 tropospheric columns with an ensemble of global and European regional air quality models, Atmospheric Chemistry and Physics, 10, 3273–3296, doi:10.5194/acp-10-3273-2010, http://www.atmos-chem-phys.net/10/3273/2010/, 2010.

35   Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R. J., Errera, Q., Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J. W., Kapsomenakis, J., Lefever, K., Leitão, J., Razinger, M., Richter, A., Schultz, M. G., Simmons, A. J., Suttie, M., Stein, O., Thépaut, J.-N., Thouret, V., Vrekoussis, M., Zerefos,

C., and the MACC team: The MACC reanalysis: an 8 yr data set of atmospheric composition, Atmospheric Chemistry and Physics, 13, 4073–4109, doi:10.5194/acp-13-4073-2013, http://www.atmos-chem-phys.net/13/4073/2013/, 2013.

Jaumouillé, E., Massart, S., Piacentini, A., Cariolle, D., and Peuch, V.-H.: Impact of a time-dependent background error covariance matrix on air quality analysis, Geoscientific Model Development, 5, 1075–1090, doi:10.5194/gmd-5-1075-2012, http://www.geosci-model-dev.net/5/1075/2012/, 2012.

Kalnay, E.: Atmospheric modeling, data assimilation, and predictability, Cambridge University Press, 2003.

Kok, J. F.: A scaling theory for the size distribution of emitted dust aerosols suggests climate models underestimate the size of the global dust cycle, Proceedings of the National Academy of Sciences, 108, 1016–1021, doi:10.1073/pnas.1014798108/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1014798108, 2011.

Koohkan, M. R. and Bocquet, M.: Accounting for representativeness errors in the inversion of atmospheric constituent emissions : application to the retrieval of regional carbon monoxide fluxes, Tellus B, 64, 19 047, http://dx.doi.org/10.3402/tellusb.v64i0.19047, 2012.

Lei, H. and Wang, J. X. L.: Sensitivities of NOx transformation and the effects on surface ozone and nitrate, Atmospheric Chemistry and Physics, 14, 1385–1396, doi:10.5194/acp-14-1385-2014, http://www.atmos-chem-phys.net/14/1385/2014/, 2014.

Lorenc, A. C.: Recommended Nomenclature for EnVar Data Assimilation Methods, in: Research Activities in Atmospheric and Oceanic Modeling, 2013.

Lorenc, A. C., Bowler, N. E., Clayton, A. M., Pring, S. R., and Fairbairn, D.: Comparison of Hybrid-4DEnVar and Hybrid-4DVar Data Assimilation Methods for Global NWP, Monthly Weather Review, 143, 212–229, doi:10.1175/MWR-D-14-00195.1, http://dx.doi.org/10.1175/MWR-D-14-00195.1, 2015.

Lu, S., Lin, H. X., Heemink, A. W., Fu, G., and Segers, A. J.: Estimation of Volcanic Ash Emissions Using Trajectory-Based 4D-Var Data Assimilation, Monthly Weather Review, 144, 575–589, doi:10.1175/MWR-D-15-0194.1, http://dx.doi.org/10.1175/MWR-D-15-0194.1, 2016.

Ma, J. and van Aardenne, J. A.: Impact of different emission inventories on simulated tropospheric ozone over China : a regional chemical transport model evaluation, Atmospheric Chemistry and Physics, 4, 877–887, doi:10.5194/acp-4-877-2004, 2004.

Mallet, V. and Sportisse, B.: Uncertainty in a chemistry-transport model due to physical parameterizations and numerical approximations: An ensemble approach applied to ozone modeling, Journal of Geophysical Research, 111, D01 302, doi:10.1029/2005JD006149, http://www.agu.org/pubs/crossref/2006/2005JD006149.shtml, 2006.

Mandel, J., Bergou, E., Gürol, S., Gratton, S., and Kasanický, I.: Hybrid Levenberg–Marquardt and weak-constraint ensemble Kalman smoother method, Nonlinear Processes in Geophysics, 23, 59–73, doi:10.5194/npg-23-59-2016, http://www.nonlin-processes-geophys.net/23/59/2016/, 2016.

Marécal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chéroux, F., Colette, A., Coman, A., Curier, R. L., van der Gon, H. A. C., Drouin, A., Elbern, H., Emili, E., Engelen, R. J., Eskes, H. J., Foret, G., Friese, E., Gauss, M., Giannaros, C., Guth, J., Joly, M., Jaumouillé, E., Josse, B., Kadygrov, N., Kaiser, J. W., Krajsek, K., Kuenen, J., Kumar, U., Liora, N., Lopez, E., Malherbe, L., Martinez, I., Melas, D., Meleux, F., Menut, L., Moinat, P., Morales, T., Parmentier, J., Piacentini, A., Plu, M., Poupkou, A., Queguiner, S., Robertson, L., Rouïl, L., Schaap, M., Segers, A., Sofiev, M., Tarasson, L., Thomas, M., Timmermans, R., Valdebenito, Á., van Velthoven, P., van Versendaal, R., Vira, J., and Ung, A.: A regional air quality forecasting system over Europe: the MACC-II daily ensemble production, Geoscientific Model Development, 8, 2777–2813, doi:10.5194/gmd-8-2777-2015, http://www.geosci-model-dev.net/8/2777/2015/, 2015.

Ménétrier, B., Montmerle, T., Michel, Y., and Berre, L.: Linear Filtering of Sample Covariances for Ensemble-Based Data Assimilation. Part I: Optimality Criteria and Application to Variance Filtering and Covariance Localization, Monthly Weather Review, 143, 1622–1643, doi:10.1175/MWR-D-14-00157.1, http://dx.doi.org/10.1175/MWR-D-14-00157.1, 2015.

Miyazaki, K., Eskes, H. J., Sudo, K., Takigawa, M., Weele, M. V., Boersma, K. F., and Bilt, D.: Simultaneous assimilation of satellite NO2, O3, CO, and HNO3 data for the analysis of tropospheric chemical composition and emissions, Atmospheric Chemistry and Physics, 12, 9545–9579, doi:10.5194/acp-12-9545-2012, http://www.atmos-chem-phys.net/12/9545/2012/, 2012.

Pedlosky, J.: Geophysical Fluid Dynamics, Springer, 1992.

Rouil, L. and the MACC team: MACC-II Report n. D113.5: Validation report for 2012, Tech. rep., https://www.gmes-atmosphere.eu/documents/maccii/deliverables/eva, 2014.

Saad, Y. and Schultz, M. H.: GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems, SIAM Journal on Scientific and Statistical Computing, 7, 856–869, doi:10.1137/0907058, http://dx.doi.org/10.1137/0907058, 1986.

Sillman, S.: The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments, Atmospheric Environment, 33, 1821–1845, doi:http://dx.doi.org/10.1016/S1352-2310(98)00345-8, http://www.sciencedirect.com/science/article/pii/S1352231098003458, 1999.

Stockwell, W. R., Kirchner, F., Kuhn, M., and Seefeld, S.: A new mechanism for regional atmospheric chemistry modeling, Journal of Geophysical Research: Atmospheres, 102, 25 847–25 879, doi:10.1029/97JD00849, http://dx.doi.org/10.1029/97JD00849, 1997.

Tang, X., Zhu, J., Wang, Z. F., and Gbaguidi, A.: Improvement of ozone forecast over Beijing based on ensemble Kalman filter with simultaneous adjustment of initial conditions and emissions, Atmospheric Chemistry and Physics, 11, 12 901–12 916, doi:10.5194/acp-11-12901-2011, http://www.atmos-chem-phys.net/11/12901/2011/acp-11-12901-2011.html, 2011.

Tang, X., Zhu, J., Wang, Z., Gbaguidi, A., Lin, C., Xin, J., Song, T., and Hu, B.: Limitations of ozone data assimilation with adjustment of NOx emissions: mixed effects on NO2 forecasts over Beijing and surrounding areas, Atmospheric Chemistry and Physics, 16, 6395–6405, doi:10.5194/acp-16-6395-2016, http://www.atmos-chem-phys.net/16/6395/2016/, 2016.

Thompson, R. L. and Stohl, A.: FLEXINVERT: an atmospheric Bayesian inversion framework for determining surface fluxes of trace species using an optimized grid, Geoscientific Model Development, 7, 2223–2242, doi:10.5194/gmd-7-2223-2014, http://www.geosci-model-dev.net/7/2223/2014/, 2014.

Tremolet, Y.: Accounting for an imperfect model in 4D-Var, Quarterly Journal of the Royal Meteorological Society, 132, 2483–2504, doi:10.1256/qj.05.224, http://dx.doi.org/10.1256/qj.05.224, 2006.

Trémolet, Y.: Model-error estimation in 4D-Var, Quarterly Journal of the Royal Meteorological Society, 133, 1267–1280, doi:10.1002/qj.94, http://dx.doi.org/10.1002/qj.94, 2007.

van der A, R. J., Allaart, M. A. F., and Eskes, H. J.: Multi sensor reanalysis of total ozone, Atmospheric Chemistry and Physics, 10, 11 277–11 294, doi:10.5194/acp-10-11277-2010, http://www.atmos-chem-phys.net/10/11277/2010/, 2010.

Verwer, J. G.: Gauss–Seidel Iteration for Stiff ODES from Chemical Kinetics, SIAM Journal on Scientific Computing, 15, 1243–1250, doi:10.1137/0915076, http://dx.doi.org/10.1137/0915076, 1994.

Weaver, A. and Courtier, P.: Correlation modelling on the sphere using a generalized diffusion equation, Quarterly Journal of the Royal Meteorological Society, 127, 1815–1846, http://onlinelibrary.wiley.com/doi/10.1002/qj.49712757518/abstract, 2001.

Wu, L., Mallet, V., Bocquet, M., and Sportisse, B.: A comparison study of data assimilation algorithms for ozone forecasts, Journal of Geophysical Research, 113, D20 310, doi:10.1029/2008JD009991, http://doi.wiley.com/10.1029/2008JD009991, 2008.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., and Baklanov, A.: Real-time air quality forecasting, Part II: State of the science, current research needs, and future prospects, Atmospheric Environment, 60, 656–676, doi:10.1016/j.atmosenv.2012.02.041, http://dx.doi.org/10.1016/j.atmosenv.2012.02.041, 2012.

Zhao, Y., Nielsen, C. P., Lei, Y., McElroy, M. B., and Hao, J.: Quantifying the uncertainties of a bottom-up emission inventory of anthropogenic atmospheric pollutants in China, Atmospheric Chemistry and Physics, 11, 2295–2308, doi:10.5194/acp-11-2295-2011, 2011.

Zyryanov, D., Foret, G., Eremenko, M., Beekmann, M., Cammas, J. P., D'Isidoro, M., Elbern, H., Flemming, J., Friese, E., Kioutsioutkis, I., Maurizi, A., Melas, D., Meleux, F., Menut, L., Moinat, P., Peuch, V. H., Poupkou, A., Razinger, M., Schultz, M., Stein, O., Suttie, a. M., Valdebenito, A., Zerefos, C., Dufour, G., Bergametti, G., and Flaud, J. M.: 3-D evaluation of tropospheric ozone simulations by an ensemble of regional Chemistry Transport Model, Atmospheric Chemistry and Physics, 12, 3219–3240, doi:10.5194/acp-12-3219-2012, 2012.