Geoscientific
Model Development

A low-order coupled chemistry meteorology model for testing online and offline data assimilation schemes: L95-GRS (v1.0)

J.-M. Haussaire and M. Bocquet

CEREA, Joint laboratory École des Ponts ParisTech and EDF R&D, Université Paris-Est, Champs-sur-Marne, France

Correspondence to: J.-M. Haussaire (haussaij@cerea.enpc.fr)

Abstract. Bocquet and Sakov (2013) introduced a low-order model based on the coupling of the chaotic Lorenz-95 (L95) model, which simulates winds along a mid-latitude circle, with the transport of a tracer species advected by this zonal wind field. This model, named L95-T, can serve as a playground for testing data assimilation schemes with an online model. Here, the tracer part of the model is extended to a reduced photochemistry module. This coupled chemistry meteorology model (CCMM), the L95-GRS (generic reaction set) model, mimics continental and transcontinental transport and the photochemistry of ozone, volatile organic compounds and nitrogen oxides. Its numerical implementation is described. The model is shown to reproduce the major physical and chemical processes being considered. L95-T and L95-GRS are specifically designed and useful for testing advanced data assimilation schemes, such as the iterative ensemble Kalman smoother (IEnKS), which combines the best of ensemble and variational methods. These models provide useful insights prior to the implementation of data assimilation methods into larger models. We illustrate their use with data assimilation schemes on preliminary yet instructive numerical experiments. In particular, online and offline data assimilation strategies can be conveniently tested and discussed with this low-order CCMM. The impact of observed chemical species concentrations on the wind field estimate can be quantitatively assessed. The impacts of the wind chaotic dynamics and of the chemical species non-chaotic but highly nonlinear dynamics on the data assimilation strategies are illustrated.

1 Introduction

Several data assimilation methods have been used in the field of atmospheric chemistry and air quality in many studies (as exemplified in the reviews of Carmichael et al., 2008; Sandu and Chai, 2011; Zhang et al., 2012; Bocquet et al., 2015a). Yet, how efficiently data assimilation schemes operate in high-dimensional and heterogeneous models such as those used in the field remains largely unclear. Indeed, atmospheric chemistry models are becoming increasingly complex, with multiphasic chemistry, size-resolved particulate matter, and possibly coupled to numerical weather prediction models. In the meantime, data assimilation methods have also become more sophisticated. Let us briefly and non-exhaustively describe this evolution. Kalman filters have been used with atmospheric chemistry models by Ménard et al. (2000). The numerical cost of this algorithm was addressed by the use of the ensemble Kalman filter and variants thereof in Segers et al. (2000), Eben et al. (2005), Hanea et al. (2007), Wu et al. (2008), and Sekiyama et al. (2011). In order to address rank deficiencies and sampling issues, localisation and inflation have been used in this context (Constantinescu et al., 2007a, b; Schutgens et al., 2010). Moreover, 3D- and 4D-Var techniques have been applied to chemical transport models (CTMs) in the wake of their success in operational meteorology (Elbern et al., 2000; Quélo et al., 2005; Errera et al., 2008; Wu et al., 2008). These methods, however, require the development of the adjoint models (e.g. Hakami et al., 2007; Henze et al., 2007) and this has led to the implementation of easier approximate adjoints (Bocquet, 2005, 2012; Koohkan and Bocquet, 2012; Singh and Sandu, 2012). Attention has been paid to the construction of the background error covariance matrix (Elbern et al., 2007; Singh et al., 2011). Getting the best of the ensemble Kalman filter and varia-

tional methods through an hybrid ensemble–variational approach is a quest recently initiated in meteorology (Buehner et al., 2010; Clayton et al., 2013; Bocquet and Sakov, 2014; Desroziers et al., 2014; Nino Ruiz and Sandu, 2016) that could be applied to atmospheric chemistry models (Bocquet and Sakov, 2013). Finally, the recent development of coupled chemistry meteorology models (CCMMs) opens the Pandora's box of data assimilation in coupled systems characterised by heterogeneous dynamics with distinct timescales, heterogeneous sources of uncertainty, and complex interactions.

Hence, it will become increasingly difficult to disentangle the merits of data assimilation schemes, of models, and of their numerical implementation in a successful high-dimensional data assimilation study. That is why we believe that the increasing variety of problems encountered in the field of atmospheric chemistry data assimilation puts forward the need for simple low-order models, albeit complex enough to capture the relevant dynamics, physics, and chemistry that could impact the performance of data assimilation schemes. Low-order models, also called toy models, are models of reduced dimension meant to capture the prominent characteristics of the dynamics of larger models, but at a much lower computational cost. They are not meant to be realistic, but their study provides insights into the larger models and their dynamics. Their low numerical cost also comes with the ability to compute reliable statistical scores in various regimes and hence to validate methods with greater confidence. Moreover, they can be distributed and used with the goal to benchmark data assimilation methods since their baseline performance can easily be reproduced.
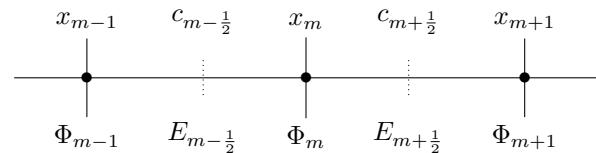
1.1 The Lorenz-95 and tracer model

The Lorenz-95 (L95) model is a very popular low-order meteorology model (Lorenz and Emanuel, 1998). It is a one-dimensional model, whose $M = 40$ state variables extend over a mid-latitude circle. It is defined by the following set of ordinary differential equations

$$\frac{dx_m}{dt} = (x_{m+1} - x_{m-2})x_{m-1} - x_m + F, \qquad (1)$$

for $m = 1, \ldots, M$. The domain is periodic (circle-like, i.e. $x_{-1} = x_{39}$, $x_0 = x_{40}$, and $x_{41} = x_1$). F is chosen to be 8 so that the dynamics is chaotic with a doubling time of about 0.42 Lorenz time units and with 13 positive Lyapunov exponents. A time step of $\Delta t = 0.05$ is meant to represent a time interval of 6 h in the real atmosphere. The L95 model has been extensively used as a test bed and benchmark for data assimilation experiments.

Bocquet and Sakov (2013) added a tracer field to the L95 model state vector. The field is discretised into 40 additional variables meant to represent 40 tracer concentrations. The 40 scalar variables of L95 are considered to be the magnitude of winds at 40 locations, their sign giving their direction.

The tracer field is advected by these winds with a simple Godunov/upwind scheme. These 80 variables are defined on the circle using an Arakawa C-grid as shown below.

The equations of the model are those of Eq. (1) together with

$$\frac{dc_{m+\frac{1}{2}}}{dt} = \Phi_m - \Phi_{m+1} - \lambda c_{m+\frac{1}{2}} + E_{m+\frac{1}{2}}, \qquad (2)$$

where $\Phi_m = x_m c_{m-\frac{1}{2}}$ if $x_m \geq 0$, $\qquad (3)$

$\Phi_m = x_m c_{m+\frac{1}{2}}$ if $x_m < 0$. $\qquad (4)$

This model will be called L95-T in the following. Because the meteorological and tracer state vectors are simulated together, it is an *online* model. The tracer is emitted over the whole domain and the emission fluxes are denoted $E_{m+\frac{1}{2}}$. It is deposited over the whole domain, using a simple scavenging scheme parametrised by a scavenging ratio λ. The reference values for those parameters in Bocquet and Sakov (2013) are $E_{m+\frac{1}{2}} = 1$ and $\lambda = 0.1$. Additional details and illustrations can be found in Bocquet and Sakov (2013). L95-T is a *one-way* coupled model in the sense that there is no physical feedback from the transport part to the meteorology part. However, when applying data assimilation to the model, information is exchanged *both ways* through covariances between the meteorological and transport subsystems. In particular, observations of tracer concentrations can in principle improve the estimation of the meteorological variables.

Hence, L95-T represents an instructive model for more ambitious CCMMs (Bocquet et al., 2015a). From a dynamical perspective, this model couples chaotic meteorology with non-chaotic transport, as any realistic online tracer model would do. Uncertainty in the meteorology comes from errors on the initial conditions, which grow due to the chaotic dynamics. Uncertainty in transport comes from the uncertainty in the emission field and from the wind uncertainty, but the dynamics being stable, there is no exponential growth of the error in the transport subsystem. This is meant to mimic CCMMs. Data assimilation techniques applied to the model, which are meant to reduce these uncertainties, should be able to control the growing error modes as done in numerical weather prediction, and also estimate forcings, typically emissions, as done with CTMs.

However, in order to develop a qualitatively representative low-order CCMM, nonlinear chemistry must be added. The primary goal of this article is to extend the L95-T model with

a simple photochemical kinetic mechanism. To that end, we will use the generic reaction set (GRS; Azzi et al., 1992). The resulting coupled model, which will be introduced in detail in Sect. 2, will be called L95-GRS. In particular, it is meant to be useful to test a state-of-the-art data assimilation and parameter estimation method with a significant potential for such coupled models with heterogeneous observations and dynamics.

1.2 The iterative ensemble Kalman smoother

Hence, the secondary goal of this work is to illustrate the usefulness of low-order CCMMs to better understand the application of specific data assimilation techniques to CCMMs and CTMs. The data assimilation method we shall use is the iterative ensemble Kalman smoother (IEnKS). It was introduced and developed by Bocquet and Sakov (2014). A short account of the scheme can be found in Bocquet and Sakov (2013), which we do not repeat here. However, the main characteristics of the method are recalled in the following and its algorithm is recalled in Appendix A. The IEnKS is an ensemble variational (EnVar) method. It solves a variational problem over a data assimilation window (DAW), as 4D-Var would do. However, because the variational problem is solved in the reduced space generated by an ensemble of state vectors, the adjoint of the model can easily be estimated without the burden of generating the full adjoint model. Indeed, one can for instance use the ensemble of perturbations within the DAW to estimate the sensitivities using finite differences. Because it can solve nonlinear variational problems, it has an edge over a state-of-the-art ensemble Kalman filter. In particular, it is known to handle parameter estimation very well, as well as nonlinear models (Bocquet and Sakov, 2013, 2014). As such, it is as good a candidate method as 4D-Var (Elbern and Schmidt, 1999, 2001) when accounting for nonlinear chemistry such as in a photochemical model. Unlike 4D-Var, a posterior ensemble is generated as the output of the analysis using techniques known in deterministic ensemble Kalman filtering. The IEnKS then propagates the updated ensemble, allowing a better transfer of the errors from an update to the next.

It was shown with the L95 and L95-T models that the IEnKS outperforms the ensemble Kalman filter (EnKF) and 4D-Var for filtering applications (i.e. present-time estimation and forecasting), especially in strongly nonlinear conditions. It was also shown to outperform the 4D-Var and the standard ensemble Kalman smoother for smoothing (i.e. reanalysis). As for any EnVar method, the toll for applying these methods to high-dimensional models is to use ad hoc techniques to regularise the error statistics obtained by empirical ensemble statistics that are prone to sampling errors. As a consequence, localisation and possibly inflation are required when implementing the IEnKS in high-dimensional systems.

If Δt is the time interval between two batches of observations, $L\Delta t$ is the DAW length over which the IEnKS vari-

ational analysis is performed. The simplest variant of the IEnKS algorithm is given in Appendix A. Note that the case $L = 0$ corresponds to the ensemble square root Kalman filter in its ensemble transform implementation (Hunt et al., 2007), while the case $L = 1$ corresponds to the iterative ensemble Kalman filter (Sakov et al., 2012).

1.3 Outline

In Sect. 2, the L95-GRS model with a reduced photochemistry module is introduced and justified. This entails numerical complications similar to what is experienced in larger numerical CTMs and CCMMs due to the numerical stiffness of the chemical reactions. Its physical and chemical relevance is also discussed. The following sections illustrate the usefulness of these models. In Sect. 3, additional experiments with the IEnKS on L95-T are described, focusing on features not discussed in Bocquet and Sakov (2013). In particular, data assimilation for the full L95-T model is compared to an offline variant where the tracer data assimilation system is operated independently from the L95 data assimilation system. We also demonstrate the importance of the emission regime for the efficiency of the data assimilation scheme and we assess the impact of the tracer observation network density. In Sect. 4, the EnKF and IEnKS are applied to the newly developed L95-GRS model, with emphasis on the precision, localisation, and parameter estimation. Conclusions are given in Sect. 5.
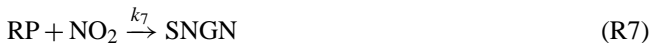
2 A low-order photochemical and transport model

In this section, we substitute the tracer part of L95-T for a reduced-order photochemical kinetic mechanism to form the low-order coupled chemistry meteorology model L95-GRS. We will first describe the resulting model, then we will evaluate its ability to reproduce major physical and chemical characteristics of the processes considered. All the parameters and equations described in the following, with additional details, are gathered in Appendix B.

2.1 Description of the model

The photochemistry module is based on the GRS of Azzi et al. (1992). GRS consists of seven chemical species meant to represent the atmospheric chemistry of ozone formation from VOC (volatile organic compound) and NO_x emissions. The chemical reactions are

$$\text{ROC} + \text{h}\nu \xrightarrow[k_1]{\text{O}_2} \text{RP} + \text{ROC} \tag{R1}$$

$$\text{RP} + \text{NO} \xrightarrow{k_2} \text{NO}_2 \tag{R2}$$

$$\text{NO}_2 + \text{h}\nu \xrightarrow{k_3} \text{NO} + \text{O}_3 \tag{R3}$$

$$\text{NO} + \text{O}_3 \xrightarrow[k_4]{\text{O}_2} \text{NO}_2 \tag{R4}$$

$$RP + RP \xrightarrow{k_5} RP \tag{R5}$$

$$RP + NO_2 \xrightarrow{k_6} SGN \tag{R6}$$

$$RP + NO_2 \xrightarrow{k_7} SNGN \tag{R7}$$

where ROC represents the reactive organic compounds, RP is the radical pool, SGN is the stable gaseous nitrogen product, and SNGN is the stable non-gaseous nitrogen product. The kinetic rate constants, taken from Venkatram et al. (1994), are as follows

$$k_1 = 10\,000 \times e^{-\frac{4710}{T}} \times k_3 \ \mathrm{min}^{-1}, \tag{5}$$

$$k_2 = 5.482 \times e^{\frac{242}{T}} \ \mathrm{ppb}^{-1}\,\mathrm{min}^{-1}, \tag{6}$$

$$k_4 = 2.643 \times e^{-\frac{1370}{T}} \ \mathrm{ppb}^{-1}\,\mathrm{min}^{-1}, \tag{7}$$

$$k_5 = 10.2 \ \mathrm{ppb}^{-1}\,\mathrm{min}^{-1}, \tag{8}$$

$$k_6 = 0.12 \ \mathrm{ppb}^{-1}\,\mathrm{min}^{-1}, \tag{9}$$

$$k_7 = 0.12 \ \mathrm{ppb}^{-1}\,\mathrm{min}^{-1}, \tag{10}$$

where T is the temperature, chosen to be constant equal to 300 K for the sake of simplicity and k_3 is the photolysis rate for NO_2 in min^{-1}, which is a function of sunlight. Details are given in Appendix B.

Since $k_6 = k_7$ and Reactions (R6) and (R7) are similar, one can merge the last two species of the scheme into a lumped species and use a kinetic rate of $0.24\,\mathrm{ppb}^{-1}\,\mathrm{min}^{-1}$ that represents the formation of all the stable nitrogen products, whether gaseous or non-gaseous, i.e.

$$RP + NO_2 \xrightarrow{k=2\cdot k_6} S(N)GN. \tag{R8}$$

To further reduce the GRS scheme and improve the efficiency of its numerical implementation, we use the quasi-steady-state approximation (QSSA) for the radical pool species RP, which is highly reactive and has the shortest lifetime among all the GRS species. This means that the radical pool is in a dynamic equilibrium, adjusting rapidly to the other species concentrations. Solving the algebraic second-order equation for the concentration of the radical pool, we obtain

$$[RP] = \frac{k_2[NO] + 2k_6[NO_2]}{2k_5}$$
$$\left(\sqrt{1 + \frac{4k_1 k_5 [ROC]}{(k_2[NO] + 2k_6[NO_2])^2}} - 1 \right). \tag{11}$$

This approximation has been validated a posteriori. There was little to no impact on the simulated concentrations, while the mean adaptive time step of the chemistry solver increased significantly. Further explanations on how Eq. (11) is obtained are given in Appendix B.

GRS is coupled to the L95 model. As for the L95-T model, the L95 variables are seen as wind speeds that advect the GRS chemical species. The objective is, therefore, to create a simplified model that is able to reproduce the temporal variability of ozone chemistry at a regional to transcontinental scale. There is a total of 40 wind variables and 200 concentration variables, namely the ROC, NO, NO_2, O_3, and S(N)GN concentrations at each of the 40 grid points defined on the circle using the C-grid. Note that the RP concentrations are obtained from Eq. (11).

The transport equations for species $[C_i]$ are consequently

$$\frac{d[C_i]_{m+\frac{1}{2}}}{dt} = \psi_m^i - \psi_{m+1}^i + R_i\left([C_j]_{j=1,\dots,6}\right)$$
$$- \lambda_i [C_i]_{m+\frac{1}{2}} + E_{m+\frac{1}{2}}^i, \tag{12}$$

$$\text{with} \quad \psi_m^i = x_m [C_i]_{m-\frac{1}{2}} \quad \text{if} \quad x_m \geq 0, \tag{13}$$

$$\psi_m^i = x_m [C_i]_{m+\frac{1}{2}} \quad \text{if} \quad x_m < 0, \tag{14}$$

where λ_i is the scavenging ratio, E^i is the emission rate and $R_i\left([C_j]\right)$ is the production term for C_i. There is no such production term for ROC ($R_{ROC} = 0$) so that ROC behaves as the tracer of L95-T. The full equations are given in Appendix B for completeness.

When L95-GRS is seen as a global low-order model, the photolysis rate constant k_3, which depends on sunlight, should vary around the domain and with the season since it is directly linked to the solar zenith angle at a given grid point. Hence, there are points on the grid where it is *nighttime*, with $k_3 = 0$, and others where it is *daytime*, with $k_3 \neq 0$. However, for the sake of simplicity, it has been chosen constant over the domain and it varies according to a uniform daily cycle. This choice does not impact the order of magnitude of the simulated concentrations. A test where the coefficient varies around the domain was performed and led to the same visual result as in Fig. 3 but with a delay around the domain: the black stripes of the figure that signal the time when the NO concentrations reach 0 are slanted instead of straight.

As ROC is not consumed in Reaction (R1), it will eventually produce enough RP to consume all the NO, NO_2, and O_3. Therefore, we have added emission fluxes for ROC and NO_x and a single scavenging ratio for all the species. The emissions are considered constant over time and uniform over the domain, even though a distinction between continent and ocean will also be made in the following. These constants have been chosen using a genuine emission inventory. Since the domain of our model is supposed to be a mid-latitude circle discretised with 40 grid points, one cell of our domain is roughly a few hundred kilometres in length. We used an emission rate for NO_x of $0.27\,\mathrm{ppb}\,\mathrm{day}^{-1}$, where NO accounts for 90 % of these emissions and NO_2 for 10 %. This corresponds to an emission of $3\,\mathrm{kg}\,\mathrm{year}^{-1}\,\mathrm{inhabitant}^{-1}$ of NO for 60 million inhabitants in a volume of $700\,\mathrm{km} \times 700\,\mathrm{km} \times 3\,\mathrm{km}$ (typically France). We have fitted the values of the ROC emission and scavenging coefficient in order to obtain concentrations of O_3 and NO_x within the range of realistic continental concentrations. Specifically, we used an emission rate
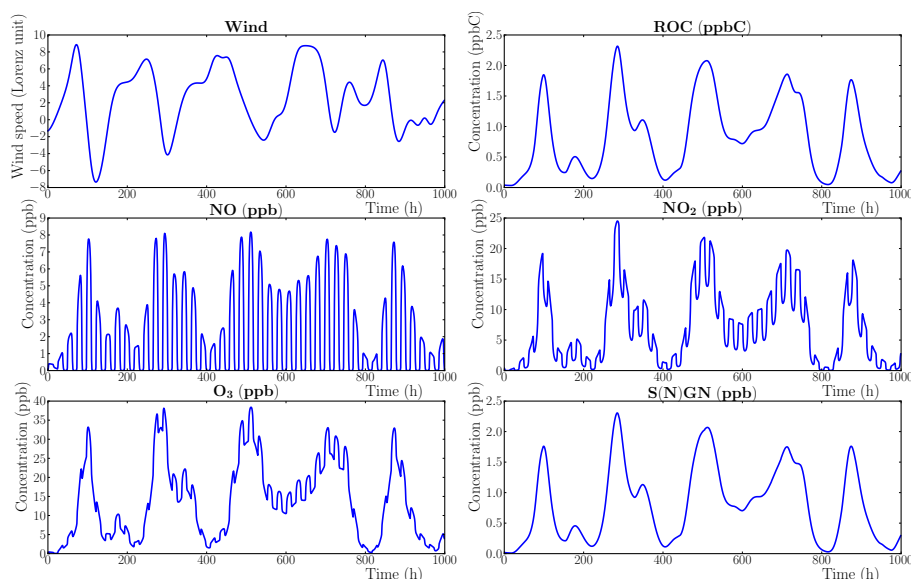
Figure 1. Time evolution of the L95-GRS variables at one grid point. The L95 variables, flagged "Wind", are shown with the original Lorenz unit, while the concentration unit is ppb (ppb C for ROC).

of $0.0235\,\text{ppb}\,\text{C}\,\text{day}^{-1}$ for the ROC species and a scavenging ratio of $0.02\,\text{day}^{-1}$. This ratio is the same as the reference value of the L95-T model, since 1 Lorenz time unit corresponds to 5 days. All parameters are listed in Appendix B.

2.2 Time integration of the model

The L95-T model is integrated in time using a fourth-order Runge–Kutta (RK4) scheme with a time step of $\delta t = 0.05$ in Lorenz time units, i.e. 6 h (Bocquet and Sakov, 2013).

Similarly, the L95 subsystem of the L95-GRS model and the transport part are integrated with the RK4 scheme. A first-order splitting of this integration and the chemistry integration is performed, integrating first the L95 and species transport part, followed by the GRS integration.

The chemical reactions of L95-GRS have a wide range of rates, which leads to numerical stiffness. Hence, the RK4 scheme is an inadequate solver to integrate the chemistry, even though it is more precise. An implicit or semi-implicit scheme is required. That is why the GRS chemical scheme is integrated with a second-order Rosenbrock method, following Hundsdorfer and Verwer (2003). This method is costly since it is based on a semi-implicit scheme that requires using the tangent linear model and solving two linear systems. This is potentially the most time-consuming operation of the whole model integration. Since the chemistry is local and because of the splitting, the Rosenbrock scheme is actually implemented block-wise, one block per grid point. The linear systems to be solved point-wise have a size equal to the number of species. Because the integration of the chemistry is block-wise, it can easily be parallelised. The tangent linear model of GRS required by the Rosenbrock scheme is simple

to derive analytically and implement given the limited number of reactions.

Furthermore, an adaptive time stepping has been implemented that adjusts the time step to the instantaneous stiffness of the reaction rates. However, it has often been proven unnecessary in the free model run (i.e. without data assimilation) in conjunction with the QSSA used for the radical pool.

The typical integration time step for the chemistry is $\delta t = 1\,\text{h}$. The L95 and transport subsystem of L95-GRS is integrated with $\delta t = 1\,\text{h}$ ($\delta t = 0.05/6$ in Lorenz units).

2.3 Qualitative analysis of the L95-GRS model

The outcome of a free run (after spin-up) at a grid point is shown in Fig. 1. One notices the daily cycle induced onto the NO, NO_2, and O_3 concentrations by the variation of the photolysis coefficient k_3, since they are directly related to the value of that coefficient through Reactions (R3) and (R4). The wind speed and orientation variations are responsible for the wave with a period of about 1 week. In real situations, O_3 and NO concentrations can drop close to zero at night (Finlayson-Pitts and Pitts Jr., 1986). In our simulation results, only NO reaches zero at night while O_3 remains at high levels. However, if the ROC emissions are sufficiently lowered or if the NO_x emissions are sufficiently increased, the opposite behaviour occurs.

This model, even if not chaotic, is highly nonlinear, exhibiting distinct chemical regimes. This can be seen in Fig. 2, which represents ozone isopleths for different mean ROC and NO_x concentrations. This feature is typical of lower troposphere ozone chemistry and is commonly known as an Em-
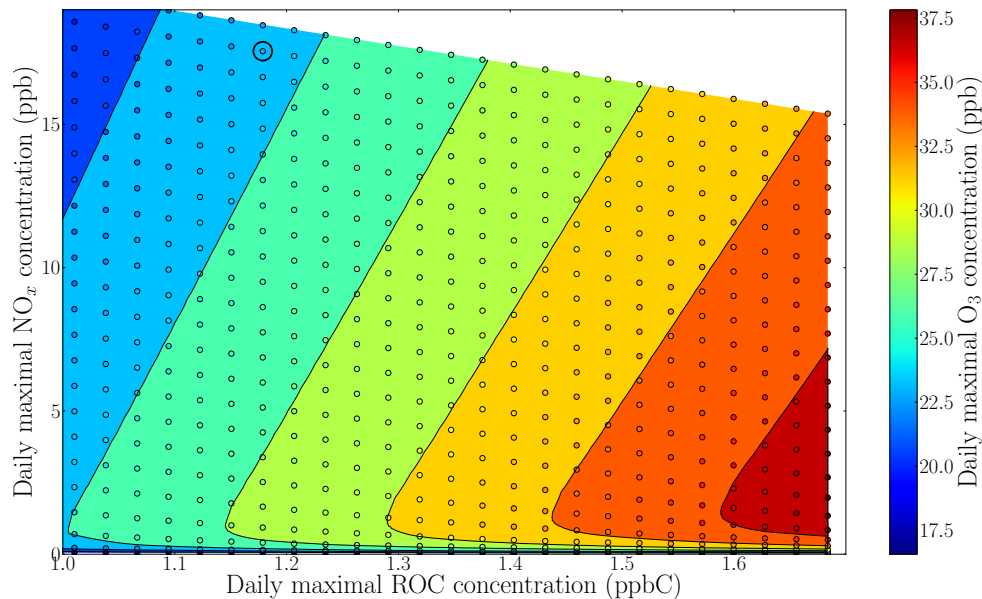
Figure 2. Maximal ozone concentration (in ppb) averaged over the domain depending on maximal averaged ROC (in ppbC) and NO_x concentrations (in ppb). Each dot corresponds to a run with a different emission rate for ROC and NO_x, leading to different maximal averaged concentrations. The reference run is circled in black.

pirical Kinetic Modeling Approach (EKMA) diagram (Dimitriades, 1977). Two different regimes are visible in this graph. The top part of the diagram, where the isopleths are steep, corresponds to a ROC-limited regime. In this regime, a reduction in the emissions of the ozone precursor NO_x leads to an increase of the ozone concentrations (as long as the regime does not change), while a diminution of ROC emissions reduces the ozone concentrations significantly. This is due to the fact that in that ROC-poor regime, RP concentrations are low and NO reacts preferentially with O_3. On the contrary, the bottom part of the diagram, where the isopleths are flat, corresponds to a NO_x-limited regime. In this regime, a reduction in the emissions of the ozone precursor NO_x leads to a strong decrease in the ozone concentrations, whereas the ROC emissions have little to no impact on the ozone concentrations. Since the black circle corresponds to our reference case, it is located in the ROC-limited regime.

In the NO_x-limited regime, the low levels of NO concentrations reduce the amplitude of the daily cycle of the ozone. Hence, the resulting concentrations rather correspond to a background ozone simulation, with very low concentrations of NO_x and little daily variability of ozone. Because GRS is meant to be used with concentrations typical of urban areas, we chose to remain in a ROC-limited regime even though a global simulation should be NO_x-limited. Nevertheless, several free runs and data assimilation experiments have been performed as well in a NO_x-limited regime and lead to one noteworthy result: ROC concentration estimations are worse than in our reference case, unlike NO_2. This makes sense because the NO_2 concentrations mainly control

the model and an error on the ROC concentrations has less impact on other species in this context.

To emphasise the impact of the transport of the chemical species by the wind in the model, an experiment was performed, where the domain was split into a *continental* zone and an *oceanic* zone. In this experiment, we set the ROC and NO_x emissions on the *continent* to $E_i > 0$ for i in $[1, 20]$ and on the *ocean* to $E_i = 0$ for i in $[21, 40]$. The results of this experiment, displayed in Fig. 3, show that puffs of ozone and its precursors can cross the *ocean*, similarly to what is witnessed over the Pacific (Lin et al., 2012) and the Atlantic (Guerova et al., 2006). Moreover, ozone concentrations are higher above the ocean in the absence of NO emissions. Note also that the tracer plumes move eastward (increasing indices), which is consistent with a positive group velocity for the L95 model, while the peaks and lows of the L95 field move westward according to the L95 negative phase velocity (Lorenz and Emanuel, 1998).

So far, the wind kinetics (amplitude and variability) has been determined by the original L95 model characteristics. In the reference experiment, the waves of the wind extend over several days. The concentrations are driven by this wind kinetics but vary within those waves according to the photochemical daily cycles. However, other types of behaviour are possible with L95-GRS by choosing differently the timescale of the L95 model. If time within the L95 model is rescaled by α and the wind variables are rescaled by β,

$$\frac{\mathrm{d}x_m}{\mathrm{d}t} = \alpha \left[\beta(x_{m+1} - x_{m-2})x_{m-1} - x_m + \frac{F}{\beta} \right], \qquad (15)$$
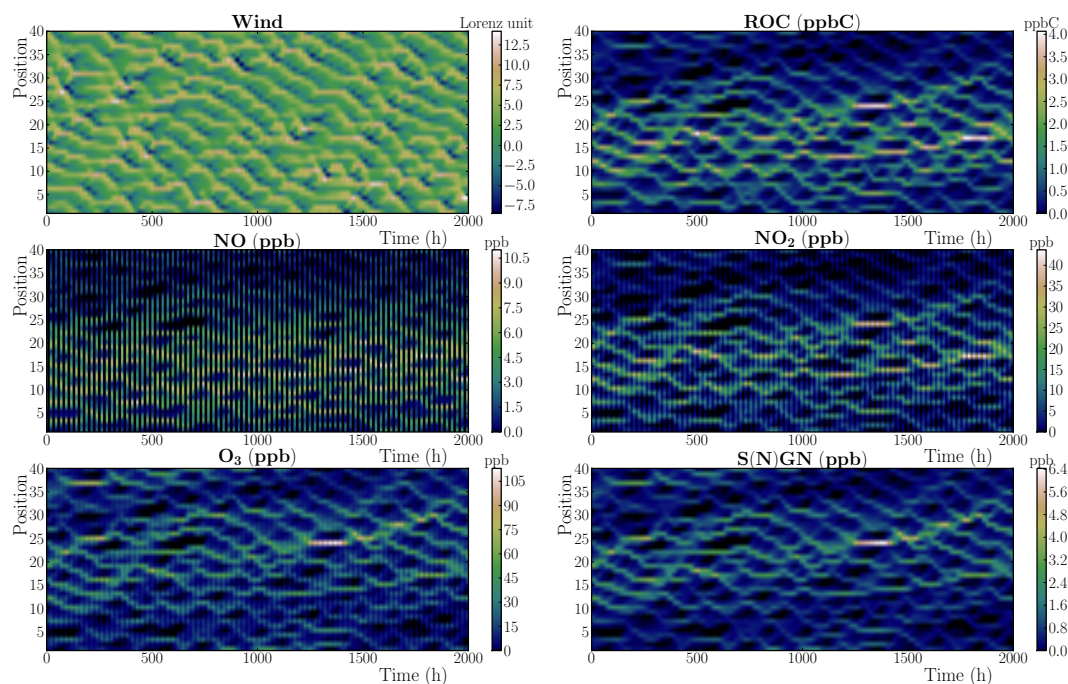
Figure 3. Time evolution of the L95-GRS variables over the whole domain in the case of a continent/ocean division. The L95 variables, flagged "Wind", are shown with the original Lorenz unit, while the concentration unit is ppb (ppbC for ROC).

it is possible to reduce the period of the wind wave and to match that of the chemical daily cycles. This way, the species concentrations are significantly modulated by the wind variations. In terms of spatial scale, this would correspond to regional modelling rather than continental to global modelling of the species fields. Figure 4 illustrates this timescale change with $\alpha = 20$ and $\beta = 1$. The winds fluctuate at a higher rate than the concentrations, quite differently from the reference configuration of Fig. 1. Adjusting β, it is also possible to rescale the amplitude of the winds in the L95 equations to match a more regional/lower troposphere transport behaviour with weaker mean wind magnitude. Note that α, β are only rescaling parameters that do not fundamentally impact the nature of the model dynamics in contrast to, e.g. Carrassi et al. (2008).

3 Numerical experiments with the L95-T model

In this section, we experiment on the use of data assimilation techniques for forecasting and reanalysis with the tracer model (L95-T) beyond the preliminary results of Bocquet and Sakov (2013). The aim is to demonstrate the advantages brought by this model to study certain data assimilation strategies. Several of the results and interpretations in this section will also apply to data assimilation systems operating with the L95-GRS model.

3.1 Definition of online and offline data assimilation systems

A typical offline model is a CTM where the meteorological fields have been generated externally and are given as an input to the model. These fields usually stem from operational meteorological prediction centres or from any independently run meteorological model. On the other hand, online models consistently process meteorology, chemistry, and transport of species all together, but at a higher numerical cost. The choice of an offline or online approach is a crucial issue as far as modelling is concerned (Zhang et al., 2012). It is even more so when data assimilation techniques are applied to offline or online models because of the fluxes of information between the two subsystems: chemistry and transport on the one hand and meteorology on the other hand (Milewski and Bourqui, 2011; Bocquet et al., 2015a).

The L95-T model stands as a well-suited simple tool to experiment on this issue. In the following, we apply the quasistatic IEnKS to L95-T using either an offline or an online approach. A distinction is made between

– The full online data assimilation system for the L95-T model. Even though the L95 subsystem of the model does not depend on the tracer subsystem, it should be kept in mind that information propagates both ways in advanced data assimilation methods, as long as the error covariance matrices are defined over both subsystems.
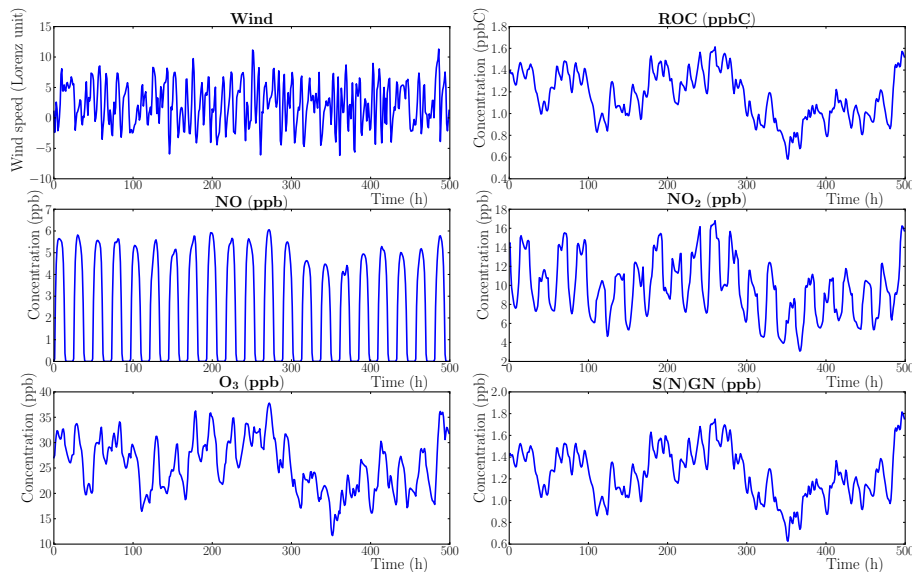
Figure 4. Time evolution of the L95-GRS variables at one grid point with a time rescaling of $\alpha = 20$ applied to the L95 model. The L95 variables, flagged "Wind", are shown with the original Lorenz unit, while the concentration unit is ppb (ppb C for ROC).

– The offline data assimilation system for the L95-T model. The L95 subsystem is run separately. The IEnKS is applied to L95 with a DAW length L_w (in units of Δt). The IEnKS is applied separately to the tracer subsystem (transport, deposition, and emission) with a DAW length L_c (in units of Δt). For advection, the winds are provided by the analyses of the independent L95 data assimilation system. Therefore, no feedback from the tracer subsystem to the L95 subsystem is to be expected. The information gained from the observations flows one way. Moreover, for the tracer subsystem, the uncertainty of the wind field constitutes a realistic and significant source of model error. We believe that this is an elegant way to create consistent model error, beyond stochastic noise or offset parameters.

We conduct synthetic data assimilation experiments, applying the IEnKS to the L95-T model. A simulation of L95-T that represents the truth is generated, with $E = 1$, $\lambda = 0.1$. Synthetic observations are generated from the truth every $\Delta t = 0.05$. The system is fully observed, on both wind and concentration variables. An unbiased Gaussian white noise is used to perturb the observations. At any observation time, the observation error covariance matrix for the wind and the tracer is in both cases $\mathbf{R} = \mathbf{I}$ the identity matrix. The analysis, output of the data assimilation system, is compared to the truth using a root mean square error (RMSE), for the meteorological subsystem as well as for the tracer subsystem. Reliable statistics are computed over runs of $10^5 \times \Delta t$ with a burn-in period of $5 \times 10^3 \times \Delta t$.

The ensemble size of the IEnKS has been chosen to be $N = 20$ in a dynamical regime of L95-T where localisation is unnecessary. Yet, sampling errors due to the finite-size of the ensemble would require the use of inflation of the errors. To avoid this issue, we use the finite-size scheme of Bocquet (2011), Bocquet and Sakov (2012), and Bocquet et al. (2015b) where no inflation tuning is necessary. Practically, it means that whenever the IEnKS is mentioned, the finite-size IEnKS (IENKS-N) has been used instead or, equivalently, that we enforced optimal inflation meant to account for sampling errors. However, note that the finite-size approach does not account for model error but rather sampling errors.

We consider several practical variants of the offline data assimilation system for the tracer model. In the first offline system, called *Offline 1a* in the following, the mean analysis wind is provided to the IEnKS of the tracer subsystem, both for the forecast step and the analysis step of the IEnKS. In this baseline case, we choose $L_w = L_c$. In a second variant, the winds are obtained through an EnKF; i.e. $L_w = 0$ and L_c is varied (experiment *Offline 1b*). In a third variant, the winds are obtained from an IEnKS with a given L_w and an EnKF is run for the tracer subsystem, i.e. $L_c = 0$ (experiment *Offline 1c*).

Because the uncertain winds are a source of model error for the offline system, we also implement a multiplicative inflation on top of the IEnKS-N. It is applied on the prior by a rescaling of the anomalies. We choose the inflation that leads to the best RMSE.

In a last variant of the offline model, called *Offline 2*, the analysis mean wind is still provided for the *analysis step* of the IEnKS applied on the tracer subsystem. Yet, the full analysis wind ensemble, rather than the mean, is provided in the *forecast step* of the IEnKS applied on the tracer subsystem. If the wind ensemble spread is representative of the wind uncertainty, it is hoped that the uncertainty in the winds

will be properly accounted for. As in the Offline 1 experiments, a multiplicative inflation is also applied for any residual model error.

The full online data assimilation system is also run for comparison (experiment *Online*), without any multiplicative inflation.

3.2 Comparison of online and offline data assimilation systems

The performance of these systems as a function of the DAW length is reported in Fig. 5. The best estimate of the present-time wind and concentration state vectors are compared to the truth leading to the *filtering* RMSE, which is a good indication of the forecasting quality in this context. Best estimates for the past state vectors (reanalysis) are also compared to the truth, leading to the *smoothing* RMSE. For a DAW of length L, a run of length N_t (both in units of Δt), and a state vector of size M, the formulas of the RMSEs are

$$\text{RMSE}^{\text{filtering}} = \frac{1}{N_t} \sum_{i=1}^{N_t} \sqrt{\frac{1}{M} ||\mathcal{M}_{L \leftarrow 0}(\boldsymbol{x}_a^i) - \boldsymbol{x}_t^{i+L}||^2}$$

and

$$\text{RMSE}^{\text{smoothing}} = \frac{1}{N_t} \sum_{i=1}^{N_t} \sqrt{\frac{1}{M} ||\boldsymbol{x}_a^i - \boldsymbol{x}_t^i||^2},$$

where $\boldsymbol{x}_a^k$ is the state analysis at time k, $\boldsymbol{x}_t^k$ is the truth at time k, and for a vector $\boldsymbol{x}$ of size M, $||\boldsymbol{x}||^2 = \sum_{j=1}^{M} x_j^2$.

First of all, the online system has a very significant edge over the offline systems because of the two-way information flows, both for the concentration variables and for the wind variables. This shows that concentration observations can significantly improve meteorological forecasts, in agreement with the results of Semane et al. (2009) obtained when assimilating real observations of lower stratospheric ozone. For all offline systems, the wind variables cannot benefit from the assimilation of concentrations, but only from the L95 observations. Therefore, from now on, we shall focus on the tracer subsystem.

The extrinsic model error due to the uncertain winds must be accounted for in the tracer subsystems. Otherwise, the ensemble of the tracer subsystem collapses (the ensemble method diverges). In the absence of any correction for model error, we observe that the estimation is close to a free run, with an average filtering RMSE of about 0.65.

Yet, as expected, accounting for model error offers better performance. Let us first consider cases 1a, 1b, and 1c that use the best estimate of the mean wind and apply multiplicative inflation to account for model error. Configuration Offline 1a, i.e. when $L_c = L_w$, offers a baseline performance for the filtering and smoothing RMSEs, which improves as the joint DAW length increases. It remains quite far from the

performance of the online system since multiplicative inflation cannot compete with a better wind estimate.

With configuration Offline 1b, where the mean wind estimate comes from an EnKF ($L_w = 0$), the average filtering RMSE does not benefit from an extended DAW for the tracer, while the smoothing RMSE only marginally benefits from short DAW (up to $L_c = 2$) before degrading. Hence a longer window for this CTM system is inefficient. We attribute this important property to the stable and linear dynamics of transport.

With configuration Offline 1c, the tracer data assimilation system is based on an EnKF, while the wind estimation gets better as L_w gets larger. Therefore, the improvement that is observed for the filtering RMSE comes from reduced model errors. In this configuration, the filtering and smoothing RMSEs coincide since the concentrations are merely estimated by an EnKF ($L_c = 0$).

In the light of these results, we understand that the improvement that is observed in configuration 1a comes from the reduced uncertainty in the wind fields in the first place. Note that as $L_c = L_w$ gets larger, the tracer analysis within the DAW of length L_c uses wind fields with lower error thanks to smoothing. This explains why the improvement in the RMSEs is more pronounced than in configuration 1b, which only benefits from the filtered winds at present time.

With configuration Offline 2, model error is addressed by not only the multiplicative inflation but also the ensemble of winds in the forecast steps. Each wind member is ascribed to a tracer member. This is similar to stochastic parametrisation where one changes the model input parameters for each member of the CTM (Wu et al., 2008). This shows much better performance. As far as filtering is concerned, the optimal inflation is an increasing function of L_c (with an inflation of 1.06 for anomalies at $L_c = 25$), whereas the absence of inflation is optimal for smoothing.

One lesson is that a variational analysis over a long DAW is useless for the *offline* transport subsystem, because of its linear dynamics; an IEnKS, or a 4D-Var analysis does not perform better than an EnKF analysis in this context. This conclusion will not necessarily hold with the L95-GRS model because of the nonlinear chemistry.

3.3 Emission/deposition regime

The scavenging ratio λ and the emission rate E control the tracer mass budget in the domain. Their ratio can lead to different regimes of the physics of the model. It can impact the performance of data assimilation. In the reference case ($E = 1$, $\lambda = 0.1$), parcels of tracer travel over large distances before deposition. Hence, an observation of the tracer concentration at a grid point gives information about the wind magnitude and direction at other grid points several time steps in the past.

A synthetic experiment where the scavenging ratio λ is varied over several orders of magnitude has been set up to
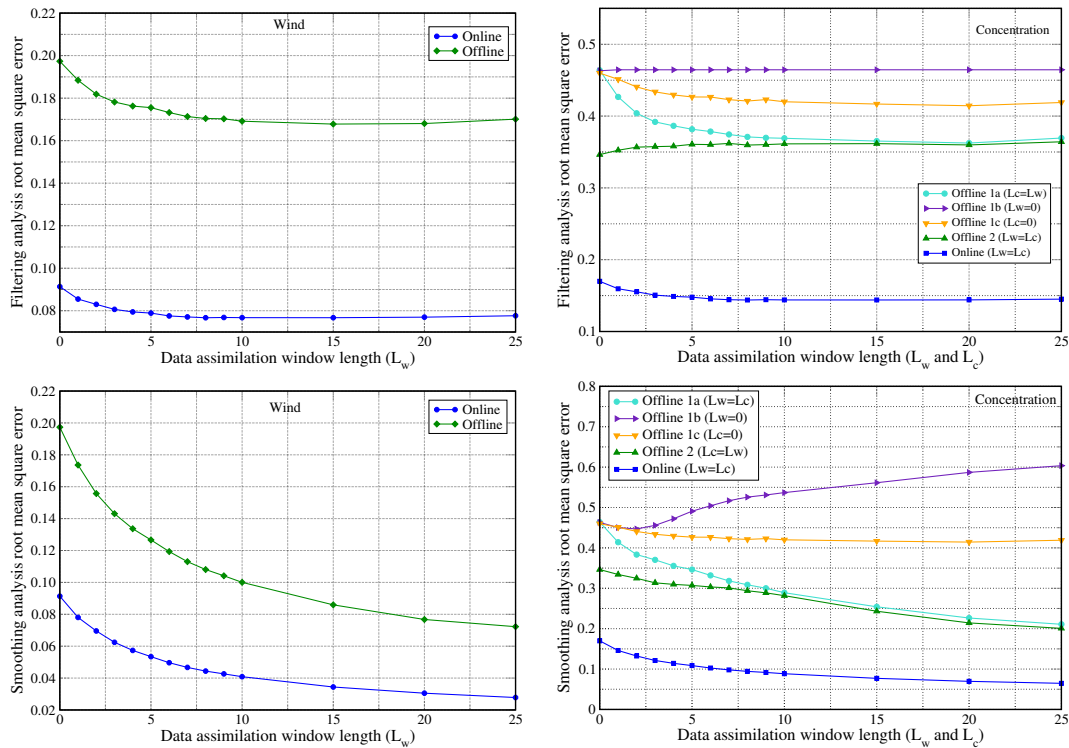
Figure 5. Average RMSEs of the L95-T data assimilation system using the IEnKS, as a function of the DAW length (in units of $\Delta t = 0.05$ of the L95 model). The two top panels show the filtering RMSEs, while the two bottom panels show the smoothing RMSEs. The scores of the wind variables are on the left, while scores of the concentration variables are on the right. The case $L = 0$ corresponds to the ensemble transform EnKF.
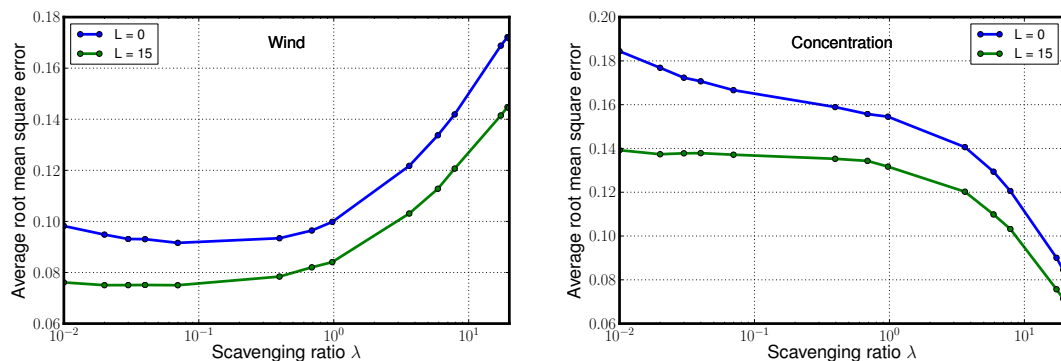

Figure 6. Average filtering analysis RMSEs of the wind variables (left) and concentration variables (right) of the L95-T, as a function of the scavenging ratio for the ensemble transform EnKF (IEnKS with $L = 0$) and the IEnKS with $L = 15$.

highlight this point. The emission flux E is tuned in order to keep the ratio E/λ constant. Thus, the order of magnitude of the concentrations is unchanged so that the relative precision of the concentration observations remains the same with an unchanged error covariance matrix. The set-up of this experiment is the same as in the previous section, but only the online data assimilation system based on the L95-T model is used.

The average filtering RMSE of the concentration variables and of the wind variables are plotted in Fig. 6 as a function of the scavenging ratio. The RMSE remains rather constant for both the concentrations and the winds for small scavenging ratios, with the same performance as in the reference case ($E = 1$, $\lambda = 0.1$). However, as soon as $\lambda > 1$, the behaviour changes. With such higher values of the scavenging ratio, the wind does not have sufficient time to transport the tracer over significant distances. The information about the wind embed-

ded in the observations of the concentrations diminishes and the wind RMSE increases. On the contrary, the absence of transport of the tracer by the wind reduces the detrimental impact of diffusion making the concentrations easier to estimate (in the fully observed configuration at least). Moreover, the benefit of using a larger DAW ($L = 15$ here) is greater with lower scavenging ratios because the tracer is advected farther in space. Indeed, pieces of information contained in concentration observations help estimate the wind back in time before the tracer was advected by that wind. These results stress that variational schemes implemented over large DAW have an advantage over the EnKF in that context.

3.4 Observation network

The performance of the EnKF and of the IEnKS is now studied with the L95-T model when the observations of the tracer concentrations are sparser. The set-up of this synthetic experiment remains unchanged except for the density of the observations. The wind variables are observed on all grid points while only some of the observations of the tracer concentrations are assimilated. The observations of the concentrations are chosen to be evenly spread. The number of observations is a divisor of 40 and belongs to $\{1, 2, 4, 5, 8, 10, 20, 40\}$. The performance of the IEnKS is studied for several DAW lengths $L = 0, 1, 2, 3, 4, 5, 10, 20$. The resulting average filtering RMSEs for the concentration variables and the wind variables are shown in Fig. 7. In both cases, the gain derived from using a longer DAW is undeniable: the gain in RMSE is greater for large L than for small L, but the marginal gain decreases with L.

4 Applying the IEnKS to the L95-GRS model

The IEnKS is now applied to the L95-GRS model introduced in Sect. 2. We showed that the model could reproduce the main physical and chemical processes of interest. The aim in this section is to show that it offers a rich playground for testing data assimilation methods. The approach is similar to the one applied in Sect. 3. Apart from an overall performance test, we will focus on specific aspects not addressed in Sect. 3 of relevance for this type of model.

Twin experiments are conducted where each chemical species is observed. The observations are drawn from the truth every $\Delta t = 6\,\text{h}$ and perturbed using a Gaussian white noise. The standard deviations of the error for the concentrations have been chosen to correspond to about 10 % of the average value of the concentration over the domain. Specifically, the observation error covariance matrix of each species is of the form $\mathbf{R} = \sigma^2 \mathbf{I}$, where $\sigma^2 = 1$ for the wind in Lorenz units, $\sigma^2 = 0.01\,\text{ppb C}^2$ for ROC, $\sigma^2 = 0.16\,\text{ppb}^2$ for NO, $\sigma^2 = 1\,\text{ppb}^2$ for NO$_2$, $\sigma^2 = 4\,\text{ppb}^2$ for O$_3$, and $\sigma^2 = 0.01\,\text{ppb}^2$ for S(N)GN. All the RMSEs shown in this section are normalised by the observational error standard deviation

of the corresponding species. All the data assimilation runs use the same set-up as used with the L95-T model. The size of the ensemble is set again to $N = 20$, except when localisation is tested.

4.1 Performance

At first, the number and distribution of observations of the concentration variables have been varied following the same set-up as in Sect. 3.4. All the chemical species are observed but only at selected grid points. The resulting average filtering RMSEs for the concentration variables and the wind variables are shown in Fig. 8. We found that with poor observability of the concentrations, the system's state estimate can be imprecise. When the concentrations are only observed at one point, the EnKF diverges from the truth. The IEnKS with $L = 1$ also fails to estimate the S(N)GN better on average over the whole domain than the standard deviation of the single observation.

To be more realistic, further experiments will assimilate sparse concentration observations. We choose to keep eight observations in the domain per species, that is to say at 1 every 5 grid points. In this context, the DAW length has been varied over a wider range of values. The time-averaged analysis filtering and smoothing RMSEs for the wind and the concentrations are shown in Fig. 9. Even though the model is strongly nonlinear, the IEnKS method can account for these nonlinearities and, therefore, it performs well and improves with L for both the filtering and smoothing RMSEs. The S(N)GN species is still the one with the worst results, probably because it is little correlated with the other species except for ROC. A misestimation of its concentration has no consequences on the other species.

4.2 Parameter estimation

In atmospheric chemistry, there is a strong dependency of the model on the values of the various forcings, such as the boundary conditions (Roustan et al., 2010), the meteorological fields (Dawson et al., 2007) or the emission rates of the pollutants and their precursors (Cohan et al., 2005). It is therefore important to estimate these inputs and data assimilation can be a powerful tool in this context. Here, we show how the L95-GRS model allows us to test parameter estimation strategies, which is illustrated using the IEnKS.

To estimate a set of model parameters $\boldsymbol{\theta} \in \mathbb{R}^P$ along with the state variables, the state vector is augmented from $\boldsymbol{x} \in \mathbb{R}^M$ to a vector

$$z = \begin{pmatrix} x \\ \theta \end{pmatrix} \in \mathbb{R}^{M+P}, \tag{16}$$

in the joint state and parameter space. It is also necessary to define a forward model for the parameters. The persistence model is chosen, i.e. $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k$.

The estimation of the main parameters of the L95-T model (forcing of the L95 and emission rate of the tracer) with
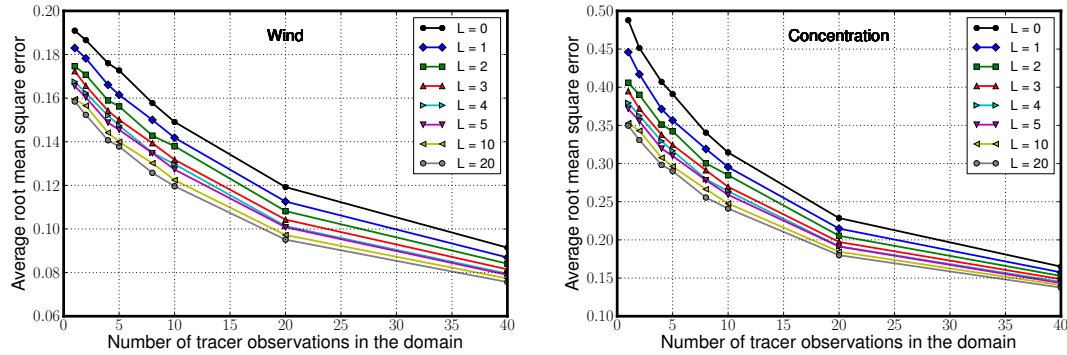
Figure 7. Average filtering analysis RMSEs of the wind variables (left) and concentration variables (right) of the L95-T, as a function of the number of concentration observations for the IEnKS with several DAW lengths. The case $L = 0$ corresponds to the ensemble transform EnKF.
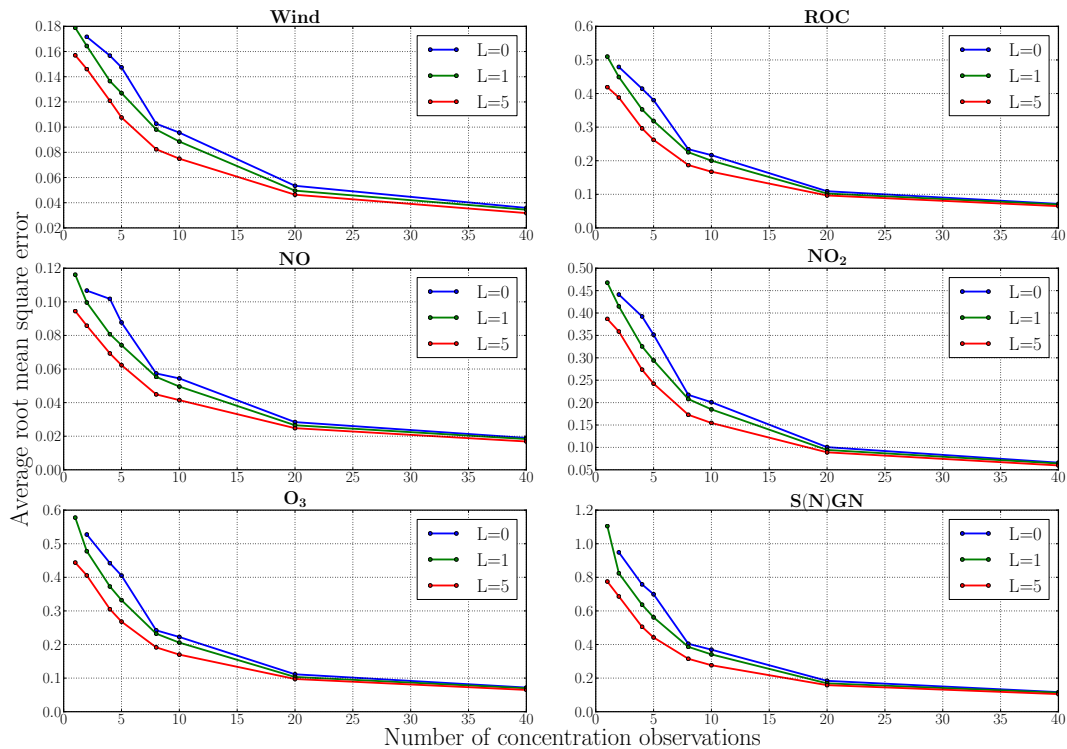

Figure 8. Average filtering analysis RMSEs of the L95-GRS variables, as a function of the number of concentration observations for the IEnKS with three DAW lengths. The case $L = 0$ corresponds to the ensemble transform EnKF. The RMSEs are normalised by the standard deviations of the observation error.

various data assimilation methods, including the IEnKS, has been experimented upon by Bocquet and Sakov (2013). Similarly, we conduct a twin experiment with the L95-GRS model where, in addition to the state variables, the F forcing of the L95 model and the emission rates of ROC and NO_x are estimated simultaneously. The state space is, therefore, augmented from a 240- to a $240 + 3$-vector of the joint state and parameter space. The parameter variables in the ensemble are set as follows.

– For the emission rates: the ensemble is initialised around the truth by adding an unbiased Gaussian noise of standard deviation 10 % of the true value.

– For F: the ensemble is initialised around the value $F = 7$ by adding an unbiased Gaussian noise of standard deviation 10 % of the true value.

Rather than the single data assimilation (SDA) version of the IEnKS presented in Appendix A, we use the multiple data assimilation (MDA) version, which is very similar, less accu-
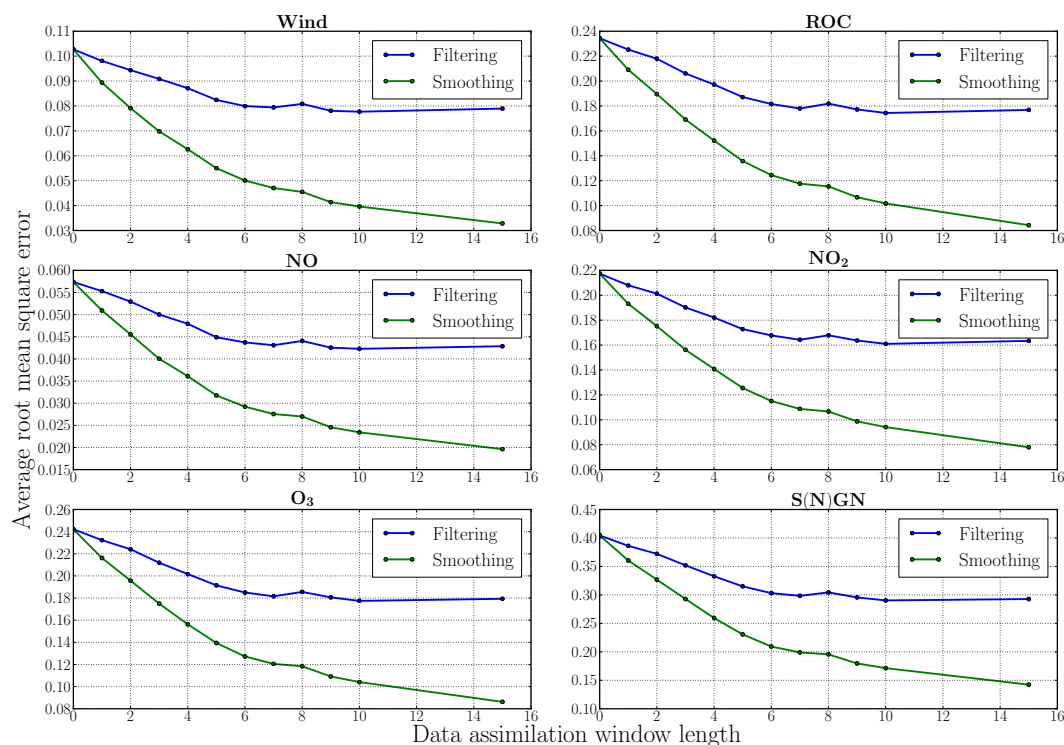
Figure 9. Average filtering and smoothing analysis RMSEs of the L95-GRS variables, as a function of the DAW length (in units of $\Delta t = 6\,\mathrm{h}$). The case $L = 0$ corresponds to the ensemble transform EnKF. The RMSEs are normalised by the standard deviations of the observation error.

rate for small L but more stable for large L. The MDA IEnKS algorithm is described in detail in Bocquet and Sakov (2014).

Let us first mention that the RMSEs of the state variables are barely changed by the joint state and parameter estimation, as well as by the use of a different variant of the IEnKS in this experiment. Hence, the results in Fig. 9 for state variables still hold. The parameter values are plotted in Fig. 10, over time intervals of different lengths. We observe that only a few days are required to converge to the right value for the forcing parameter of the meteorology F, while a few dozens of days are necessary for the emission rates to stabilise. This is due to the high sensitivity of both wind and concentration variables to the forcing F of the wind model, which controls its chaotic behaviour, as well as to the fact that there is a bias on the initial value of F. At the end of a long data assimilation run, the algorithm has converged to the right values with a precision of less than a percent. The use of a long DAW improves the estimation of the parameters and the smoothness of the results. The case $L = 1$ shows that the method can sometimes converge to a biased value for a long period of time.

It could be possible to estimate chemical reaction rates, for instance, the ROC photolysis rate. However, our experiments have shown that the filter diverges. It probably happens because this rate is equal to 0 at night. Hence, it is imperative to set a prior distribution on this type of parameter to avoid

divergence when the model becomes insensitive to the parameter. But, this is outside the scope of this work.

4.3 Localisation

Estimating covariances from a limited size ensemble of state vectors produces spurious long distance correlations between variables. This degrades the estimation of the error statistics and can lead to divergence in ensemble data assimilation methods. To address this issue, localisation is used in high-dimensional systems implementing ensemble methods. There are two main localisation methods known as *covariance localisation* and *local/domain analysis* (Sakov and Bertino, 2011, and references therein). The first method consists in tapering the empirical covariance matrix using a Schur product with short-range correlation function. The second method performs the analysis in a local domain around each grid point, using only nearby observations. Both localisation methods have been tested with success with EnKF techniques applied to the L95-GRS model. Here, we provide an illustration of such experiments.

We tested covariance localisation on the L95-GRS model using the DEnKF data assimilation method from Sakov and Oke (2008). This method has the advantage of being computationally efficient while allowing a straightforward use of Schur localisation. The RMSEs are shown in Fig. 11, as a function of the ensemble size, with optimally tuned infla-
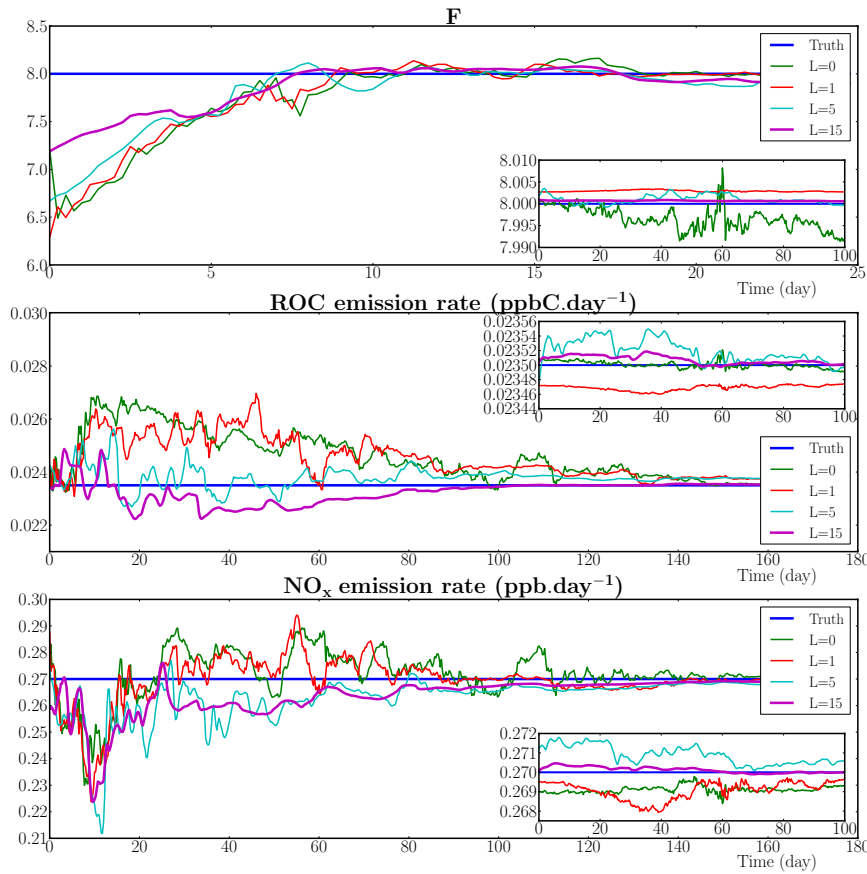
Figure 10. Time evolution (days) of the parameter variables for several DAW lengths without spin-up (main) or after a long time (inset). The case $L = 0$ corresponds to the ensemble transform EnKF. F is shown with the original Lorenz unit, while the emission rate unit is $ppb\,C\,day^{-1}$ or $ppb\,day^{-1}$.

tion and localisation radius. For comparison, the results without localisation are shown as well. We see how localisation allows us to reduce the size of the ensemble below 18 without diverging, even though it leads to a degradation of the scores for very small ensemble sizes ($N < 10$).

5 Conclusions

The aim of this article is to introduce low-order models on which to test advanced data assimilation methods in order to gain insights on some of the many difficulties encountered in data assimilation applied to meteorology and atmospheric chemistry. Amongst them, the questions of inflation, localisation for ensemble methods, model error, online and offline modelling, or nonlinearities have been addressed.

Building on the L95-T model, where the transport of a tracer is coupled to the L95 model, we introduced a new model, L95-GRS, where the tracer part is replaced with a simplified ozone chemistry. The L95-GRS model shows important peculiarities typical of tropospheric ozone chemistry. It has been adjusted to simulate pollutant concentra-

tions of realistic magnitude. Ozone precursors can experience long-range transport by the meteorology and lead to ozone episodes far from the pollutant sources. It is possible to tune the wind magnitude in order to modify the time- and space scale of the model. Moreover, it has stiff equations that require the use of the same numerical tools as high-dimensional CTMs. Last but not least, it shows a nonlinear response to the emission rates of the ozone precursors. It thus includes several of the hardships of high-dimensional chemistry models without the high numerical cost. As such, it can be used to experiment upon and validate new data assimilation methods in the context of atmospheric chemistry modelling and coupled chemistry meteorology modelling.

To illustrate the use of advanced data assimilation methods on these models, and specifically ensemble variational methods, we first performed new experiments on the L95-T with the iterative ensemble Kalman smoother (including the ensemble Kalman filter). We showed that this model is suitable to test online and offline strategies for data assimilation, as well as to emulate model error stemming from a meteorological field, or an ensemble forecast of meteorological fields.
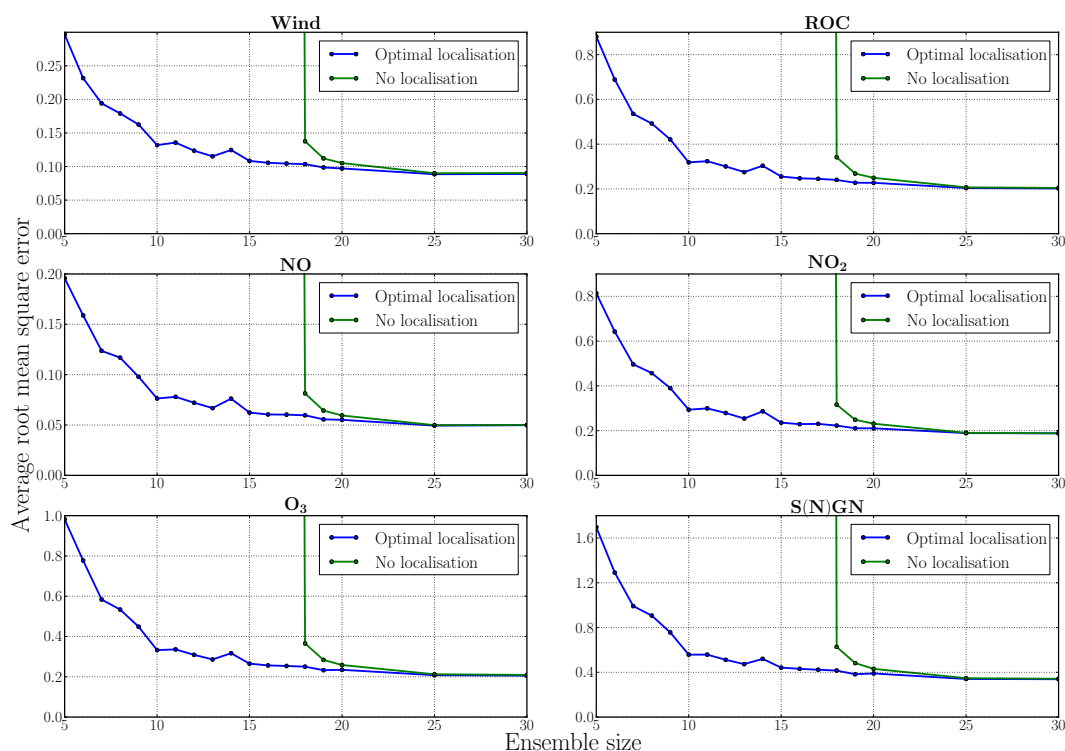
Figure 11. Average filtering analysis RMSEs of the L95-GRS variables, as a function of the ensemble size for the DEnKF without localisation or with optimally tuned localisation radius. The L95 variables, flagged "Wind", are shown with the original Lorenz unit, while the concentration unit is ppb (ppb C for ROC).

More specifically, we experimented on the offline version of the L95-T model, where the meteorology and the tracer subsystems are integrated and assimilated separately. This decoupling introduces model error into the tracer subsystem. In this context, having an ensemble of analyses from a data assimilation on the meteorology as an input to the tracer subsystem gives us a representative sample of this model error. By doing so, we have avoided the use of inflation and obtained optimal performance. We noticed as well that, for data assimilation purposes, the coupling of the two subsystems is only relevant when they have similar evolution timescales. In the case where the tracer subsystem evolves too quickly or too slowly compared to the meteorology, the coupling of these two parts fails to improve the results of the data assimilation compared to an offline case.

The use of data assimilation methods was also illustrated with the L95-GRS model. The iterative ensemble Kalman smoother performs well despite the nonlinearities of the model and even if the observation network is sparse. In particular, the model can help testing parameter estimation techniques with multiple parameters usually met in CCMMs and CTMs. The use of localisation was also successfully tested with L95-GRS. By making this wide range of experiments, we concluded that the L95-GRS model is suitable to test advanced data assimilation schemes.

A broad class of models could be developed by exchanging the L95 meteorological part with another low-order model. The L95 model has anti-correlations in space and time that are not observed in more realistic models. It could be replaced by its continuous extension, the Lorenz 2005-II model (Lorenz, 2005), which could offer a more complex set-up for testing localisation, but still suffers from similar correlation patterns. Alternatively, the L95 meteorological part could be exchanged with the multiscale Lorenz 2005-III model to explore the impact of subgrid-scale model error. To study space–time intermittent behaviours, the model of Kuramoto–Sivashinsky (Vannitsem and Nicolis, 1994, and references within) could be implemented. Alternatively, the Burgers equation could be used to study the impact of a front (concentration/rarefaction) on the chemistry. If such continuous meteorological models are used, the choice of the advection scheme could be revised as well, and a more accurate higher-order one compared to the upwind scheme could be used, alongside with a flux limiter.

The L95-GRS model depends on several key species-dependent chemical and physical parameters that introduce many time- and space scales in the data assimilation system and that impact its observability and controllability. These parameters are likely to be representative of those of realistic CCMMs and CTMs. We have only investigated the impact of a few of those parameters, fixing the others. But a more gen-

eral parameter-wise exploration of data assimilation systems built on L95-GRS is desirable.

Finally, following this study, we are planning to test the IEnKS on the Polair3D CTM of the research and operational Polyphemus modelling platform (Mallet et al., 2007; Sartelet et al., 2012) building on the experience acquired with the L95-T and L95-GRS low-order models.

Code availability

The code for the models L95-T and L95-GRS can be downloaded from the following website: http://cerea.enpc.fr/l95-grs/

Appendix A: Algorithm of the IEnKS

The following pseudo-algorithm specifies a variant of the IEnKS, called the single data assimilation (SDA) IEnKS (Bocquet and Sakov, 2014). The minimisation of the nonlinear cost function is based on a Gauss–Newton scheme (other schemes such as BFGS or Levenberg–Marquardt are possible). The sensitivities are computed by a finite difference scheme (bundle variant, as in Bocquet and Sakov, 2012). They could also be computed using the propagation of the ensemble (ensemble transform variant, as in Sakov et al., 2012) as well. The data assimilation window is of length $L\Delta t$. In the algorithm, t_L is present time. The forward model resolvent from t_k to t_{k+1} is $\mathcal{M}_{k+1\leftarrow k}$; the observation operator at t_k is H_k. ϵ, e, and $j_{\max}$ are algorithmic parameters with values taken to be 10^{-4}, 10^{-3}, and 20, respectively. The outcome is largely independent of these parameters. $\mathbf{E}_0$ is the ensemble matrix at t_0, whose columns are the ensemble vectors: $\boldsymbol{x}_1^0, \boldsymbol{x}_2^0, \ldots, \boldsymbol{x}_N^0$. $\boldsymbol{y}_k$ is the observation vector at t_k. λ is the inflation factor. $\mathbf{U}$ is an orthogonal matrix in $\mathbb{R}^{N\times N}$ satisfying $\mathbf{U1} = \mathbf{1}$, where $\mathbf{1} = (1, 1, \ldots, 1)^{\mathrm{T}}$. In between two updates, the ensemble is forecast over $S\Delta t$. The configuration $L = 0, S = 1$ corresponds to the ensemble transform Kalman filter; $L = 1, S = 1$ corresponds to the iterative ensemble Kalman filter; and for any L, $S = 1$ corresponds to the generic quasi-static IEnKS.

Algorithm 1 A cycle of the lag-L / shift-S / SDA / bundle / Gauss-Newton IEnKS.

1: $j = 0$, $\boldsymbol{w} = \mathbf{0}$
2: $\boldsymbol{x}_0^{(0)} = \mathbf{E}_0 \mathbf{1}/N$
3: $\mathbf{A}_0 = \mathbf{E}_0 - \boldsymbol{x}_0^{(0)} \mathbf{1}^{\mathrm{T}}$
4: **repeat**
5:     $\boldsymbol{x}_0 = \boldsymbol{x}_0^{(0)} + \mathbf{A}_0 \boldsymbol{w}$
6:     $\mathbf{E}_0 = \boldsymbol{x}_0 \mathbf{1}^{\mathrm{T}} + \epsilon \mathbf{A}_0$
7:     **for** $k = 1, \ldots, L$ **do**
8:        $\mathbf{E}_k = \mathcal{M}_{k\leftarrow k-1}(\mathbf{E}_{k-1})$
9:        **if** $k \in [L-S+1, L-1]$ **then**
10:           $\overline{\boldsymbol{y}}_k = H_k(\mathbf{E}_k)\mathbf{1}/N$
11:           $\mathbf{Y}_k = (H_k(\mathbf{E}_k) - \overline{\boldsymbol{y}}_k \mathbf{1}^{\mathrm{T}})/\epsilon$
12:        **end if**
13:     **end for**
14:     $\overline{\boldsymbol{y}}_L = H_L(\mathbf{E}_L)\mathbf{1}/N$
15:     $\mathbf{Y}_L = (H_L(\mathbf{E}_L) - \overline{\boldsymbol{y}}_L\mathbf{1}^{\mathrm{T}})/\epsilon$
16:     $\nabla\widetilde{\mathcal{J}} = (N-1)\boldsymbol{w} - \sum_{k=L-S+1}^{L} \mathbf{Y}_k^{\mathrm{T}}\mathbf{R}_k^{-1}(\boldsymbol{y}_k - \overline{\boldsymbol{y}}_k)$
17:     $\widetilde{\mathcal{H}} = (N-1)\mathbf{I}_N + \sum_{k=L-S+1}^{L} \mathbf{Y}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{Y}_k$
18:     Solve $\widetilde{\mathcal{H}}\Delta\boldsymbol{w} = \nabla\widetilde{\mathcal{J}}$
19:     $\boldsymbol{w} := \boldsymbol{w} - \Delta\boldsymbol{w}$
20:     $j := j + 1$
21: **until** $\|\Delta\boldsymbol{w}\| \le e$ **or** $j \ge j_{\max}$
22: $\mathbf{E}_0 = \boldsymbol{x}_0 \mathbf{1}^{\mathrm{T}} + \sqrt{N-1}\mathbf{A}_0\widetilde{\mathcal{H}}^{-\frac{1}{2}}\mathbf{U}$
23: $\mathbf{E}_S = \mathcal{M}_{S\leftarrow 0}(\mathbf{E}_0)$
24: $\overline{\boldsymbol{x}}_s = \mathbf{E}_S\mathbf{1}/N$
25: $\mathbf{E}_S := \overline{\boldsymbol{x}}_S\mathbf{1}^{\mathrm{T}} + \lambda\left(\mathbf{E}_S - \overline{\boldsymbol{x}}_S\mathbf{1}^{\mathrm{T}}\right)$

Appendix B: Full equations and parameters of L95-GRS

Equations for the Lorenz variables x_m, $m \in [1, M]$:

$$\frac{\mathrm{d}x_m}{\mathrm{d}t} = (x_{m+1} - x_{m-2})x_{m-1} - x_m + F.$$

Equations of the chemical species concentrations:

$$\frac{\mathrm{d}[\mathrm{ROC}]_{m+\frac{1}{2}}}{\mathrm{d}t} = \psi_m^{\mathrm{ROC}} - \psi_{m+1}^{\mathrm{ROC}} - \lambda[\mathrm{ROC}]_{m+\frac{1}{2}} + E_{m+\frac{1}{2}}^{\mathrm{ROC}}, \quad \text{(B1)}$$

$$\frac{\mathrm{d}[\mathrm{RP}]_{m+\frac{1}{2}}}{\mathrm{d}t} = \psi_m^{\mathrm{RP}} - \psi_{m+1}^{\mathrm{RP}} - \lambda[\mathrm{RP}]_{m+\frac{1}{2}} + k_1[\mathrm{ROC}]_{m+\frac{1}{2}}$$
$$- [\mathrm{RP}]_{m+\frac{1}{2}}\left(k_2[\mathrm{NO}]_{m+\frac{1}{2}} + 2k_6[\mathrm{NO_2}]_{m+\frac{1}{2}} + k_5[\mathrm{RP}]_{m+\frac{1}{2}}\right), \quad \text{(B2)}$$

$$\frac{\mathrm{d}[\mathrm{NO}]_{m+\frac{1}{2}}}{\mathrm{d}t} = \psi_m^{\mathrm{NO}} - \psi_{m+1}^{\mathrm{NO}} - \lambda[\mathrm{NO}]_{m+\frac{1}{2}} + E_{m+\frac{1}{2}}^{\mathrm{NO}}$$
$$+ k_3[\mathrm{NO_2}]_{m+\frac{1}{2}} - [\mathrm{NO}]_{m+\frac{1}{2}}\left(k_2[\mathrm{RP}]_{m+\frac{1}{2}} + k_4[\mathrm{O_3}]_{m+\frac{1}{2}}\right), \quad \text{(B3)}$$

$$\frac{\mathrm{d}[\mathrm{NO_2}]_{m+\frac{1}{2}}}{\mathrm{d}t} = \psi_m^{\mathrm{NO_2}} - \psi_{m+1}^{\mathrm{NO_2}} - \lambda[\mathrm{NO_2}]_{m+\frac{1}{2}} + E_{m+\frac{1}{2}}^{\mathrm{NO_2}}$$
$$+ k_4[\mathrm{NO}]_{m+\frac{1}{2}}[\mathrm{O_3}]_{m+\frac{1}{2}}$$
$$+ k_2[\mathrm{NO}]_{m+\frac{1}{2}}[\mathrm{RP}]_{m+\frac{1}{2}} - [\mathrm{NO_2}]_{m+\frac{1}{2}}\left(k_3 + 2k_6[\mathrm{RP}]_{m+\frac{1}{2}}\right), \quad \text{(B4)}$$

$$\frac{\mathrm{d}[\mathrm{O_3}]_{m+\frac{1}{2}}}{\mathrm{d}t} = \psi_m^{\mathrm{O_3}} - \psi_{m+1}^{\mathrm{O_3}} - \lambda[\mathrm{O_3}]_{m+\frac{1}{2}} + k_3[\mathrm{NO_2}]_{m+\frac{1}{2}}$$
$$- k_4[\mathrm{NO}]_{m+\frac{1}{2}}[\mathrm{O_3}]_{m+\frac{1}{2}}, \quad \text{(B5)}$$

$$\frac{\mathrm{d}[\mathrm{S(N)GN}]_{m+\frac{1}{2}}}{\mathrm{d}t} = \psi_m^{\mathrm{S(N)GN}} - \psi_{m+1}^{\mathrm{S(N)GN}} - \lambda[\mathrm{S(N)GN}]_{m+\frac{1}{2}}$$
$$+ 2k_6[\mathrm{NO_2}]_{m+\frac{1}{2}}[\mathrm{RP}]_{m+\frac{1}{2}}, \quad \text{(B6)}$$

with the following definitions of the fluxes

$$\psi_m^{\mathrm{C}} = x_m[\mathrm{C}]_{m-\frac{1}{2}} \quad \text{if } x_m \ge 0$$
$$\psi_m^{\mathrm{C}} = x_m[\mathrm{C}]_{m+\frac{1}{2}} \quad \text{if } x_m < 0; \quad \text{(B7)}$$

the constants

$$M = 40, \qquad F = 8, \qquad \lambda = 0.02\,\mathrm{day}^{-1}; \quad \text{(B8)}$$

the emission rates (in ppbC day^{-1} for ROC or ppb day^{-1} for NO$_x$)

$$E_{m+\frac{1}{2}}^{\mathrm{ROC}} = 0.0235, \quad E_{m+\frac{1}{2}}^{\mathrm{NO}} = 0.27 \times 0.9 = 0.243,$$
$$E_{m+\frac{1}{2}}^{\mathrm{NO_2}} = 0.27 \times 0.1 = 0.027; \quad \text{(B9)}$$

and the kinetic rates computed at $T = 300\,\mathrm{K}$ (all in ppb^{-1} min^{-1}, except for k_1 in min^{-1})

$$k_1 = 0.00152 \times k_3, \quad k_2 = 12.3, \quad k_4 = 0.275,$$
$$k_5 = 10.2, \quad k_6 = 0.12. \quad \text{(B10)}$$

Table B1. Hourly values of k_3 in min^{-1}.

t	k_3	t	k_3	t	k_3	t	k_3
00 h	0	06 h	0.1972314	12 h	0.622824	18 h	0.00675528
01 h	0	07 h	0.3910734	13 h	0.611526	19 h	0
02 h	0	08 h	0.5074326	14 h	0.5755002	20 h	0
03 h	0	09 h	0.5755002	15 h	0.5074326	21 h	0
04 h	0	10 h	0.611526	16 h	0.3910734	22 h	0
05 h	0.00675528	11 h	0.622824	17 h	0.1972314	23 h	0

k_3 is the photolysis rate of NO_2 in min^{-1}, which is a function of solar radiation. It was computed using Fast-JX (Voulgarakis et al., 2009). We then took the value on 21 March at the Equator and used it repeatedly without attenuation. If k_3 is required between two hours, a linear interpolation is performed. Specifically, hourly values of k_3 used are reported in Table B1.

The quasi-steady-state approximation (QSSA) consists in replacing Eq. (B2) by diagnosing the concentration of RP at each grid point assuming steady-state for a given time step. This means that there is a dynamical equilibrium between the chemical production and decay of the RP, which implies

$$0 = \frac{d[RP]}{dt} = k_1[ROC] - [RP](k_2[NO] + 2k_6[NO_2] + k_5[RP])$$

$$\Leftrightarrow [RP] = \frac{k_2[NO] + 2k_6[NO_2]}{2k_5}$$

$$\left(\sqrt{1 + \frac{4k_1 k_5[ROC]}{(k_2[NO] + 2k_6[NO_2])^2}} - 1 \right). \tag{B11}$$

The model is integrated with an autonomous second-order Rosenbrock method with a time step of $\delta t = 1$ h. With this time step, an adaptive time step is unnecessary with the QSSA version of the model but is required in the non-QSSA case.

References

Azzi, M., Johnson, G. M., and Cope, M.: An introduction to the generic reaction set photochemical smog mechanism, Proceedings of the Eleventh International Conference of the Clean Air Society of Australia and New Zealand, 5–10 July 1992, Brisbane, Qld., Australia, 451–462, 1992.

Bocquet, M.: Reconstruction of an atmospheric tracer source using the principle of maximum entropy, I: Theory, Q. J. Roy. Meteor. Soc., 131, 2191–2208, doi:10.1256/qj.04.67, 2005.

Bocquet, M.: Ensemble Kalman filtering without the intrinsic need for inflation, Nonlin. Processes Geophys., 18, 735–750, doi:10.5194/npg-18-735-2011, 2011.

Bocquet, M.: Parameter field estimation for atmospheric dispersion: Application to the Chernobyl accident using 4D-Var, Q. J. Roy. Meteor. Soc., 138, 664–681, doi:10.1002/qj.961, 2012.

Bocquet, M. and Sakov, P.: Combining inflation-free and iterative ensemble Kalman filters for strongly nonlinear systems, Nonlin. Processes Geophys., 19, 383–399, doi:10.5194/npg-19-383-2012, 2012.

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

Bocquet, M. and Sakov, P.: An iterative ensemble Kalman smoother, Q. J. Roy. Meteor. Soc., 140, 1521–1535, 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, Atmos. Chem. Phys., 15, 5325–5358, doi:10.5194/acp-15-5325-2015, 2015a.

Bocquet, M., Raanes, P. N., and Hannart, A.: Expanding the validity of the ensemble Kalman filter without the intrinsic need for inflation, Nonlin. Processes Geophys., 22, 645–662, doi:10.5194/npg-22-645-2015, 2015b.

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 I: Description and single-observation experiments, Mon. Weather Rev., 138, 1550–1566, 2010.

Carmichael, G. R., Sandu, A., Chai, T., Daescu, D. N., Constantinescu, E. M., and Tang, Y.: Predicting air quality: Improvements through advanced methods to integrate models and measurements, J. Comp. Phys., 227, 3540–3571, 2008.

Carrassi, A., Vannitsem, S., and Nicolis, C.: Model error and sequential data assimilation: A deterministic formulation, Q. J. Roy. Meteor. Soc., 134, 1297–1313, 2008.

Clayton, A. M., Lorenc, A. C., and Barker, D. M.: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office, Q. J. Roy. Meteor. Soc., 139, 1445–1461, 2013.

Cohan, D. S., Hakami, A., Hu, Y., and Russell, A. G.: Nonlinear response of ozone to emissions: Source apportionment and sensitivity analysis, Environ. Sci. Technol., 39, 6739–6748, 2005.

Constantinescu, E. M., Sandu, A., Chai, T., and Carmichael, G. R.: Ensemble-based chemical data assimilation. I: General approach, Q. J. Roy. Meteor. Soc., 133, 1229–1244, 2007a.

Constantinescu, E. M., Sandu, A., Chai, T., and Carmichael, G. R.: Ensemble-based chemical data assimilation. II: Covariance localization, Q. J. Roy. Meteor. Soc., 133, 1245–1256, 2007b.

Dawson, J. P., Adams, P. J., and Pandis, S. N.: Sensitivity of $PM_{2.5}$ to climate in the Eastern US: a modeling case study, Atmos. Chem. Phys., 7, 4295–4309, doi:10.5194/acp-7-4295-2007, 2007.

Desroziers, G., Camino, J.-T., and Berre, L.: 4DEnVar: link with 4D state formulation of variational assimilation and different possible implementations, Q. J. Roy. Meteor. Soc., 140, 2097–2110, doi:10.1002/qj.2325, 2014.

Dimitriades, B.: Oxidant control strategies. Part I. Urban oxidant control strategy derived from existing smog chamber data, Environ. Sci. Technol., 11, 80–88, 1977.

Eben, K., Juruş, P., Resler, J., Belda, M., Pelikan, E., Krueger, B. C., and Keder, J.: An ensemble kalman filter for short-term forecasting of tropospheric ozone concentrations, Q. J. Roy. Meteor. Soc., 131, 3313–3322, 2005.

Elbern, H. and Schmidt, H.: A four-dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling, J. Geophys. Res., 104, 18583–18598, 1999.

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

Elbern, H., Schmidt, H., Talagrand, O., and Ebel, A.: 4D-variational data assimilation with an adjoint air quality model for emission analysis, Environ. Model. Softw., 15, 539–548, 2000.

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmos. Chem. Phys., 7, 3749–3769, doi:10.5194/acp-7-3749-2007, 2007.

Errera, Q., Daerden, F., Chabrillat, S., Lambert, J. C., Lahoz, W. A., Viscardy, S., Bonjean, S., and Fonteyn, D.: 4D-Var assimilation of MIPAS chemical observations: ozone and nitrogen dioxide analyses, Atmos. Chem. Phys., 8, 6169–6187, doi:10.5194/acp-8-6169-2008, 2008.

Finlayson-Pitts, B. J. and Pitts Jr., J. N.: Atmospheric chemistry: Fundamentals and experimental techniques, John Wiley and Sons, New York, NY, 1986.

Guerova, G., Bey, I., Attié, J.-L., Martin, R. V., Cui, J., and Sprenger, M.: Impact of transatlantic transport episodes on summertime ozone in Europe, Atmos. Chem. Phys., 6, 2057–2072, doi:10.5194/acp-6-2057-2006, 2006.

Hakami, A., Henze, D. K., Seinfeld, J. H., Singh, K., Sandu, A., Kim, S., Byun, D., and Li, Q.: The adjoint of CMAQ, Environ. Sci. Technol., 41, 7807–7817, 2007.

Hanea, R. G., Velders, G. J. M., Segers, A. J., Verlaan, M., and Heemink, A. W.: A Hybrid Kalman Filter Algorithm for Large-Scale Atmospheric Chemistry Data Assimilation, Mon. Weather Rev., 135, 140–151, 2007.

Henze, D. K., Hakami, A., and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem, Atmos. Chem. Phys., 7, 2413–2433, doi:10.5194/acp-7-2413-2007, 2007.

Hundsdorfer, W. and Verwer, J. G.: Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations, Springer-Verlag Berlin Heidelberg New York, 2003.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D, 230, 112–126, 2007.

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, 19047, doi:10.3402/tellusb.v64i0.19047, 2012.

Lin, M., Fiore, A. M., Horowitz, L. W., Cooper, O. R., Naik, V., Holloway, J., Johnson, B. J., Middlebrook, A. M., Oltmans, S. J., Pollack, I. B., et al.: Transport of Asian ozone pollution into surface air over the western United States in spring, J. Geophys. Res., 117, doi:10.1029/2011JD016961, 2012.

Lorenz, E. N.: Designing Chaotic Models, J. Atmos. Sci., 62, 1574–1587, 2005.

Lorenz, E. N. and Emanuel, K. A.: Optimal sites for supplementary weather observations: Simulation with a small model, J. Atmos. Sci., 55, 399–414, 1998.

Mallet, V., Quélo, D., Sportisse, B., Ahmed de Biasi, M., Debry, É., Korsakissok, I., Wu, L., Roustan, Y., Sartelet, K., Tombette, M., and Foudhil, H.: Technical Note: The air quality modeling system Polyphemus, Atmos. Chem. Phys., 7, 5479–5487, doi:10.5194/acp-7-5479-2007, 2007.

Ménard, R., Cohn, S. E., Chang, L.-P., and Lyster, P. M.: Assimilation of stratospheric chemical tracer observations using a Kalman filter. Part I: Formulation, Mon. Weather Rev., 128, 2654–2671, 2000.

Milewski, T. and Bourqui, M. S.: Assimilation of Stratospheric Temperature and Ozone with an Ensemble Kalman Filter in a Chemistry–Climate Model, Mon. Weather Rev., 139, 3389–3404, 2011.

Nino Ruiz, E. D. and Sandu, A.: A derivative-free trust region framework for variational data assimilation, J. Comput. Appl. Math., 293, 164–179, 2016.

Quélo, D., Mallet, V., and Sportisse, B.: Inverse modeling of NO_x emissions at regional scale over northern France: Preliminary investigation of the second-order sensitivity, J. Geophys. Res., 110, D24310, doi:10.1029/2005JD006151, 2005.

Roustan, Y., Sartelet, K. N., Tombette, M., Debry, E., and Sportisse, B.: Simulation of aerosols and gas-phase species over Europe with the POLYPHEMUS system. Part II: Model sensitivity analysis for 2001, Atmos. Environ., 44, 4219–4229, 2010.

Sakov, P. and Bertino, L.: Relation between two common localisation methods for the EnKF, Comput. Geosci., 15, 225–237, 2011.

Sakov, P. and Oke, P. R.: A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters, Tellus A, 60, 361–371, 2008.

Sakov, P., Oliver, D. S., and Bertino, L.: An iterative EnKF for strongly nonlinear systems, Mon. Weather Rev., 140, 1988–2004, 2012.

Sandu, A. and Chai, T.: Chemical data assimilation – An overview, Atmosphere, 2, 426–463, 2011.

Sartelet, K. N., Couvidat, F., Seigneur, C., and Roustan, Y.: Impact of biogenic emissions on air quality over Europe and North America, Atmos. Environ., 53, 131–141, 2012.

Schutgens, N. A. J., Miyoshi, T., Takemura, T., and Nakajima, T.: Sensitivity tests for an ensemble Kalman filter for aerosol assimilation, Atmos. Chem. Phys., 10, 6583–6600, doi:10.5194/acp-10-6583-2010, 2010.

Segers, A. J., Heemink, A. W., Verlaan, M., and van Loon, M.: A modified RRSQRT-filter for assimilating data in atmospheric chemistry models, Environ. Model. Softw., 15, 663–671, 2000.

Sekiyama, T. T., Deushi, M., and Miyoshi, T.: Operation-oriented ensemble data assimilation of total column ozone, SOLA, 7, 41–44, 2011.

Semane, N., Peuch, V.-H., Pradier, S., Desroziers, G., El Amraoui, L., Brousseau, P., Massart, S., Chapnik, B., and Peuch, A.: On the extraction of wind information from the assimilation of ozone profiles in Météo-France 4-D-Var operational NWP suite, Atmos. Chem. Phys., 9, 4855–4867, doi:10.5194/acp-9-4855-2009, 2009.

Singh, K. and Sandu, A.: Variational chemical data assimilation with approximate adjoints, Comput. Geosci., 40, 10–18, 2012.

Singh, K., Jardak, M., Sandu, A., Bowman, K., Lee, M., and Jones, D.: Construction of non-diagonal background error covariance matrices for global chemical data assimilation, Geosci. Model Dev., 4, 299–316, doi:10.5194/gmd-4-299-2011, 2011.

Vannitsem, S. and Nicolis, C.: Predictability experiments on a simplified thermal convection model: The role of spatial scales, J. Geophys. Res., 99, 10377–10385, 1994.

Venkatram, A., Karamchandani, P., Pai, P., and Goldstein, R.: The development and application of a simplified ozone modeling system (SOMS), Atmos. Environ., 28, 3665–3678, 1994.

Voulgarakis, A., Savage, N. H., Wild, O., Carver, G. D., Clemitshaw, K. C., and Pyle, J. A.: Upgrading photolysis in the p-TOMCAT CTM: model evaluation and assessment of the role of clouds, Geosci. Model Dev., 2, 59–72, doi:10.5194/gmd-2-59-2009, 2009.

Wu, L., Mallet, V., Bocquet, M., and Sportisse, B.: A Comparison Study of Data Assimilation Algorithms for Ozone Forecasts, J. Geophys. Res., 113, D20310, doi:10.1029/2008JD009991, 2008.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., and Baklanov, A.: Real-Time Air Quality Forecasting, Part I: History, Techniques, and Current Status, Atmos. Environ., 60, 632–655, doi:10.1016/j.atmosenv.2012.06.031, 2012.