Accounting for model error in air-quality forecasts: an application of 4DEnVar to the assimilation of atmospheric composition using QG-Chem 1.0 . Reply to referee # 1

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

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

Correspondence to: Emili (emili@cerfacs.fr)

1   Reply to general comments

We thank the reviewer for his comments, which helped to improve the manuscript. The detailed replies follow:

1. *Some of the assumptions made for the model are rather imbalanced leading to very unrealistic conditions. In the one hand, you choose a detailed gaseous chemistry, which is fine. But on the other hand you neglect deposition which is very unrealistic to obtain an interesting dynamical equilibrium of the chemical species (typical of air pollution in the boundary layer), but which is not difficult to model.*

   Answer:

   We originally built QG-Chem with the main scope of testing data assimilation algorithms on a system that reproduces the relatively fast (hourly) dynamics of reactive gases in the boundary layer. The longest DA experiment in the manuscript was of 3 days (length of the analysis plus forecast period in Section 4.2.3) with the model set-up reproducing the "Flux" experiment described in Crassier et al. (2000) on the central part of the QG-Chem domain. The experiment of Crassier et al. (2000) was meant to simulate the build-up of a typical summer pollution episode in Paris. Since time scales of dry deposition are typically longer [1], the authors also neglected this process in their study. Indeed, the experiments n. 1 and n. 3 (Section 4.2) of our manuscript were designed to examine the transient phase from a relatively unpolluted atmosphere to a locally polluted one. We do not agree completely with the reviewer about considering these conditions very unrealistic, because at hourly time scales the chemical system is most of the time out of the equilibrium due to the rapid variability of radiation, emissions and boundary layer height.

   However, we agree with the reviewer about the fact that over longer integration periods (several weeks), deposition terms become important in the chemical balances of trace gases, with time/space-averaged concentrations satisfying a chemical equilibrium between sources and sinks. Neglecting deposition in our original manuscript was mainly affecting the initial concentrations for the polluted experiment (n. 2), which started after a spin-up period of 20 days (instead of 1 day for

---

[1]If we consider for example a typical dry deposition velocity of $3.5 \cdot 10^{-3}$ m s^{-1} for ozone over agricultural surfaces (Hardacre et al., 2015) and a boundary layer height (BLH) of 1.2 km, we obtain an e-folding time scale of about 4 days.

the experiments n. 1 and 3). It also limits the usability of QG-Chem for DA experiments that last longer than the ones presented in the manuscript. Therefore we decided to include dry deposition in the QG-Chem model and repeat all the experiments presented in the original manuscript. This lead, as expected, to different average concentrations for some species in case n. 2 and slightly different daily cycles in some cases, but all DA main results and conclusions remained the same. Details about the dry deposition mechanism are given now in the manuscript and some additional discussion of the differences between previous and new results can be found in detailed answers n. 20-21-24-25. We upgraded the manuscript with the new results.

2. *At some point, advection is neglected. This assumption is really too strong since it leads to a collection of 0/1-D box/column-models, whereas the purpose of this paper is more on the 4D-EnVar aspects where advection is critical.*

Answer:

The purpose of this study was twofold: i) presenting a new toy model that can be useful to test DA algorithms for atmospheric chemistry and ii) exploring the advantages of 4DEnVar method and a possible treatment of model errors. We agree on the fact that advection is a critical aspect to be examined in a DA testing framework. For this reason we considered a passive tracer in all experiments (CO_2) and we gave final conclusions based on several cycles of DA with both model error and advection, with changing wind field (Section 4.3, Table 6). However, aspects related to 4DEnVar and advection were already discussed in previous studies using meteorological models (Lorenc et al., 2015; Desroziers et al., 2014; Fairbairn et al., 2013). On the other hand, the performances of ensemble/variational DA on the very different dynamics of reactive gases are less documented in the literature. For example, DA issues related to the non-linearities of the chemical system were only recently pointed out by Tang et al. (2016). Some authors focused exclusively on 0-D box models to gain deeper understanding of the behavior of Ensemble Kalman filters (Tang et al., 2016) or chemical adjoint models (Hamer et al., 2015). It can be tricky to isolate possible issues within DA when the full system is simulated and chemical fields result from a complex interaction of all the underlying processes (chemistry, physics and transport). For this reason, we believe that a simplified set-up where only chemical processes are activated, can be a useful intermediate step when testing a novel technique, as for example the bias estimation procedure of this study. Finally, conditions of weak winds and local accumulation of pollutants can also occur in real cases, for which a 0-D model can provide useful insights. Therefore, we prefer to keep this experiment in the manuscript, which also shows some interesting effects of local emissions error.

In the revised manuscript, we better clarify the intent of the experiments without advection. The following sentences has been added in the introductory part of Section 4.2:

"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 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"

3. *The claim of not localizing in time is actually partially wrong; that is only an approximate statement. The dynamics of a consistent space-localization operator within the time window of 4D-EnVar was explored in Bocquet (2016), from which it can be understood that in the absence of time-localization the localization operator satisfies a Liouville equation.*

   Answer:

   We were aware of some recent studies that explored alternative approaches to apply a coherent localization operator in a 4D framework, like Bocquet (2016) or Desroziers et al. (2016). However, in this study we employed the approximation of a static covariance localization (Desroziers et al., 2014), which represented the simplest choice for the initial implementation of our toy model, it was already available in the DA library and provided satisfactory results with the DA windows used in our experiments (24 hours maximum). A deeper analysis of the impact of a static localization in our system and inclusion of more recent methodologies from latest studies will be considered in future work.

   We slightly updated the text in Section 3.2 to avoid possible misunderstandings:

   "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 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"

4. *You are occulting the fact that the IEnKS has a fully nonlinear variational analysis which you don't have (this remark also applies to the current implementations of 4D-EnVar in meteorology). This is important because this could prove an asset when dealing with highly non-linear air quality models.*

   Answer:

   We agree with the reviewer about the importance of addressing nonlinear behavior of chemistry models within DA and we stated it several times in the manuscript (Pages 33 lines 1-3, Page 38 lines 25-26). We did not want to deliberately occult any information about the properties of IEnKS, which has been already proven to be effective in such conditions. Similarly to what was done for localization, we used here a basic set-up (one outer loop) to examine other aspects of chemical DA and left a thoughtful explorations of strong departures from the linearity for a future work.

   The capacity of IEnKS to provide a fully non-linear analysis has been highlighted in the revised manuscript. The following lines are added in the introduction, when describing the IEnKS:

   "A major asset of IEnKS for chemistry applications is that it can also account for strong non-linearities of the forecast model."

   and at the end of Sec. 4.2.2:

"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)."

5. *Actually 4D-EnVar should be compared to the EnKF. This is pretty obvious for assimilators using both ensemble filter and variational methods. Your standpoint looks like a biased one from specialists only used to variational methods. I do not encourage you to redo everything replacing 3D-Var with the EnKF. But several of your statements should be mitigated.*

Answer:

We acknowledge that a comparison with EnKF would have been probably more fair and provided more robustness to the conclusions of the study. However, previous studies already compared 4DVar and EnKF for air-quality applications and found similar results of the two schemes (Wu et al., 2008). Several state-of-the-art European air-quality operational models are based either on 3D-Var and EnKF and extensive comparisons of yearly chemical reanalyses showed very similar performances for both algorithms in practice (Rouil and the MACC team, 2014). For these reasons we considered that 3DVar represents a good-enough reference for evaluating the 4DEnVar scheme proposed in this study. Finally, implementing EnKF would have required a considerable amount of additional work and will be considered for a future study.

We included the following discussion at the end of Sec. 4.3:

"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."

6. *From my understanding, the bias removal technique that you propose is actually a parametrized one since you pick up the perturbations that you apply to the ensemble. This is just a stochastic variant of the parameter estimation technique. If I am correct, you should substantially mitigate your claim on this point.*

Answer:

The proposed bias estimation is based on the implicit treatment of model errors done in the 4DEnVar state estimation. The analogy between 4DEnVar and the weak constraint 4D-Var (Desroziers et al., 2014) suggests that there is no particular limitation to the representation of model error that can be used. In the manuscript we analyzed the case of model errors that result from chemical emissions, which is a typical source of errors in chemistry models. Since we can identify emissions with a model parameter, the bias estimation has a strong analogy with the parameter estimation typically done within the EnKF framework. The main difference is that the bias estimation is done here only in the model state space,

4

without any update of model emissions (parameters). Therefore we agree with the reviewer on the fact that in some sense we "pick up the perturbations that we apply to the ensemble".

Nevertheless, the proposed method, not being linked to a particular parameter selection, should allow to correct biases as well as in more general cases where the ensemble is not generated by perturbing a particular set of parameters. We could think for example of stochastic ensemble perturbations or even multi-model ensembles (Marécal et al., 2015), where each model has its own chemical mechanism. Therefore, even if in certain cases the analogy might be strong, we prefer not to associate this method to a "parameter estimation technique", since fundamentally it does not address parameters estimation. Still, to obtain a reliable bias estimation, the model error dynamics has to be correctly represented by the ensemble, whatever the source of uncertainty (parametric or not). This is a crucial step that deserves further investigation using real air-quality models and observations, as already stated at page 38 lines 23-24.

7. *To shorten the paper, I would suggestion to get rid of the first configuration. It is unrealistic and I believe a bit detrimental to the paper anyway.*

   Answer:

   Since the conclusions of the DA experiments did not show significant differences among the configurations n. 1 and n. 2, we followed the suggestion of the reviewer and removed the first one from the manuscript (Sections 4.1 and 4.3) . However, we kept the model error experiment in section 4.2, which represents a pollution build-up case corresponding to the former configuration n. 1 (see also reply n. 2). The following text has been added in the introductory part of Section 4.2 to clarify this:

   "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 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."

8. *Please number all your equations. Avoiding excessive numbering is usually reserve for books, not for articles especially meant for peer-reviews.*

   Answer:

   All equations have been numbered in the revised manuscript.

Additional remarks on the revised manuscript

Doing again all the experiments after the inclusion of chemical deposition in QG-Chem let us discover two minor issues affecting the results presented in the original manuscript:

1. The correlation length of the initial perturbation used for the CO_2 experiments with a perfect model (Sec. 4.1) was erroneously set to 1000 km instead of 750 km as for the other species. This explained the different sensitivity of the CO_2

4DEnVar analysis to the size of the ensemble and the localization scale (Fig. 8 and 9 of the original manuscript), which was a bit surprising. New results (Fig. 5 and 6 of the revised manuscript) show now similar sensitivities for CO_2 as for other species. Therefore, the discussion dedicated to the CO_2 case in Sec. 4.1.3 (Page 26, lines 14-15 and 27) has been removed.

2. The ensemble of forecasts used to compute the 4DEnVar analyses with both the initial perturbation and model error enabled (Sec. 4.2, Fig. 11 bottom plot, Fig. 12 top right plot, Fig. 13, 14 and 15) were erroneously done with the advection enabled. This introduced an incoherence between the ensemble and the other simulations (truth, forecast, observations), which produced some biases in the model error estimation as well (e.g., the negative bias in the estimated model error in Fig. 11 and 12). The new results (Fig. 8, 9, 10, 11, 12 in the revised manuscript) are not anymore affected by this problem. The discussion has been revised where needed (e.g. Page 30, lines 19-24) and the color scale in Fig. 12 and 14 (Fig. 9 and 11 in the revised manuscript) has been modified to better display the positive model error values found for all experiments.

2 Reply to specific comments

1. *Page 1, line 2-3, "using a reduced-order chemical transport model based on quasi-geostrophic dynamics": this statement is ambiguous since we do not understand whether it is a CCMM or a CTM. Please clarify.*

   Answer:

   Even tough QG-Chem is a one-way coupled system, since the meteorology is not affected by the chemistry, we changed the sentence to "... using a reduced-order coupled chemistry-meteorology model based on quasi-geostrophic dynamics...". This should clarify the ambiguity. However, as already explained in the text our manuscript does not deal with aspects concerning the one-way coupling between meteorology and chemistry. A CTM-like configuration is actually adopted for the reported experiments. Aspects concerning coupled DA could be investigated in the future with QG-Chem.

2. *Page 1, line 4: "to a generic software library for data assimilation": which one? This meant to become a GMD paper. That type of info should be mentioned even in the abstract.*

   Answer:

   We changed the text with "to the software library for data assimilation OOPS (Object Oriented Prediction System)" . A reference to OOPS is now given in the manuscript.

3. *Page 1, line 12: "analysys" -> "analysis".*

   Answer: Done

4. *Page 1, line 12-13: "A comparison with results of 3D-Var, widely used in operational centers, shows that, for some species, analysys and next day forecast errors can be halved when model error is taken in account": 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 4D-EnVar method) on the low-order online tracer model L95-T was improved by 25% to 50%.

Answer:

The suggested sentence and the corresponding reference is now added when discussing the reanalysis results of the repeated assimilation cycles (Sec. 4.3)

5. *Pages 3, line 6: "next day..." -> "Next day...".*

Answer: Done

6. *Page 3, line 13-14: Haussaire and Bocquet (2016) also did a similar experiment simultaneously estimating gaseous concentrations and fluxes using a nonlinear 4D-EnVar.*

Answer:

We were aware of the above-mentioned study. However, in this paragraph, we cited only studies that used more standard algorithms (variational or EnKF) since the introduction of EnVar methods follows later. The reference to Haussaire and Bocquet (2016) concerning the application to the emissions estimation in a chemistry model has been reported again later in the introduction (see also reply to question n. 11.)

7. *Page 3, line 15: "Results seems promising but still relies on the assumption that the model is perfect, i.e. that there are no additional sources of uncertainties in the model forecast other than the controlled variables (i.e. the initial state and the selected emissions)": This is a rather biased statement. The model per see is not assumed perfect anymore, but only part of it, such as the dynamical part. This should be explained in a less biased way. If you estimate several parameters of the parametrization of a CTM, you are morally assuming that the model is imperfect. It is just that model error is parametrized.*

Answer:

We were referring in particular only to the two previously cited studies (Elbern et al., 2007; Hamer et al., 2015), which make use of 4D-Var in the strong constraint formulation. We agree with the reviewer about the fact that addressing parameter estimation, even with 4D-Var, corresponds morally to address in part the model error issue. However, this approach becomes rapidly unpractical when the possible sources of model error can invest the several thousands parameters of complex CTM (e.g. time and geographically varying emissions, photo-chemistry coefficients, deposition velocities, wind forcing fields etc.). In this sense we can probably state that the strong constraint 4D-Var can address air-quality DA within an "almost"-perfect model hypothesis. We revised the text in the manuscript in the following way:

Results seems promising but still relies on the assumption that the model is *almost* perfect, i.e. that there are no additional sources of uncertainties in the model forecast other than the controlled variables (i.e. 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.

8. *Page 4, line 4-5: "The IEnKS (Bocquet and Sakov, 2014) is a fixed-lag ensemble Kalman smoother formulated under perfect model assumptions, which can also be used to estimate erroneous model parameters through an augmented state formalism (Haussaire and Bocquet, 2016)." No! the IEnKS is an iterative ensemble Kalman smoother. This is quite different from the standard "fixed-lag ensemble Kalman smoother"! It is better described as a nonlinear 4D-EnVar method. Please correct. For instance: "The iterative ensemble 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 (Haussaire and Bocquet, 2016)." Moreover, the augmented state formalism applied to the IEnKS was first demonstrated in Bocquet and Sakov (2013).*

Answer: The text has been revised according to the reviewer suggestion.

9. *Page 4, line 8-10: "These type of 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)." - Instead of "These types of approaches..." it would be much better to write "These approaches", to avoid any ambiguity in the rightful claim that the IEnKS, the 4D-Var-EnKS and the 4D-EnVar are ensemble variational techniques. - What you call the "hybrid" methods are in fact called EDA standing for ensemble of data assimilation methods.*

Answer: Done

10. *Page 4, line 18: "the need in IEnKS to select a number of model parameters among all the possible erroneous parameters in complex CTMs": yes, but the EnKS and the IEnKS can account for stochastic perturbations in the integration step of the ensemble as already demonstrated in Haussaire and Bocquet (2016), section 3.2, configuration Offline 2. Please amend your statement.*

Answer: To avoid possible misunderstandings we removed this sentence from the manuscript.

11. *Page 4, line 23-25: "To the knowledge of the authors, EnVar type methods have not yet been implemented in air-quality or atmospheric chemistry models and no previous study has already examined the potential of 4DEnVar for chemical DA": Without any ambiguity, they have been tested in Haussaire and Bocquet (2016). I do not see any problem in recognizing that fact. Please correct.*

Answer:

We were originally referring to the 4DEnVar algorithm described in Desroziers et al. (2014). Given the conceptual similarity with IEnKS we modified this sentence as follows:

"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 chemical DA (Haussaire and Bocquet, 2016)."

12. *Page 4, line 26: "Then" -> "Than".*

   Answer: Done

13. *Page 5, line 6: "This allow" -> "This allows".*

   Answer: Done

14. *Page 5, line 14: "under a generic library for data assimilation": Again, which one? Please be specific.*

   Answer: The reference to OOPS has been added.

15. *Page 6, line 18-19: "For all the experiments presented in this study a coarse resolution of 16x8 grid points has been used.": That is quite a low resolution and in contradiction with one of the early promises: "This choice permits to examine the behavior of DA in presence of complex gradients of wind fields and vorticity". Please revise or tune down your promises.*

16. *Page 6, line 21-22: "The only desired property is to obtain wind fields that exhibit typical patterns of the complex atmospheric circulation." which you don?t have with such a resolution. I am not criticizing your choice but the claims that are not matched by what you present. Please rephrase.*

   Answer to 15 and 16: We do not fully agree with the reviewer on this particular point. Within the QG model, typical patterns of Rossby waves can develop also with low horizontal resolutions, such as the ones that has been used in the study. The fields in Figure 1 of this document have been obtained changing only the spatial resolution of QG-Chem from the original 750 km (16 by 8 grid points) to 150 km (80 by 40 grid points). We can observe, that both meteorological and chemical fields are not significantly more complex than the ones that were used for the study (Figure 3 in the older manuscript, Figure 2 in the newer). We consider that the main requirement of reproducing vorticity fields that look like those typical of large scale circulation are satisfied also with the lower resolution. Higher resolutions would be beneficial to better examine aspects of DA such as the impact of the observational network, the spatial localization operator or interactions between small and large scale in chemistry. However, this was not the main objective of the current study.

17. *Page 7, line 5: "(Cariolle D, personal communication)": Daniel Cariolle is the third author. Remove this or give more details.*

   Answer:

   The reference has been removed. Some details of the method are given afterwards in the manuscript.

18. *Page 7, line 6: "which has a special treatment of the Jacobian matrix". This is too vague. Please be more specific.*

Answer:

A manuscript on the ASIS solver is currently in preparation. We recall here only the main characteristics of the ASIS method that are given later in the paragraph. We removed this too vague sentence.

19. *Page 7, line 13: "The meridional boundary conditions for chemical species are set to climatological values.": Ok, but what type of numerical boundary conditions? That is important.*

    Answer:

    The text has been revised as follows:

    The chemical field is set equal to climatological values on the N-S boundaries, which corresponds to Dirichlet type conditions.

20. *Page 7, line 14-15: "Moreover, no physical removal process for the chemical species has been included in the model so far.": This is both problematic (because this a key process of air pollution modeling) and odd (because it is not so really difficult to implement). Actually accounting for removal processes here is as important as having a fine, realistic meteorology. This is quite a weakness of your paper. The budget of all species is strongly affected. This also leads to an unbalanced photochemistry (induced by a wrong ratio of precursors).*

    Answer:

    With the inclusion of dry deposition mechanism in the revised manuscript, these lines are removed.

21. *Page 7, line 21: How does this configuration relate to regional air pollution modeling? Please elaborate.*

    Answer:

    Lines 25-27, page 7 have been replaced by:

    This configuration relates to regional air pollution modeling mainly concerning the type and amplitude of chemical emissions, spatial hetereogenity of sources and presence of boundaries.

22. *Page 8, Fig. 1: Did you show the grid? If not, could you please do so. That would help.*

    Answer:

    The model grid is now displayed in Figure 1 of the new manuscript.

23. *Page 10, table 3: Please give the extended name of each species in a another column.*

    Answer: Done in the revised manuscript

24. *Page 12, Figure 2: Haussaire and Bocquet (2016) have more realistic values of ozone for regional air pollution with a simplified CCMM than with your model! Please discuss this. Besides, the absence of a clear daily cycle for ozone is worrying and cast doubts.*

25. *Page 13, Figure 3: The magnitude and variation of ozone concentration is not realistic. I would have thought it should for such a toy-model.*

Answer to 24 and 25:

After the implementation of dry deposition in QG-Chem, the magnitude of the average ozone concentration in former Fig. 3 (Fig. 2 in the revised manuscript) is now closer to typical values encountered during summer periods in Europe. As discussed previously (see reply n. 1), the hourly variability was not much affected by the introduction of dry deposition. Figure 2 in this document shows that the typical variability of surface ozone in real models and observations is not too far from the one simulated with QG-Chem.

26. *Page 13, "by the variance:": You mean by the covariance matrix?*

Answer:

The text has been modified as follows: 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:

27. *Page 14, lines 3-9: What is the point in using the diffusion equation trick to obtain $\mathbf{C_x}$ for such toy-model and such 1D-correlation function? And why not for $\mathbf{C_y}$?*

28. *Page 14, lines 3-9: Why did you use a 2D correlation function C x,y ? There, using the diffusion equation would have been more meaningful(?).*

Answer to 27 and 28:

We built QG-Chem on top of the original operators that can be found in the OOPS distribution for the QG model. We have verified the code and found that the horizontal correlation operator $\mathbf{C_x}$ was actually not using a diffusion model. It is based on a Fast Fourier Transform approach to apply a filter in the spectral space. Since the domain is not periodic in N-S direction, $\mathbf{C}_y$ uses instead a symmetric positive-definite matrix filled with Gaussian correlations. We corrected the text as follows:

$\mathbf{C}_x$ and $\mathbf{C}_y$ are isotropic homogeneous correlation operators providing Gaussian spatial structures.

29. *Page 14, line 25: "The 4DEnVar algorithm is meant to solve the main drawbacks of 3D-Var": Of course that is not its primary purpose. Please replace "is meant to" with "can".*

Answer: Done

30. *Page 14, line 28: "written as :" -> "written as:"*

Answer: Done

31. *Page 14, 31-32: "The cost function is computed for an assimilation window that can span several hours or days": How long is the data assimilation window in your study? That is a key value that must be mentioned and discussed, including in the paper at this point.*

Answer:

The main parameters of the 4DEnVar specific to our study are discussed later in the manuscript (Sec. 4). In particular the choice of the assimilation windows and sub-windows was discussed at page 18, lines 7-11. Since we wanted to provide here only a general description of the DA method, we removed also the sentence concerning the length of the sub-windows.

32. *Page 15, line 10: "an hybrid 3D-Var": What is an hybrid 3D-Var for you? What you have described does not look like what is usually understood as an hybrid algorithm.*

    Answer:

    We consider a 3D-Var to be hybrid when the background error covariance is a weighted sum of a static and an ensemble derived term (Wang et al., 2013). At line 10 we described a particular case that corresponds to a zero weight on the static term. We think that this could still be a particular case of an hybrid 3D-Var. However, to avoid possible misunderstandings, we decided to remove the sentence.

33. *Page 15, line 29: "...no time localization is applied and the same 3D (and multivariate) correlation operator C is used for all 4DEnVar sub-windows": As I explained, this is not really a "no localization" condition. Please mention this.*

    Answer:

    The requested clarification has been made (see reply n. 3) and "no time localization is applied" has been removed.

34. *Page 15, line 32: "Hence, for the experiments presented in this study, we could use the covariance operator described in (6) by setting the variance terms to one.": The sentence is confusing. I would write "Hence, in order to specify C, we could use the covariance operator described in (6) by setting the variance terms to one."*

    Answer: The sentence has been modified according to the suggestion.

35. *Page 16, line 1: "is an ongoing research topic (Bocquet et al., 2015)" should be "is an ongoing research topic (Bocquet, 2016)"*

    Answer: Done. A second recent reference has also been included (Desroziers et al., 2016).

36. *Page 16, lines 19-23: There is an approximation here. These are not strict inequalities. You should use at least one $\simeq$ symbol.*

37. *Page 16, line 26: Again, you should use the $\simeq$ symbol here.*

    Answer: The approximate equality symbol is used now whenever linearization is used.

38. *Page 17, before section 4: How are the perturbations generated? This is a critical part of the EnVar schemes, rigorously derived in the IEnKS (Bocquet and Sakov, 2014), and approximately so in other EnVar systems (so far).*

    Answer:

The choice of the perturbations used in the numerical experiments is already detailed in the results section (Sec. 4.1 and 4.2).

39. *Page 18, line 1: "and by adding a normally distributed error": What did you do with the negative measurements?*

    Answer:

    Negative observations did not occur in the numerical experiments, since the used standard deviation was significantly smaller than the truth concentrations. Other error distributions (e.g. log-normal) could probably be adopted when exploring cases with larger departures from the truth.

40. *Page 18, line 5-6: "The meteorology is never observed neither perturbed.": so this is a CTM-like experiment. This should be emphasized since this is critical.*

    Answer:

    We clarify this also in the introductory paragraph of section 2 by adding the following sentence:

    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.

    and extending the previous one to:

    The meteorology is never observed nor perturbed, which corresponds to use QG-Chem in a CTM-like mode.

41. *Page 18, line 22: "gaussian" -> "Gaussian".*

    Answer: Done

42. *Page 18, line 22: 1 in the Gaussian distribution should actually be the identity matrix.*

    Answer: The formula has been updated.

43. *Page 19, line 1: "ensemble localization" -> "localization of covariances".*

    Answer: Done

44. *Page 19, Table 4: That would be good to have the relative value of the stds, i.e. divided by a standard concentration value.*

    Answer:

    We added the relative value as percentage of the average chemical concentrations given in Figure 3 (Figure 2 in the revised manuscript).

45. *Page 19, line 5: "All vertical terms of the covariance or localization matrix are always set to zero in this study, since only the bottom layer of the QG-Chem model is considered": I do not understand the justification. Could you please clarify?*

Answer:

Even if QG-Chem has two vertical layers, in this study we limited our attention to the bottom layer, to focus on boundary layer chemistry and 2D transport. Moreover, there is no chemical coupling between the two layers and no vertical transport of chemical species. Assimilated observations are also located in the bottom layer and the observation operator does not interpolate between levels. Therefore, setting the vertical correlation to zero in this study has only two effects: i) it provides vertically uncorrelated errors on the initial condition and ii) it prevents spurious 4DEnVar increments to spread in the top layer. Anyhow, this does not have any influence on the presented results, which are only displayed only for the bottom layer.

The sentence has been modified as follows for sake of clarity:

Since we use QG-Chem 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.

46. *Page 19, line 11: "Fig. 4 " -> "Figure 4".*

    Answer: Done

47. *Page 19, line 20-21: "The memory of the initial condition is rapidly lost for O3, as it was also demonstrated within regional air-quality models (Jaumouillé et al., 2012).": I doubt Jaumouillé et al. (2012) were the first to show/discuss this. Please give an earlier reference in addition to yours unless I am mistaken.*

    Answer:

    The following reference is also given: Wu et al. (2008)

48. *Page 20, line 4: "...which stays...": what is "which" referring to? Please clarify.*

    Answer:

    Since the comparison of chemical trajectories between experiments 1 and 2 is not presented anymore in the revised manuscript, lines 2-4, Page 20 have been removed.

49. *Page 21, line 1: "Fig. 5 a,d Fig. 7" -> "Figures 4 and 7".*

    Answer: Done

50. *Page 21, line 5: This definition is a bit confusion since Eq.(17) has a normalization and Eq.(18) has not. Please clarify or use non-confusing notations.*

    Answer:

    To improve the readability of the results we provided values of the RMSE gain of DA always as percentage of the average chemical concentration. Equation (18) is only used to define the RMSE (in absolute value) that is used to derive the RMSE gain (Eq. 17). The clarification has been included in the revised manuscript.

51. *Page 24, line 10: "satisfactory" is inappropriate. Be more specific. Per se the RMSE do not mean much because of what you wrote. Only the comparison of the RMSEs between the 3D-Var and 4D-EnVar is relevant. This comparison yields satisfactory results.*

Answer:

We clarified the sentence as follows:

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.

52. *Page 25, line 25: "If this is not the case, a larger ensemble size allows in principle less severe localization (Ménétrier et al., 2015).": The reference to Ménétrier et al., 2015 is inappropriate here (fully justified later on). This is very well known in ensemble data assimilation (especially for the EnKF) for 15 years. Cite an EnKF paper instead, or nothing since this is common knowledge.*

Answer:

The citation has been removed.

53. *Page 25, line 33: "The ensemble size is the main limiting factor in operational forecast centers...": This is not always true. For instance in Meteorology Environment and Climate Change Canada is using large ensembles, an option that they prefer (to better deal with inflation and localization).*

Answer:

The sentence has been changed to:

The ensemble size can be one of the main limiting factors in operational forecast centers...

54. *Page 26, line 16: "In few cases..." -> "In a few cases...".*

Answer: Done

55. *Page 26, line 21-22: "We remind that the objective of this study is to demonstrate the applicability of a DA algorithm that outperforms currently implemented methods in operational centers..." A biased statement since other ensemble-based methods are ignored.*

Answer:

The sentence has been modified as follows:

We remind that the objective of this study is to demonstrate the applicability of a DA algorithm that could outperform currently implemented methods in operational centers...

The same has been done at Page 5, line 12.

56. *Page 27, Fig. 8: Unfortunately the average RMSE of case 2, which I consider as the most enlightening indicator in this figure cannot be seen very easily because of the larger bars of the maximum. I suggest that you multiply by 5 the average values or, alternatively, provide a second y-axis on the right.*

    Answer:

    The average gain has been multiplied by 10 in the revised figures.

57. *Page 28, line 4: "...i.e. a case that cannot be addressed using 3D-Var or strong constraint 4D-Var." Indeed, but you could address it with an EnKF, and it has already been. Please mitigate your statement.*

    Answer:

    The sentence has been modified as follows:

    However, the main interest of 4DEnVar for atmospheric chemistry arises when the model is not perfect, i.e. a case that is more easily addressed using ensemble-based methods.

58. *Page 29, line 4: "A multiplicative factor of 2.35 has been sampled for the NO emissions of the forecast simulation...": I do not understand "has been sampled"; please clarify.*

    Answer:

    The has been modified as follows:

    A multiplicative factor of 2.35 has been used for the forecast simulation

59. *Page 29, line 5-6: "The advection has been deactivated in this set of experiments to better focus on the impact of emissions uncertainty on chemistry.": This seems too strong an assumption to me! the model becoming 1D on the vertical. Some of your conclusions are based on these experiments which have a limited scope because of this assumption. Advection is of course critical for qualitatively realistic atmospheric chemistry modeling.*

    Answer: See reply n. 2

60. *Page 32, line 12: "during night" -> "during the night".*

    Answer: Done

61. *Page 33, Fig. 13: the useful subtitles of each panel, as seen in Fig. 11, have disappeared. Please add them.*

    Answer:

    Results from only one experiment and different chemical species are shown in Fig. 13 (Fig. 10 in the revised manuscript), whereas 3 different experiments were displayed in Fig. 11 (Fig. 8 in the revised manuscript). For this reason no subtitles were given. We revised the legend of the figure as follows:

    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.

62. *Page 33, line 3: "This difficulty could be overcome by introducing external loops within 4DEnVar": this has been partly addressed in Haussaire and Bocquet (2016). Please mention it.*

   Answer:

   The reference has been included

63. *Page 36, line 1: The title of section 4.3 "Validation on multiple DA cycles" is very misleading, since you are not cycling the scheme, but only repeating one-cycle experiment, i.e. gathering statistics. Please rephrase.*

   Answer:

   The title of the section has been changed to:

   Statistical comparison between 4DEnVar and 3D-Var

64. *Page 38, line 6: "The justification of using an ensemble method for operational air-quality DA, which is significantly more costly than 3DVar, has been demonstrated when model errors were introduced": This has been demonstrated long before by many teams using RRSQRT, EnKF for air quality. Please rephrase.*

   Answer:

   The sentence has been rephrased as follows:

   "Clear advantages of using an ensemble method, which is significantly more costly than 3DVar, have been found when model errors were introduced"

   This mitigates the original statement. However, to the knowledge of the authors, if we limit us to studies using real observations, only two studies found reduced RMSE in ozone forecasts when including parametric model errors in DA experiments (Constantinescu et al., 2007b; Tang et al., 2011). Moreover, results of Tang et al. (2011) were somehow mitigated by the fact that better ozone forecasts corresponded to worse NO_x ones (Tang et al., 2016). Studies that strictly compared strong constraint 4D-Var and ensemble methods (mostly variants of EnKF) with the same chemical model and without estimating model errors did not show significant advantages of using ensembles (Skachko et al., 2016, 2014; Wu et al., 2008; Constantinescu et al., 2007a). We also did not find other studies than Jaumouillé et al. (2012) which compare ensemble based methods to the quite cheap 3D-Var scheme with static error covariances. Therefore, we think that additional comparisons are needed to draw robust conclusions on the delicate performance to cost ratio in chemical DA, using both realistic and simplified air-quality models.

65. *Page 38, line 7-16: Many of the statements there should be mitigated: systematically recall that these results have been obtained in the context of a simplified model (with unrealistic features). Generalized statements for 4D-EnVar cannot be made.*

Answer:

We already recall that these results have been obtained in the context of a simplified model few lines later (Lines 23-27) and that application to DA experiments using real observations is needed. Within lines 7-16 we just summarize the results that have been obtained in this study. In our opinion, this makes already clear that the given statements are not generalized to all possible applications of 4DEnVar.

66. *Page 38, line 15: "...on multiple cycles of DA...": this is very misleading, as discussed before! Strictly speaking this is wrong. Please remove this statement.*

     Answer:

     We changed the phrase to:

     This has been tested with success on several independent DA windows ...

67. *Page 38, line 18: "We conclude that 4DEnVar provides a practical and powerful algorithm for chemical DA.": Again, you have to mitigate this statement. "In the context of a low-order/simplified model, we conclude". Also powerful is too much.*

     Answer:

     The sentence has been changed to:

     We conclude that 4DEnVar is potentially of high interest for chemical DA.

68. *Page 38, line 26: "QG-Chem will represent a useful tool for this type of studies." -> "QG-Chem could represent a useful tool for this type of studies."*

     Answer: Done

References

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.

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.

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.

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.

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.

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.

Gaubert, B., Coman, A., Foret, G., Meleux, F., Ung, A., Rouil, L., Ionescu, A., Candau, Y., and Beekmann, M.: Regional scale ozone 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, 10645–10667, doi:10.5194/acp-15-10645-2015, http://www.atmos-chem-phys.net/15/10645/2015/, 2015.

Hardacre, C., Wild, O., and Emberson, L.: An evaluation of ozone dry deposition in global scale chemistry climate models, Atmospheric Chemistry and Physics, 15, 6419–6436, doi:10.5194/acp-15-6419-2015, http://www.atmos-chem-phys.net/15/6419/2015/, 2015.

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.

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.

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.

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,

19

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.

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.

Skachko, S., Errera, Q., Menard, R., Christophe, Y., and Chabrillat, S.: Comparison of the ensemble Kalman filter and 4D-Var assimilation methods using a stratospheric tracer transport model, Geoscientific Model Development, 7, 1451–1465, doi:10.5194/gmd-7-1451-2014, 2014.

Skachko, S., Menard, R., Errera, Q., Christophe, Y., and Chabrillat, S.: EnKF and 4D-Var Data Assimilation with a Chemistry Transport Model, Geoscientific Model Development Discussions, pp. 1–30, doi:10.5194/gmd-2016-95, http://www.geosci-model-dev-discuss.net/gmd-2016-95/, 2016.

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.

Wang, X., Parrish, D., Kleist, D., and Whitaker, J.: GSI 3DVar-Based Ensemble–Variational Hybrid Data Assimilation for NCEP Global Forecast System: Single-Resolution Experiments, Monthly Weather Review, 141, 4098–4117, doi:10.1175/MWR-D-12-00141.1, http://dx.doi.org/10.1175/MWR-D-12-00141.1, 2013.

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.
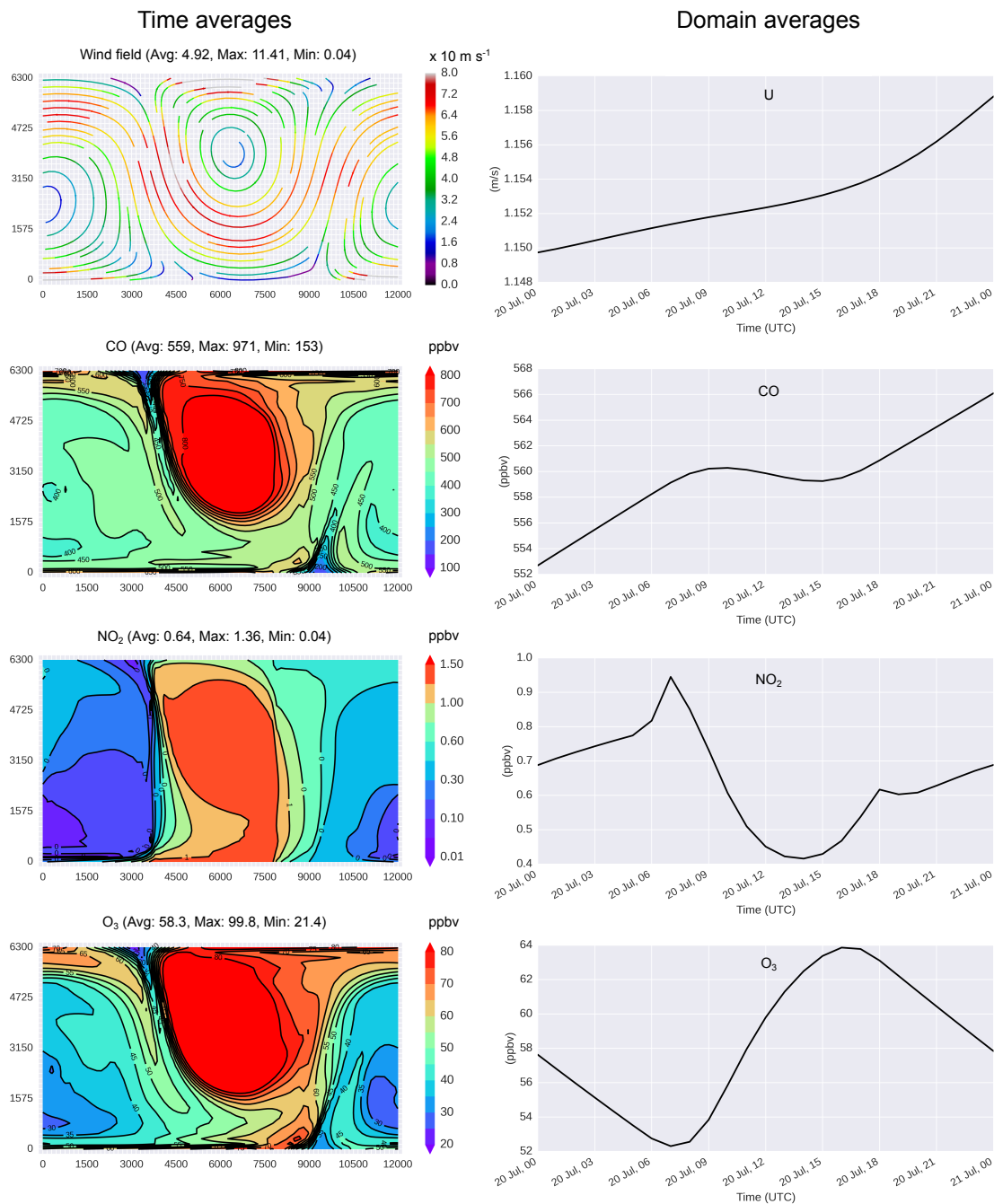
Figure 1. Meteorological and chemical fields of QG-Chem (with deposition included) on day 20 computed using an horizontal grid of 80 x 40 points (resolution of $\sim$ 150 km). Time-averaged fields for a 24 hours period on the left, time series of domain-averaged values for the same 24 hours period on 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.
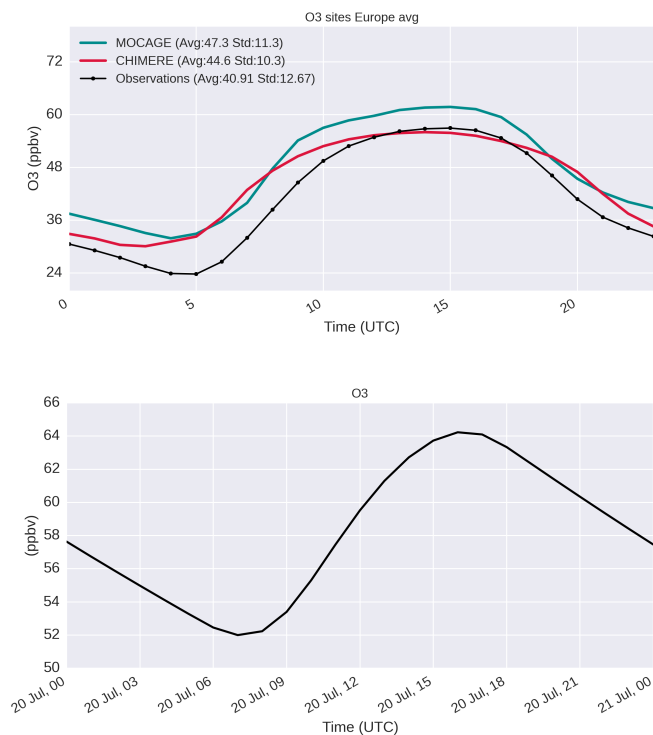
Figure 2. Top: average daily cycle of surface ozone during the period from 5-7-2010 to 21-7-2010 at European AIRBASE measurement sites. Measured ozone in black (averaged over 254 measurement sites), corresponding simulated ozone from two French operational air-quality models in red (CHIMERE) and green (MOCAGE). Bottom: average ozone modeled with QG-Chem on day 20 (corresponding to bottom-right plot in figure 2 of the revised manuscript, with deposition included)