Geosci. Model Dev. Discuss., 8, C2880–C2891, 2015
www.geosci-model-dev-discuss.net/8/C2880/2015/
© Author(s) 2015. This work is distributed under
the Creative Commons Attribute 3.0 License.

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "A low-order coupled chemistry meteorology model for testing online and offline data assimilation schemes" *by* J.-M. Haussaire and M. Bocquet

J.-M. Haussaire and M. Bocquet

jean-matthieu.haussaire@cerea.enpc.fr

Received and published: 12 November 2015

We would like to thank the referee for his/her time, his/her useful input on the manuscript, and his/her interest in our work. Please find below the response to your comments and how the manuscript was modified accordingly.

- *1. Page 7350 last line. You might want to say that the direction is given by the sign of the variable.*

  We agree with your comment. We have therefore modified the sentence as follows: "The 40 scalar variables of L95 are considered to be the magnitude of

winds at 40 locations, their sign giving their direction."

- *2. Page 7352 lines 5-9. This part may need improvements. There is a mix between the origin of the uncertainty (initial condition error in the meteorology and parametric error in the transport part) and the mechanism of increasing uncertainty (i.e. the exponential grow in the chaotic part). Uncertainty is inherent to the initialization procedure and is later increased via exponential growth. I understand that the uncertainty on x_m affects c_m via ϕ_m (Eq. 3 and 4), but it is not clear what do the author mean by "...grow within the transport subsystem". Furthermore, have you perfect initial conditions in the concentration ? I understand that this may be a minor issue if that part of the model dynamics is stable, but this should be clarified here.*

  Indeed, "does not significantly grow within the transport subsystem" means that this part of the model is stable, hence there is no exponential growth of the error in that subsystem. We have changed the paragraph into the following:

  "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."

  The initial concentrations are not generated as perfect, but a noisy representation of the truth.

- *3. Last paragraph of Section 1.1. You might want to specify here in which section L95-GRS will be described.*

  We have modified line 16 of page 7352 as follows: "... and the coupled model, which will be introduced in detail in Sect 2., will be called L95-GRS"

- *4. Page 7353 lines 1-4. This fact and its consequence may not be clear for a general reader without expertise in data assimilation. It is not just the reduced*

GMDD
8, C2880–C2891, 2015

size of the ensemble that matters, but the mere fact that an ensemble exist and that the way how ensemble perturbations evolve is used to estimate the linear and adjoint dynamics.

We agree with your remark. We have added the following to the paragraph: “Indeed, one can for instance use the ensemble of perturbations within the DAW to estimate the sensitivities using finite differences.”

- *5. Page 7353 lines 5-6. Do you expect this in the present study ? In the present form this is not clear in the text, and given that results already exist in the literature that shows the ability of IEnKS to estimate parameters in nonlinear model, the "expected" can be confusing.*

  Instead we wrote in the revised manuscript that the IEnKS is known to handle well parameter estimation.

- *6. Page 7353 lines 8-12. These two sentences can be condensed into one by saying that, under its conditions of use, the IEnKS solves the full Gaussian estimation smoothing problem, that is to say it provides an updated ensemble which is used to compute the covariances. Also, what is the difference between 4D-Var and "standard" 4D-Var for you?*

  For the sake of clarity, we have kept both sentences, but removed the second “Unlike standard 4D-Var”. We found the use of "standard 4D-Var” sometimes necessary because current operational 4D-Vars (such as those in the ECMWF or Meteo-France) are evolutions of the basic 4D-Var, which partly propagate the uncertainty. The expression has disappeared in the revised manuscript.

  The text has been modified into: “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.”

GMDD

8, C2880–C2891, 2015

- *7. Page 7354 line 16. The use of the verb "replace" here can be confusing, given that the resulting L95-GRS model still has a transport part which is structurally similar to L95-T, but it also has an additional GRS component.*

  The transport part (as in the advection equations) is indeed similar in the two models. However, the tracer species is replaced by the set of species of the GRS. The fact that the ROC species behaves like the tracer of the L95-T is a rather fortunate but involuntary coincidence. We would not say here that the tracer part is enhanced by adding the rest of the chemistry of the GRS since the ROC species, even if it numerically behaves like the unreactive tracer of the L95-T, is considered like a reactive species of the whole chemical scheme. We changed “replace” by “substitute” in the revised manuscript.

- *8. Page 7355 line 1. Do you mean “ROC” instead of "VOC"?*

  No, we do mean VOC here, even though we agree that the difference is subtle. As explained by Venkatram et al. (1994) at page 3666, VOCs are emitted and ROC is a surrogate for all the products of the oxidation of these emitted VOCs. There is therefore more to ROC than simply VOC in a chemical sense. Therefore, when explaining in a general manner the purpose of the GRS model, we would rather talk about the VOC emissions.

- *9. Equation 11. More explanation and details on how Eq. (11) is obtained via the QSSA approximation are required. Also, I suggest to say at the very beginning of Section 2.1 that full details are provided in Appendix B. In the present form the reader may think that only specific points are highlighted therein.*

  We agree with you to warn the reader early of the content of Appendix B, where more details and justification about the QSSA are given. We first corrected the equation of $\frac{\mathrm{d}[\mathrm{RP}]_{m+\frac{1}{2}}}{\mathrm{d}t}$, which should have been numbered (B2), and which should have been :

GMDD

8, C2880–C2891, 2015

CC BY

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

Moreover, we have changed the paragraph page 7376 line 14-17 into : " 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

$$0 = \frac{\mathrm{d}[\mathrm{RP}]}{\mathrm{d}t} = k_1[\mathrm{ROC}] - [\mathrm{RP}]\left(k_2[\mathrm{NO}] + 2k_6[\mathrm{NO_2}] + k_5[\mathrm{RP}]\right)$$
$$\Leftrightarrow [\mathrm{RP}] = \frac{k_2[\mathrm{NO}]+2k_6[\mathrm{NO_2}]}{2k_5}\left(\sqrt{1+\frac{4k_1k_5[\mathrm{ROC}]}{(k_2[\mathrm{NO}]+2k_6[\mathrm{NO_2}])^2}} - 1\right).$$

- *10. Page 7357 lines 21-23. Note that when defining the value of k_3, the reference at the equator has been taken (Appendix B). I guess this will not lead to significant quantitative difference, but it is worth mentioning why it was not used a mid-latitude reference in that case.*

We agree that it would have been more logical to choose a coefficient at mid-latitude, since the wind is supposed to be of mid-latitude as well. This will indeed not lead to significant differences. As can be seen in the figure attached at the end of this comment, the value of the photolysis coefficient at mid-latitude (50^oN) on 21 March (blue line) is of the same order of magnitude than at the equator (green line). Moreover, this coefficient at mid-latitude in summer (21 June, red line) reaches values as high as the one that we used in our reference run.

GMDD
8, C2880–C2891, 2015

- *11. Page 7357 lines 26-28. Do you mean that you have tuned E^{ROC} and λ^{ROC} in the model to fit the real value of ROC ?*

  Since ROC is a surrogate compound for COV and other radical producers, it is not straightforward as to what "real values" of ROC are. However, we roughly know what the "real values" of O_3 or NO_X are and we tried to fit λ (unique and identical for each species) and E^{ROC} to match these values. One should bear in mind that λ is kept like in the L95-T model.

- *12. I think all figures need to be improved, in particular in their labels and titles size. In the printed version they are barely readable. I understand this may be fixed at a later stage, but it has to be done.*

  We agree with you on this point and we have made our best to improve the readability of the pictures.

- *13. Page 7359 line 10. By looking at the O_3 panel, I would not say that O_3 remains at high levels, but just that it remains small but not zero.*

  Even if in the panel, the O_3 concentrations can get close to 0, it is however not happening particularly because of the day/night cycle, but rather because of the transport. The impact of the night, which is to consume O_3 and NO through Eq. (R4) does not influence the O_3 concentration to the extent of bringing it down to 0. This fact can actually be seen on the Fig. 3, where there are some clear black stripes on the NO concentrations, unlike for the O_3, meaning it does remain at "high levels", regardless of the fact that some cells in the grid do have close to zero concentration levels.

- *14. Section 2.3, 2^{nd} paragraph. How do you define the maximal ozone concentration ? Is it the maxima in space and in time over a specified long simulation ? Then, if my understanding is correct, you are doing an analysis of the model behavior (the maximal ozone concentration) by changing two of its parameters*

GMDD

8, C2880–C2891, 2015

(the emission of ROC and NO_X). This type of analysis, which is of course worth doing and much informative, does not automatically teach us about the nonlinear behavior of the model in terms of the relation between state-variables, but rather about the model phase diagram, that is to say how the structural properties of the model change (for instance from periodic to aperiodic) by changing parameters.

As we tried to explicit in the caption, the maximum is calculated as an "average over the domain". By averaging in space like this, we get a response (after the transitory regime) of O_3, NO_X or ROC which are merely daily oscillations which amplitudes are solely determined by the emission factors. This way, we show that there is a nonlinear response of the ozone concentration to its precursor concentrations. We invite you to read as well the answer to the first comment of the referee #1, Stéphane Vannitsem. The EKMA model can also be considered as a phase diagram, whose structure can only be obtained with a significantly nonlinear chemistry.

- *15. Section 2.3, last paragraph. It should be said that changing β the forcing and advection terms are changed, so that a modification of the stability properties of the L95 model is obtained in terms of, for instance, the number and amplitude of positive Lyapunov exponents or the Kolmogorov entropy. The consequence for the performance of the data assimilation methods are significant. For some value of β L95 may no longer be chaotic. See Carrassi, Vannitsem and Nicolis (2008, Q.J.Roy.Meterol.Soc.) for a more extensive analysis of the L95 properties for different values of the forcing, dissipation and advection, and in relation with another data assimilation strategy.*

  The α and β parameters that are considered are merely unit change factors. Technically, defining new variables $T = \frac{1}{\alpha}t$ and $X = \frac{1}{\beta}x$, one can write :

$$\frac{\mathrm{d}X_m}{\mathrm{d}T} = \frac{\alpha}{\beta}\frac{\mathrm{d}x_m}{\mathrm{d}t} = \frac{\alpha}{\beta}\left[(x_{m+1} - x_{m-2})x_{m-1} - x_m + F\right]$$

GMDD
8, C2880–C2891, 2015

CC BY

$$= \alpha \left[\beta(X_{m+1} - X_{m-2})X_{m-1} - X_m + \frac{F}{\beta} \right] \quad (1)$$

which is the equation number (15). Considering this new equation, changing β should only mean converting the unit of x into another one. The dynamics of a wind variable and the stability of its underlying model should not be impacted by the choice of the unit in which it is expressed. We believe that changing the advection term and inversely proportionately changing the forcing one should then not impact the properties of the L95 model. We made the following change in the revised manuscript to account for your remark. "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."

- *16. Page 7362 lines 10-13. I would not just say with "state-of-the-art data assimilation methods", cause this may one think on methods that are operational in prediction centers. On the other hand those methods do not necessarily (and indeed almost never) make possible the propagation of information across model compartments. For this to happen one has to use global error covariance matrices defined over the full system, as you do here.*

  Thank you for pointing out to this potential confusion. We have consequently amended the paragraph as follows:

  "... both ways in advanced data assimilation methods, as long as the error covariance matrices are defined over both subsystems."

- *17. Page 7363 line 5. How do you select the dynamical regime ? Or am I getting wrong on what you mean by this.*

  The dynamical regime is the one stated in the parameters of the model (F=8, $\alpha = \beta = 1$). In this regime, the unstable subspace being of size 14, a set of members of size 20 is sufficient not to need localisation. There is nothing hidden behind this statement.

GMDD
8, C2880–C2891, 2015

CC BY

- *18. Page 7363 line 12. What do you mean by "extrinsic model error".*

  The statement could be confusing, so that we clarified the sentence and changed it into “However, note that the finite-size approach does not account for model error but sampling errors.”

- *19. Page 7363 line 13. Typo "offlne" ⇒ "offline".*

  We do not notice, in the printable version of the manuscript, the typo you are referencing.

- *20. Page 7363 line 17. Do you mean "second" instead of "first" ?*

  It is indeed the second offline system, but the first variation from the baseline. We changed the text for the sake of clarity.

- *21. Page 7364 lines 5-10. You might want to add the explicit formulas for* $RMSE^{filter/smoother}$

  Following this very relevant remark, we have added the following explanations in the revised manuscript:

  “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

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

  and

$$\text{RMSE}^{\text{smoothing}} = \frac{1}{N_t}\sum_{i=1}^{N_t}\sqrt{\frac{1}{M}||\boldsymbol{x}_{\mathbf{a}}^{\boldsymbol{i}} - \boldsymbol{x}_{\mathbf{t}}^{\boldsymbol{i}}||^2}$$

  where $\boldsymbol{x}_{\mathbf{a}}^{\boldsymbol{k}}$ is the average of the updated ensemble at time k, $\boldsymbol{x}_{\mathbf{t}}^{\boldsymbol{k}}$ is the truth at time k and for a vector $\boldsymbol{x}$ of size M, $||\boldsymbol{x}||^2 = \sum_{j=1}^{M} x_j^2$ ”

GMDD

8, C2880–C2891, 2015

CC BY

- *22. Page 7368 lines 1-7. Can you comment on the choice of the values of observational standard deviations for the different variables ? Also, I suggest to state clearly that the normalization is done using the observational error standard deviation.*

  The concentrations of each species have been averaged over the domain, leading, as explained in the answer to the comment 14, to a daily oscillation curve of fixed amplitude. 10% of the maximum of this oscillation is taken as the observational error standard deviation. For ozone, this ratio of 10% is reasonable and in accordance with usual observational standard deviations used in data assimilation experiments. For instance, Wu et al.(2008) used a standard deviation of $10\ \mu g.m^{-3}$ while noting a mean of ozone observations of about $70\ \mu g.m^{-3}$. We assume this choice to be valid for other species as well.

  Regarding, the second point of your comment, we have changed the sentence line 4-5 into "All the RMSEs shown in this section are normalised by the observational error standard deviation of the corresponding species."

- *23. Page 7368 lines 9-16. It is not clear how the observations are distributed. Are they evenly distributed as in the results of Fig. 7 ?*

  Yes they are. We assumed the first sentence of the paragraph, stating "as in Sect 3.4", where Fig 7 is presented, was explicit enough. We changed the text to make it clearer: "At first, the number and distribution of observations of the concentration variables have been varied following the same setup as in Sect. 3.4."

- *24. Page 7369 line 20. Do you "40" instead of "240" ?*

  As explained in the p.7356 L.19-21, the whole state is composed of 240 variables, which are the wind and 5 species at 40 grid points. So here, we do mean "240".

- *25. Page 7369 lines 24-25. Why do you chose to have an unbiased initial en-*

GMDD

8, C2880–C2891, 2015

semble for the emission rate (i.e. centred around the truth) and a biased one for the forcing ? You comment on this a bit later in the text, but it would be better to say something about this choice when you present it.

In the article of Bocquet and Sakov (2013), the setup of the experiment of parameter estimation with the L95-T chooses to have a biased initial ensemble around $F = 7$ but has an unbiased initial ensemble for the emission rate of the tracer species. We just kept the same experimental setup as this reference. If the two emission rates would have been centered around the true value minus 10% (instead of the true value), the convergence speed of the parameters would have been slightly faster.

- *26. Page 7375. The numbering of equation seems to be incorrect. (B3) should move one line forward.*

  You are right. Thank you very much for noticing this detail! This has been corrected.

- *27. Caption of Fig. 8. I would change "several" into "three", and I would better say "observational error" instead of just "...observations." at the end of the caption.*

  This has been taken into account and modified. Thank you for the suggestion.

Interactive comment on Geosci. Model Dev. Discuss., 8, 7347, 2015.

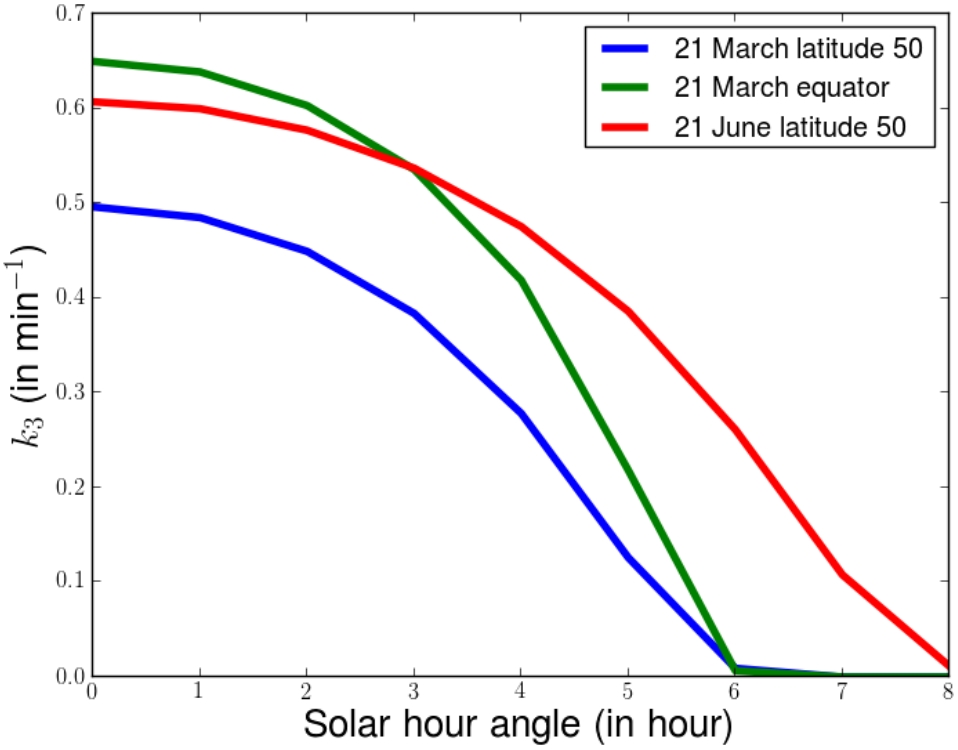

Fig. 1. Photolysis coefficient

GMDD
8, C2880–C2891, 2015