Geosci. Model Dev. Discuss., 6, C1764–C1777, 2013
www.geosci-model-dev-discuss.net/6/C1764/2013/

Geoscientific Model Development Discussions
Open Access

Interactive comment on "Regional scale ozone data assimilation using an Ensemble Kalman Filter and the CHIMERE Chemical-Transport Model" *by* B. Gaubert et al.

B. Gaubert et al.

gaubert@lisa.u-pec.fr

Received and published: 28 October 2013

1 General comments

Among the Airbase ozone data, they selected randomly about a half for the analysis and the other half for validating their simulation (without assimilation) and their analyses with various configurations. The different configurations they proposed aimed at providing insight of the sensitivity of their EnKF to the background and observation error covariance matrices. The evaluation of their analyses is provided for a period of 10 days during summer 2009 when high surface ozone concentrations were observed. The discussion is well conducted by separating the results of the day from the results of

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

the night or by looking at the station types. Nevertheless they assumed that the weak sensibility of their system to the background and observation error statistics proves the robustness of their system. But they did not discuss why they have such robustness. Is it due to the large amount of assimilated observation? Or is it due to some property of the EnKF?

Answer: Observations were objectively selected and a large density of stations is found, particularly over Western Europe. In this case, it is well known that sensibility to analyses is less sensitive to a change in observation/background error variance. In this case many observation sites are present within the localisation radius (250 km) and within the correlation length with respect to a given model grid cell. Then, in the assimilation procedure, the weight of the observations will be large independently of "not too big" variations of the background and observation errors, the weight of the background will be small. This makes the system less sensitive to the background and observational error formulation.The text has been modified in the conclusions: "Stations used for assimilation or evaluation are spatially close, where the ozone observations network is spatially dense. Thus many observation sites are present within the localisation radius (250 km) and within the correlation length with respect to a given model grid cell. Then, in the assimilation procedure, the weight of the observations will be large independently of "not too big" variations of the background and observation errors, the weight of the background will keep small. Thus, in terms of accuracy measures, only small changes in performance statistics are found among experiments even for substantial changes (up to a factor of two) in the model and observation errors. However, for a dense network correlations between assimilated observations cannot be excluded. A step further would be to perform an a posteriori diagnostic of the observation error correlation and if necessary to take it into account in the assimilation procedure"

As they conclude, they demonstrated that their system is robust enough and suitable for implementation in operational systems supported by the European FP7 MACC project.

GMDD

6, C1764–C1777, 2013

My main concern is to know if a 10 days period is enough to demonstrate the ability of the system? The chosen period furthermore corresponds to an episode of high ozone concentration when the model already perform well without assimilation. It would have been useful to have a winter period as well in the study even if this period of the year has less media exposure (the ozone concentrations being generally lower).

Following the referee's suggestion, in the revised version of the paper an evaluation of the assimilation system was added for a whole three months summer period in one specific configuration. The following paragraph is added: "As the evaluation of analyses is provided for a period of 10 days only (in August 2009), an analysis of 90 days has been conducted in order to validate the performance statistics. The period covers three months of the summer 2009 (JJA 2009) which allows the evaluation of rather different meteorological conditions. The configuration of the data assimilation system is similar to the PAR_MOD_DESR experiment and is called SUMMER_NEW. An ensemble of 20 members is used; the observation error variance is set to 25 ppb^2; the covariance inflation factor is tuned every day from the hourly diagnostics of the background error and several model parameters are perturbed (Table 1) but not corrected in the assimilation process. The set of assimilated as well as evaluation stations is unchanged with respect to the previous experiments. We employed the same evaluation method based on accuracy statistics (Table 5) and on analysis of the diurnal ozone and error cycle (Fig. 7b). A similar improvement is found for the SUMMER_NEW analysis for the whole summer as for the 10 days period: the averaged hourly correlation coefficient of 0.87 and a RMSE around 7 ppb (with respect to 0.72 and 10.2 ppb for the reference simulation). These errors exhibit a similar diurnal cycle with an average value of 8 ppb during the night and 6 ppb in the daytime (Fig. 7b). These results indicate that the data assimilation system is suitable for reanalysis of longer periods. However it should still be tested in other seasons even if ozone pollution is most important during the summer season."

Figure 1 (see below): Average diurnal cycle of ozone and the associated RMSE (from

GMDD
6, C1764–C1777, 2013

the 1st June to the 30rd of August 2009 at evaluation stations) for the CHIMERE reference run and the analysis. Separated panels are shown for the suburban stations (left, N=172), rural stations (middle left, N=102), and for background stations with an altitude below 300m a.s.l (middle right, N=17) and above (right N=27).

Accuracy measures	Simulation name	TOT	Background	Rural	Suburban	Background (> 300m)
Bias (ppb)	REFERENCE	-2.15	-0.82	-1.27	-2.81	4.9
	SUMMER_NEW	0.0	1.14	0.94	-0.67	6.64
RMSE (ppb)	REFERENCE	10.2	8.73	9.89	10.53	12.92
	SUMMER_NEW	7.09	7.55	7.16	7.0	11.51
Correlation	REFERENCE	0.72	0.55	0.68	0.76	0.47
	SUMMER_NEW	0.87	0.7	0.84	0.89	0.7
RMSE (8hmax, ppb)	REFERENCE	7.7	6.97	7.47	7.9	10.86
	SUMMER_NEW	4.92	6.09	4.9	4.81	9.45
RMSE (daily max, ppb)	REFERENCE	10.16	8.97	10.19	10.26	15.64
	SUMMER_NEW	7.12	8.31	7.4	6.83	13.93

Table 1. Comparison of the accuracy measures (hourly bias, RMSE and correlation coefficient, daily maximum and maximum of the daily eight-hour ozone mean RMSE) for the CHIMERE run (reference) and for the analysis over the evaluation set for suburban stations (N=187), rural stations (N=112), and for background stations that have altitude below 300m asl. (N=18) and above (N=27). These statistics are computed at evaluation stations for the period from the 1st June to the 30rd of August 2009.

My other concern about this 10 days period is to know if the restriction to 10 days is due or not to computational costs? If not, why the model is evaluated with a period of 3 months and not the analysis? If yes, it is an issue to operationally run the system in MACC?

Answer: The restriction to 10 days was chosen in order to perform a certain number of different assimilation exercises (sensitivity tests) which are presented in the paper. The assimilation algorithm is well suited in an operational framework, an analysis of one day (including the control run) takes 5 hours on a 12 CPUs node where only CHIMERE

GMDD
6, C1764–C1777, 2013

simulation are parallelized (but they are launched sequentially).

I finally strongly suggest the author to pay more attention at the structure of the paragraphs (even of the document itself sometime) and at the English language. See few examples in specific comments hereafter. Concerning the structure of the paper, section 2 could be divided into two parts, one on the EnKF and one for the model description (currently section 4.1) as the model is a part of the assimilation system. The current section 4.2 could be include in section 5. Section 5 should also contain a subsection with a better description of the period chosen for the study as well as the repartition of the stations (by type and by purpose - assimilation or validation) for these periods. Note that this repartition is already described in the first paragraph of section 5.1.

Answer: The section 2 is now dedicated to the CHIMERE-EnKF data assimilation setup and diagnostics with a part dedicated to the diagnostics and error modelling and another one to the CHIMERE description. The (new) section 4 contains now all the results with a first point made on the period chosen. The paper has been carefully corrected for English language.

2 Specific comments page 3034 lines 2 to 4: "The Ensemble Kalman Filter is ... analysis step." This kind of sentences are useless in an abstract.

P3034: The sentence has been removed. The abstract has been modified.

page 3034 lines 6 to 9: "The analyzed ozone field is evaluated ...". There are three ideas in this sentence: 1. the comparison with consistent set of observations (do you mean "ozone ground measurements" by observations?); 2. the reduction of the quadratic error and the improvement of the hourly correlation coefficient ; 3. an improvement despite a low ensemble size. Please split the sentence to highlight these 3 ideas.

P3034: The sentence has been modified, it is now “The analyzed ozone field is evalu-

GMDD
6, C1764–C1777, 2013

ated using a consistent set of ozone observations. This comparison shows a reduction of the RMSE by about a third and an improvement of the hourly correlation coefficient. These achievements are obtained despite of a low ensemble size designed for operational purposes. ”

page 3034, introduction: you write that ozone impacts human health and vegetation growth. You have then one sentence to explain the impact on "human health" but nothing for the impact on the "vegetation growth". Please think of adding a sentence to explain this impact.

P3034: The following sentence has been added: “Cumulative ozone uptake through leaf stomata over a given threshold causes injury to vegetation (Fowler et al. 2009).”

page 3035 line 1: GMES is not focused only on "ozone and other pollutants" as suggested by your sentence. GMES services are indeed Land Monitoring, Marine Monitoring, Atmosphere Monitoring, Emergency Management, Security and Climate Change. Please correct.

P 3035: The sentence has been modified and starts with: “Among others tasks, the GMES (Global Monitoring for Environmentand Security) programs foster. . .”

page 3035 lines 9 and 10: What is the difference between "direct model simulation" and "forecast"? What do you mean by performing a forecast in synergy with observations?

P 3035: The term forecast has been removed. Observations are generally used for validation or to correct the simulation thanks to simulation post processing or data assimilation.

page 3035 line 16: Please precise here what you mean by "analysis". Is it a result of a data assimilation system? In your document, does "analysis" always refer to as the result of a data assimilation system?line 21: What is the "analysis product"?

P 3035: The analysis always refers to a result of a data assimilation system. The sentence has been modified, it is “In addition, analyses resulting from data assimilation

GMDD
6, C1764–C1777, 2013

of observations in near real time provide the best representation of the surface pollutant concentrations."Thus, the term 'analysis product' has been removed.

page 3035 line 23: What are these "reporting activities"? On which analysis are they based? CHIMERE one? page 3035 line 26: What in the link between the first sentence of this new paragraph and the following sentence?

P 3035: The aim of the PREV'AIR platform is to provide daily air quality forecasts and maps resulting from numerical simulations on different spatial scales. It is based on the CHIMERE model for European and France simulations. However, threshold exceedances are calculated for the analyzed field. A new sentence has been added: "One of the challenges of these air quality modelling chains is to provide uncertainties and errors associated to the modelling results. These uncertainties are estimated by the comparison of simulations obtained by different models and the same sets of observations."

page 3036 line 9: Are Courtier et al. (1998) and Houtekamer and Mitchell (1998) really the right references for pioneering work in data assimilation for numerical weather predictions? Without going to Kalman (1960), pioneering works were realised by Talagrand and Courtier (1987) and Courtier and Talagrand (1987). Moreover, it looks like Elbern et al. (1997) was a study done before the pioneering work for numerical weather predictions you assumed to be done in 1998! page 3036 line 27: In Geer et al. (2006) for example you can see that there is more than only two major strategies that are employed to obtain an accurate 4D-analysis of ozone concentrations. Please rephrase. Moreover, in your sentence, you forgot to mention that the 4D-Var analysis allows the correction of the initial condition.

P 3036: These references (Courtier et al. 1998 and Houtekamer and Mitchell 1998) show the complete achievement and the evaluation of the implementation of some data assimilation methods in a near operational framework. The sentence is now "After pioneering work in numerical weather predictions such as the 4D-Var (Talagrand and

GMDD

6, C1764–C1777, 2013

Courtier 1987, Courtier and Talagrand 1987), methods already developed in this field have been successfully applied to air quality simulations."

page 3040 line 16: What do you mean by "these" in the sentence "These are derived from ..."? Do you mean that in your study the ensemble of perturbed ozone concentration fields are derived from a two-dimensional Gaussian distribution? In this case, why do you add the reference to Evensen et al.(2003)? And you also should specify what you are using as a variance and spatial correlation. page 3040 lines 21 to 22: Diagnostics from Desroziers et al. (2005) allow to derive the forecast or the background error variance (in the observation space) not the model error variance. Please correct here and all over the document if needed to better differentiate background error and model error. page 3040 line 24: The whole paragraph (and the section as well) deals with the BECM and suddenly you give an equation for the observation error. You could introduce this equation later within a paragraph on the observation error for example.

P 3040: L16 Yes, "these" denotes perturbations. More equations and explanation have been added in the method section. The sentence is now: "The noise is derived from a two-dimensional Gaussian distribution with some fixedspatial characteristics namely zero mean and unitary variance (Evensen1994, 2003). Pseudo-random fields (η) are generated with a fixed horizontal decorrelation length of 200 km (Boynard et al. 2011, Coman et al., 2012).This parameter is close to the value of 270 km used in several other studies (Chai et al. 2007, Constantinescu et al. 2007c, Frydendall et al. 2009) and in any case our results are similar with both values. These perturbations are only added to the analysed ozone state. As suggested in Sandu and Chai (2011), the same noise is applied for all vertical layers inside the calculated boundary layer and thus induces a vertical correlation in the background error. The noise q_(k-1) is the product of a spatially correlated noise η and a tunable coefficent of relative standard deviation (sd)." The method used to generate these fields is explained in the appendix E of Evensen (2003). L21-22, all over the text, the term 'model error' has been replaced by "background error". L24 In the new version,all the diagnostics are in a section dedicated to

GMDD

6, C1764–C1777, 2013

a posteriori diagnostics and include also a description of the RCRV statistics.

page 3043 line10 : Is the "daily average ozone profile" also averaged over the whole domain? If yes, please specify.

P 3043: Indeed, we calculated an average over all the considered station types. The new sentence is “We plotted on the figure 2 the daily ozone diurnal cycle obtained for summer 2009 averaged over each type of observations.” page 3044 line 20: What are the horizontal and vertical resolutions of meteorological variables from ECMWF? What are you using as spatial interpolation?

P 3044: For the meteorological variables provided by the ECMWF, the horizontal resolution is 0.25°x0.25° (T799)and the vertical resolution is 60 levels from the surface to 0.1hPa. These meteorological data are bi-linearly interpolated on the CHIMERE grid.

page 3045 line 7: The selection of the period you propose to study is an important information. I found it strange to find the information for the period for which assimilation is performed in parenthesis. And I found it strange to find this information in the section of the evaluation of the reference simulation. I suggest you to create a new section (or subsection) in which you could describe the period you studied and the reasons why you have chosen that one. page 3045 line 10: Fig. 6a displays the ozone fields for 15 August but you introduced it in a sentence where you discussed the period between 19 and 21 August. page 3045 line 21: Table 4 does not support the fact that you have a minimum of the daily maximum during the episode. page 3045 line 24: I guess that Tab. 3 contains statistics only for the assimilation period and not for the whole summer. Please make it clear that you are not discussing anymore about statistics over the whole summer (as suggested by the first sentence of your paragraph).

P 3045: L7 a new section is dedicated to the choices of assimilation and simulation periods, parenthesis are removed. L10 The figure displays the ozone fields for the 20th, it was a mistake in the figure caption. L21 This statement makes references to figure 3, the sentence has been corrected. L24 In the new version, statistics and

GMDD
6, C1764–C1777, 2013

description of the model evaluation period is separated into two sections (one for the summer and another one for the assimilation period).

page 3048 line 11: What do you mean by "observation error of 8ppb at 60km horizontal resolution"? page 3048 line 25: I suggest you add the reference to your analysis, i. e. "the analysis from REF_ASSIM experiment". As you gave names to your analyses experiments, please always refer to these names in your text.

P 3048: L11 the sentence is now: “Using observations from the AIRNOW database in the USA, Chai et al. (2006) evaluated the ozone standard deviation inside a model grid cell (60 km horizontal resolution) and got on average a daily range between 5 ppb and 13 ppb (at night). Finally, in their assimilation framework, they assumed an observation error of 8 ppb.” L25 corrected.

page 3049 line 17: I found it difficult to see "the ability of the ensemble to extend innovations along with the ozone flow" with the figures you proposed. For instance, the differences in Spain for example between Fig. 6a and Fig. 6b can come from the transport from France as the ozone concentrations should have been modified in France the previous assimilation cycles due to the numerous stations there. For me, you did not demonstrate the "clear advantage of the Kalman filter". Page 3049 line 23: Does the "8-h mean maxima" is the same as the "daily maximum of the running eight hour mean" of line 2 page 3050? If they are the same, please explain its role ("indicator for the impact of ozone exposure to human health") before in the text. This could enter in a section "evaluation method".

P3049: We precise that these corrections are transported during the forecast step, this means that correction can be found away from the localization window. The new sentence is “However, the spatial shape of the corrections, for instance over the North Sea illustrates the ability of the sequential assimilation to extend innovations along with the ozone flow (in the north-west direction) during the forecast step. L23 Yes, corrected, indication of its role has been added.

GMDD

6, C1764–C1777, 2013

page 3051 line 28: It is difficult to see differences between Figs. 6b and c.

P3051: We do not well understand this remark, most of the changes can be found in the southern part of Europe (i.e. in Spain, Greece and in the south of Italy).

Page 3052 line 26: You present the RCRV you have already introduced in the same section but earlier (page 3050 line 27). You could introduce it once and provide the information in a section "evaluation method".

P 3052: In the new version, the RCRV appears in a dedicated section “A posteriori diagnostics and covariance modeling”.

Page 3056 line 6: What sdmeas stands for?

P 3056 line 6: This part has been removed.

Page 3056 Eq. 8: Please explain the coefficient 1.2 in the equation.

P 3056: This part has been removed.

Page 3070 Tab 3: Why are you changing the the name of your REF ASSIM experiment with ANA just for this table? Please keep REF ASSIM and do not introduce a new reference.

P 3070:Corrected

page 3071 Tab. 4: Please specify in the caption the period you used to compute the statistics. Be careful you did it in each caption of each table or figure.

P3071:The period is specified.

page 3075 Fig. 3: Please rephrase the caption. Moreover, please add the label (a) and (b) on the figure.

P3075: There is only one figure left (for the all summer period; the results for the assimilation part were removed).

GMDD

6, C1764–C1777, 2013

Page 3079 Fig. 7: Please specify the analysis (REF ASSIM?).

P3079: Done

3 Minor revisions/comments

Page 3034 line 13: I would suggest "a posteriori diagnostics" instead of "a posteriori Desroziers diagnostics" as "Desroziers diagnostics" is not a generic term. But further in the text, I would suggest to keep "Desroziers diagnostics" if you precise clearly that by this you mean the diagnostics from Desroziers et al. (2005). Page 3034 line 18: "parameters" instead of "parameter".

P 3034: Done

Page 3038 Eq. (2): Please add a transpose sign on the last right parenthesis.

P 3038: Done

Page 3040 line 21: "diagnostics" instead of "diagnostic".

P 3040: Done

Page 3041 lines 2 to 3: I suggest to rewrite the sentence with "These quantities have to be computed a posteriori because they require the analysis ya (in the observation space), the observations yo and the ensemble mean of the forecast yf (in the observation space)". Moreover, link this sentence with the following or start a new paragraph.line 10: Please provide the number of the section where this "has been demonstrated".

P3041: L2-3 the sentence is now: “Because they require the analysis ya (in the observation space), the observations yo and the ensemble mean of the forecast yf (in the observation space), these quantities have to be computed a posteriori.” L10 it is in the article of Li et al. (2009b)

Page 3042 line 18: Please rephrase. I suggest: "In this paper, we choose this approach

GMDD

6, C1764–C1777, 2013

because it has the advantage of being directly targeted on the pollutant species of interest (ozone) and it does not require any other metadata."

P3042: Done

Page 3044 line 17: As the aerosols are not included in our simulation, I suggest to remove the reference Bessagnet et al. (2004).

P3044: Done

Page 3071 Tab 4: "averaged" instead of "average"

P3071: Done

page 3073 Fig 1: Could you please remove the background grid.

P3073: Done

page 3083 Fig. 11: The legends on the right of the figure and on the top left are difficult to read.

P3083: Done

Interactive comment on Geosci. Model Dev. Discuss., 6, 3033, 2013.

GMDD
6, C1764–C1777, 2013

Interactive Comment

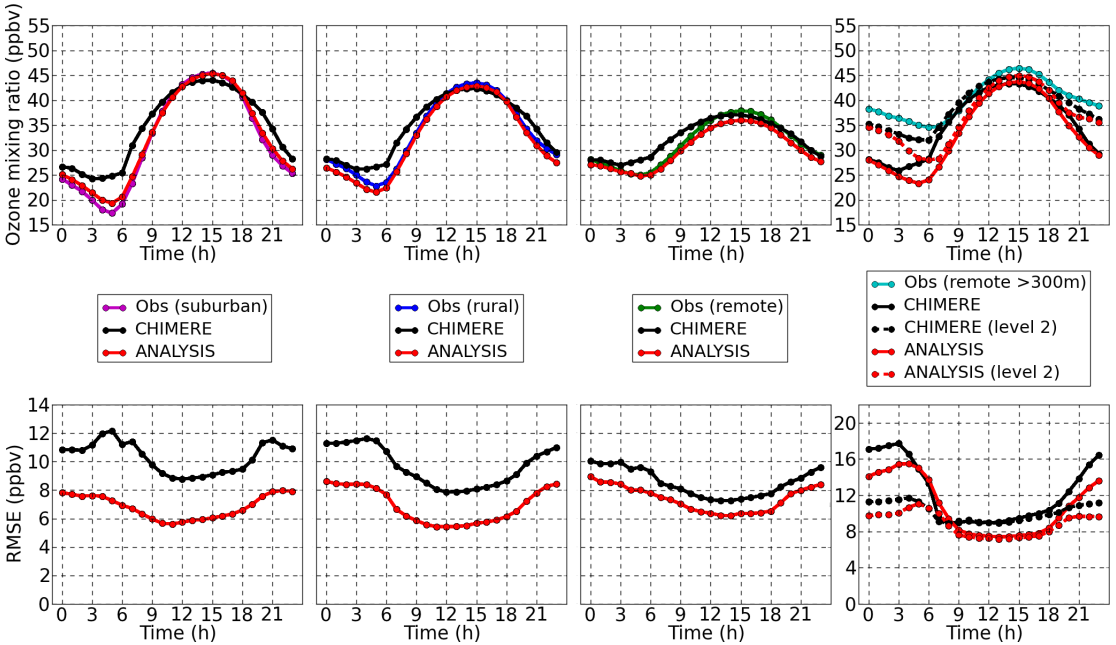

Fig. 1.

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper