

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 comment

The authors test basically only the ability of the Kalman Filter to interpolate dense surface observations in the spatial domain. Use of the analysis as initial conditions for forecast is not considered in the paper. One could argue that mapping algorithms such as Kriging or Optimum Interpolation would have done a similar job to interpolate these observations. The presented study somehow fails to convince why a 4D data assimilation method was applied to the problem, i.e. the interpolation of a rather dense network of ozone observations to 0.5x0.5 grid. The study period (10 days) is too

C1744

short for air-quality assessment according to EU guidelines. The authors compare the ozone surface analysis with ozone surface observation, which were not assimilated, but these observations are still very close to the assimilated observations and probably highly correlated, in other words, they are not really independent. This might have consequences for the assumption that observation errors are not correlated. See for example Liu and Rabier (2001) for a discussion of the problem. I also think that the large number of assimilated observations is actually not helpful for a key point of the study i.e. to test the impact of different formulations of the observation and background errors. The presented ozone analysis is constrained by the large number of observations and not sensitive to changes in the observations errors. It would have been helpful to thin the assimilated observations further and further to see at what point the analysis quality starts to deteriorate and if at this point changes in the background observation error formulation make a difference. The question would be can a Kalman-Filter make better analyses with a thinned network than an interpolation method? Answer: First of all, we must recall the main objectives of the paper. Our purpose was to present a new assimilation system (ENKF) for the surface ozone assimilation coupled to the state of the art CHIMERE model. We wanted to prove that this system was robust and allowed producing significantly more accurate ozone analysis than the reference mode. More specifically, the context of these developments was to prepare a system that could be used for operational applications. We have set-up experiments that in the frame of such applications, especially using a large number of surface observations. Moreover to render the study more robust, authors have chosen to put additional efforts on two specific points: 1/ to develop an objective station classification (based on the work of Flemming et al, 2005) to better evaluate the performances of the reference model and the assimilation system; 2/ to conduct sensitivity tests on the covariance matrix formulation (for the model and observation errors) and to evaluate its impact on the performance of the system. As it has been mentioned by the reviewers the results presented in this work are raising interesting questions to analyze more deeply the system and then prepare future improvement. While we can answer many questions in our re-

C1745

ply and in the revised paper, others correspond to important research questions which deserve specific studies and are thus beyond the scope of this paper. Concerning the duration of the experiment, we have tried to follow the reviewers recommendation and we present the results for a one summer assimilation experiment in section 4.7 of the new version of the paper. About the comparison of the system with simpler assimilation systems such like the Optimal Interpolation and to evaluate the gain of using a full 4D assimilation system, we do believe that this is interesting but we have planned to do such a work separately from this study that is already dense. Our working hypothesis is also that the ENKF system is on a long term the most versatile system for example for multi-component data assimilation. We have added a sentence in the conclusion to mention the interest of such a future comparison. The problem of thinning the data that referee raises is an interesting academic problem. However, as mentioned earlier, we were interested to use the full available set of surface observations, because the assimilation system is intended to work later on in an operational framework in order to produce the “best” possible analysis. So the starting point of our paper was to address the question how to obtain the densest possible network of “suitable” sites. This is why we applied a specific classification procedure, in order to include as many as possible sites potentially including also those labeled as “urban” based on criteria used for the Airbase data set. It turned out that this framework was not helpful to make evident the sensitivity of the assimilation system to its error formulation as the referee states. Yet, this small sensitivity is a result from the operational point of view, even a very positive result. Given the dense network, the danger of observational error correlation certainly exists. However, the characterization of observation error correlation is beyond the scope of this study. To our knowledge, this point has not yet been specifically addressed for ozone observations, and such a study should be performed in a separate work. If validation stations are too close to assimilation stations, such a potential correlation can also be an issue for the evaluation of the system. Nevertheless, we follow the common way to evaluate operational air quality systems so this way of doing things is consistent with our goals. To complete this, we can add that our results indicate that

C1746

observations belonging to separate classes do not seem to be correlated; for instance the assimilation of only suburban stations improves the suburban stations but not other classes (not shown). Considering a single class of station type the density is lower (for example mean distance of 70 km between pairs of suburban stations and 112 km between pairs of remote/rural stations) reducing partly this issue of correlations. These elements are only indications and a robust answer to this question, again, would be given in a specific study.

The model and the assimilation show greater errors during the night. It could be related to the vertical resolution of the model. This aspect is hardly discussed in the paper. The strong vertical gradients of ozone during the night might not be resolved. This has consequences for the vertical representativeness of the observation during the night. Answer: It is true that the number of vertical levels does not allow resolving the strong ozone gradient during the night. The analysis is only performing in the boundary layer, this vertical localization leads to an assimilation of nighttime observations generally only in the first or the two first model levels. One conclusion of the paper is that representativeness issues are the main error sources in this assimilation system, particularly at night. It is rather impossible to distinguish between horizontal and vertical representativeness errors. Also following some specific comments below, elements concerning the vertical representativeness have been added.

Please differentiate more clearly between the “model error” (no assimilation) and the “background error” (forecast started from the previous analysis). The text would have greatly benefited from further proofreading both for the English language, consistency of wording and structure of the document. For example, headings 4.2 and 5.1 have the same title “Evaluation of reference simulation”. Answer: the term “model error” has been replaced by “background error” all along the paper. Headings have been changed. The term “model error” had been used because most of background errors are due to an error growth during the forecast step (i.e. the model error) and because the tuned covariance inflation factor is used to model this error growth. 2 Specific com-

C1747

ments P 3034 L1 The abstract should avoid general statements but should clearly state the purpose of the paper and summarise methods and results. I would omit the first sentence but say what the purpose of the study is. L7 “quadratic error” or RMSE ? – the latter is more common L9 please make clear that you use an existing method for the classification L26, “which increases : :” omit – it is obvious and the relation is not linear over all possible ozone concentrations. P 3034: L1 The abstract has been reformulated following the referee’s recommendations. L7 The RMSE terms has been chosen. L9 The reference to the Flemming et al. (2005) is now clearly stated: “Based on statistics from Flemming et al. (2005), observation sites of the European ozone monitoring network have been classified using criteria on ozone temporal variability.” L26 the word is omitted. P 3035 L1 provide reference for GEMS, L3 “ and in-situ ..” L6 delete “the” before “monitoring” L10 What is a “modelling platform”, please clarify , Do you mean a model ? L16 please clarify that by “analysis” you mean the result of a data assimilation method in the remainder of the test L23, why are the analyses “a fundamental result” – Consider reformulating the whole sentence P 3035: L1 The reference for GEMS is given at the end of the sentence (i.e. Hollingsworth et al. 2008). L3, L6, done. L10 A modelling platform is an operational software which performs model simulations and gives a post-processing of the results; observations are generally employed as soon as they are available. L16 The analysis always refers to a result of a data assimilation system. L23 The sentence is now: “In addition, analyses resulting from data assimilation of observations in near real time provide the best representation of the surface pollutant concentrations.” P 3036 L16 please clarify the difference between model error statistics and background error statistics L28 There are variational methods (4DVAR) and sequential Extended Kalman Filter methods. The Ensemble Kalman filter is just one example of the possible Kalman Filter approaches. Both methods have been used to correct emission rates. Brunner et al. 2012 and Miyazaki et al. 2012 are important examples which should be mentioned in the paper. P 3036: L16 In an ensemble Kalman Filter algorithm, the analysis error statistics are given through the analysis procedure. During the forecast step, errors are increased due to the model

C1748

error. The background errors are due to the model error and the analysis error of the previous time step. We employed a covariance inflation factor which consists in the perturbation of the ozone initial condition in order to finally get a suitable background error. But considering the low analysis error statistics, the background error is mainly increased by the presence of model error. The following sentences help to clarify: “As mentioned above, with finite and generally small ensemble sizes, and due to significant model errors, the EnKF estimate of the BECM generally underestimates the true analysis error covariance matrix. In addition, errors are increased due to the model error during the forecast step. Thus, a particular strategy must be employed for inflating the model error term Q in Eq. (3) by treating model and analysis errors contributions in the same framework. Finally, the ensemble design must reflect the background error and generate adequate error correlations.” L28 The sentence has been corrected: “In order to obtain an accurate 4D-analysis of ozone concentrations, a strategy consists in the correction of indirect or possibly unobserved quantities such as emissions rates or even wind fields using variational (Elbern et al. 2007, Semane et al. 2009) or sequential methods (Brunner et al. 2012, Miyazaki et al. 2012) such as Kalman Filters.” P 3037 L18 It would be good to mention the purpose of the study much earlier, i.e. at the beginning of the introduction. P 3037: In the new version of the paper, this purpose is exposed more clearly in the abstract section. P 3038 L2 Please clarify if you mean all Extended Kalman Filters or just the specific Ensemble Kalman Filter implementation. L14 Clarify what is x in your study, only the surface ozone field? Eq 2 for Pf is not a notation that leads to the covariance matrix. It is only the variance. One has to use the transpose and/or a second index.

P 3038: L2 The description is dedicated to the Ensemble Kalman Filter, although some equations are common with Extended Kalman Filters. L14 The following sentence has been added: “Here $x_{i,k}^f$ represents the vector of forecasted ozone concentrations”. The transpose index has been added in the equation.

P 3039 Eq 2 to 4 are the equations for the optimum interpolation. The forecast of the

C1749

state vector and P with the model M are important steps of any Kalman Filter algorithm. If the forecast of P with the model M is not part of your EnKF, please make this clear. The forecast of P usually contains a noise term (Model error), which makes sure that P does not deflate. L22. In all data assimilation approaches the observations have a random error – please clarify what you mean.

P 3039: Description of the noise perturbations and equations are added to illustrate the estimation of Pf, the full model is used for each ensemble member and Pf is calculating from the ensemble covariance: “An ensemble of 20 perturbed model states is created using Monte-Carlo methods and evolves forward in time in order to obtain a forecast $x_{i,k}^f$ from the time step (k-1) to the time step k.

$$x_{i,k}^f = M(x_{i,k-1}^a + q_{i,k-1})x_{i,k-1}^a \text{ where } q_{i,k-1} = sd * \eta_{i,k-1}(1)$$

$\eta_{i,k-1}(1)$

Here $x_{i,k}^f$ represents the vector of forecasted ozone concentrations where the subscript i indicates the ensemble number and M is the CHIMERE model. The noise is derived from a two-dimensional Gaussian distribution with some fixed spatial characteristics namely zero mean and unitary variance (Evensen 1994, 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 coefficient of relative standard deviation (sd). The ensemble mean value over the N ensemble members is defined in

C1750

Eq. (2):

$$\overline{x_k^f} = 1/N \sum_{i=1}^N x_{i,k}^f \quad (2)$$

At the analysis step, the BECM is approximated by the ensemble spread over the N realizations of the model at a given time:

$$P_k^f = 1/N \sum_{i=1}^N (x_{i,k}^f - \overline{x_k^f})(x_{i,k}^f - \overline{x_k^f})^T \quad (3)$$

L22, it means that observations are perturbed explicitly in the procedure. P 3040 L2 unclear sentence, also please correct “measurement’s perturbations” L4 again, this a property of all Extended Kalman Filters. Do you use Ensemble Kalman Filter as synonym for Extended Kalman Filter?

P 3040: Although only the random part of the observational error is taken into account, contrary to other algorithms, observations are not explicitly perturbed in our case. This subsection is dedicated to discuss among others differences in analysis algorithm in Ensemble Kalman Filters, although there are some similarities and common equations between Extended and Ensemble Kalman Filter. P 3041 L5 Please explain what the Desroziers Diagnostics tell us, i.e. what they mean. L10 What “assimilation exercises”? Please provide reference. It is not clear in the whole section if you make statements based on references in the literature, or if you want to test these assumption later in the paper. L16 How do you know it is reliable for a dense network, please provide reference.

P3041: L5 Two sentences to better explain Desroziers diagnostics have been added into the text: “The difference in these two equations is represented by the first term of the product. Diagnosed background or observation errors will be small if the analysis

C1751

estimate is close to the background or observations, respectively.” L10 These assimilation exercises have been made in Li et al. 2009b, to make this clear the new sentence is: “The ability of the tuned ensemble to represent more accurate error statistics has been demonstrated in assimilation exercises that take into account different ranges of true model errors (Li et al. 2009b).” L16 We consider that the background error can be sampled in observations space only in a case of a sufficiently representative network. This has been shown for the estimation of background error correlation in Schwinger and Elbern 2010. P 3042 L14 It is not clear how the classification (either based on observation time series or meta data) is used to quantify the representativeness, which is needed for data assimilation. A rural station might have a larger area of representativeness but the actual values (100, 50 or 10 km?) are not provided by the classification. L26, Why are you using a new label for the air-quality regimes in FLEM05. Later in the text you only use the term “urban”, “rural” etc. Please be precise and consistent.

P 3042: L14 These classifications are only used to qualify pollution regimes, they cannot directly provide the representativeness distances. Only in the study of Henne et al. (2010) they estimate a spatial footprint of the observations by the use of a transport model, but such an approach is out of the scope of this paper. The representativeness area which is targeted in data assimilation is the size of the model grid cell (which is around 50km in this study) and is therefore relative to the model configuration. However, in this paper, we get indeed an estimation of the representativeness error by the use of a posteriori diagnostics. L26 The labels for stations types have been homogenized in the whole document using the terms remote, rural, and suburban. P 3043 L2 Do you calculate P50DA and P50DV from the annual values or only for the JJA season (Table 1 seems to indicate that). L6 all urban stations according to FLEM05 or to Airbase? How do you justify this assumption that they are not representative for the 0.5 x 0.5 grid. The urban emissions should be part of the emissions in any grid box containing urban stations. L18 what “variability” – daily variability??

P3043: L2 the two indices (P50DA and P50DV) are calculated for the entire summer

C1752

(JJA) period. In L6, we mean urban stations according to FLEM05 classification. Urban stations are not representative for the 0.5 x 0.5 grid, because of the large variability of emissions inside a grid cell at urban locations. For the case of NO emissions, this directly affects ozone concentrations via the O₃+NO titration reaction. Thus, within urban locations, ozone is expected to show a large variability within a grid cell, and thus urban sites are not representative for the whole grid cell. L18 Here, we mean daily variability (the text has been corrected).

P 3044 L15 not sure what GEMS means here. Please provide vertical extent of the eight layers in m. In particular the height of the surface layer is important for the interpretation of the study L18 please rephrase “mandatory” L20 “Analyses” use plural L24 Is it MOZART 3.5 ? please double check, also the reference for MOZART 3.5.

P3044: L15 GEMS is the name of the domain, it has been removed. The heights of the top of the vertical layers are given: “Our simulations cover the European continental domain (Fig. 1) for 8 hybrid (σ , p) vertical levels from 995hPa to 500hPa (the height at the top of each box is on average: 42m, 115m, 240m, 455m, 838m, 1520m, 2820m and 5500m). L18 The word ‘mandatory’ has been removed. L24 The reference for MOZART 3.5 is now Kinnison et al., 2007.

P 3045 L24 Is it a temporal correlation of the hourly values - please specify. L27 Please discuss this also w.r.t to the vertical resolution of the model, i.e. the depth of the surface layer.

P3045: L24 Indeed, we mean temporal correlation of the hourly values. L27 The sentence is now: “These errors can be typically attributed to the limited model resolution in both horizontal and vertical direction. It does not allow a good estimation of the subgrid processes such as vertical turbulent transport and spatial variability of anthropogenic emissions (Valari and Menut, 2010).” P 3046 L5 Why did you not choose an appropriate model level for the mountain stations? Would have a different level with a better match during the night led to worse results during the day? L8 Please consider also

C1753

the NO gradients in the nocturnal PBL. L19 -24. I believe this short introduction of the chapter is not helpful and can be deleted. The titles of the 2nd level headings should be clear enough. Please keep explanation of Table 2.

P3046: The ozone and the averaged diurnal error have been calculated for both the first and the second levels and are added in figures 3 and 7 of the new document. Observations are better simulated for the second level, particularly in nighttime. Only slight differences are found in the daytime, but the second level is still better. The following sentences constitute the end of the paragraph: "Simulation of remote stations above 300m a.s.l. exhibits a much stronger positive bias during nighttime, as these stations are generally more representative of higher model layers which are less affected by dry deposition and NO emissions. Then, we also plot on Fig. 3 the average simulated ozone values and RMSE for the second level which corresponds to a height varying between 115 m and 240m. It shows that nighttime values are better simulated with the second level for both the ozone mean and RMSE." L8 Done. L19-24 The short introduction has been deleted.

P 3047 L1 Please check the title "Evaluation of the reference :" – it should be "setup of assimilation experiment " L5 - It would be good to give some indication of the observation density, i.e. something like the average distance to the nearest neighbor station, both for the total observation set and the split one. Even after splitting in the observations in evaluation stations and assimilating stations one can not assume (as it is sometimes implicitly done) that the evaluation stations are truly independent of the assimilated stations. L13. The PBL heights (I suggest to use this term rather BLH) are a crucial point in your method and needs more explanation. How is PBL height diagnosed? How is it defined during night time conditions, when there is no mixed layer? What happens if the PBL height is smaller than the height of the surface level? L14 please correct "tri-dimensional" L17 please give value (in ppb) of the magnitude of the added noise, How does it compare to the observation error standard deviation.

P 3047: L1 The title has been modified to "setup of assimilation experiment". L5 The

C1754

following sentence has been added: "Finally the average nearest distance is around 37km for the entire set of selected observations and around 61km for the subset of assimilated observations." L13 The PBL heights parameterization is described in Menut et al. (2013), for the stable case, it is diagnosed following a K diffusion approach (Troen and Mahrt 1986) and a thermal plume approach is used for an unstable case (Cheinet and Teixeira 2003)." During the analysis step, the selected levels are (in the vertical localization) up to the one just above (i.e., that contains) the PBL height. If the PBL height is lower than the first level, observations are therefore assimilated only in the surface level. L14 this has been corrected. L17 The magnitude of the added noise is close to the noise standard deviation, the value (around 5 ppb) has been added in the text. P 3048 L9 Above you say that the classification of the station regime is used to define different representativeness error – now all stations get the same value (5pp). It contradicts a bit the statement before. Do you distinguish between observation error (instrument error) and representativeness error. What is the assumed resolution for the estimate of the representativeness error in Flemming (2004). In your case you also need to discuss the representativeness in the vertical direction, in particular during the night. The vertical ozone gradient can be quite pronounced. The vertical representativeness also raises the issue if the error is only "random". I would argue it is more bias, which needs to be corrected before the assimilation. L27. The caption of the Figure say 3.00 UTC – here you say 15 UTC!

P 3048: The classification is used to qualify the representativeness of observations. It allows the use of part of urban observations from the Airbase database, and to reject those labeled "urban" in the new classification. From the existing literature, no clear values for the representativeness errors for different types of stations could be obtained (some studies have been presented in section 5.1). In Flemming et al. (2004), a median estimation of an observation error standard deviation of 5 ppb is found for the different types of stations using the indirect Hollingsworth and Lönnberg method at 0.25° resolution. In this way, all random errors (representativeness and instrument) are considered. Here, we estimate these values from a posteriori diagnostics in section

C1755

5.4. The vertical representativeness is also unknown, to avoid this problem, we use a severe localization which consists in correcting only levels that contains the PBL. Finally, observations are generally assimilated only in the first level during nighttime due to the low PBL height. It cannot be excluded, that these errors are not only random, and lead to a bias. However, we do not dispose of any observations allowing confirming or even quantifying such a bias. L27 Corrected. P 3049 L1 What about the low values over the North Sea in the analysis. Are they realistic or a consequence of the extrapolation of urban stations in the UK? Night time values over the sea are in general higher than over land because of the reduced dry deposition. L8 How far away is the nearest assimilated station to Odense, what type is it ?

P 3049: L1 The figure 6 show the simulated and analyzed values for the 20th August at 15 UTC. At this time, low values are associated to a cold front that rapidly moves eastward. Comparison with surely only few observations in the UK (but all observations types are represented) show a general improvement. L8 The closest station near Odense is also an evaluation station (DK0054A, at 76km distance), its type is background remote (in our classification). The nearest station used for assimilation (situated at 85km) is a rural (in our classification).

P 3049 L18 I am not convinced that OI (i.e. a Kalman Filter run without update of P) which includes a model would not be able to transport the information from the observations to other area. Is there evidence that the change over the North Sea comes from the analysis step in the KF and not from the model forecast started from an analysis. (The latter would also work with OI). Finally, as said above, it is not clear if the changes over the North Sea are actually an improvement.

P 3049: It is true that changes in the North Sea away from coastal stations are due to the transport during the forecast step. The sentence has been corrected, it is now: "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." P 3050 L7 "error

C1756

profile" is perhaps misleading, "diurnal cycle of error" is a better way of calling what is shown in Fig. 5 L7 "globally" ? use "overall" or else L15 It is not the "model error" but the something like the "background error covariance description" L20 again it is not the "model error" but the "background error", i.e. the error of the model started from the analysis one hour earlier L24 a bit more detail of what the RCRV is required at this point (or later)

P 3050: L7 corrected. L15/L20 The term "Model error" is replaced by "background error". L24 An introduction to the RCRV as well as other diagnostics is presented in a dedicated section (2.2) called "A posteriori diagnostic and error modeling" in the new version. P 3051: L1 it would be interesting to see graph of the original BECM diurnal cycle (perhaps the diagonal elements average) and the modified one of MOD_DESR. One would like to know if the BECM changes a lot or only a little. L10 The following discussion seems interesting but is difficult to follow. Perhaps you should provide formulas for "diagnosed errors", "ensemble standard deviation" – is it before or after the analysis step. L12 "model error" ? Do you mean "background error" ? L13 "shape of the diagnostics" more realistic than the "model error" – I don't understand this L16 Please use a consistent Figure order in the text. L19 Do you want to say that the BECV of MOD_DESR are better (because the analyses have a smaller error?) during evening and morning than the standard case. If yes, are the analysis errors smaller during this time?

P 3051: L1 The original (REF_ASSIM) background error standard deviation is plotted on Fig. 8 (left panel). L10 Explanation and references to equations have been added. The ensemble standard deviation is calculated before the analysis step. L12 it is background error. L13 the text is now: "The comparison with the forecast RMSE indicates that the daily variation of the background error diagnostic is now more realistic than that of the initial background error." L16 the order of figure has been modified, it reflects the chronology as closely as possible. L19 The background error is more appropriate during the daytime, this statement is demonstrated thanks to the use of

C1757

the RCRV diagnostics. P 3052 L8 Please mention in the text what parameters you perturb. Reference to the supplement is no sufficient. Please include the table in the main paper. L9 Could not find reference for Hanna et al. 2001 L26 – please remove brackets. It would be better to introduce RCRV this point and not already on page 3050 without explanation

P 3052: L8 The list of parameters and the associated table are added in the new section (2.2, “A posteriori diagnostic and error modeling”). L9 the references are added. L26 corrected

P 3053 L11 – please check sentence. L12 Do you mean that biases are not taken into account ? L19 “fading” ?, now you use the term PBL without explanation , before it was BLH.

P 3053: L11-19 corrected. L12 Biases are not explicitly taken into account. I mean that the EnKF algorithm is supposed to correct random errors. P 3054 L9 “error budget” perhaps better “behaviour of error statistics” L21 I don’t understand this. You said the spread of the ensemble needs to be increased before, now it is “preferable to reduce the perturbations”

P 3054: L9 Done. L21 Because of the existing model error, perturbations are necessary. However, it is preferable to avoid large perturbations which accumulate in regions without assimilation stations. Moreover, a posteriori diagnostics suggest that model errors were overestimated in the daytime.

P 3055 L1 If you change the observation error for the assimilated stations, this also needs to be reflected for the evaluation. Hence, a larger observation error (that is what we assume to be the correct value) should also mean that a certain RMSE or bias of the analysis is less problematic. One can not conclude that a smaller analysis error is a confirmation that the chosen observation error is more correct. L6 please discuss also the vertical representativeness L28, RMSE, bias, Correlation are no “skill scores” - they are accuracy measures. Skill scores compare accuracy measures against a reference

C1758

(see the textbook by Wilks “Statistical Methods in the Atmospheric Sciences”)

P 3055:L1 It is only a change in our estimate of observations errors statistics, so it should affect the analysis error. However, the real observation error keep unchanged, thus a better RMSE with respect to observations means a better analysis. L6 It has been changed to “This latter error depends on the horizontal and vertical model resolution.” L28 It has been changed to accuracy measures P 3056 L1 I believe the little impact can be explained with the large number of assimilated observations. They dominate the analysis and the error description is of minor importance. L2 The evaluation stations are not independent of the assimilated stations because they come from the same network. A smaller analysis error (if we assume the observation error is larger) is not an indication that the analysis is better.

P 3056:L1 Yes the authors agree with this statement, it explains why errors are decreased only against background (isolated) stations. L2 The evaluated observation error is close to the RMSE between analyses and evaluation stations. It thus suggests that these values are on average the lower bound of analyses RMSE in our spatial resolution.

P 3057 L10 Again the numbers of the assimilated stations are high. One can perhaps not expect too much. L13 “new” I understand that the Kalman Filter was already developed and used in Coman et al. 2012. So please be specific about what is new. L13 “chain” is unclear L22 Please discuss possible causes for the night-time problem. L25 Please mention how many observations sites you assimilate and what the average density of the assimilated stations is.

P 3057: L10 some elements have been added in the conclusion to this effect. L13 the terms “new” and “chain” have been removed. The sentence is now “In this paper we present a data assimilation system based on the rCTM CHIMERE in an EnKF framework and using surface ozone observations provided by the European Airbase database.” L22 We added the following sentence: “This error cycle is caused by the

C1759

reduced spatial representativeness due to nighttime physical processes such as vertical mixing which are not resolved at the chosen model resolution of 0.5 degrees.” L25 Done P 3058 L9 Please discuss the differences between your approach and Hanea et al. 2004 in more detail. Both papers assimilates surface ozone in Europe using a Kalman Filter. This should also be mentioned in section 1 (Introduction) L13 The sentence “Considering ..” does not make sense to me. L15 please describe the RCRV briefly with half a sentence or so. L24 How were the observation errors estimated? Give value of 5 ppb for the standard deviation. L24 Please mention that the RCRV diagnostics have indicated higher errors for the rural stations. This could be an important finding of the paper L27 Again they are no “skill scores” but accuracy measures. See textbook by Wilks.

L24 P 3058: L9 The following sentence has been added in the conclusion: “These improvements are similar to EnKF analyses performed in Hanea et al. (2004) where background errors were estimated by the correction of the LOTOS-EUROS model parameter.” L13 the sentence has been removed L15 The following sentence is added “By the comparison of the innovations statistics and prescribed errors, RCRV statistics allow the evaluation of both weighted bias and errors prescription.” L24 It is diagnosed a posteriori. The sentence is now “Generally, the diagnostic indicates a large contribution of the observational error, higher than expected especially for rural stations.” L27 all the terms “skill scores” have been removed.

P 3059 L7 Not just the robustness of the system but the very high number of stations. If you reduce the stations numbers you will see a much larger effect of the error statistics on the results. L13 Please mention references of papers on this topic. (Brunner et al, 2012, Miyazaki et al. 2012) L18 Is it also used for NRT forecasting? It would be interesting to discuss this in the paper.

P 3059:L7 the sentence is now “Stations used for assimilation or evaluation are spatially close, where the ozone observations network is spatially dense. Thus, in terms of evaluation scores, only small changes in performance statistics are found among ex-

C1760

periments even for substantial changes (up to a factor of two) in the model and observation errors. This suggests that the high number of observations exerts a significant constraint on the analyzed fields. 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.”L13 References have been added. L18 No, the system has not been used yet for NRT forecasting. Table1 Please spell out MOU, RUR, SUB. It is confusing that you introduce new abbreviations for the FLEM05 classification, and sometimes only the adjectives “rural” etc. Please be consistent throughout the paper. Are the mean and P50DV valid for the whole year or only JJA?

Table 1: Done, it is only for JJA.

Table 2 Spell out OECM and PECM What means fixed (= constant)? Why profile ? Provide indicative value in ppb for all variances.

Table 2: The range of error standard deviation has been added for OECM. However, these hourly average profiles are shown on Fig. 8 and Fig. 11.

Table 3 Not skill scores – it is better called accuracy measures

Tables 3 and 5 and also 4 should be merged into on table.

Tables 3/4/5: Done

Table 4 – why no discrimination of different regimes?

Table 4: Done

Table 5 “Accuracy measures” - Explain MOU, RUR SUB etc.

Table 5: Done

Fig1 I found it difficult to see the cyan square for Odense.

Fig1: Done

Fig2 Explain MOU, RUR, SUB

C1761

Fig2 : Done

Fig3 Is the top CHIMERE identical to Fig 2 ? Perhaps Fig 2 can be omitted. Why is SUB obs in red and the rest not ? Mention length of assimilation period.

Fig3: Done

Fig 4. What is “prescribed noise”. Green and blue colours are difficult to discriminate. Consider changing the colours.

Fig 4: Done

Fig 5. It seems that some of it is already shown in Fig. 3.

Fig 5: Done

Fig 6. The shapes are difficult to distinguish. The text says the time is 15 UTC.

Fig 6: Done

Fig 9. Please spell out RCRV. Fig 9: Done Cheinet, S. and Teixeira, J.: A simple formulation for the eddy diffusivity parameterization of cloud-topped boundary layers, *Geophys. Res. Lett.*, 30, 1930, doi:10.1029/2003GL017377, 2003.

Flemming, J., Van Loon, M. and Stern R.: Data assimilation for CTM based on optimum interpolation and Kalman filter, In: Borrego, C., Incecik, S. (Eds.), *Air Pollution Modelling and its Application XVI*. Kluwer Academic/Plenum Publishers, New York, pp. 373–383, 2004.

Li, H., Kalnay, E. and Miyoshi T.: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society* 135(639): 523-533, 2009b.

Liu, Z.-Q. and Rabier, F.: The interaction between model resolution, observation resolution and observation density in data assimilation: A one-dimensional study. *Q.J.R. Meteorol. Soc.*, 128: 1367–1386. doi: 10.1256/00359000232, 2002.

C1762

Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J.-L., Pison, I., Siour, G., Turquety, S., Valari, M., Vautard, R., and Vivanco, M. G.: CHIMERE 2013: a model for regional atmospheric composition modelling, *Geosci. Model Dev.*, 6, 981–1028, doi:10.5194/gmd-6-981-2013, 2013.

Schwinger J. and Elbern H.: Chemical state estimation for the middle atmosphere by four-dimensional variational data assimilation: A posteriori validation of error statistics in observation space, *Journal of Geophysical Research* 115(D18): D18307-D18307, 2010. Troen, I. and Mahrt, L.: A simple model of the atmospheric boundary layer: Sensitivity to surface evaporation, *Bound.-Lay. Meteorol.*, 37, 129–148, 1986.

Wu, L., Mallet, V., Bocquet, M., and Sportisse, B.: A comparison study of data assimilation algorithms for ozone forecasts, *Journal of Geophysical Research*, 113 (D20310), doi:10.1029/2008JD009991, 2008.

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

C1763