Geosci. Model Dev. Discuss., 5, C932–C940, 2012
www.geosci-model-dev-discuss.net/5/C932/2012/

**Geoscientific
Model Development
Discussions**

Interactive comment on "Quantifying the model structural error in carbon cycle data assimilation systems" *by* S. Kuppel et al.

S. Kuppel et al.

sylvain.kuppel@lsce.ipsl.fr

Received and published: 21 November 2012

The authors would like to thank Ian Enting for the very helpful comments and suggestions. The comments have been taken into consideration in the revised manuscript. We answer all of them individually in the following.

Questions about methodology

The assumption that error is stationary in time (i.e. characterized by a single lag time) is questionable for a system dominated by strong seasonal variation. The authors should at least acknowledge (and maybe briefly discuss) the limita-

tions/implications of this assumption.

The last sentence of the last paragraph in section 2.3 has been modified to:

'Appendix B describes the application of both diagnoses at each flux measurement site, where for simplicity and to increase statistical significance we consider errors to be stationary in time, i.e. characterized by a single time lag. We acknowledge that this is a strong assumption for modeled carbon fluxes affected by significant seasonal variations, and this limitation should be kept in mind throughout the analysis of the results.'

P2271, L 21: As one of the authors of the Kaminski et al paper, I don't understand the claim of 'equivalence'. Kaminski et al are discussing a truncation of the 'model space', while this paper is based on a truncated observation space.

This sentence was meant to point out the analogous nature of these two errors in terms of truncation, although indeed the truncated spaces are different. In the revised manuscript, we have modified the sentence in order to be clearer:

'This term is analogous to the aggregation error that has been rigorously described in atmospheric inversions (Bocquet et al., 2011; Kaminski et al., 2001; Thompson et al., 2011), in that it arises from truncating a given space of variables.'

Presentation of notation

This is a demanding body of work, and to help the reader I would suggest additional assistance with the notation involved in the estimation.

A table of notation would seem to be appropriate. I give a few possible comments below (assuming that my interpretation of the paper is correct).

We have added the following sentence in the revised manuscript at the beginning of section 2.3:

'In this study, notations for vectors and matrices are according to the following convention : subscripts refer to contributions (e.g., model, measurements), superscripts refer to contexts (e.g., prior state, posterior state), and hats ('$\hat{}$') refer to estimates.'

In the figure captions, giving the mathematical expression of what is being plotted (as well as the verbal description) would be helpful.

The figures captions have been changed as follows:

'Figure 1. All-site median of the autocorrelation of the residuals from the prior ORCHIDEE model (black, corresponding to $\mathbf{D}$ in Eq. 1) and of the prior-parameter error projected in the flux space (purple, corresponding to $\mathbf{HBH}^T$ in Eq. 1), as functions of the time lag, for daily NEE.'

'Figure 2. All-site median of the autocorrelation of the observation error (i.e., model error + measurement error) $\mathbf{R}$, estimated at each site with three methods: prior diagnostics with the linear assumption (orange, $\hat{\mathbf{R}}^{prior}$ from Eq. 1), prior diagnostics with ensemble simulations (blue, $\hat{\mathbf{R}}^{prior}$ from Eq. 1), and posterior diagnostics (grey, $\hat{\mathbf{R}}^{post}$ from Eq. 3).'

'Figure 3. Distance correlogram of the observation (model+measurement) error $\hat{\mathbf{R}}^{prior}$ estimated from Eq. (1), using pairs of distant sites for a same time. The value represented by each blue diamond includes all the common years of one site pair. The thick black line represents the overall median using 400-km bins, and the dotted line an exponential decay with an e-folding length of 500 km.'

My understanding is that the b superscript, used for the 'prior', comes from Desroziers et al. where it refers to 'background' (i.e. the forecast from the previous cycle). The authors should note this meaning, or maybe consider whether symbol this is appropriate given that the calculation is a 'batch' calibration of the model, rather than (cf CarbonTracker) an assimilation system where the 'state' is being progressively updated with time.

In the revised manuscript, we have replaced the 'b' superscript by the more general 'prior' designation and the 'a' superscript by the more general 'post' (for posterior) : $\mathbf{x}^b \rightarrow \mathbf{x}^{prior}$, $\mathbf{d}^{o-b} \rightarrow \mathbf{d}^{o-prior}$, $\mathbf{x}^a \rightarrow \mathbf{x}^{post}$, $\mathbf{d}^{o-a} \rightarrow \mathbf{d}^{o-post}$. It is also now more consistent with notation used for the $\mathbf{R}$ matrix (e.g. $\hat{\mathbf{R}}^{prior}$)

P2264, L11: 'prior residuals' → 'residuals from the prior model' P2264, L11: more seriously, $\mathbf{D}$ (as expressed by eqn (1)) is the covariance for the distribution of these residuals – the covariance of the actual set of residuals is (or can be) only an estimate of this distribution.

The sentence has been modified in the revised manuscript, it is now:

'. . . and $\mathbf{D}$ is the covariance matrix for the distribution of the residuals from the prior model $\mathbf{d}^{o-prior}$ (i.e., the observation-minus-model mismatch), defined as. . . '

P2264, L20: shouldn't this be: 'to directly derive an estimate of R' and then have Eqn (3) as something like $\hat{\mathbf{R}}^{empirical} = \mathbf{F}$ (Clearly, cf P2269, L 12, the authors understand this distinction, but to help others, they need to be more thorough in distinguishing (unknown) distributional quantities from the estimates of such quantities.) P2264, L19: Shouldn't the parameters (that are to be optimized) be denoted $\mathbf{x}$. (with $\mathbf{x}^b$ being the (prior) estimate).

These two remarks have been taken into account; the sentence and Eq. (3) are now:

'An additional diagnosis makes use of the optimized model state, i.e. the NEE fluxes after the optimization of the model parameters $\mathbf{x}$ *(see the inversion procedure in Sect. 2.4) to directly derive an estimate of* $\mathbf{R}$ *(Desroziers et al., 2005):'*

$$\hat{\mathbf{R}}^{post} = \mathbf{F}$$

Also, if the superscripts a, b, o are abbreviations for words 'analysis', 'background', 'observations' rather than mathematical variables, then these superscripts should be in an upright font to distinguish them from variables.

These superscripts were modified from upright to italic font by the automatic typesetting procedure, according to the rules of Copernicus Publications. We will try to emphasize this point during the submission of the revised manuscript.

Since the Desroziers paper is quite complex, and the present paper only uses one aspect of that analysis, it may be worth summarising (in Appendix A) the relevant relation(s) from Desroziers to show why $\mathbf{F}$ **gives an estimate of** $\mathbf{R}$**.**

We have added to revised manuscript a new Appendix, which demonstrates Eq. (3):

*'**Appendix A: Demonstration of Eq. (3) based on Desroziers et al. (2005)***

Using the common linear assumption, the optimized state xpost can here be decomposed as follows:

$$\mathbf{x}^{post} = \mathbf{x}^{prior} + \delta\mathbf{x}^{post} = \mathbf{x}^{prior} + \mathbf{KHd}^{o-post}, \quad (1)$$

where the optimization increment $\delta\mathbf{x}^{post}$ *can be expressed from the residuals of the*

prior model $\mathbf{d}^{o-prior}$ *using the gain matrix of the optimization* $\mathbf{K}$ *(Talagrand, 1997):*

$$\mathbf{K} = \mathbf{BH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1} \quad (2)$$

Using the relation of Eq. (A1) in Eq. (4), do-post is given by

$$\begin{aligned} \mathbf{d}^{o-post} &= \mathbf{y}^o - H(\mathbf{x}^b + \mathbf{Kd}^{o-prior}) \\ &= \mathbf{y}^o - H(\mathbf{x}^b) - \mathbf{HKd}^{o-prior} \\ &= \mathbf{d}^{o-prior} - \mathbf{HKd}^{o-prior} \\ &= (\mathbf{I} - \mathbf{HK})\mathbf{d}^{o-prior} \end{aligned}$$

where $\mathbf{I}$ *is the identity matrix. Then, the use of Eq. (A2) gives*

$$\begin{aligned} \mathbf{d}^{o-post} &= \mathbf{I} - \mathbf{HBH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1}\mathbf{d}^{o-prior} \\ &= ((\mathbf{HBH}^T + \mathbf{R})(\mathbf{HBH}^T + \mathbf{R})^{-1} - \mathbf{HBH}^T(\mathbf{HBH}^T + \mathbf{R})^{-1})\mathbf{d}^{o-prior} \\ &= \mathbf{R}(\mathbf{HBH}^T + \mathbf{R})^{-1}\mathbf{d}^{o-prior}, \end{aligned}$$

then the covariance between the distribution of do-post and do-prior, F, can be expressed as

$$\mathbf{F} = \mathbf{R}(\mathbf{HBH}^T + \mathbf{R})^{-1}\mathbf{D},$$

which, from Eq. (1), simplifies to'

$$\mathbf{F} = \mathbf{R}.$$

My understanding of the notation

$\mathbf{R}$ **: Covariance matrix for observation error. Subscripts refer to contributions to R, superscripts refer to contexts, "hat" refers to estimates.**

This interpretation is correct. As mentioned earlier in this document, we have added a presentation of the notations at the beginning of section 2.3.

Other comments

P 2261, L 1. My understanding is that the information from the model equations is spread in space as well as time.

Indeed, and we have added to this the revised manuscript:

'Their model prognostic equations can also spread the observational information well beyond the temporal and spatial cover of the measurements (Scholze et al., 2007; Rayner et al., 2011).'

P 2261, L2-5. This sentence is somewhat clumsy and might benefit from being restructured (and perhaps noting that the 'strong constraint' approach is a choice, not a necessity).

The sentence has been modified, it now reads:

'Indeed, the choice of a process-based approach imposes these equations as a strong constraint to the inversion, even though they are an imperfect representation of the biophysical and ecophysiological mechanisms that drive terrestrial ecosystems.'

Minor points

P 2261, L 6: replace 'Bayes theory' by 'Bayes theorem' OR 'Bayesian estimation'

'Bayes theory' has now been replaced by 'Bayes theorem'

P 2261, L 25: parameters → parameter

It has been corrected

P 2266, L14: 20121 → 2012 (as date for Crisp et al paper)

It has been corrected

P 2273, L10: 'occult → 'fail to capture'

It has been corrected

P2273, L 23: 'all the more that' → 'particularly since'

It has been corrected

References

- Bocquet, M., Wu, L., and Chevallier, F.: Bayesian design of control space for optimal assimilation of observations. Part I: Consistent multiscale formalism, Quarterly Journal of the Royal Meteorological Society, 137, 1340-1356, Doi 10.1002/Qj.837, 2011.
- Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society, 131, 3385-3396, Doi 10.1256/Qj.05.108, 2005.
- Kaminski, T., Rayner, P. J., Heimann, M., and Enting, I. G.: On aggregation errors in atmospheric transport inversions, Journal of Geophysical Research-Atmospheres, 106, 4703-4715, 2001.

- Rayner, P. J., Koffi, E., Scholze, M., Kaminski, T., and Dufresne, J. L.: Constraining predictions of the carbon cycle using data, Philosophical Transactions of the Royal Society a-Mathematical Physical and Engineering Sciences, 369, 1955-1966, Doi 10.1098/rsta.2010.0378, 2011.
- Scholze, M., Kaminski, T., Rayner, P., Knorr, W., and Giering, R.: Propagating uncertainty through prognostic carbon cycle data assimilation system simulations, Journal of Geophysical Research-Atmospheres, 112, Artn D17305, Doi 10.1029/2007jd008642, 2007.
- Talagrand, O.: Assimilation of observations, an introduction, Journal of the Meteorological Society of Japan, 75, 191-209, 1997.
- Thompson, R. L., Gerbig, C., and Rodenbeck, C.: A Bayesian inversion estimate of N(2)O emissions for western and central Europe and the assessment of aggregation errors, Atmospheric Chemistry and Physics, 11, 3443-3458, DOI 10.5194/acp-11-3443-2011, 2011.

Interactive comment on Geosci. Model Dev. Discuss., 5, 2259, 2012.