Geosci. Model Dev. Discuss., 7, C3364–C3366, 2015
www.geosci-model-dev-discuss.net/7/C3364/2015/


Geoscientific Model Development Discussions
Open Access



# *Interactive comment on* "Forecast error covariance structure in coupled atmosphere–chemistry data assimilation" *by* S. K. Park et al.

**S. K. Park et al.**

spark@ewha.ac.kr

Received and published: 18 February 2015

We appreciate the positive and constructive comments by Referee 2 (Dr. E. Nino). The referee seems to fully understand the contents of our research. An item-by-item reply to the referee's questions is provided below:

**1. How well your idea scales regarding the dimension of the model? Observations?**

*In principle, there is nothing in the described methodology that would pre-*

*vent its use with larger dimensions of model or observations. Please note that the actual experiment specifications in our paper include* $132 \times 147 \times 28 \approx 0.5 \times 10^6$ *grid points and 11 control variables (10 of which are 3-dimensional), making the actual state vector dimension about* $0.5 \times 10^7$*, and thus the error covariance being a* $0.5 \times 10^7$ *by* $0.5 \times 10^7$ *matrix. We address high dimensionality of state by using the ensemble-based square-root covariance matrix, which has dimensions* $0.5 \times 10^7 \times 32$ *(for 32 ensemble columns), which can be processed column-by-column if the computer memory is restrictive. Although the high dimensional observations were not used in this manuscript due to our aim to describe the structure of the error covariance, this issue is generally addressed within the ensemble data assimilation algorithm. In our case, this is the Maximum Likelihood Ensemble Filter (MLEF – Zupanski 2005; Zupanski et al. 2008). As in similar ensemble filters without the perturbed observations (e.g., square-root ensemble filters), the high dimensional observations are processed in the low-dimensional local ensemble subspace. The practical advantage is that the matrix inversion is done in ensemble space, with dense and well-conditioned matrices. Therefore, the anticipated scaling of computing due to dimension change is likely sub-linear. For the error covariance the scaling is probably linear since an increase of the column dimension implies a proportionally longer I/O, and the required matrix-vector product has also a proportionally more term to calculate. For observation processing, although there is a linear scaling of the observation vectors and the involved calculations, the cost is ultimately governed by the ensemble dimension, implying a negligible cost increase due to increased observation dimensions.*

**2. What happen in the context of realistic scenarios? For your reference, here I cite two: * SPEEDY Model: http://www.ictp.it/research/esp/models/speedy.aspx * QGCM Model: http://www.q-gcm.org/**

*Please note that the WRF-CHEM model is itself a very complex coupled model, with complex chemical and atmospheric interactions. We believe that the use of ensemble-based error covariance is a huge advantage for describing complex and typically unknown correlations in a coupled system. The only required input is a set of (nonlinear) ensemble forecasts used to construct the square-root forecast error covariance, which automatically produces the most complex correlations. Although this is a flow-dependent error covariance, its relevance is quite important even if the processes are changing slowly with time, because the structure of cross-component covariance brings a wealth of dynamically-based correlations as represented by the employed model. The only important restriction is that a covariance localization is implemented due to a low-dimensional ensemble space, as is in our system. Therefore, using an atmosphere-ocean model (such as the mentioned QGCM), or the SPEEDY model, would automatically reveal all the relevant structure of correlations in such a modeling system.*

---

Interactive comment on Geosci. Model Dev. Discuss., 7, 8757, 2014.