

“Spectral diagonal ensemble Kalman Filter”

Recommendation :

Major revision

General comments :

The aims of the paper is to document the use of a parametric formulation for the background covariance matrix in the Ensemble Kalman Filter. The trick is to replace the direct sample estimation damped with the Schur product, by a diagonal assumption within a prescribed orthogonal basis or more generally within a frame (Pannekoucke et al., 2007): in particular, orthogonality is not essential.

Theoretical framework that supports this diagonal strategy is related to the variance of the sampling error for a given ensemble size (Mallat, 1999, Chap. X ; Furrer and Bengtsson, 2007). The derivation proposed by the authors takes into account the sampling error of the mean that does not change the core of the above result: it is equivalent to a sample estimation from an ensemble size $N-1$ when the mean is assume to be known, in relation with the property that covariance and average estimators Eq.(4) are unbiased.

This work is the EnKF counterpart of the strategy followed within variational framework. As well known, the 3D-Var set with the appropriate background covariance matrix at the analysis time, is the variational solution of the BLUE while practical implementation of the EnKF analysis step provides an alternative construction of the same BLUE. Of course, the problem in variational implementation is to provide such an appropriate background matrix, and this is partly done as follows: (a) an ensemble of perturbed analysis is moved forward to provide an ensemble of background error and (b) then used to set the parametric formulation of the day (Berre et al. 2007). Hence, the approach proposed in the present paper is not really new, and the perspective of using covariance model in EnKF has been recognised to be promising at the early stage of the wavelet formulation filtering property. This remark constitutes my major point and I will try to provide some help to precise the real place of this contribution.

The present work contributes to orient the EnKF community toward the variational community in order to take advantage of the knowledge accumulated in background error covariance matrix modelling. I think it could be suitable for publication if appropriate modification are done, following the major comments. For this reason I ask a major revision.

Major comments:

- 1) The derivation of theorem 1 (p5) is obtained through the spectral decomposition of the matrix, but the present derivation can be obtained directly from the computation in an arbitrary basis since Eq.(10) and Eq.(11) can be rewritten using the intrinsic operator of trace as

$$\mathbb{E} [\|\mathbf{mC} - \mathbf{mC}^{\wedge N}\|^2_F] = 1/(N-1) \text{Trace}(\mathbf{mC}^{\wedge 2}) + 1/(N-1) \text{Trace}(\mathbf{mC})^2$$

$$\mathbb{E} [\|\mathbf{mC} - \mathbf{mD}^{\wedge N}\|^2_F] = 2/(N-1) \text{Trace}(\mathbf{mC}^{\wedge 2})$$

independent of the basis and directly related to the spectrum. Hence, the comment p6, l 164 “the analysis in Furrer and Bengtsson (2007) is in the physical domain rather.. “ can be suppressed. From my view, this theorem is not really new, and references to Mallat and Furrer & Bengtsson should be enough. If the authors really want to put something here (to make the manuscript self contained), then they should mention the Wick formula that helps to compute general formulation of Gaussian moments. I think enough the derivation when

the average is assume known equal to 0.

- 2) As mentioned in the general comments, the method proposed here is not strictly new since in variational data assimilation, algorithms are existing that estimate the covariance of the day, introducing the flow dependence (equivalent to the EnKF) within the cost function minimising process ; with quantification of real impact in operational NWP ! Hence, it is important to mention this point in the manuscript (e.g. Buehner, 2005 ; Berre et al., 2007 ; Varella et al., 2011): if the strategy of resolution of the BLUE is different in the hybrid 3D-Var and the EnKF, the idea to model the covariance matrix to benefit of a noise less matrix is the same. In the major comment (6) below, I provide you elements to precise this point in the introduction and conclusion of the work. Note also that hybrid formulation comes from EnKF community with the work of Hamill and Snyder (2000) that also have introduced a spectral diagonal assumption (see their Eq.(3)) : this should be specified in the introduction. Of course, with the diagonal assumption in spectral space, since only homogeneous and isotropic correlations can be represented, there is no need to update the diagonal at each analysis step and climatological estimation is better. This is no more true with other formulations as encountered with the wavelets (and frame) that are able to produce heterogeneous correlation function where the spatio-temporal evolution makes sens.
- 3) The formulation of the background error covariance model using the diagonal assumption following a product of linear operator should mention all the work done in variational literature that intensively relies on this trick to build covariance matrix in huge dimension (Courtier et al. , 1998; Fisher and Andersson 2001; Weaver and Courtier, 2001). In particular, this should be specified in line 180 where operator transforms (FFT, DWT) are mentioned.
- 4) From this link with the variational community, some perspectives of the present contribution must be precise. In particular, from the history about covariance modelling in variational algorithm, the next steps of the work can be drawn as follows: construction of non-separable formulation (Courtier et al., 1998; Fisher and Andersson, 2001 ; Pannekoucke, 2009), representation of balances between variables in order to obtain a more realistic multivariate formulation (Derber and Bouttier, 1999 ; Fisher, 2003 ; Weaver et al. 2005), representation of heterogeneity using a physical/spectral localised formulation (non-separable wavelet formulation for Fisher and Anderson, 2001 ; separable formulation based on diffusion operator for Weaver and Courtier, 2001 or recursive filter for Purser et al. 2003 ; non-separable formulation based on hybridation diffusion/wavelets Pannekoucke, 2009) ... In particular, even if formulations as the diffusion operator or the recursive filter are not diagonal assumptions, they lead to an approximation of the covariance matrix free of sampling noise, and objectively parametrised from ensemble estimation (Pannekoucke and Massart, 2008 ; Michel, 2013 ; Pannekoucke et al. 2014). Along this route, filtering strategies can be employed to improve the estimation and the design of covariance formulations using results on the estimation of variances and length-scales (Berre et al, 2007 ; Raynaud et al. 2009 ; Raynaud and Pannekoucke 2013 ; Menetrier et al. 2015).
- 5) p11, l312: From the above major comments, saying “The paper provide a new technology for data assimilation” is too much and risks to appear arrogant while considering all the work that has been done for each community. However you right that until now very few person have try to seriously consider covariance model in EnKF, the main reason is that it require to build covariance matrix parameterisation, this represents a real cost in terms of technology investment for NWP codes. You should mention this in the introduction of the paper: “The idea of using covariance model in EnKF to benefit of sample noise reduction effect is known (Hamill and Snyder, 2000; Buehner, 2005), but as far as we know no reference has been published to document the real advantage of this method. In terms of practical implementation of the BLUE, one of the reason could be the relative distance existing between the EnKF and the variational to resolve an equivalent analysis step. However, the employ of forecast ensemble has been tested with success in the variational

framework (Buehner 2005, Berre et al. 2007, Varella et al., 2011)”. For the conclusion, I guess you can replace the sentence “The paper provide..” as follows: “The paper provide a preliminary test, within an academic setting, of the employ of parametric covariance in pure EnKF strategy, while the reverse strategy is existing in variational framework (hybrid formulation)”.

- 6) p11, l314 : “because an implementation only needs an orthogonal transformation” orthogonality is not a necessary condition for diagonal assumption that can be considered within a frame as detailed in Pannekoucke et al. (2007). Of course this have an influence for the representation of the observational error covariance matrix in the “spectral space” that is no more diagonal (as specified in p7, l 179). Note that in Pannekoucke (2008, appendix D) wavelet packets are used to take advantage of the orthogonal basis dictionary it provides ; for this, the problem is then to connect horizontal sheet along the vertical in a 2D/3D formulation, in direct 3D formulation this could be used without the connection issue.

Minor comments :

- 1) p2, l54: Add references for the wavelets formulation: Fisher and Andersson, 2001; Deckmyn and Berre, 2005 ;
- 2) p6, before sec. 6: You should mention that the parametric formulation does not converge toward the “true” covariance matrix as the ensemble size increases toward infinity.
- 3) It is not direct that Σ defined p10, l283 is positive, please provide a proof of it ;
- 4) P9, l250: 's' is meaningless here since 't' is a pseudo-time (Lorenz'96 is not related to a physical model but only an academic framework, nice to play with), hence replace “0.01s” by \$0.01\$ and \$1s\$ by \$1\$;
- 5) p9, l 252: “The the ensemble ..” → “Then the ..” ;
- 6) p10, l297 : “The relax the ensemble members the model .. “ → must be rephrased ;
- 7) p11, l 310 : “experimens” → experiments ;
- 8) p11, l 311 : “shown that the method that the analysis” → must be rephrased ;
- 9) p13, l 345 : the parenthesis are not well positioned leading to ambiguities, please write something as $E[(\sum ..)^2]$, this appears at many times in the derivation of the proof.
- 10) Some work exists concerning the balance in EnKF that should be mention in the manuscript see Kepert (2009).

References:

Berre, L.; Pannekoucke, O.; Desroziers, G.; Stefanescu, S.; Chapnik, B. & Raynaud, L. ECMWF (Ed.) A variational assimilation ensemble and the spatial filtering of its error covariances: increase of sample size by local spatial averaging, *ECMWF Workshop on Flow-dependent aspects of data assimilation, 11-13 June 2007.*, **2007**, 151-168

Buehner, M. Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting *Quarterly Journal Royal Meteorological Society*, **2005**, *131*, 1013-1043

Courtier, P.; Andersson, E.; and J. Pailleux, W. H.; Vasiljević, D.; Hamrud, M.; Hollingsworth, A.; Rabier, F. & Fisher, M. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quarterly Journal Royal Meteorological Society*, **1998**, *124*, 1783-1807

Deckmyn, A. & Berre, L. A wavelet approach to representing background error covariances in a LAM. *Monthly Weather Review*, **2005**, *133*, 1279-1294

- Derber, J. & Bouttier, F. A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus A*, **1999**, 51, 195-221
- Fisher, M. & Andersson, E. Developments in 4D-Var and Kalman Filtering *ECMWF*, **2001**. **Tech. Rep. 347**
- Fisher, M. ECMWF (Ed.) Background error covariance modelling *Proc. ECMWF Seminar on "Recent developments in data assimilation for atmosphere and ocean"*, **2003**, 45-63
- Hamill, T. M. & Snyder, C. A Hybrid Ensemble Kalman Filter-3D Variational Analysis Scheme *Monthly Weather Review*, **2000**, 128, 2905-2919
- Kepert, J. Covariance localisation and balance in an Ensemble Kalman Filter *Quarterly Journal Royal Meteorological Society*, **2009**, 135, 1157-1176
- Mallat, S. A Wavelet Tour of Signal Processing, *Academic Press* , **1999**, 620
- Menetrier, B.; Montmerle, T.; Michel, Y. & Berre, L. Linear Filtering of Sample Covariances for Ensemble-Based Data Assimilation. Part I: Optimality Criteria and Application to Variance Filtering and Covariance Localization *Mon. Wea. Rev.*, **2015**
- Michel, Y. Estimating deformations of random processes for correlation modelling in a limited area model *Quarterly Journal Royal Meteorological Society*, **2013**, 139, 534-547
- Pannekoucke, O. Modélisation des structures locales de covariance des erreurs de prévision à l'aide des ondelettes, *Univertisté de Toulouse, France*, **2008**
- Pannekoucke, O. & Massart, S. Estimation of the local diffusion tensor and normalization for heterogeneous correlation modelling using a diffusion equation. *Quarterly Journal Royal Meteorological Society*, **2008**, 134, 1425-1438
- Pannekoucke, O. Heterogeneous Correlation Modeling Based on the Wavelet Diagonal Assumption and on the Diffusion Operator *Monthly Weather Review*, **2009**, 137, 2995-3012
- Pannekoucke, O.; Emili, E. & Thual, O. Modeling of local length-scale dynamics and isotropizing deformations *Quarterly Journal Royal Meteorological Society*, **2014**, 140, 1387-1398
- Raynaud, L.; Berre, L. & Desroziers, G. Objective filtering of ensemble-based background- error variances *Quarterly Journal Royal Meteorological Society*, **2009**, 135, 1177-1199
- Raynaud, L. & Pannekoucke, O. Sampling properties and spatial filtering of ensemble background-error length-scales *Quarterly Journal Royal Meteorological Society*, **2013**, 139, 784-794
- Varella, H.; Berre, L. & Desroziers, G. Diagnostic and impact studies of a wavelet formulation of background-error correlations in a global model *Quarterly Journal Royal Meteorological Society*, **2011**, 137, 1369-1379
- Weaver, A. & Courtier, P. Correlation modelling on the sphere using a generalized diffusion equation *Quarterly Journal Royal Meteorological Society*, **2001**, 127, 1815-184
- Weaver, A.; Deltel, C.; Machu, E.; Ricci, S. & Daget, N. A multivariate balance operator for variational ocean data assimilation *Quarterly Journal Royal Meteorological Society*, **2005**, 131,

3605-3625