# Hybrid ensemble Var for a limited area model

## Monika Krysta1, Susan Rennie1, Shaun Cooper1, Peter Steinle1, Xudong Sun1, Jin Lee1 and Jim Fraser1

*1Environment and Research Division, Bureau of Meteorology, Melbourne*

*monika.krysta@bom.gov.au*

**Motivation**

Data assimilation aims at providing an optimal state of the atmosphere by weighting contribution from prior knowledge about the state of the atmosphere and from observations. Prior knowledge is usually represented by a past forecast valid at the time of optimisation and its contribution to the optimised state of the atmosphere is weighted by an error covariance matrix, commonly called a background error covariance matrix in the context of variational data assimilation. Grid point-wise specification of the covariances in this matrix is not tractable, let alone due to storage requirements it would entail. Traditionally, therefore, in variational data assimilation, these statistics have been modelled (Lorenc et al., 2000) outside of a cycling numerical weather prediction (NWP) system. The parameters intervening in covariance modelling are estimated using samples of atmospheric states as produced by NWP simulations over selected past periods and transferred to a cycling system where they remain static over long periods of time. The covariance model constructed in this way accounts for climatological features of the flow and, in addition, its model is based on strong hypotheses of isotropy and homogeneity. While the covariance modelling approach made the problem of using high-dimensional matrices of error statistics tractable, in presence of strong gradients this approach leads to sub-optimal estimates of the state of the atmosphere. At present, ever finer improvements to the numerical forecasting are sought rendering the climatological formulation of the background error statistics insufficient. Hence, the rest of this abstract focuses on how to remedy this problem with a contribution of an ensemble prediction system (EPS).

**Hybrid ensemble 4D-Var**

In parallel to the approach of modelling of the background error covariance matrix for variational data assimilation, in the realm of ensemble methods developed around Kalman filtering, reduced rank covariance matrices which are based directly on members of the EPS are employed. Contemporary NWP cycling systems reflect spatial and temporal variations in the observation network via the construction of an atmospheric analysis and they subsequently propagate error structures using the full nonlinear model in an EPS forecast step. Hence, the error covariance matrix **Be** based on the EPS reflects the present state of the atmosphere and accounts for flow dependencies. Consequently, a remedy to the problem of a lack of representation of the errors of the day in variational data assimilation could be to replace the climatological background error covariance matrix, **Bc**with **Be**. One striking characteristic of the error covariance matrix derived from the EPS is that it is rank deficient, having at most the rank equal to the ensemble size. Another problem which it brings alongside resides in spurious correlations appearing due to a limited size of the sample, which is counterintuitive as the errors at distant regions should be uncorrelated resulting in a larger number of independent structures than those described by a raw **Be**. Hence, an idea of a hybrid method emerged (Hamill and Snyder, 2000) which advocates forming a linear combination of two covariances

so that qualities of one will be complemented with the qualities of the other. However, even the hybrid approach requires additional treatment of the ensemble covariances commonly encountered in Kalman filtering, namely localisation.

Variational data assimilation optimizes the state of the atmosphere by minimizing a cost function and its practical implementation requires preconditioning of the problem so that the quadratic term in the cost function is represented by a unit matrix. A sequence of three transforms was therefore devised to remove inter- and intra-variable correlations in model variables so that they could be replaced with uncorrelated control variables of unit variance. First, a parameter transform, **Tp**, transfers model variables to the space of control variables, which are further decorrelated via geometrical transforms, a vertical one, **Tv**, projecting onto a set of uncorrelated empirical vertical modes and a horizontal one, **Th**, projecting, in case of a limited area model, on Fourier modes. Consequently, following (Clayton et al., 2003), the analysis increment in model space, δ**w**, is mapped to a control variable, **v**, using a sequence of these three transforms

while the (approximate) inverses of the aforementioned operators, **Up**, **Uv** and **Uh**, respectively, are used to define an inverse transform, mapping the control variable **v** to the model increment δ**w**,

These transforms are directly related to the climatological background error covariance as they form its square root, **Bc**=**UUT** = **UpUvUhUhTUvTUpT**. Similarly, a practical implementation of the ensemble covariance matrix, makes use of its readily available square root, **W**,

which is a rectangular matrix whose columns are formed by K scaled error modes **wk’**.Therefore, one can use **W** to transform model increment, δ**w**, into a control variable, in the ensemble context referred to as alpha control variable,

where

is such an alpha control variable. It is here that the localization of the error modes intervenes. Physically, it results in disconnecting spuriously connected error structures and, mathematically, in increasing the rank of the error covariance matrix. Technically, the localized background error covariance matrix, **Be** is thus represented as **L**○**WWT** where **L** is a localization matrix and ○ represents a Schur, element by element, product of two matrices. The model increment can now be written as

by generalizing scalars αks to fields of **αk**s. Also the localization matrix **L** needs to be factored in terms of its square root **Uα**, **L**=**UαUαT**, which is subsequently split into horizontal and vertical transforms

so that the mapped alpha control variable is now **vkα**

and the full alpha control vector reads

Implementing hybrid in this set-up (Clayton et al, 2013) results in an analysis increment of the form

where the extended (Lorenc, 2003) control variable vector **v** is now a concatenation of the climatological and alpha control variables

While localization constitutes a remedy to the problem of unphysical spurious correlations it also brings along a set of new challenges as it is affecting the dynamical structure of the error modes and hence requiring quite a subtle treatment. The hybrid method performs localization in the control space spanned by the stream function ψ, velocity potential χ, geostrophically-unbalanced pressure pA and humidity µ. Localisation in model space allows not only to better account for non-local observations and but also renders increments which preserve geostrophic balance to a larger extent than it would be the case if localization was performed in the model space. A new form of the localized increment (Clayton et al, 2013)

exhibits the action of the parameter transform **Tp**. This transform discards the balanced fraction of the pressure increment form the error mode **wk’**. Subsequently, the remaining unbalanced pressure is localized and after that the balanced fraction of the pressure increment is brought back to the analysis increment via the **Up** transform. In this way the geostrophic portion of the analysis increment remains geostrophic even after localisation.

Localisation needs to be applied in both horizontal and vertical directions. In horizontal, a Gaussian form is implemented while vertical localization is more challenging for two reasons. Firstly, evidence was found that the correlation does not monotonically decrease with vertical distance (Ingleby, 2001). Secondly, it is not clear with respect to which vertical coordinate the correlation length scales could be constant across the levels. To circumvent these problems, vertical localization is performed in correlation space (Clayton et al, 2013) and is based on globally averaged values from **Bc** in such a way that if cavg is the average correlation between two levels, the corresponding localized correlation is

where R is a parameter.

**Implementation**

The upcoming BoM global operational system, G3/GE3, will run a hybrid variational data assimilation as a default mode. The global system will be followed by a city system, referred to as C3/CE3. The latter, however, does not have the hybrid capability rendering the present coupling between the two systems one-way. In fact, C3 provides high resolution analysis to form a high resolution initial condition for the CE3 system. However, no feedback from the CE3 system to C3 is brought at present. A hybrid ensemble Var will ultimately constitute a system where such a feedback to C3 is provided and a CE3 ensemble would be used to build an ensemble-based contribution to the background error covariance which will be employed in variational data assimilation in a C3system, albeit both likely in a different version at that time.

At present, BoM is actively working on porting a hybrid capability for a limited area model as developed at the KMA (Korea Meteorological Administration) to BoM computing environment. This initial implementation will only loosely be connected to the C3/CE3 system, however, the operating environment and the system components are shared so that ultimately a two-way feedback system could be developed. To begin with, the hybrid system will run in an off-line mode using the ensemble members as precalculated by CE3. Still, an option for an on-line generation of the ensemble members will also be available. Both hourly 3D-Var and hourly 4D-Var will be supported.

# References

Clayton, A., M., Lorenc, A., C., and D., M., Barker, 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office, Quart. J. Roy. Meteorol. Soc. **139**, 1445-1461

Hamill, T., M., and C., Snyder, 2000: A Hybrid Ensemble Kalman Filtrer-3D Variational Analysis Scheme, Mon. Wea. Rev., **128**, 2905-2919

Ingleby, N., B., 2001: The statistical structure of forecast errors and its representation in The Met. Office Global 3-D Variational Data Assimilation Scheme, Quart. J. Roy. Meteorol. Soc. **127**, 209-231

Lorenc, A., C., Ballard S., P., Bell R., S., Ingleby N., B., Andrews P., L., F., Barker D., M., Bray J., R., Clayton A., M., Dalby T., Li D., Payne T., J., and F., W., Saunders, 2000: The Met. Office global three-dimensional variational data assimilation scheme, Q. J. R. Meteorol. Soc. **126**, 2991–3012

Lorenc A., C., 2003: The potential of the ensemble Kalman filter for NWP: a comparison with 4D-Var, Q. J. R. Meteorol. Soc. **129**, 3183–3203.