Advances in
Science & Research
Open Access Proceedings

Modelling static 3-D spatial background error covariances – the effect of vertical and horizontal transform order

M. A. Wlasak and M. J. P. Cullen

Met Office, FitzRoy Road, Exeter, Devon, EX1 3PB, UK

Correspondence to: M. A. Wlasak (marek.wlasak@metoffice.gov.uk)

Abstract. A major difference in the formulation of the univariate part of static background error covariance models for use in global operational 4DVAR arises from the order in which the horizontal and vertical transforms are applied. This is because the atmosphere is non-separable with large horizontal scales generally tied to large vertical scales and small horizontal scales tied to small vertical scales. Also horizontal length scales increase dramatically as one enters the stratosphere. A study is presented which evaluates the strengths and weaknesses of each approach with the Met Office Unified Model.

It is shown that if the vertical transform is applied as a function of horizontal wavenumber then the horizontal globally-averaged variance and the homogenous, isotropic length scale on each model level for each control variable of the training data is preserved by the covariance model. In addition the wind variance and associated length scales are preserved as the scheme preserves the variances and length scales of horizontal derivatives. If the vertical transform is applied in physical space, it is possible to make it a function of latitude at the cost of not preserving the variances and length scales of the horizontal derivatives.

Summer and winter global 4DVAR trials have been run with both background error covariance models. A clear benefit is seen in the fit to observations when the vertical transform is in spectral space and is a function of total horizontal wavenumber.

1 Introduction

For the last couple of decades variational data assimilation schemes have been the preferred method of determining the appropriate initial conditions for global atmospheric forecasts. The schemes need to represent approximately the error statistics of the Bayesian prior which is normally the error associated with the previous forecast. A typical forecast model can have 10^8 or more degrees of freedom. To get a reasonably accurate full-rank background error covariance without making any underlying assumptions as an approximation to the prior, one would need to use an unworkably large number of samples. In this study the input data is taken from ensembles of forecasts and forecast differences are considered to be proxies for forecast error. Even if it were possible to get enough samples the data storage requirements of this covariance matrix would be also be impractical. To define any full-rank covariance matrix of degree n one needs $n(n+1)/2$ numbers.

The majority of these schemes include a background error covariance model - a mathematical model that uses physical, dynamical and statistical insight to reduce the amount of information needed to determine the background error covariance. One form of covariance model that is getting increasingly popular approximates the background error covariance matrix by a low rank sample covariance computed from an ensemble and then removes (localises) spurious correlations caused by sampling error. Such models have variances and correlations that change for every data assimilation cycle and are outside the scope of this paper. This paper considers full-rank covariance models. Such models are designed to be full-rank by using statistics that are averaged in space and/or time by resorting to assumptions like homogeneity, isotropy and stationarity.

Most covariance models have a multivariate step, normally called the parameter transform, which typically involves a reduction of the number of atmospheric variables considered, to a smaller set called control variables that are assumed to be uncorrelated with one another. The standard control variables used at the Met Office are streamfunction ψ, velocity potential χ, ageostrophic pressure Ap and humidity μ. For each uncorrelated "control variable" a spatial decomposition is done to precondition the problem and to separate structures that hold large background error variances from those that do not. This paper focuses on this spatial decomposition.

The three dimensional aspects of forecast errors vary with horizontal and vertical position and cannot be represented completely without further approximation. The standard approximation that is used is to assume that the errors are homogenous and isotropic on either some horizontal model level in some vertical co-ordinate or some empirical vertical mode basis. This paper looks at the differences between these two approaches within the Met Office's global data assimilation system.

The rest of this paper is organised as follows. Section 2 describes the current and proposed static background error covariance model used in global 4DVAR at the Met Office and lists their corresponding strengths and weaknesses. Section 3 gives details of the forecast trials and shows the effect of the proposed static background error covariance model. Section 4 gives the conclusions.

2   A comparison between current and proposed static background error covariance model

The two static covariance models can be viewed in terms of their implied background error covariances $\mathbf{B}_{s1}$ $\mathbf{B}_{s2}$ and their respective parameter, horizontal and vertical transforms:

$$\mathbf{B}_{s1} = \mathbf{U}_p\mathbf{U}_v\mathbf{U}_h\mathbf{U}_h^T\mathbf{U}_v^T\mathbf{U}_p^T \tag{1}$$

$$\mathbf{B}_{s2} = \mathbf{U}_p\mathbf{U}_h\mathbf{U}_{vs}\mathbf{U}_{vs}^T\mathbf{U}_h^T\mathbf{U}_p^T \tag{2}$$

where $\mathbf{U}_p$ is the parameter transform from control variables to model variables, $\mathbf{U}_h$ is a spectral horizontal transform from spectral space to grid point space, $\mathbf{U}_v$ is a vertical transform applied in grid-point space and $\mathbf{U}_{vs}$ is a vertical transform applied in spectral space.

The covariance model associated with $\mathbf{B}_{s1}$ is the current static background error covariance model as described in Lorenc et. al. (2000). Since the Unified Model (UM) is a grid-point based forecasting model, it is natural to consider applying the vertical transform in grid-point space. The vertical transform from model levels to empirical vertical modes is constructed by calculating the eigenvectors of a horizontally-area-averaged vertical covariance matrix which has been weighted in vertical. This vertical scaling for ψ is a pressure weighting that takes account of the mass of each layer.

The key feature involves a vertical mode basis defined by the eigenvectors that are the same across the whole globe, but whose variance in vertical mode space varies as a function of latitude. A latitude-dependent vertical regression is also applied between the full pressure increment and a geostrophic pressure increment derived from a variant of the linear balance relation. This regression makes ψ and Ap uncorrelated with each other.

One of the advantages of the operational covariance model is that it can to a degree take into account the latitude dependence of the covariance statistics. However, since a single vertical covariance matrix is used for each control variable, the scheme is unable to preserve the variances of horizontal derivatives that are needed to get the wind variances and associated length scales right as a function of model level. Instead of using the actual calibration statistics for ψ and χ, the rotational kinetic energy (RKE) and divergent kinetic energy (DKE) are used. These substitutes have flatter power spectra and as a result each has a single global vertical covariance that is better at representing the vertical structure across the full range of horizontal wavenumbers. This choice also shortens the length scales of ψ and χ. Ingleby (2001) shows that this makes the implied pressure and temperature variances smaller than what is observed in the training data.

Strategically it makes sense to have a covariance model that preserves faithfully some key properties of the training data. It would allow a shift in emphasis towards getting the training data right instead of concentrating on tuning the background error covariance model as has been the case in the past.

To this end we have developed an alternate simpler covariance model that is represented by $\mathbf{B}_{s2}$. This model has features that are similar to the old ECMWF covariance system (Courtier et al., 1998). While the old ECMWF system assumed that the correlations of the control variables were homogenous and isotropic on each model level, $\mathbf{B}_{s2}$ assumes that the covariances are homogenous and isotropic. The main difference between $\mathbf{B}_{s2}$ and the current operational Met Office covariance model is as follows. Instead of storing a square root of the vertical covariance matrix and a power spectrum for each vertical mode, the order of the horizontal and vertical transforms is reversed and the square root of each vertical covariance matrix for each total horizontal wavenumber is stored. Also a global vertical regression is stored between the full pressure increment and the balanced pressure increment in order to get Ap to be uncorrelated with ψ. By not allowing latitudinal variation in the vertical regression, the horizontal-average variance estimate of the vertical derivative of the training data is preserved. This is needed as the hydrostatic equation is used to calculate temperatures from the vertical derivative of pressure. By preserving the horizontal-average variance estimate of the vertical derivatives, the horizontal-average estimate of the temperature variance is preserved on each model level.

The advantage of this approach is that the horizontally-averaged variance estimate for each model level is preserved for the observed winds, temperatures and pressures. Also it preserves the variance for each horizontal wavenumber and thus preserves homogeneous and isotropic length scale information of the training data for each model level. By preserving the homogeneous and isotropic length scales of the derivatives of the control variables on each model level, the homogeneous and isotropic length scales of the training data for the RKE and DKE will also be preserved as well as the horizontal-averaged variance estimate of the wind.

The control variable ψ determines the balanced pressures and temperatures. By preserving the length scales of ψ, we preserve also the balanced pressure and temperatures.

The mathematical mechanism can be seen most easily if we consider ψ, χ in spectral space. The streamfunction can be written in terms of its spectral coefficients ψ_l^m as

$$\psi = \sum_{l=0}^{L} \sum_{m=-l}^{l} \psi_l^m Y_l^m(\lambda, \sin\varphi), \tag{3}$$

where L is the order of the triangular truncation, l and m are the total and meridional wavenumbers respectively, and $Y_l^m(\lambda, \sin\varphi)$ is the spherical harmonic function for latitude φ and longitude λ.

One property of the gradient of a spherical harmonic is

$$\frac{1}{4\Pi} \int_{-1}^{1} \int_{0}^{2\Pi} \nabla Y_s^t[\lambda, \sin\varphi] \cdot \nabla Y_l^{m*}[\lambda, \sin\varphi] \partial\lambda \partial\sin\varphi$$
$$= l(l+1)\delta_s^l \delta_t^m, \tag{4}$$

where $*$ denotes the adjoint operator, s and l are total wavenumbers, t and m are meridional wavenumbers, ∇ is the gradient operator, δ is the Dirac delta function and $\cdot$ is the scalar product operator between two vectors.

We first consider the variance of the divergent wind, which is given by the variance of the gradient of the velocity potential. The divergent wind can be calculated by the sum of the product of the spectral coefficients of the velocity potential and the derivative of the spherical harmonics. The above Eq. (4) tells us that the gradient of the spherical harmonics has an orthogonal basis for each total wavenumber. So provided the total variance of the spectral coefficients of velocity potential is kept for each total wavenumber then we will be able to conserve the full variance of the divergent wind.

A similar argument holds for streamfunction spectral coefficients and the variance of the rotational wind, noting that $(\mathbf{k} \times \nabla Y) \cdot (\mathbf{k} \times \nabla Y) = \nabla Y \cdot \nabla Y$, where $\mathbf{k}$ is the unit vector in the vertical and $\times$ is the vector product of two vectors.

This argument is not limited to variances but extends to cross-covariances between different model levels as we do not need the product of the spectral coefffficients to be positive.

Table 1. Comparison of covariance models. Strengths denoted by (+), weaknesses denoted by (−).

Current static operational covariance model	Proposed static covariance model
Latitude dependence built into vertical and parameter transform (+)	Latitude dependence built into parameter transform only (−)
Power spectra not conserved for model variables (−)	Global variances and power spectra conserved for model variables on model levels (+)

If the streamfunction and velocity potential power spectra do not vary much with model level then covariance model $\mathbf{B}_{s1}$ would sufffice. This was indeed the case when the atmospheric model was predominantly limited to the troposphere. However as the atmospheric model now includes the stratosphere and the mesosphere, the characterisation of the power spectra varies with height with successively longer length scales seen in the upper regions. Also there are huge differences in the vertical correlations across horizontal wavenumbers, with much broader vertical correlations for ψ at low wavenumbers.

A weakness of this approach is that the covariance statistics can only take account of latitudinal dependency through the parameter transform. Also both methods suffer from the degree to which homogeneity and isotropy assumptions are valid.

A summary of the strengths and weaknesses of the two methods is given in Table 1.

3 Results

3.1 Trial results

Two forecast trials were run to compare the strengths and weaknesses of the two covariance model formulations. 4DVAR non-hybrid trials were used for this comparison so that the behaviour of the two models could be seen clearly and not obscured by the effects of the errors of the day from the ensemble. The UM (version 7.9) was used at N320 resolution. The data assimilation VAR system (version 29.2.1) was run in a two step process with the first step running at N108 resolution, giving Hessian eigenvectors for a subsequent N216 resolution minimisation.

The background error statistics were generated using the same set of training data – 300 samples of forecast errors from a spun up early version of the ECMWF ensemble of analyses system that was propagated for 6 h by the UM. The 6 h run ensures compatibility with the UM while not changing the statistics much. The period from which the ensemble of analyses was extracted was 1–17 October 2006.

The first trial was a summer trial from 28 June to 31 July 2012. The overall measure used at the Met Office for verification is the NWP index for fit to observations and
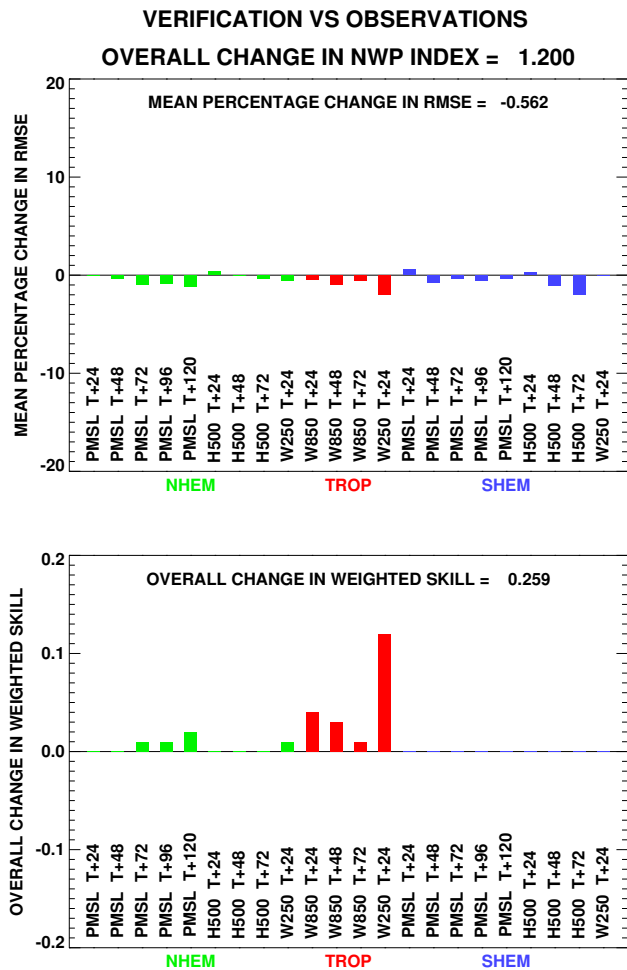
Figure 1. Verification results for winter trial 2012: new covariance model versus current covariance model. Upper plot shows reduction in RMSE. Lower plot shows how the change in RMSE affects NWP observation index.

own analyses. Verification against analyses is only relevant when making small changes to the nature of the analysis; this verification measure is not meaningful when changing the covariance model in a profound way. For this reason we stick to presenting only the verification scores for fit to observations.

The NWP index is a basket of scores of different forecast parameters. Detail of the NWP index is found in the Appendix of paper Rawlins et al. (2007). There was a marked improvement in skill in fit to observations of +0.7. This value far exceeds what is typically considered to be the standard threshold of +0.3 which indicates a statistically significant positive result at the 95 % confidence level (P. Jermey, personal communication, 2014). The benefit was seen mainly in the southern hemisphere in the surface pressure, 500 hPa geopotential height and 250 hPa winds. Also there was a very large improvement in the fit to 250 hPa winds in the tropics.

In addition a winter trial was run from 18th January 2012 to 28 Febuary 2012. The results for this trial were even more favourable with an improvement in the NWP index fit to observations of +1.2. As shown in Fig. 1, the benefit was seen for most measures in north, south and tropical regions.

3.2　Analysis increments

The biggest difference seen between the analysis increments from the two systems is in the horizontal scale and variance of the pressure and potential temperature. The new covariance model has doubled the variance of the pressure and potential temperature with much more variance at lower horizontal wave numbers due to preserving the characteristics of the training data. The larger length scale also manifests itself in the first Hessian eigenvector.

4　Conclusions

The paper shows that a simple covariance model designed to preserve global length scales and variances for each model level of each model variable can beat a long standing operational covariance model. Getting the global length scales and variances right on each model level matters.

This study has limited itself to two static covariance models. When including a non-static component to the covariance model as is described in Clayton et al. (2013) we see a neutral signal. Further work on tuning the non-static model is required.

It is noted that homogeneity and isotropy are quite harsh approximations. It would be good to identify an intermediate vertical grid that varies with latitude such that for each transformed model level the assumption of homogeneity of the variance holds more strongly.

References

Clayton, A. M., Lorenc, A. C., and Barker, D. M.: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office, Q. J. Roy. Meteor. Soc., 139, 1445–1461, doi:10.1002/qj.2054, 2013.

Courtier, P., Andersson, E., Heckley, W., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., Fisher, M., and Pailleux, J.: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation, Q. J. Roy. Meteor. Soc., 124, 1783–1807, doi:10.1002/qj.49712455002, 1998.

Ingleby, N. B.: The statistical structure of forecast errors and its representation in The Met. Office Global 3-D Variational Data

Assimilation Scheme. Q.J.R. Meteorol. Soc., 127: 209–231. doi:10.1002/qj.49712757112, 2001.

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 Saunders, F. W.: The Met. Office global three-dimensional variational data assimilation scheme. Q.J.R. Meteorol. Soc., 126: 2991–3012. doi:10.1002/qj.49712657002, 2000.

Rawlins, F., Ballard, S. P., Bovis, K. J., Clayton, A. M., Li, D., Inverarity, G. W., Lorenc, A. C., and Payne, T. J.: The Met Ofiice Global 4-Dimensional Data Assimilation System, Q. J. Roy. Meteor. Soc., 133, 347–362,doi:10.1002/qj.32, 2007.