Met Office

Deterministic data assimilation developments — July 2017

Forecasting Research Technical Report No: 625

2 March 2018

G.W. Inverarity (editor), M.A. Wlasak, M. Jardak and A.C. Lorenc

Executive summary

We report on recent developments in hybrid data assimilation (DA) that inform the global DA package choices for parallel suite 40 (PS40). The main components are the use of wavebands for horizontal localisation, the use of Ménétrier's vertical localisation scheme, time lagging and shifting, alternative hybrid covariance weights and new covariance statistics mainly based on Met Office Global and Regional Ensemble Prediction System - Global (MOGREPS-G) ensemble training data. Two package recommendations are made for PS40, one using the new covariance statistics and the other based on the existing covariance statistics calibrated using European Centre for Medium-Range Weather Forecasting (ECMWF) training data, evolved using the Unified Model (MetUM).

1 Introduction

Hybrid four-dimensional variational data assimilation (4DVar) blends two descriptions of the forecast error statistics to generate an optimal analysis of atmospheric conditions used as initial conditions for each cycle's forecast. The stationary covariances are generated using the Covariances and VAR Transforms (CVT) software that is part of the VAR code repository. These are produced by applying a covariance model whose parameters are fitted from ensemble training data representing typical forecast errors over a calendar year. The stationary covariance statistics are blended with ensemble forecasts from the two preceding forecast cycles valid at the start of the current cycle's assimilation window to add a flow-dependent element to the description of the forecast error statistics. The current results are linked to the effort to replace the current ensemble initialisation technique that uses an Ensemble Transform Kalman Filter (ETKF) (Bishop *et al.*, 2001; Bowler *et al.*, 2008) with an ensemble of four-dimensional ensemble variational data assimilations (En-4DEnVar) (Bowler *et al.*, 2017a,b).

Five distinct enhancements are reported on:

1. waveband scale-dependent horizontal localisation (Buehner, 2012);
2. Ménétrier vertical localisation (Ménétrier and Auligné, 2015; Ménétrier *et al.*, 2015a,b);
3. using time lagging and shifting to increase the effective size of the ensemble (Lorenc, 2017);
4. adjusting the hybrid covariance weights β_c and β_e used to blend the static and ensemble covariance components ($\mathbf{B}_c$ and $\mathbf{B}_e$, respectively) to give $\mathbf{B} = \beta_c^2\mathbf{B}_c + \beta_e^2\mathbf{B}_e$;
5. calibrating static covariance statistics using training data generated by the Met Office En-4DEnVar ensemble initialisation system under development (Bowler *et al.*, 2017b).

These are described in Sections 2–6, respectively. The two packages of combined changes proposed for PS40 and the aspects that were subsequently implemented operationally are described in Sections 7 and 8, respectively, while future work is described in Section 9.

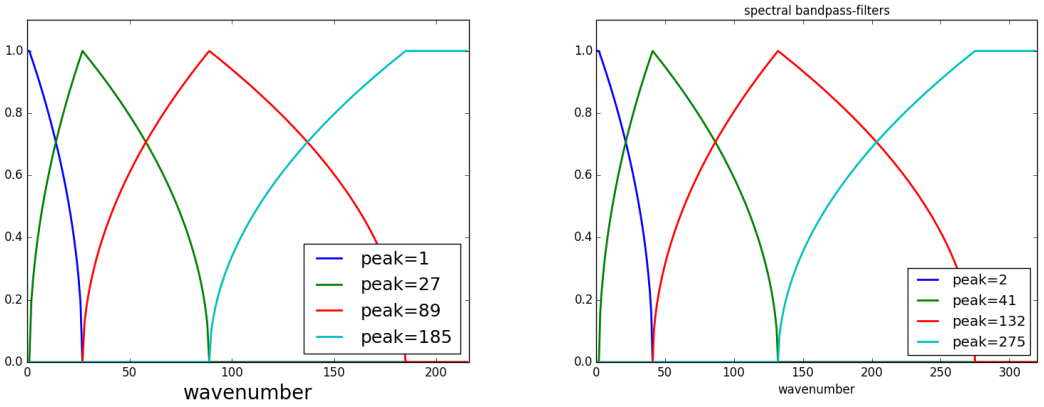


Figure 1: Spectral filters defining wavebands (Lorenc, 2017). The left figure shows the wavebands used in the experiments reported here while the right figure shows the wavebands that were implemented in PS40, and subsequently operational suite OS40, due to a setup error.

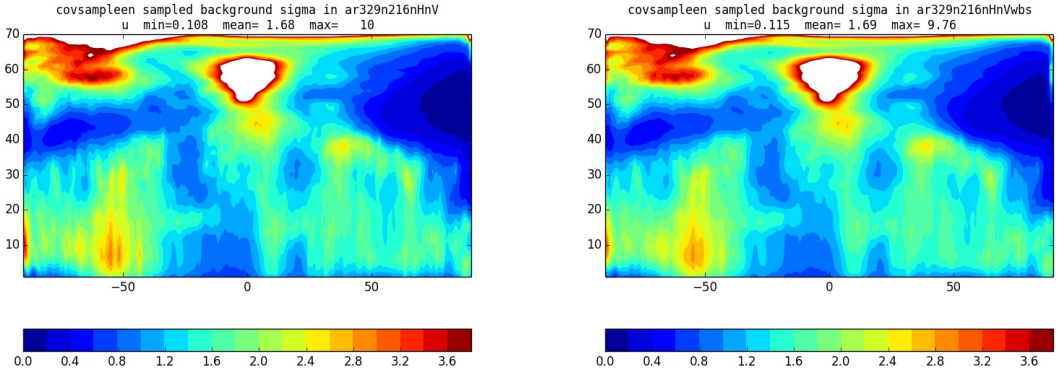


Figure 2: Zonal mean cross-section of ensemble spread in zonal wind at 0900 UTC on 25 September 2015 for a 44-member ensemble similar to the operational MOGREPS-G. Results are shown for the raw ensemble (left) and the waveband-filtered ensemble, consisting of the sum of the waveband components in Figure 3 (right).

2 Waveband horizontal localisation

Waveband localisation increases the effective number of degrees of freedom of the ensemble component of the background-error covariances, allows us to remove the high-pass filter currently applied to the ensemble forecast perturbations and, most importantly, introduces some scale-dependence into the ensemble component of the background-error covariances. We used the same four wavebands as Lorenc (2017) (Figure 1), which were chosen to roughly partition the ensemble variance (Figures 2–3). The direct effect of the waveband localisation is a slight smoothing; this is more visible in the variance maps (not shown). Neither the waveband localisation nor time lagging and shifting (section 4) had a significant effect on the variances — their main effect is to reduce sampling noise in the covariances.

Introducing wavebands allows us to apply a range of horizontal localisation lengthscales when

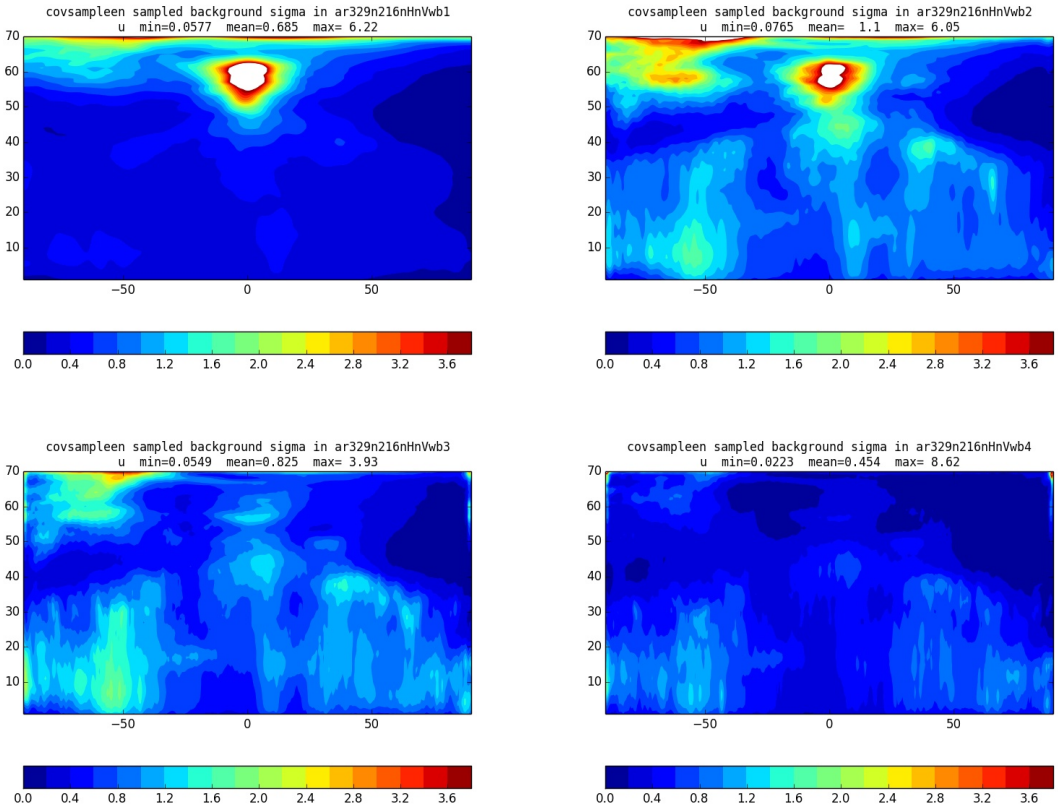


Figure 3: The contributions from wavebands 1 (top left), 2 (top right), 3 (bottom left) and 4 (bottom right) to the right plot of Figure 2. Pressure has more spread in waveband 1 while moisture has more spread in waveband 4 (not shown).

filtering the ensemble covariances. We use a Gaussian-shaped function for horizontal localisation, specified by a horizontal scale. Lorenc (2017) developed a simple method for choosing the relative scales for different bands in an idealised model then applied it to the full forecast model. The localisation scales used in this report are given in row 2 of Table 1, noting that the value used for the first waveband is actually 6241 km following a transcription error.

This scale-dependent localisation brings some other benefits:

- Because less localisation is applied to large scales, there is no need for the high-pass filter, which was introduced by Clayton *et al.* (2013) to avoid aliasing of global-scale ensemble perturbations. This means that the ensemble covariance should be closer to the actual background-error covariance, allowing it to be given a greater weight (β_e^2).
- The ensemble perturbations project onto different bands in different regions (Figures 2–3) and each band can have different localisation and other settings. This captures much of the spatial variation of optimal hybrid weight choices; Lorenc (2017) found much more consistency between different regions in experiments using scale-dependent localisation.

It would be impossible to test all combinations of horizontal localisation for each band (Ta-

Method	WB1	WB2	WB3	WB4
Idealised model & judgment (Lorenc, 2017)	8115	665	230	120
Original `hybrid-diag`	6214	919	389	256
Revised `hybrid-diag`	3039	1091	548	339

Table 1: Horizontal localisation scales (km) for a 44-member ensemble used in the trials for each waveband (WB). Note that large scales imply less localisation, with less modification to the raw ensemble covariances. The values in row 2 were calculated for each control variable on each model level, after which a weighted mean was taken over all control variables and every tenth model level starting from level 1 for each waveband.

ble 1), vertical localisation (Section 3) and hybrid weights (Section 5) individually. Ménétrier and Auligné (2015) and Ménétrier *et al.* (2015a,b) have proposed a method for diagnosing these and have provided `hybrid_diag` software for implementing it, which we have implemented and tested. `hybrid_diag` promises to be very useful, with the caveat that it is based on some theoretical assumptions and there are practical difficulties:

1. The method calculates a filtering (e.g., localisation) of the ensemble covariance such that it gives the best fit to the covariance from a similar infinite ensemble, so it is assuming that the ensemble is perfect. In practice the optimal filtering may compensate for deficiencies in the ensemble but the method cannot allow for this. It cannot be applied to our high-pass filtered ensemble, for example.

2. The fit is measured using the Frobenius norm (the root mean square (RMS) of a matrix's elements) applied to the difference between two matrices. This may not be a good measure of how well the estimated covariances perform in data assimilation applications.

3. The method relies on the ensemble members being independent. It gives incorrect results for a lag/shift ensemble, with localisations not tending to zero at large distances (e.g., Figure 5).

4. The software calculates different values for each variable and level, whereas our data assimilation software currently needs one value for each waveband. The values can differ widely (Figure 4) and some results are expected to be unreliable (e.g., moisture near the top of the model).

5. The software does not allow for waveband localisation. We have to diagnose localisation settings for each waveband in isolation and assume that they are still valid when the wavebands are used together.

6. Although the software can diagnose hybrid weights, these only apply in a simple case with a perfect ensemble; it does not work for wavebands.

Because of point 3 we gave up applying `hybrid_diag` to lag/shift ensembles after running some early experiments. Similarly, based on point 2 we did not use horizontal localisations estimated for streamfunction or velocity potential. Finally, to overcome point 4 we took median values when estimating the "average" horizontal localisation scales shown in row 3 of Table 1.

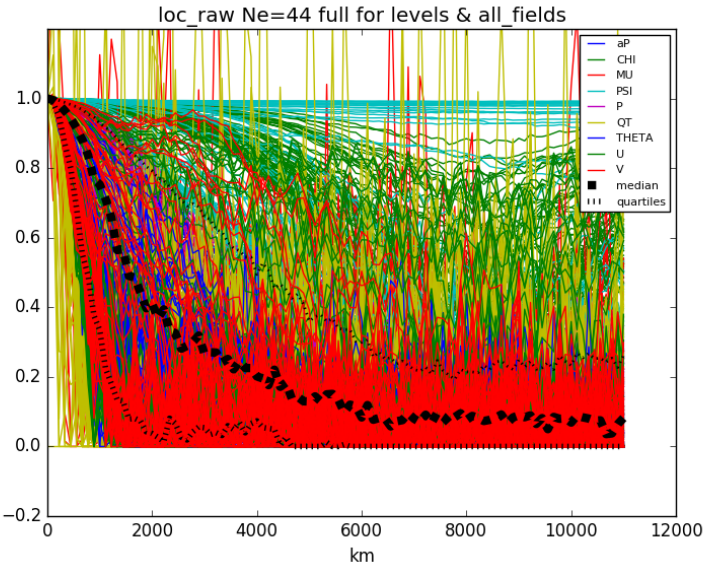


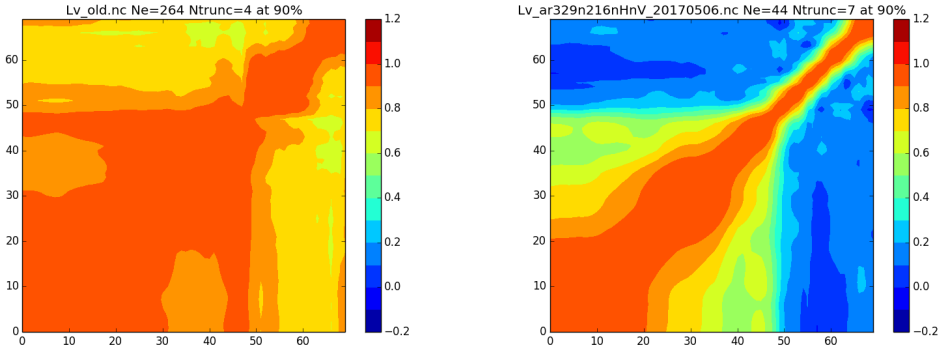
Figure 4: Raw horizontal localisation curves deduced by `hybrid_diag` from each level and variable. The heavy dots indicate the median — this ignores extreme values. The different control variables are: unbalanced pressure (aP), velocity potential (CHI), humidity (MU), streamfunction (PSI), pressure (P), total specific humidity (QT), potential temperature (THETA), zonal wind (U) and meridional wind (V).

Optimal scales are expected to increase with ensemble size. This perhaps explains why the original row 2 lengthscales perform slightly better than the revised row 3 lengthscales when we are applying 44-member results to a larger lag/shift ensemble because we expect lengthscales to then be underestimated.

3 Ménétrier vertical localisation

The current operational vertical localisation scheme uses the vertical covariance of the streamfunction from the static covariance model, converts it into a correlation matrix and then calculates the leading vertical modes that explain the most variance, which are then used as vertical basis functions for the alpha control variable. Smoothing is applied to the modes to avoid spurious spikes in implied temperatures. An alternative Gaspari-Cohn vertical localisation (Gaspari and Cohn, 1999) scheme, which is a Gaussian-like vertical localisation using the logarithm of pressure as its vertical coordinate (Clayton, 2014), was used in the experiments of Bowler *et al.* (2017a).

At the moment the vertical localisation used is the same for all control variables. The experiments of Section 7 use seven or eight modes while Bowler *et al.* (2017a) used 16 modes with Gaspari-Cohn vertical localisation. Increasing the number of vertical modes increases the effective number of degrees of freedom of the ensemble part of the background-error covariance matrix. However, it also increases the size of the control vector used in the minimisation, which has an associated input/output cost when reading and writing these vectors for Hessian preconditioning (Payne, 2011).



(a) Early version.

(b) Recommended version.

Figure 5: Mean vertical localisation matrix derived by `hybrid_diag`, truncated to the number of eigenmodes that account for 90% of the variance. (a) Early version from a lag/shift ensemble and stored in file `Lv_old.nc`. Subsequent work showed that this gave erroneously large values at long separations. (b) Currently recommended version from a normal ensemble, taking the mean from all available variables stored in file `Lv_ar329n216nHnV_20170506.nc` ("Mv2" in Tables 6–7). There is relatively little localisation within the troposphere but strong localisation across and above the tropopause (near level 50).

Ménétrier vertical localisation uses a vertical localisation matrix stored in NetCDF format that is derived using a method that identifies the degree of sampling error present in an ensemble (Ménétrier and Auligné, 2015; Ménétrier *et al.*, 2015a,b). The `hybrid_diag` results currently considered as the best estimates are shown in Figure 5(b).

4 Time lagging and shifting

Time-lagged ensembles (i.e., valid at the correct time, from an earlier forecast) have been used in our ensemble post-processing. Time-shifted ensemble perturbations (i.e., valid at a different time) are used in climatological covariance estimation. Time shifting also has the advantage of allowing a larger range of phase differences to be present in the forecast perturbations, which can be helpful in the presence of timing errors in the propagation of weather systems.

The lag/shift scheme introduced by Lorenc (2017) for 4DEnVar becomes simpler for hybrid-4DVar. We want to represent the forecast error over a single assimilation window (which is currently six hours long). We take ensemble perturbations from the previous cycle's ensemble forecasts, stored every three hours, and use the perturbations valid at the start of the current assimilation window, together with those shifted forwards by three and six hours (two time shifts). This set of ensemble forecast perturbations is then augmented by an equivalent set from the last-but-one cycle (so each forecast is six hours longer) valid at the start of the assimilation window and shifted backwards by three and six hours (one lag). This arrangement gives a symmetrical distribution of offset validity times, ranging from six hours in the past to six hours in the future (see Figure 12 of Lorenc (2017)) and multiplies the number of forecast perturbations by six. The extra perturbations

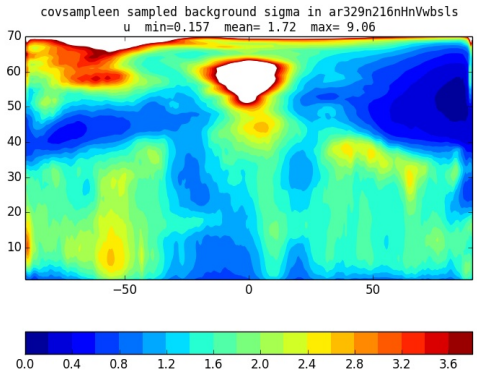


Figure 6: As for Figure 2 but showing the additional impact of time lagging and shifting on the waveband filtered results.

are not truly independent, though; Lorenc (2017) found it was only as effective as a real factor of two in independent members. Figure 6 shows the impact of this lag/shift configuration when used in addition to the waveband localisation of Figure 2.

The lag/shift system is cost effective by virtue of being significantly cheaper than running a larger ensemble. However, in early July it was found that the one lag and two shift setup was too slow for operational implementation due to slow input/output of the control vectors used to store the perturbation information so lagging and shifting was ruled out as a candidate for PS40. Nevertheless, subsequent recoding has substantially increased the speed to make this a viable option for a future parallel suite. Results of this configuration have therefore been included in Section 7 to show the potential benefit of time lagging and shifting.

5 Hybrid covariance weights

The localisation changes of Sections 2–3 have been particularly beneficial in the tropics. The operational hybrid covariance weights have been $\beta_c^2 = 0.9$ and $\beta_e^2 = 0.0$ in the upper atmosphere, $\beta_c^2 = 0.7$ and $\beta_e^2 = 0.3$ in the lower atmosphere's extratropics and $\beta_c^2 = 0.6$ and $\beta_e^2 = 0.3$ in the tropics since operational suite 37 (OS37), which ran from 15 March until 8 November 2016. The improvements in the localisation settings and also in MOGREPS-G since OS37 motivate a re-examination of the appropriate weights to apply.

In this section we seek the optimal weights using the operational static covariance statistics introduced in OS37 together with various horizontal and vertical localisation options. The experiments use the OS37-based low-resolution global research suite **u-ag802** as a control, running in an uncoupled configuration that takes the ensemble perturbation forecasts used in the ensemble covariance component of the hybrid covariances from a stored archive (Table 2). The results are summarised in Table 3.

Giving dominant weight to the ensemble with $\beta_c^2 = 0.3$ and $\beta_e^2 = 0.7$ everywhere without making

Table 2: Experimental setup for low-resolution boreal winter 2015 uncoupled trials, running from 25 December 2015 to 25 March 2016. For vertical localisation, ψ refers to the operational scheme based on streamfunction, while GC refers to the Gaspari-Cohn approach used by Bowler *et al.* (2017a).

Suite ID	Hybrid weights						Vert. loc.		Horiz. loc.	
	Extratropics		Tropics		Upper atmos.			#		
	β_c^2	β_e^2	β_c^2	β_e^2	β_c^2	β_e^2	Type	modes	Type	Scale (km)
u-ag802	0.7	0.3	0.6	0.3	0.9	0.0	ψ	8	high pass	1200
u-ah979	0.3	0.7	0.3	0.7	0.3	0.7	ψ	8	high pass	1200
u-ai514	0.3	0.7	0.3	0.7	0.3	0.7	GC	16	high pass	600
u-ai729	0.3	0.7	0.5	0.5	0.3	0.7	GC	16	high pass	600
u-ai840	0.3	0.7	0.3	0.7	0.5	0.5	GC	16	high pass	600
u-aj444	0.7	0.3	0.6	0.3	0.9	0.0	GC	16	high pass	600
u-aj446	0.3	0.7	0.7	0.3	0.7	0.2	GC	16	high pass	600

Table 3: Results for the uncoupled boreal winter experiments described in Table 2. The absolute changes in the NWP index measured against observations and own analyses are shown together with breakdowns of the percentage RMS difference change for the northern hemisphere (NH), tropics (TR) and southern hemisphere (SH).

Suite ID	#	Observation index				Analysis index			
		Index	RMS Diff (%)			Index	RMS Diff (%)		
	days	change	NH	TR	SH	change	NH	TR	SH
u-ah979	37	–0.52	0.5	1.1	0.0	–0.18	–0.1	2.1	–1.3
u-ai514	37	0.36	0.0	–2.0	0.3	1.72	–1.6	–0.9	–3.7
u-ai729	37	0.41	–0.3	0.7	–0.4	1.03	–1.4	0.7	–3.0
u-ai840	37	0.44	–0.3	0.7	–0.6	1.76	–1.7	–1.5	–3.5
u-aj444	37	0.17	–0.3	0.0	–0.3	0.65	–0.7	–0.8	–0.8
u-aj446	31	0.74	–0.4	0.1	–0.7	1.03	–1.3	0.8	–3.1

other changes is seen to be detrimental (**u-ah979**). On the other hand, the benefit of changing to the Gaspari-Cohn localisation with more vertical modes and also halving the horizontal localisation lengthscale is apparent from experiment **u-aj444** so further experiments use this configuration. With this change, using $\beta_c^2 = 0.3$ and $\beta_e^2 = 0.7$ everywhere now gives overall benefit (**u-ai514**). The remaining experiments explore the benefits of varying the hybrid weights in the three regions.

6 New covariance statistics

6.1 Training data sets

CVT is used to calibrate VAR's underlying covariance model based on training data chosen to represent typical six-hour forecast errors over the calendar year. The latter point is important because there are seasonal differences in the weather and hence different forecast error characteristics in each hemisphere, with the different behaviour of the southern hemisphere storm tracks in summer and winter being a key example. The choice of training data is consequently an important consideration. We have therefore reprocessed the training data used to produce the operational covariance statistics using a recent version of CVT and also used data produced by ETKF and En-4DEnVar

ensembles. For the new training data we have used 880 samples in each case to ensure that the sampling errors are comparable, using T+3 ensemble forecasts from each of the 44 ensemble members and subtracting the ensemble mean on each cycle to obtain the ensemble forecast perturbations read into CVT.

ECMWF The training data for the operational covariance model come from an old ECMWF training data set from 2006 produced by a ten member ensemble of 4DVars. Further details of this training data set are given in Piccolo (2011).

ETKFSummer Data are taken from the 0000 and 1200 UTC cycles from 8–17 June 2015 produced by ETKF suite **mi-aq520**.

ETKFAllYear Data from the 0000 UTC cycles between 1–10 June 2015 produced by suite **u-ad369** are combined with data from the 0000 UTC cycles between 1–10 December 2013 produced by suite **u-ad839**.

EnVarSummer Data are taken from En-4DEnVar suite **mi-aq897**, using the same cycles as **ETKFSummer**.

EnVarAllYear The 0000 UTC training data used to calibrate **EnVarSummer** are combined with 0000 UTC training data from 21–30 January 2016 produced by suite **mi-as462**.

6.2 Latitude-banded Gp regression development

The operational covariance file uses a single geostrophic pressure (Gp) vertical regression matrix to make the unbalanced pressure on average uncorrelated with the streamfunction. It does this by adjusting the geostrophic pressure field slightly so that the updated balanced pressure is statistically uncorrelated with the difference between the hydrostatic pressure and the updated balanced pressure, which we call unbalanced ageostrophic pressure.

The problem with using a single Gp vertical regression matrix is that there is a degree of geostrophic balance that is strongly dependent on latitude, with far less geostrophy near the equator than near the poles (Figure 7). A single vertical covariance matrix cannot take account of this variability.

The unbalanced pressure increment is currently defined as

$$\delta\mathbf{p}^{\mathsf{u}} = \delta\mathbf{p}^{\mathsf{h}} - \mathbf{F}\delta\mathbf{p}^{\mathsf{g}}, \tag{1}$$

where $\delta\mathbf{p}^{\mathsf{h}}$ and $\delta\mathbf{p}^{\mathsf{g}}$ are vertical columns of perturbations to hydrostatic pressure and geostrophically balanced pressure and the vertical regression matrix is

$$\mathbf{F} = \mathbf{B}_{\mathsf{v}}^{\mathsf{hg}}(\mathbf{B}_{\mathsf{v}}^{\mathsf{gg}})^{-1}, \tag{2}$$

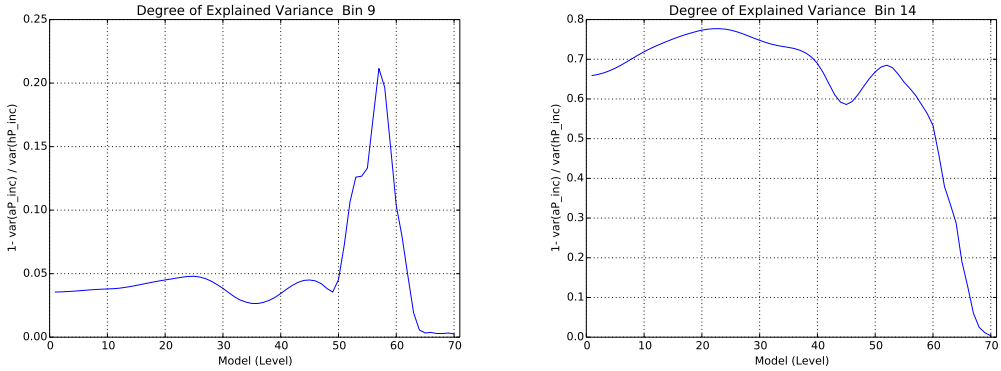


Figure 7: The degree of balance, measured as the difference between the unbalanced and hydrostatic pressure variances scaled by the variance of the hydrostatic pressure, in **ETKFAllYear** training data at the equator (left) and 60 degrees north (right). Zero indicates that the pressure is entirely unbalanced while one represents balanced pressure.

written in terms of the balanced pressure vertical covariance matrix $\mathbf{B}_{\mathrm{v}}^{\mathrm{gg}}$ and the vertical cross-covariance between hydrostatic pressure and balanced pressure $\mathbf{B}_{\mathrm{v}}^{\mathrm{hg}}$.

The latitude-banded Gp regression uses N vertical regression matrices calculated using overlapping latitude-banded weighting functions w_i. In this case, equation (1) becomes

$$\delta\mathbf{p}^{\mathrm{u}} = \delta\mathbf{p}^{\mathrm{h}} - \frac{\sum_{i=1}^{N} w_i \mathbf{F}_i \delta\mathbf{p}^{\mathrm{g}}}{\sum_{i=1}^{N} w_i}, \qquad \mathbf{F}_i = (\mathbf{B}_{\mathrm{v}}^{\mathrm{hg}})_i (\mathbf{B}_{\mathbf{v}}^{\mathbf{gg}})_i^{-1}. \tag{3}$$

This latitude-banded Gp regression has been designed to be smooth and slowly varying in latitude (by multiplying a triangular function by the cosine of latitude) using 18 overlapping bands.

This approach was first applied in a non-hybrid 4DVar context using **ETKFAllYear** training data (Section 6.1). A number of key points came out of these experiments:

1. Latitude-banded Gp regression is needed when the training data have a significant multivariate component (as seen in ETKF and En-4DEnVar training data). Note that the **ECMWF** training data set did not really need this as it was predominantly unbalanced.

2. It is possible to use an ensemble training data set to generate a covariance file with latitude-banded Gp regression that is comparable in terms of verification to that of the current operational covariance file in a non-hybrid uncoupled context (Figure 8).

3. ETKF training data have shorter horizontal length scales, much more multivariate coupling and broader vertical correlations compared to the **ECMWF** training data. Running VAR with the static covariance file leads to significantly larger condition numbers (and hence more minimisation iterations) compared to running with the operational **ECWMF**-based covariance file.

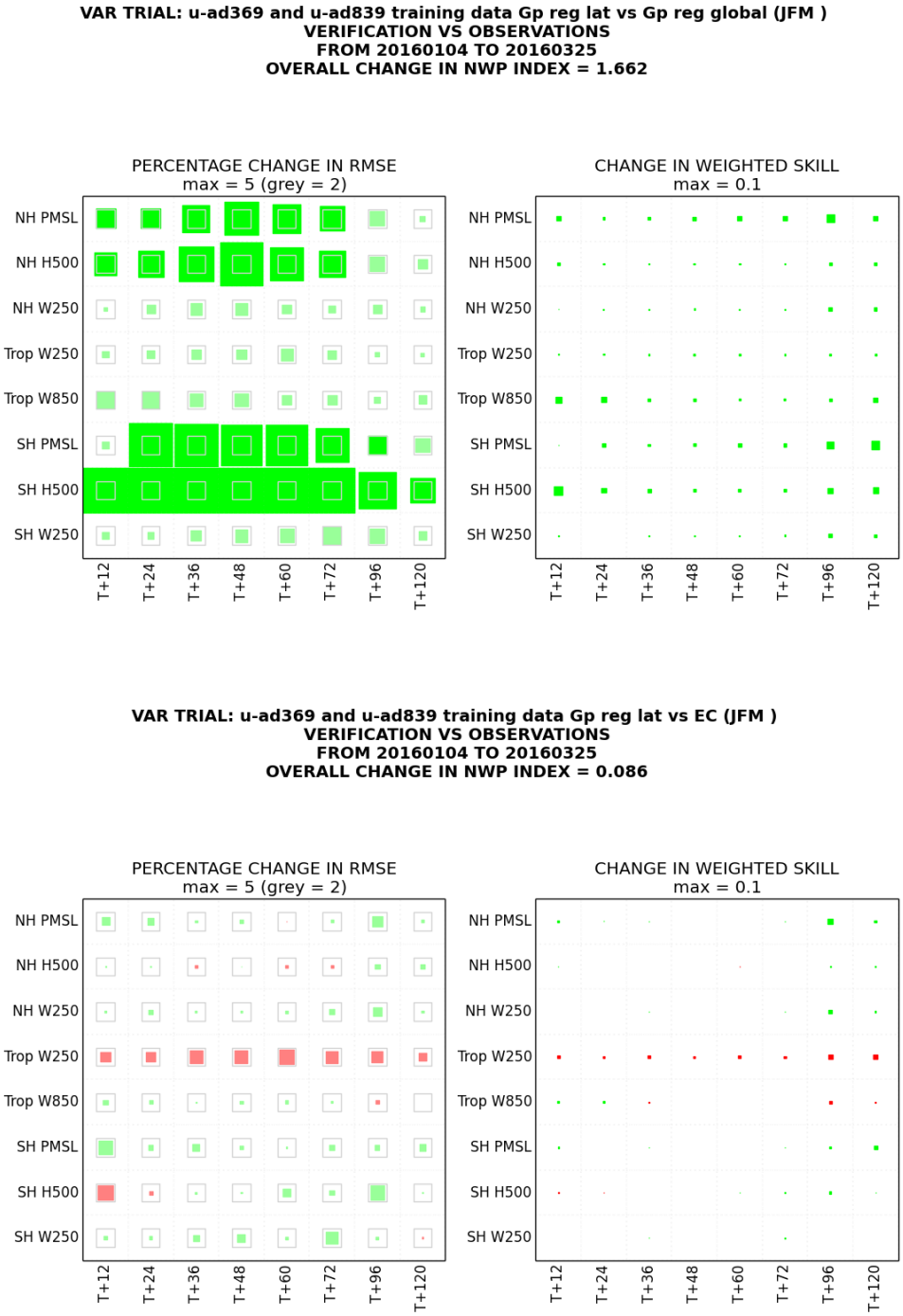


Figure 8: The top row shows the improved forecast verification versus observations when using latitude-banded Gp regression compared to the standard Gp regression technique, generating covariance files from **ETKFAllYear** training data. The bottom row shows that the covariance file using latitude-banded Gp regression is comparable to the performance obtained using the operational **ECMWF** based covariance file. The graphs show the contribution to the NWP index from pressure at mean sea level (PMSL), geopotential height (H) and wind (W) at 850, 500, 250 hPa altitudes in the Northern Hemisphere (NH), Tropics (Trop) and Southern Hemisphere (SH) for lead times given on the horizontal axis. A hollow grey square denotes a 2% difference while the biggest difference that can be plotted is 5% when fully coloured. The squares are green if the first experiment is better and red if the second experiment is better. The figure captions show the combined NWP index changes across all components.

Table 4: Comparison of consistent experiments where the same ensemble source is used to calibrate the static covariance file and provide the flow-dependent ensemble forecast perturbations for hybrid 4DVar. Performance is measured relative to the enhanced control **mi-ar938**. The "EnVar aP0" data source refers to an experiment where the wavenumber zero unbalanced pressure vertical covariance matrix is transplanted from the operational **ECMWF** derived covariance file into the **EnVarSummer** derived covariance file (Section 6.4).

Suite ID	Data source		Observation index				Analysis index			
		# days	Index change	RMS Diff (%) NH	TR	SH	Index change	RMS Diff (%) NH	TR	SH
mi-ar781	**ETKFSummer**	43	–0.22	–0.2	0.5	–0.1	1.68	–2.3	–4.3	–2.6
mi-ar796	**EnVarSummer**	43	0.28	–0.3	0.9	–0.3	2.37	–3.2	–5.5	–3.7
mi-as248	**EnVar aP0**	43	0.76	–0.6	–0.1	–0.8	2.87	–3.6	–5.8	–4.2

6.3 ETKF versus En-4DEnVar training data

In this section we concentrate on "consistent experiments" where the same ensemble setup is used to calibrate the static covariance file and provide the flow-dependent ensemble perturbations used in hybrid VAR. One way to do this is to use the forecast perturbations midway through an ensemble experiment to calibrate the static covariance file. By using this approach we can identify which aspects of the ensemble provide benefit.

The control suite **mi-ar938** is based on the OS37 suite but using Gaspari-Cohn vertical localisation with 16 empirical vertical modes for the alpha control variable and a horizontal localisation length scale of 600 km following Bowler *et al.* (2017a). The experiments use the same localisation scheme and beta weights (the same as suite **u-aj444** in Table 2); only the ensemble forecast perturbations and the static covariance files differ.

We see in Table 4 that the experiment based on **EnVarSummer** training data (suite **mi-ar796**) performs better overall in terms of the overall NWP index score compared to our enhanced control and a trial based on **ETKFSummer** training data (suite **mi-ar781**). Almost all the observations have smaller innovations and the increments are substantially more balanced, as measured by the final value of the Jc gravity wave penalty term (not shown). Nevertheless, Figures 9–10 show that there are still regions where the performance is degraded, particularly in the tropics, compared to the enhanced control. This is addressed in Section 6.4.

The ETKF derived covariance statistics led to a much more poorly conditioned VAR configuration, as measured by the leading Hessian eigenvalue on each cycle (not shown). The broader vertical correlations of the ETKF derived covariances are believed to be responsible for this effect. This problem will only get worse when running with the operational N144 / N320 VAR grids rather than the N108 / N216 grids used in **mi-ar781**, increasing the computational cost and further increasing the risk of VAR minimisation failures. Suite **mi-ar781** also suffered a UM forecast failure (bi-conjugate-gradient convergence failure) on the 1800 UTC cycle of 15 June 2015.

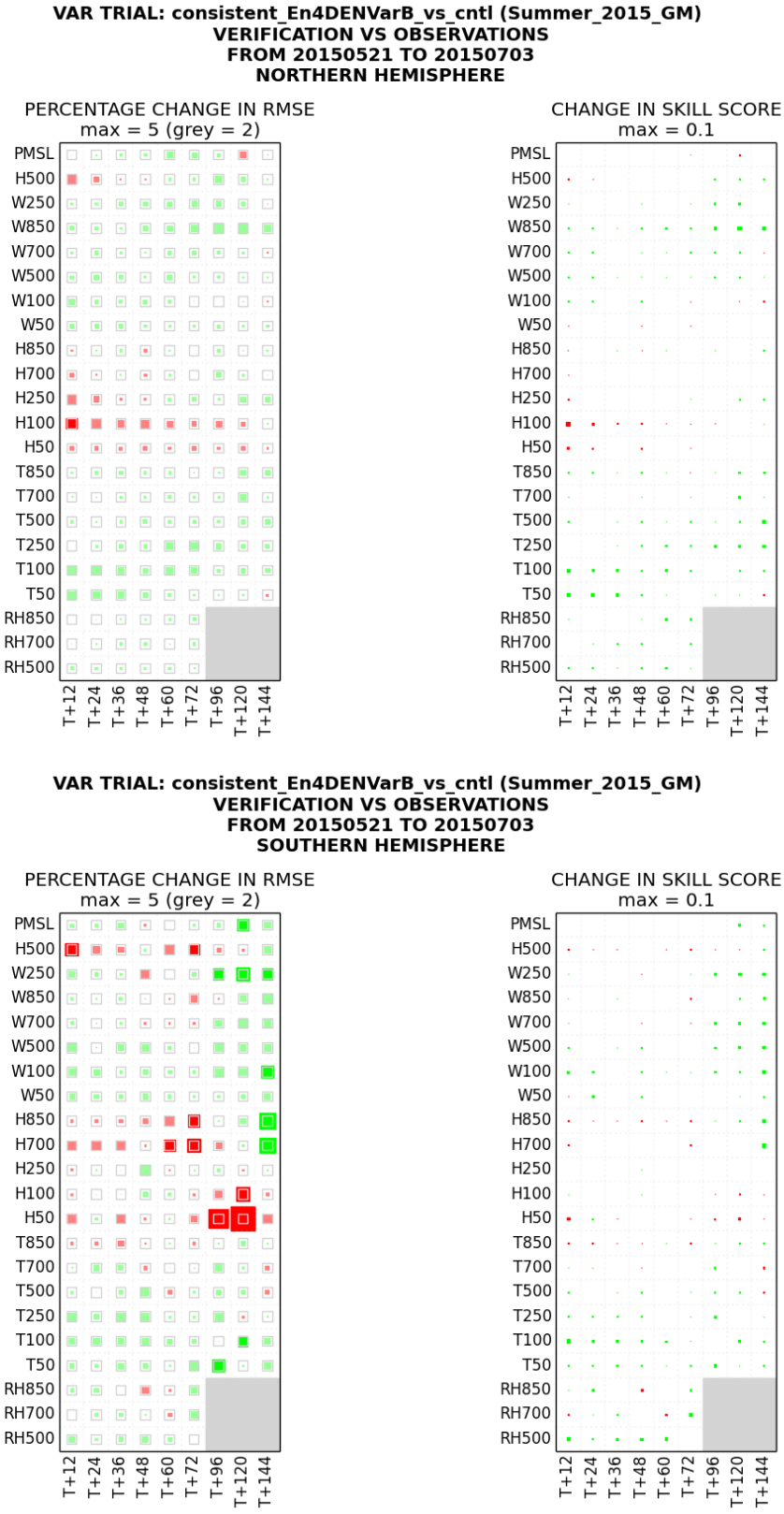


Figure 9: Verification versus observations for En-4DEnVar experiment **mi-ar796** compared to the enhanced control **mi-ar938** for the northern (top) and southern (bottom) hemisphere extratropics. The figures follow the same format as Figure 8 with the addition of fields at 700, 100 and 50 hPa together with temperature (T) and relative humidity (RH).

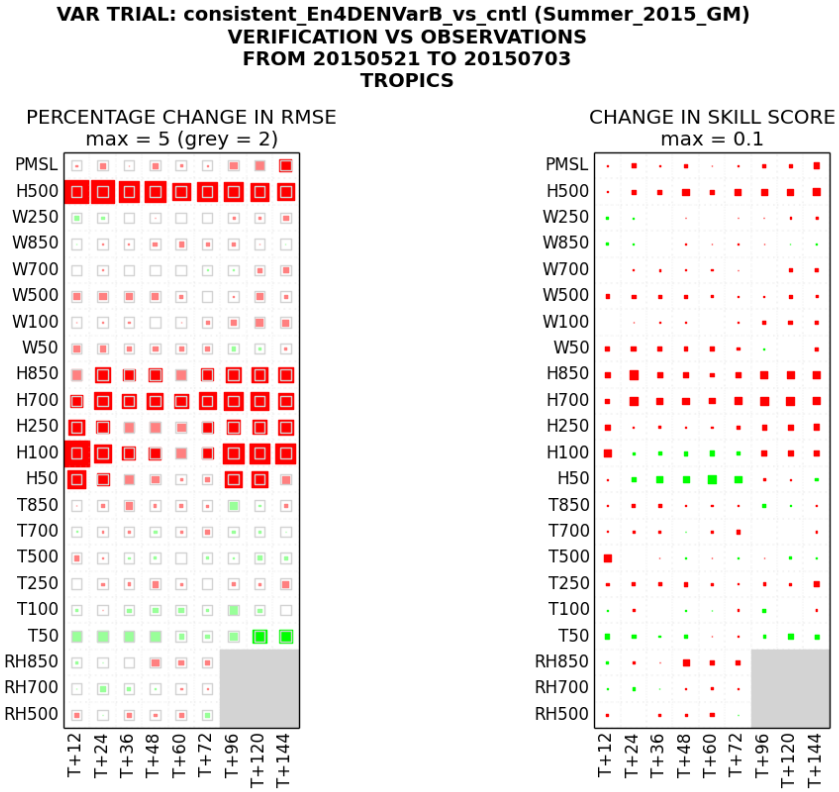


Figure 10: As for Figure 9 but in the tropics.

6.4 Constraining variational bias correction

In this section we further examine the weaker performance of the En-4DEnVar consistent experiment compared to the enhanced control as shown by Figures 9–10. Temperature, wind and surface pressure show signs of improvement in the extratropics but the geopotential height is notably degraded, suggesting a problem with the pressure covariance structure. Further study of the final mean values of the first Ap segment of the control vector in VAR job.out files (not shown) shows that the analysis increments in minimisation space are trying to correct a bias in the lowest wavenumbers of the unbalanced pressure.

The structure of the unbalanced pressure horizontal wavenumber zero vertical covariance matrix in the covariance file derived from **EnVarSummer** training data is fundamentally different to that of the **ECMWF** derived operational covariance file (Figure 11). The vertical covariance from the **ECMWF** training data peaks at the surface, while the **EnVarSummer** training data set peaks near model level 30. This motivated an experiment where the vertical covariance matrix for this mode was replaced by the corresponding matrix in the ECMWF derived covariance file — this file surgery is easily done with the nco tools for manipulating NetCDF files. The benefits of making this change are evident in Table 4 and Figures 12–13. Innovations and residuals are generally smaller, as well as Jc and variational bias correction (VarBC, Lorenc (2012)) penalties (not shown).

Variational bias correction is designed to distinguish between different sources of bias. Different

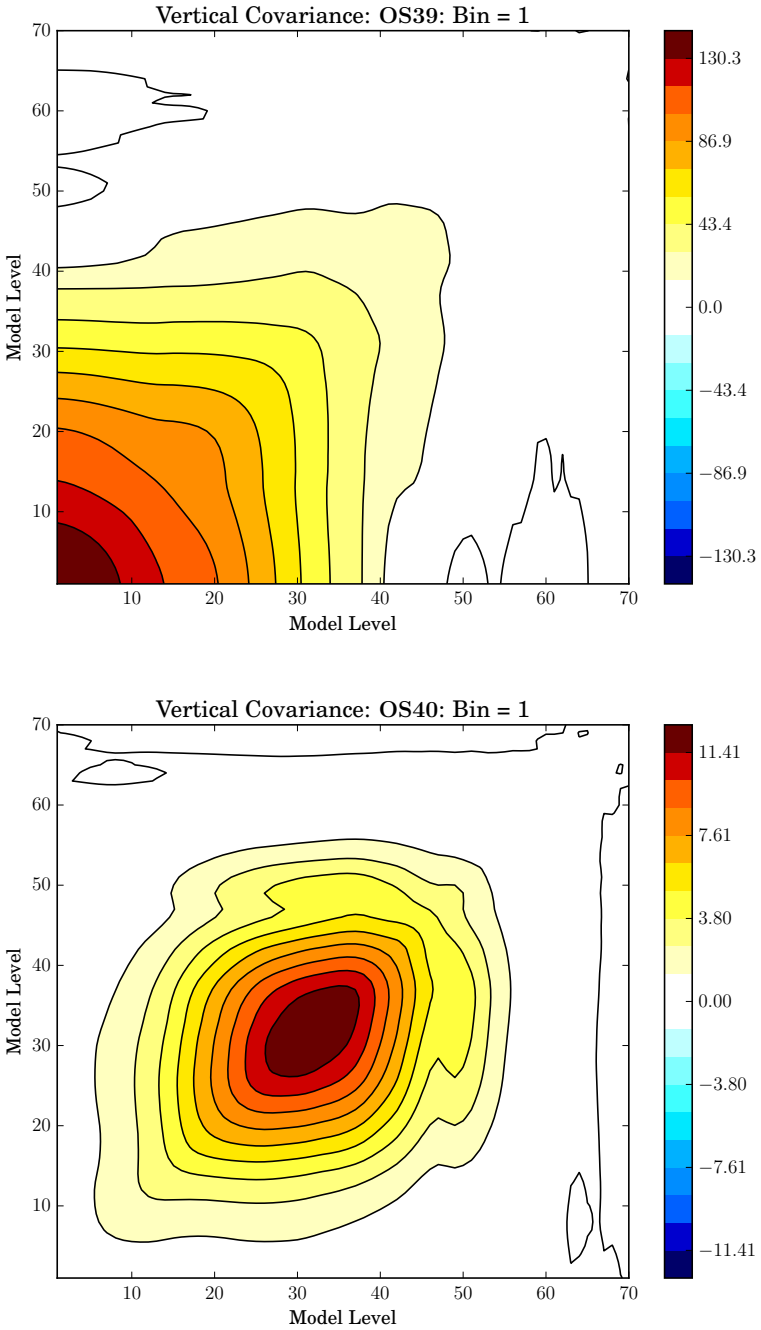


Figure 11: The vertical covariance matrices for the horizontal wavenumber zero component of the unbalanced pressure, where the horizontal and vertical axes are labelled in terms of model levels. The top plot shows vertical covariances from the operational **ECMWF** derived covariance file while the bottom plot shows the vertical covariance from the **EnVarSummer** derived covariance file. Note the different legend ranges.

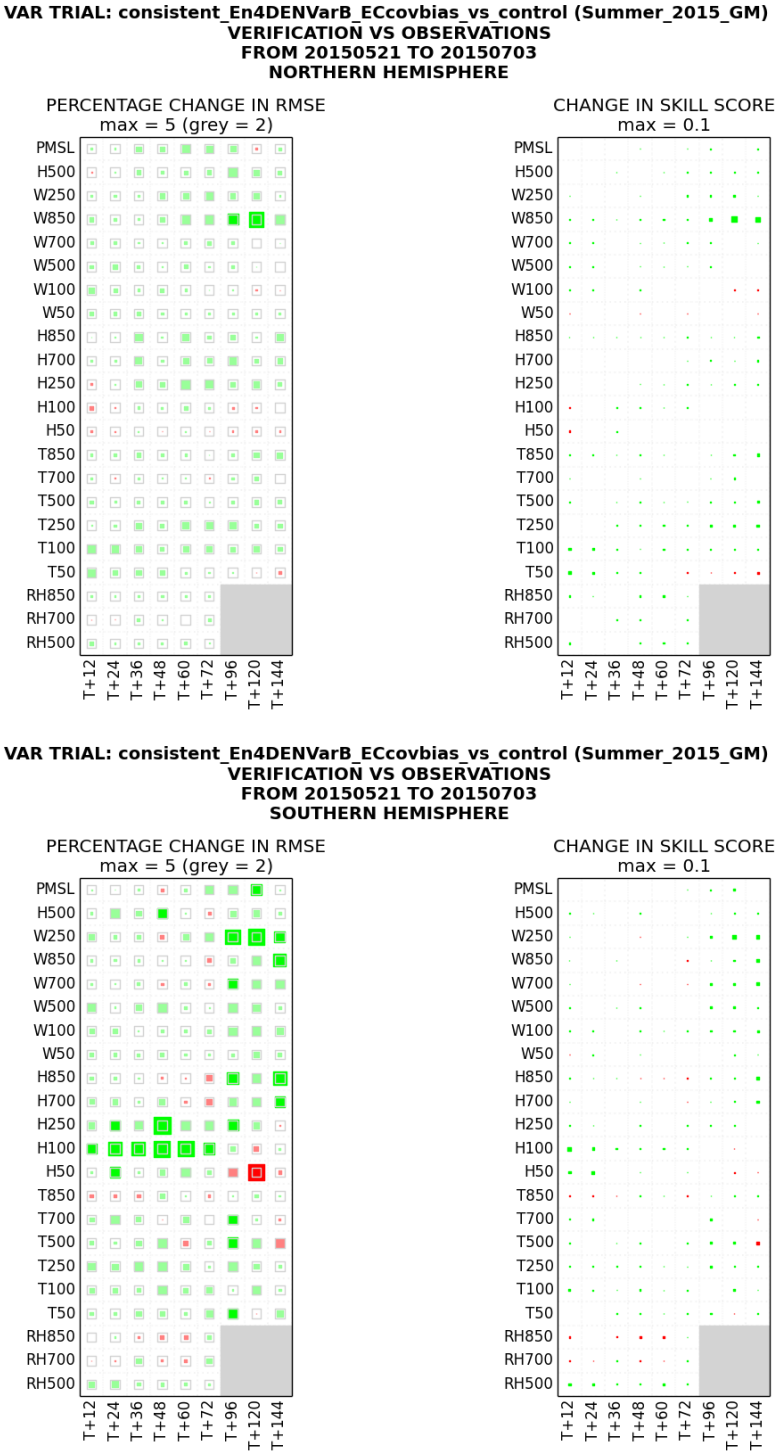


Figure 12: As for Figure 9 but after replacing the unbalanced pressure wavenumber zero vertical covariance matrix in the **EnVarSummer** derived covariance file with the corresponding matrix from the operational **ECMWF** derived covariance file.

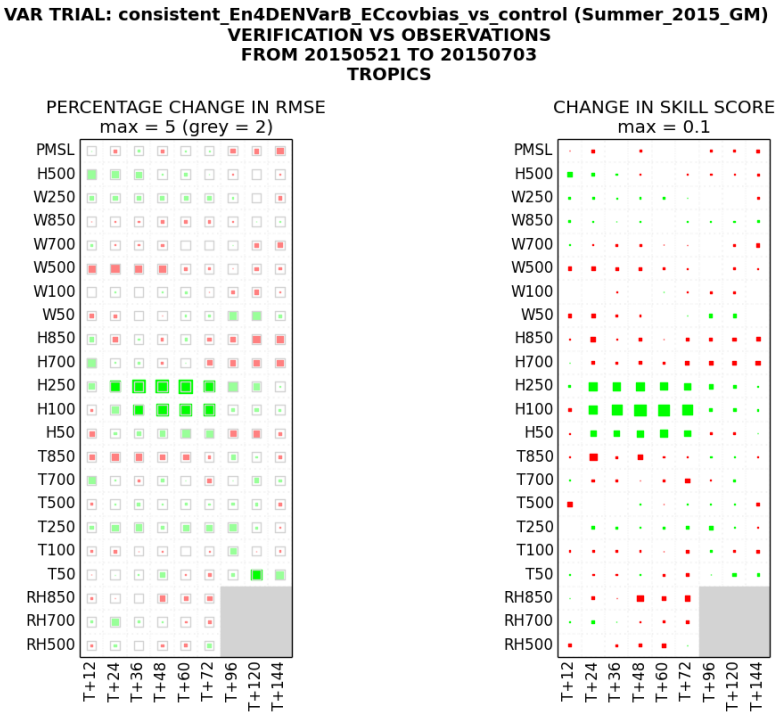


Figure 13: As for Figure 10 but after replacing the unbalanced pressure wavenumber zero vertical covariance matrix in the **EnVarSummer** derived covariance file with the corresponding matrix from the operational **ECMWF** derived covariance file.

reference observation types are used to establish whether a bias is in the model or the observation under consideration. For instance, if a consistent bias is seen for different types of observations then it could be considered as a model bias. However, this can only be corrected if the background-error covariance matrix allows the increment to have a sufficiently large value in the region of interest, which will not happen if the scaled innovation lies in the null-space of the background-error covariance matrix. For this reason anchor observations are key and need to be given sufficient weight in preference to observations whose bias is controlled by VarBC. The **ECMWF** derived covariance file for the unbalanced pressure in wavenumber zero achieves this by virtue of having larger background-error variances near the surface, where most anchor observations reside, compared to the **EnVarSummer** derived covariance file.

6.5 Surface temperature impacts

The impacts of the combined changes of horizontal waveband localisation, Ménétrier vertical localisation and varying hybrid weights (Sections 2, 3 and 5) have been combined with the modified **EnVarSummer** covariance file tested in Section 6.4 for a winter experiment using the hybrid weights $\beta_c^2 = \beta_e^2 = 0.5$ everywhere (suite **u-al830**). The control is the OS39-based low-resolution coupled suite **u-al877**, described in Table 6.

Despite the impressive scores in Table 5, the degradation in verification of surface temperatures,

Table 5: The "EnVar aP0" covariance file is the modified **EnVarSummer** covariance file of Section 6.4. The winter control suite **u-al877** ran for 66 days.

Suite ID	Data source	# days	Observation index				Analysis index			
			Index change	RMS Diff (%) NH	TR	SH	Index change	RMS Diff (%) NH	TR	SH
u-al830	EnVar aP0	66	1.46	–0.6	–0.2	–1.2	3.64	–3.8	–2.9	–4.5

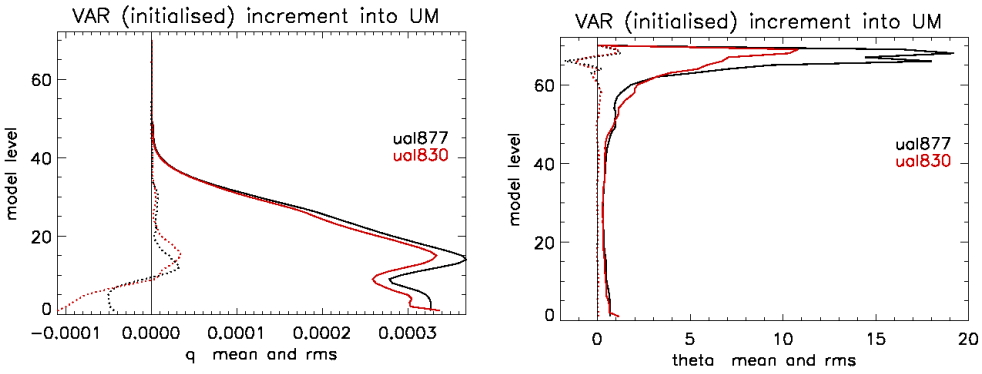


Figure 14: The average vertical profiles for the specific humidity analysis increment (left) and the potential temperature analysis increment (right). The dotted lines show the mean values of the increments, while the full lines are the RMS values. The red lines describe the trial using the modified **EnVarSummer** file of Section 6.4 and the black lines denote the standard winter control using the operational **ECMWF** derived covariance file.

especially over land, is too large to be a viable contender. For instance, the UK index: Northern Hemisphere Land (10111) results (not shown) indicate that the RMS error of surface temperatures verified against observations is 8.4% higher. Figure 14 shows that the averaged mean specific humidity analysis increment is twice as large in the experiment as in the control near the surface while the averaged RMS potential temperature increment is much smaller in the stratosphere.

The **EnVarAllYear** training data were used to create a new covariance file with the unbalanced pressure wavenumber zero vertical covariance matrix taken from the **ECMWF** covariance file in order to make the covariances more applicable all year round. Having done this, we compared the humidity wavenumber zero vertical covariance matrix from the resulting covariance file with that of the operational **ECMWF** derived covariance file (Figures 15–16). The **EnVarAllYear** covariance file has a much stronger coupling from the troposphere to the surface in wavenumber zero than the **ECMWF** derived covariance statistics, which are more dominated by stratospheric error. This allows mid-tropospheric observations to influence surface temperature and specific humidity. To counter this, the humidity wavenumber zero vertical covariance matrix was also replaced with the corresponding matrix from the operational covariance **ECMWF** derived covariance file to produce a new covariance file, which is tested in Section 7.

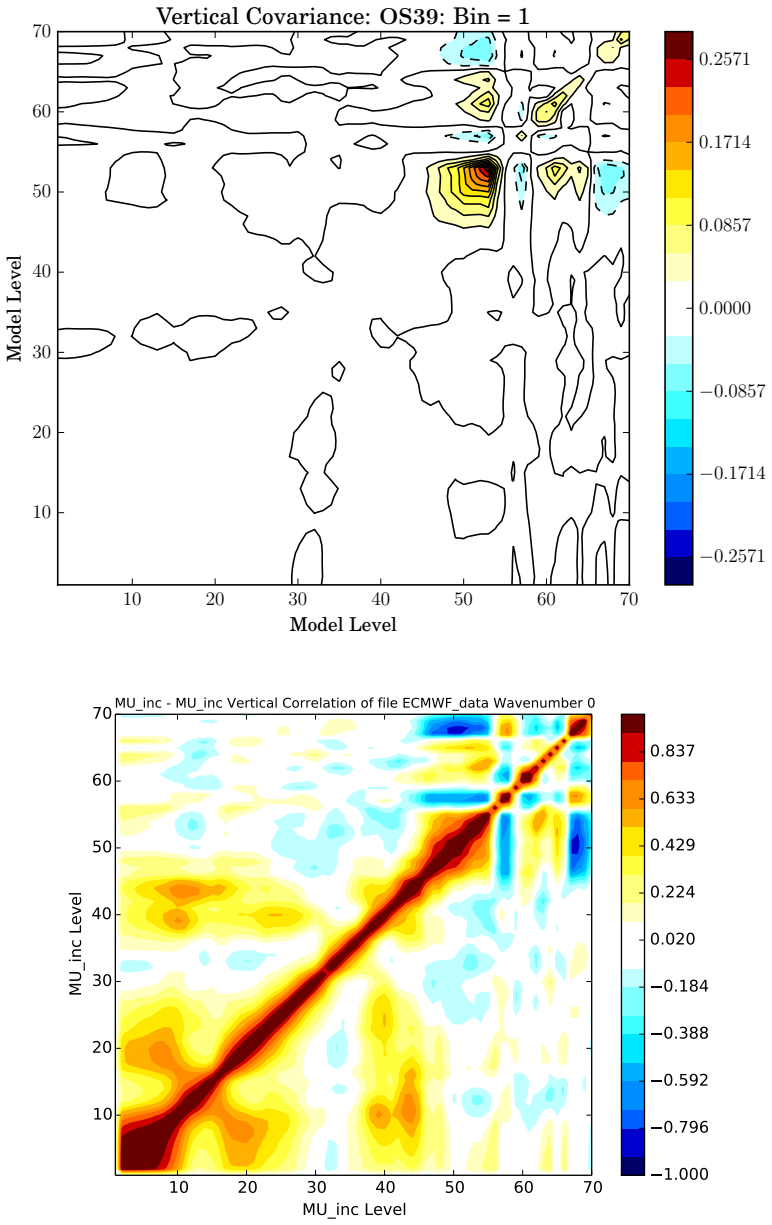


Figure 15: Vertical covariances (top) and correlations (bottom) for the horizontal wavenumber zero component of the humidity control variable taken from the operational **ECMWF** derived covariance file.

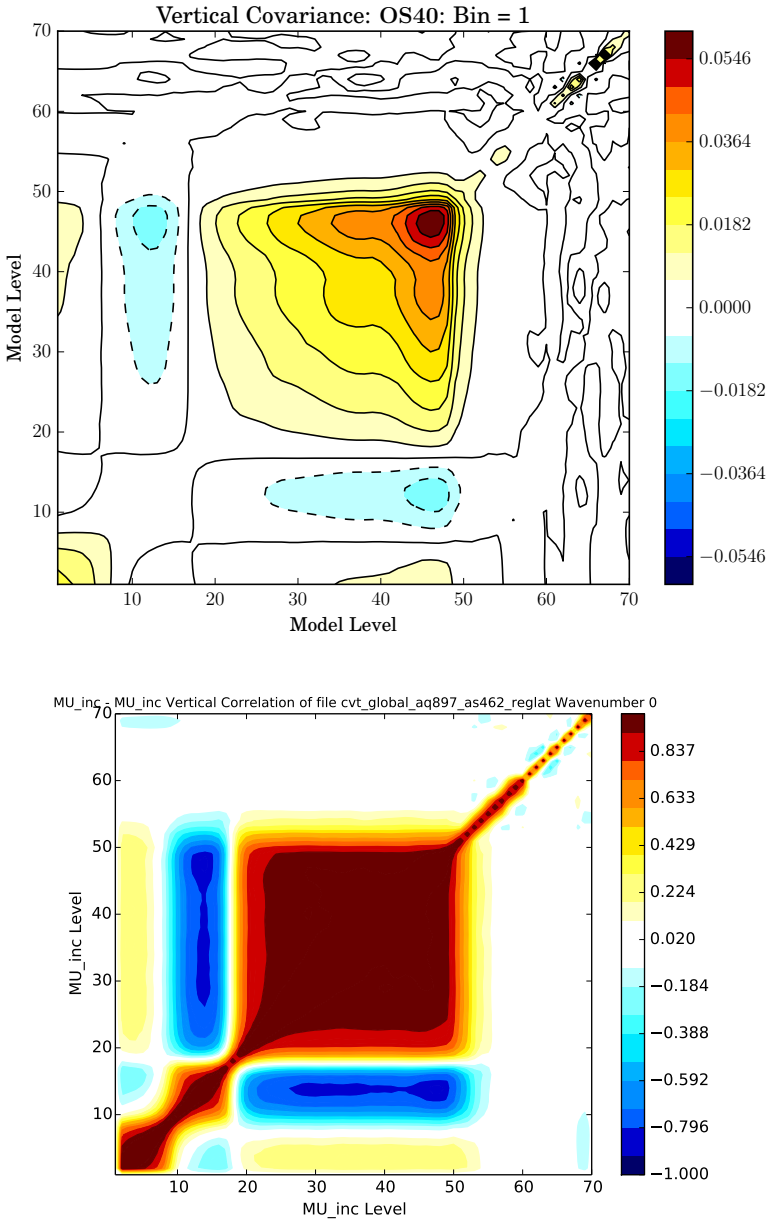


Figure 16: As for Figure 15 but using the **EnVarAllYear** covariance file (Section 6.5). Note the different legend range in the top figure compared to Figure 15.

7 Component trials

After some initial experiments two component trial configurations have been chosen. The first is based on the operational covariance file introduced in OS37, calibrated using **ECMWF** training data, while the other uses the covariance file of Section 6.5, calibrated using the **EnVarAllYear** training data with the wavenumber zero vertical covariance matrices for unbalanced pressure and humidity taken from the OS37 covariance file. The control suites used for these trials were **u-al877** (15 November 2016 to 20 February 2017, 67 days) and **u-al878** (1 July to 30 September 2016, 92 days). Both controls are run in coupled mode, running both the deterministic data assimilation and the ETKF MOGREPS-G ensemble, but at low resolution (N320 deterministic forecast, N216 ensemble forecasts, N108 VAR preconditioning step and N216 VAR update step). Faced with the need to provide evidence for a 4 July meeting to decide on the PS40 package trials two compromises were made. Firstly, only short ensemble forecasts to T+9 were run for the purpose of cycling the data assimilation. Secondly, an early version of the OS39-based global research suite **u-ak375** was used. The controls and experiments described in Tables 6–7 have then been updated while they have been running to incorporate later changes to the research suite as it evolved to its final configuration. This has meant that changes have been applied to the controls and experiments on different cycles, depending on how far they have progressed. An exception to this was a decision to keep the same station lists in the controls and experiments throughout the trials. Another issue adversely affected the running of the control and experiments. The global research suite is configured to trigger a short step forecast whenever an ensemble forecast fails, for whatever reason. The low-resolution configuration's execution times had not been tuned when we started running our suites so several short step forecasts have run purely as a result of the execution time being set too low. Even when an optimal setting had been found, however, slow running on the Cray meant that more short step forecasts were subsequently triggered. Finally, some experiments had to be stopped early in order to release computing resources for the PS40 package trials.

Tables 6–7 show an overview of the main winter and summer low-resolution coupled trials that have been run over the PS40 trialling periods, with verification results summarised in Tables 8– 9.

As alluded to in Section 2 it takes too long and is too computationally expensive to exhaustively tune all the options in Tables 6–7. An initial attempt (experiment OperCov1) was made to see whether the horizontal and vertical localisation choices used by Bowler *et al.* (2017a) together with the operational covariance file would allow more weight to be given to the ensemble in the lower atmosphere — some weight was given to the ensemble in the upper atmosphere but experience suggested that it would unwise to allow it to be the dominant component there. The results of Tables 8–9 show that this is indeed beneficial, although the detailed verification plots (not shown) indicate some detrimental impact in the tropics, particularly when verifying against own analyses. The remaining experiments adopted the horizontal waveband localisation and Ménétrier vertical localisation choices of Sections 2–3, noting that two different vertical localisation matrices (labelled

Table 6: Experimental setup for low-resolution winter 2016 trials, running from 15 November 2016 to 20 February 2017. Full coupling between the ensemble and data assimilation has been used. For vertical localisation, ψ refers to the operational scheme based on streamfunction, GC refers to the Gaspari-Cohn approach used by Bowler *et al.* (2017a) while Mv1 and Mv2 (the latter plotted in Figure 5(b)) represent two different vertical covariance matrices derived using Ménétrier's `hybrid_diag` package (Section 3). The experiments identified as "OperCov" use the covariance file that has been operational since OS37 while those identified as "EnVarCov" use the covariance file of Section 6.5, calibrated using the **EnVarAllYear** training data with the wavenumber zero vertical covariance matrices for unbalanced pressure and humidity taken from the OS37 covariance statistics.

Experiment	Suite ID	Hybrid weights						# lags	# shifts	Vert. loc.		Horiz. loc.	
		Extratropics		Tropics		Upper atmos.				Type	# modes	Type	Scale (km)
		β_c^2	β_e^2	β_c^2	β_e^2	β_c^2	β_e^2						
Control	u-al877	0.7	0.3	0.6	0.3	0.9	0.0	0	0	ψ	8	high pass	1200
OperCov 1	u-am209	0.3	0.7	0.7	0.3	0.7	0.2	0	0	GC	16	high pass	600
OperCov 2	u-am479	0.5	0.5	0.5	0.5	0.5	0.5	1	2	Mv1	8	waveband	6241, 919, 389, 256
OperCov 3	u-an218	0.5	0.5	0.5	0.5	0.5	0.5	1	0	Mv2	7	waveband	6241, 919, 389, 256
OperCov 4	u-an276	0.5	0.5	0.5	0.5	0.5	0.5	0	0	Mv2	7	waveband	6241, 919, 389, 256
EnVarCov 1	u-am652	0.7	0.3	0.7	0.3	0.63	0.27	1	2	Mv1	8	waveband	6241, 919, 389, 256
EnVarCov 2	u-an268	0.7	0.3	0.7	0.3	0.63	0.27	1	0	Mv2	7	waveband	6241, 919, 389, 256
EnVarCov 3	u-an269	0.7	0.3	0.7	0.3	0.63	0.27	0	0	Mv2	7	waveband	6241, 919, 389, 256

Table 7: As for Table 6 but for the summer trials, running from 1 July to 30 September 2016.

Experiment	Suite ID	Hybrid weights						# lags	# shifts	Vert. loc.		Horiz. loc.	
		Extratropics		Tropics		Upper atmos.				Type	# modes	Type	Scale (km)
		β_c^2	β_e^2	β_c^2	β_e^2	β_c^2	β_e^2						
Control	u-al878	0.7	0.3	0.6	0.3	0.9	0.0	0	0	ψ	8	high pass	1200
OperCov 1	u-am210	0.3	0.7	0.7	0.3	0.7	0.2	0	0	GC	16	high pass	600
OperCov 2	u-am480	0.5	0.5	0.5	0.5	0.5	0.5	1	2	Mv1	8	waveband	6241, 919, 389, 256
OperCov 3	u-an217	0.5	0.5	0.5	0.5	0.5	0.5	1	0	Mv2	7	waveband	6241, 919, 389, 256
OperCov 4	u-an275	0.5	0.5	0.5	0.5	0.5	0.5	0	0	Mv2	7	waveband	6241, 919, 389, 256
EnVarCov 1	u-am651	0.7	0.3	0.7	0.3	0.63	0.27	1	2	Mv1	8	waveband	6241, 919, 389, 256
EnVarCov 2	u-an266	0.7	0.3	0.7	0.3	0.63	0.27	1	0	Mv2	7	waveband	6241, 919, 389, 256
EnVarCov 3	u-an267	0.7	0.3	0.7	0.3	0.63	0.27	0	0	Mv2	7	waveband	6241, 919, 389, 256

Table 8: Results for the coupled winter experiments described in Table 6, following the format of Table 3.

Experiment	Suite ID		Observation index				Analysis index			
		#	Index	RMS Diff (%)			Index	RMS Diff (%)		
		days	change	NH	TR	SH	change	NH	TR	SH
OperCov 1	u-am209	83	0.61	–0.5	0.1	–0.6	0.94	–1.8	0.3	–2.6
OperCov 2	u-am479	90	0.61	–0.3	–0.2	–0.5	0.86	–0.8	1.0	–1.5
OperCov 3	u-an218	51	0.51	0.0	–0.2	–0.7	0.69	–0.6	0.7	–1.9
OperCov 4	u-an276	81	0.11	0.2	–0.1	–0.5	0.39	–0.5	0.9	–1.6
EnVarCov 1	u-am652	90	0.93	–0.4	–0.2	–0.9	3.05	–3.3	–3.4	–4.4
EnVarCov 2	u-an268	90	0.80	–0.5	–0.2	–0.7	2.99	–3.4	–3.6	–4.4
EnVarCov 3	u-an269	90	0.70	–0.4	–0.1	–0.7	2.94	–3.3	–3.5	–4.4

Table 9: Results for the coupled summer experiments described in Table 7, following the format of Table 3.

Experiment	Suite ID		Observation index				Analysis index			
		#	Index	RMS Diff (%)			Index	RMS Diff (%)		
		days	change	NH	TR	SH	change	NH	TR	SH
OperCov 1	u-am210	66	0.23	–0.3	0.1	–0.4	0.21	–1.9	0.4	–2.2
OperCov 2	u-am480	83	0.98	–0.6	–0.5	–0.6	1.08	–1.7	0.2	–0.9
OperCov 3	u-an217	48	0.77	–0.5	–0.2	–0.7	0.87	–1.5	0.5	–1.3
OperCov 4	u-an275	71	0.71	–0.6	–0.3	–0.2	1.02	–1.9	0.3	–0.5
EnVarCov 1	u-am651	83	1.26	–1.0	–0.7	–0.8	3.13	–4.2	–4.2	–3.1
EnVarCov 2	u-an266	83	1.10	–0.9	–0.7	–0.7	2.83	–4.1	–4.4	–3.0
EnVarCov 3	u-an267	83	1.27	–1.0	–0.6	–0.7	2.98	–4.2	–4.3	–3.1

"Mv1" and "Mv2") have been used.

We start by discussing the experiments without time lagging and shifting, namely OperCov4 and EnVarCov3, both configurations using the Mv2 vertical localisation matrix recommended in Section 3. The OperCov4 configuration seeks to take advantage of the improved consistency obtained when using wavebands (Section 2) to adopt the same hybrid weights in all regions. Only a single set of weights, giving equal emphasis to the static and ensemble covariance components, has been tested and the number of vertical eigenmodes was kept at seven for consistency with the lagging and shifting experiments, where the number of modes was reduced to keep the computational cost lower. Even with these judicious choices and compromises, Tables 8–9 show that this configuration is still beneficial, although Figures 17–18 indicate some drop in performance at longer lead times.

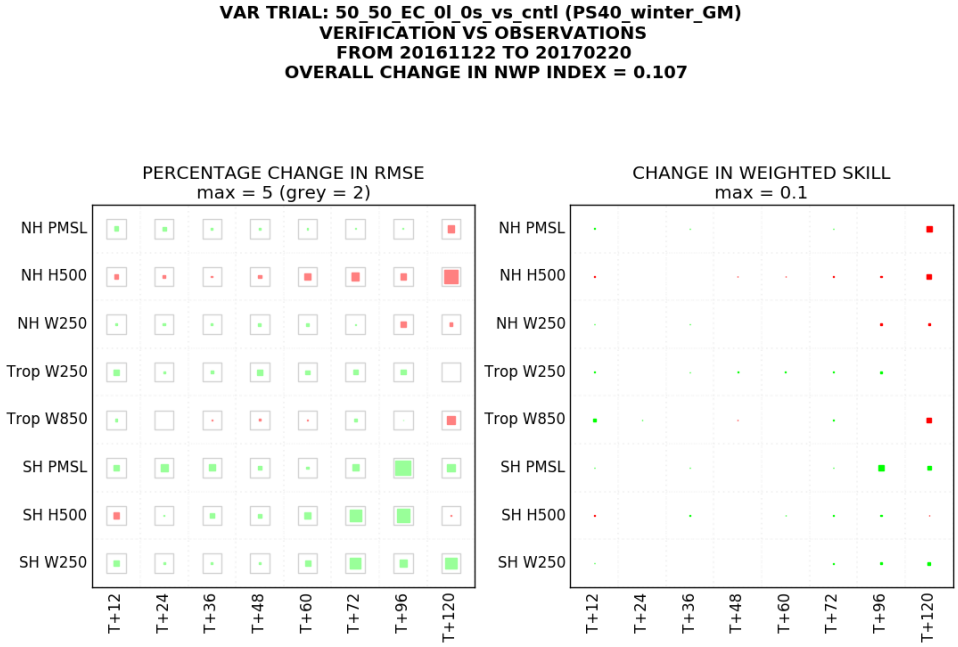
The new covariance file of Section 6.5 has benefited from some hybrid weight tuning, as a result of which it was decided to stay close to the operational weights by giving greater emphasis to the static covariances. Having said this, it was found that adopting slightly reduced static covariance weights in the upper atmosphere was beneficial. Nevertheless, this set of weights is a large step towards the ambition of having more-consistent weights in all regions. The remaining settings are the same as those of the OperCov4 configuration. Tables 8–9 show that the EnVarCov3 configuration produces the best overall performance for the experiments without lagging and shifting. This is further reinforced by Figures 19–20.

The remaining experiments have been included to show the potential benefit that might be expected from a future implementation of time lagging and shifting. Configurations OperCov2 and EnVarCov1 both outperform configurations OperCov4 and EnVarCov3, respectively, although it is worth noting that the performance change with and without lagging and shifting is lower when using the En-4DEnVar derived covariance file. Configurations OperCov3 and EnVarCov2, which use one lag and no shifting and so only double the size of the effective ensemble compared to the factor six increase of configurations OperCov2 and EnVarCov1, were included as a less costly setup but are still delivering levels of performance close to their more costly counterparts.

Based on these experiments, the OperCov4 and EnVarCov3 configurations were chosen as candidates for PS40 package trials.

8 Operational implementation

Following parallel suite trialling, PS40 went operational as OS40 on 13 February 2018. All aspects of the EnVarCov3 data assimilation package described in Table 6 were implemented without lagging and shifting. Unfortunately, due to a setup error the wavebands were implemented incorrectly, with Figure 1 showing the difference between the intended and implemented wavebands.



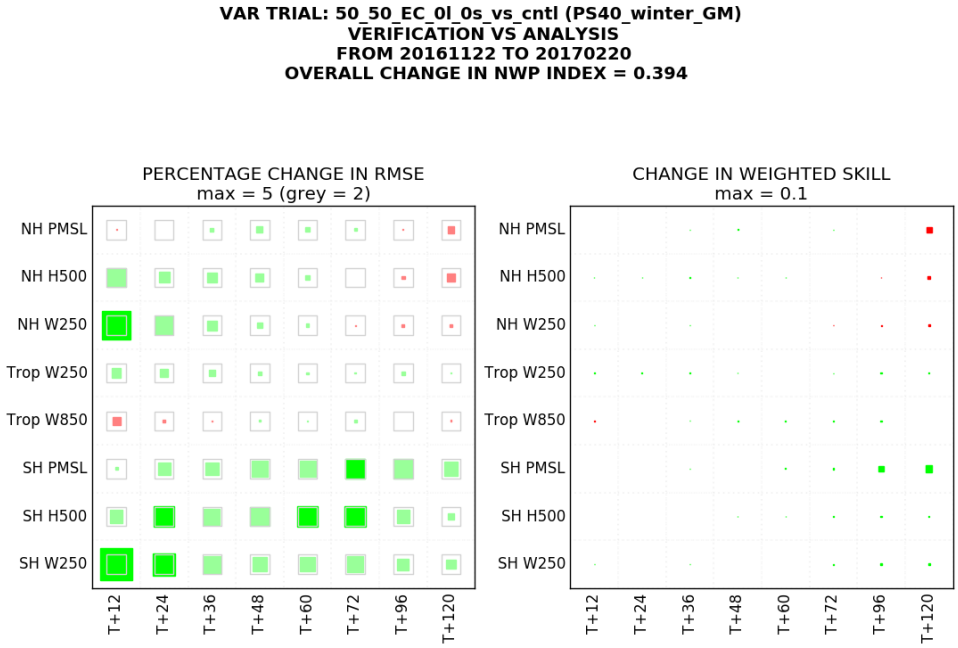


Figure 17: Verification against observations (top) and own analyses (bottom) for configuration OperCov4 in boreal winter.

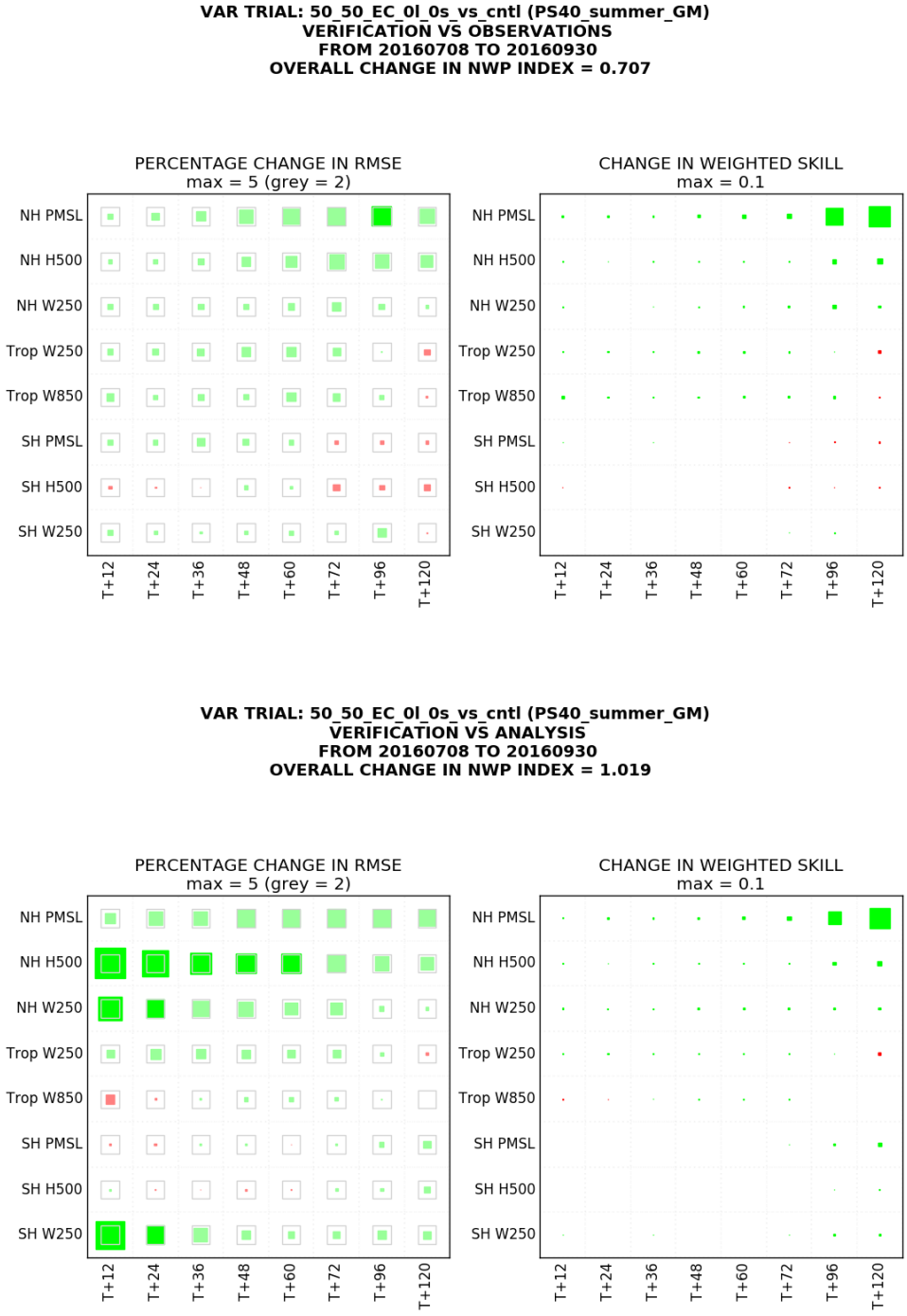


Figure 18: As for Figure 17 but in boreal summer.

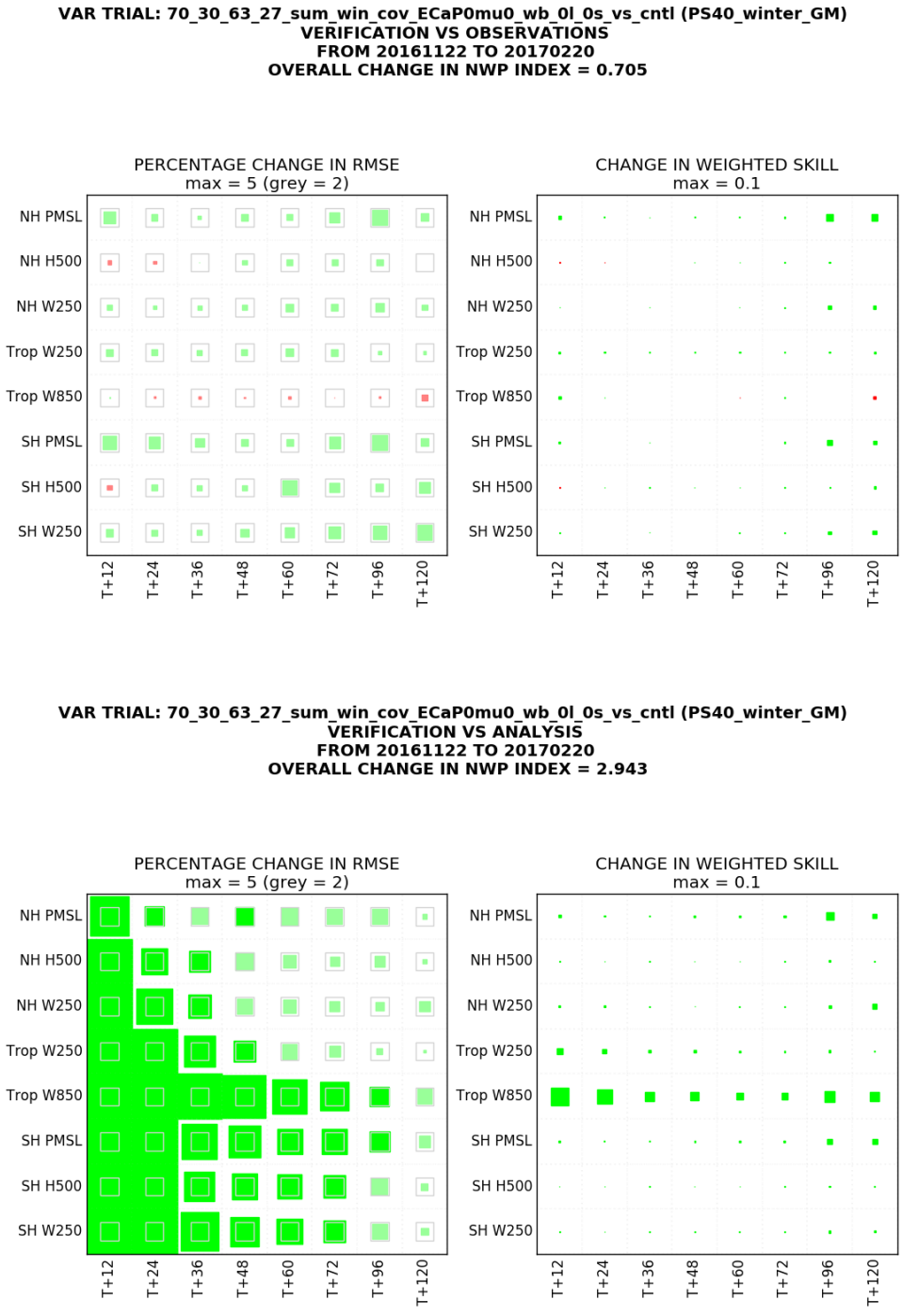
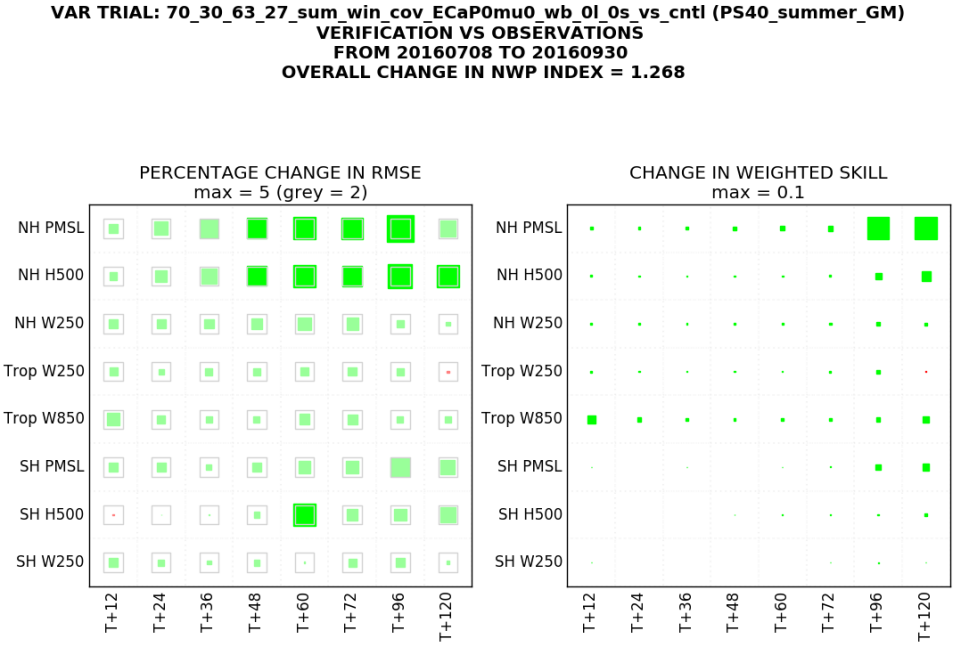


Figure 19: Verification against observations (top) and own analyses (bottom) for configuration EnVarCov3 in boreal winter.



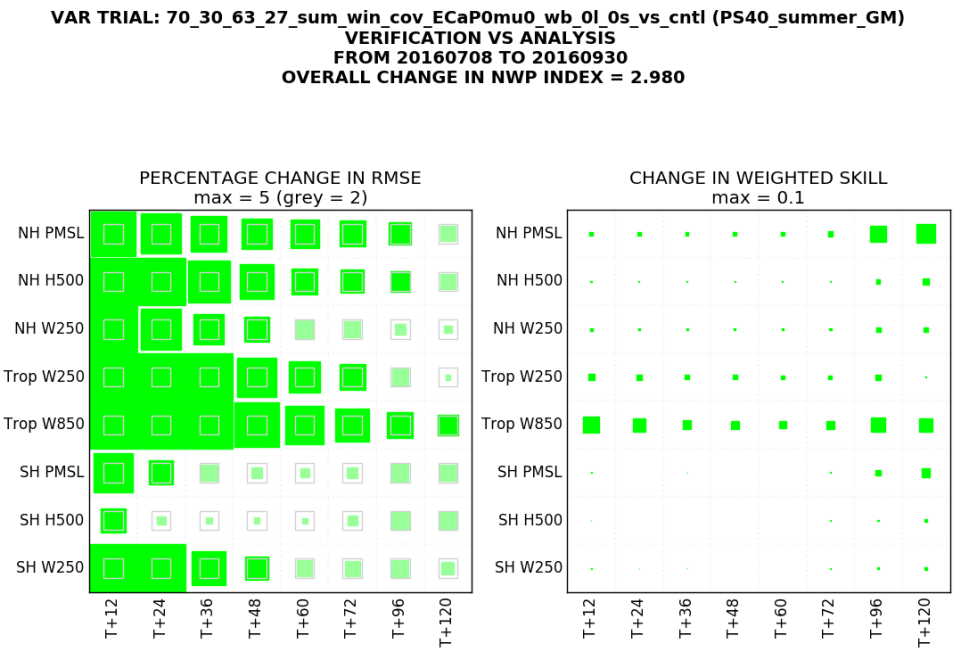


Figure 20: As for Figure 19 but in boreal summer.

9 Future work

The approach of taking the unbalanced pressure and humidity wavenumber zero vertical covariance matrices from the ECMWF derived covariance file is expedient in the short-term. In the long term a better way needs to be found to allow surface anchor observations to have sufficient influence.

The next global data assimilation and ensembles priority is the replacement of the ETKF with En-4DEnVar in MOGREPS-G. Time lagging and shifting remains an option following the operational implementation of En-4DEnVar, which is planned for late 2018 or early 2019 and which will use the intended wavebands shown in the left plot of Figure 1.

References

Bishop CH, Etherton BJ, Majumdar SJ. 2001. Adaptive sampling with the ensemble transform Kalman filter. Part 1: Theoretical aspects. *Mon Weather Rev* **129**: 420–436, doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Bowler NE, Arribas A, Mylne KR, Robertson KB, Beare SE. 2008. The MOGREPS short-range ensemble prediction system. *QJR Meteorol Soc* **134**: 703–722, doi:10.1002/qj.234.

Bowler NE, Clayton AM, Jardak M, Jermey PM, Lorenc AC, Wlasak MA, Barker DM, Inverarity GW, Swinbank R. 2017a. The effect of improved ensemble covariances on hybrid variational data assimilation. *QJR Meteorol Soc* **143**: 785–797, doi:10.1002/qj.2964.

Bowler NE, Clayton AM, Jardak M, Lee E, Lorenc AC, Piccolo C, Pring SR, Wlasak MA, Barker DM, Inverarity GW, Swinbank R. 2017b. Inflation and localisation tests in the development of an ensemble of 4D-ensemble variational assimilations. *QJR Meteorol Soc* **143**: 1280–1302, doi:10.1002/qj.3004.

Buehner M. 2012. Evaluation of a spatial/spectral covariance localization approach for atmospheric data assimilation. *Mon Weather Rev* **140**: 617–636, doi:10.1175/MWR-D-10-05052.1.

Clayton A. 2014. Use of ensemble-derived background error covariances in 3D and 4D-Var. VAR Scientific Documentation Paper 26, Met Office, Exeter, U.K.

Clayton AM, Lorenc AC, Barker DM. 2013. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *QJR Meteorol Soc* **139**: 1445–1461, doi:10.1002/qj.2054.

Gaspari G, Cohn SE. 1999. Construction of correlation functions in two and three dimensions. *QJR Meteorol Soc* **125**: 723–757, doi:10.1002/qj.49712555417.

Lorenc A. 2012. Variational bias correction of observations. VAR Scientific Documentation Paper 32, Met Office, Exeter, U.K.

Lorenc AC. 2017. Improving ensemble covariances in hybrid-variational data assimilation, without increasing ensemble size. *QJR Meteorol Soc* **143**: 1062–1072, doi:10.1002/qj.2990.

Ménétrier B, Auligné T. 2015. Optimized localization and hybridization to filter ensemble-based covariances. *Mon Weather Rev* **143**: 3931–3947, doi:10.1175/MWR-D-15-0057.1.

Ménétrier B, Montmerle T, Michel Y, Berre L. 2015a. Linear filtering of sample covariances for ensemble-based data assimilation. Part I: Optimality criteria and application to variance filtering and covariance localization. *Mon Weather Rev* **143**: 1622–1643, doi:10.1175/MWR-D-14-00157.1.

Ménétrier B, Montmerle T, Michel Y, Berre L. 2015b. Linear filtering of sample covariances for ensemble-based data assimilation. Part II: Application to a convective-scale NWP model. *Mon Weather Review* **143**: 1644–1664, doi:10.1175/MWR-D-14-00156.1.

Payne T. 2011. Minimization and convergence. VAR Scientific Documentation Paper 5, Met Office, Exeter, U.K.

Piccolo C. 2011. Growth of forecast errors from covariances modeled by 4DVAR and ETKF methods. *Monthly Weather Rev* **139**: 1505–1518, doi:10.1175/2010MWR3182.1.

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

Tel: 0370 900 0100
Fax: 0370 900 5050
enquiries@metoffice.gov.uk
www.metoffice.gov.uk

Produced with the Met Office LaTeX template v2.0.0