School of Mathematics, Meteorology
and Physics

University of Reading

Department of Mathematics

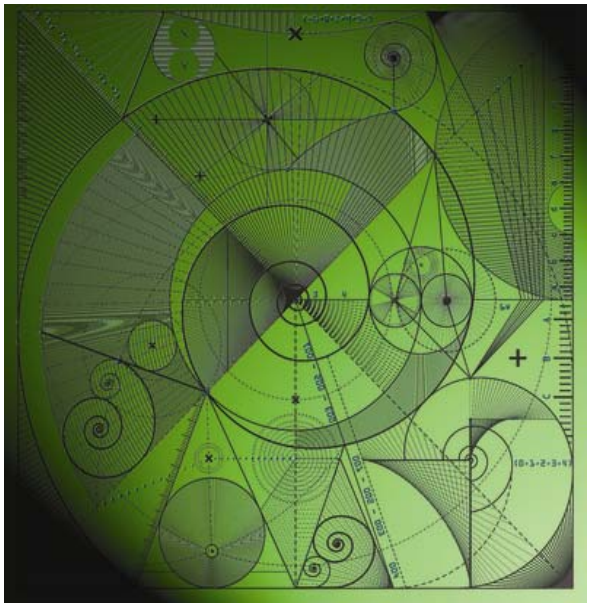
Preprint MPS_2010-10

20 April 2010

Breakdown of hydrostatic balance at convective scales in the forecast errors in the Met Office Unified Model

by

Sanita Vetra-Carvalho, Mark Dixon, Stefano Migliorini, Nancy K Nichols, Susan P Ballard



Breakdown of hydrostatic balance at convective scales in the forecast errors in the Met Office Unified Model

*Sanita Vetra-Carvalho *, Mark Dixon [#], Stefano Migliorini [%], Nancy K. Nichols *, Susan P Ballard [#]*

** Department of Mathematics, University of Reading, UK*

[#] *JCMM, Met Office, UK*

[%] *Department of Meteorology, University of Reading, UK*

Abstract

For data assimilation in numerical weather prediction, the initial forecast error covariance matrix P^f is required. For variational assimilation it is particularly important to prescribe an accurate initial matrix P^f, since P^f is either static (in the 3D-VAR case) or constant at the beginning of each assimilation window (in the 4D-VAR case). At large scale the atmospheric flow is well approximated by hydrostatic balance and this balance is strongly enforced in the initial matrix P^f used in operational variational assimilation systems such as that of the Met Office. However, at convective scales this balance does not necessarily hold any more. Here we examine the extent to which hydrostatic balance is valid in the vertical forecast error covariances for high resolution models in order to determine whether there is a need to relax this balance constraint in convective scale data assimilation. We use the Met Office Global and Regional Ensemble Prediction System (MOGREPS) and a 1.5 km resolution version of the Unified Model for a case study characterized by the presence of convective activity. An ensemble of high-resolution forecasts valid up to three hours after the onset of convection are produced. We show that at 1.5 km resolution hydrostatic balance does not hold for forecast errors in regions of convection. This suggests that in the presence of convection the covariance matrix used for variational data assimilation at this scale should not enforce hydrostatic balance. Finally, we give a measure of the balance present in the forecast perturbations as a function of the horizontal scale (from 3 km to 90 km) using a set of diagnostics.

1 Introduction

Due to a continuing increase in computer power, it has become possible for meteorological centres to run high resolution models. These models are expected to produce more realistic forecasts because of their better representation of small-scale forcing from orography and land use, as well as their explicit representation of convection. Modelling

at high resolution should thus lead to more accurate forecasting of high impact weather events such as flooding, with potential social and economic benefits.

Another important advantage of high-resolution models is that these may be used to assimilate high-resolution observations, e.g. from radar. To do so, it is important that the assumptions made for assimilation are still valid at these scales. For example, data assimilation systems used by operational meteorological centres such as the Met Office assume that the errors in the state variables of the model are hydrostatically balanced. While this assumption is reasonable for large scale models, it does not necessarily hold for perturbations at convective scales. It is then of interest, and an aim of this paper, to determine whether (and to what extent) forecast errors are hydrostatically balanced for a 1.5 km horizontal resolution version of the Met Office Unified Model (UM).

To this end, a set of high-resolution forecasts is generated for a case study on 27/07/2008 using the Met Office ensemble prediction system and the 1.5 km Unified Model over convective and non-convective regions of the southern UK for a period up to three hours after the onset of convection. The validity of hydrostatic balance in the forecast errors is examined for four selected vertical columns in the domain of the high resolution 1.5 km model. In addition, the degree to which the hydrostatic balance holds in the forecast perturbations is investigated as a function of the horizontal scale, by coarsening the 1.5 km data to produce 3 km up to 90 km resolution data sets.

The ensemble prediction system at convective scales used for this work is briefly described in section 2. The derivation of the hydrostatic balance relations for perturbations of the relevant fields is presented in section 3. Finally, in section 4 the results of the statistical analysis of the comparison between hydrostatically balanced and actual perturbations are discussed in terms of their vertical correlation structures, explained variances and root-mean-square (RMS) errors.

2 Ensemble prediction system at convective scales

Over the past few years, the Met Office has developed an ensemble prediction system, the Met Office Global and Regional Ensemble Prediction System (MOGREPS), which employs an ensemble transform Kalman filter (ETKF) to produce a set of initial conditions for generating predictions using the UM [3]. In this paper, the 24 km horizontal resolution initial conditions from the operational North Atlantic European (NAE) version of MOGREPS are interpolated to a 1.5 km resolution grid and used to integrate forward for three hours a high-resolution version of the UM. In this section, descriptions of the ETKF algorithm, the convective scale version of the UM and the interpolation strategy used to downscale the initial condition ensemble are provided.

2.1 The Ensemble Transform Kalman Filter (ETKF)

The ensemble Kalman filter (EnKF) [7] is a useful approximation to the Kalman filter (KF) [8] as it represents the errors in the state of a given system – composed, in the case of operational numerical weather prediction (NWP) models, of a number of variables of the order of 10^{7-8} – with a basis spanning only a much lower dimensional sub-space of the state space. As in the case of the Kalman filter, the EnKF assumes all errors to

be Gaussian. In the case of forecast errors of nonlinear models, such as those used for NWP, this assumption may often not be valid. Nevertheless, there are indications that the EnKF behaves well even when the errors are not Gaussian [3].

In the EnKF, the ensemble matrix $\mathbf{X}$ is defined as

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N), \tag{1}$$

where $\mathbf{x}_i \in \mathcal{R}^k$ is a state vector representing the i-th ensemble member, k is the number of components in the state vector and N is the number of ensemble members (equal to 24 in the version of MOGREPS used in this study). The ensemble mean is represented by $\overline{\mathbf{x}}$ and the ensemble perturbation matrix is given by

$$\mathbf{X}' = (\mathbf{x}_1 - \overline{\mathbf{x}}, \mathbf{x}_2 - \overline{\mathbf{x}}, \ldots, \mathbf{x}_N - \overline{\mathbf{x}}). \tag{2}$$

Superscripts f, a are used as in $\mathbf{X}^f, \mathbf{X}^a$ to mean forecast and analysis, respectively. Thus, from the ensemble forecasts the background-error covariance can be estimated by

$$\mathbf{P}^f = \frac{\mathbf{X}'^f (\mathbf{X}'^f)^T}{N-1}. \tag{3}$$

The ensemble mean analysis is updated by

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{K}\left(\mathbf{y} - H(\overline{\mathbf{x}}^f)\right), \tag{4}$$

where $\mathbf{y}$ are the observations and H is the observation operator. The Kalman gain $\mathbf{K}$ is given by

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T \left(\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R}\right)^{-1}, \tag{5}$$

where $\mathbf{R}$ is the observation error covariance matrix and $\mathbf{H}$ is a linearized version of H.

The ETKF is very closely related to the EnKF, allowing for a rapid calculation of the analysis perturbations [3]. The analysis ensemble perturbations in the ETKF are given by

$$\mathbf{X}'^a = \mathbf{X}'^f \mathbf{T} \Pi_m, \tag{6}$$

where $\mathbf{T}$ is a transform matrix and Π_m is an inflation factor (discussed below) for forecast cycle m. The transform matrix is computed according to [10] as

$$\mathbf{T} = \mathbf{C}\left(\mathbf{\Gamma} + \mathbf{I}\right)^{-\frac{1}{2}} \mathbf{C}^T, \tag{7}$$

where $\mathbf{C}, \mathbf{\Gamma}$ are matrices composed of the eigenvectors and eigenvalues respectively of the matrix

$$\mathbf{E} = \frac{\left(\mathbf{R}^{-\frac{1}{2}}\mathbf{Z}^f\right)^T \left(\mathbf{R}^{-\frac{1}{2}}\mathbf{Z}^f\right)}{N-1}. \tag{8}$$

The columns of $\mathbf{Z}^f$ are given by $\mathbf{z}_i^f = H(\mathbf{x}_i^f) - H(\overline{\mathbf{x}}^f)$, which are the ensemble perturbations in observation space for each ensemble member i and are found using the non-linear observation operator.

The variance in an ensemble generated by an EnKF is often smaller than required, due to an insufficient ensemble size [3]. To overcome this problem a variable inflation factor is

used, which seeks to ensure that the perturbation spread matches the RMS of the mean forecast,

$$\Pi_m = \Pi_{m-1}\sqrt{\frac{((tr(\mathbf{d}_m\mathbf{d}_m^T) - tr(\mathbf{R}))tr(\mathbf{S}_{m-1}))^{\frac{1}{2}}}{tr(\mathbf{S}_m)}}, \tag{9}$$

where $\mathbf{d}_m = \mathbf{y} - \overline{H(\mathbf{X}^f)}$ is the 'average' innovation vector and $\mathbf{S}_m = \mathbf{H}\mathbf{P}^f\mathbf{H}^T$ is the spread of the forecast ensemble in observational space for forecast cycle m. (Here $H(\mathbf{X}^f)$ denotes the matrix with columns given by $H(\mathbf{x}_i^f)$ and $\overline{H(\mathbf{X}^f)}$ denotes the vector average of these columns.)

2.2 The Met Office Unified Model

The 1.5 km version of the UM used in this work uses non-hydrostatic deep atmosphere equations with a hybrid height/terrain-following vertical coordinate [4]. The model has staggered grids in the horizontal and the vertical. The Arakawa C-grid is used for horizontal staggering where the zonal velocity component u is east-west staggered, and the temperature and the meridional velocity component v are north-south staggered. The Charney Phillips grid is used for vertical staggering, where potential temperature is on the same levels as the vertical velocity.

The high resolution model has a grid length of 1.5 km with 360 grid points in latitude and 288 in longitude covering Southern England and Wales. The model has a grid with 70 vertical levels, where only about 50 lowermost levels are affected by orography. The 24 km NAE model has 360 grid-points in latitude, 215 in longitude and 38 vertical levels.

The current data assimilation (DA) system for the 1.5 km model is similar to that used with the operational UK 4 km Met Office model, which is discussed in detail in [6]. In summary, the DA combines a 3D-Var scheme, used to assimilate the conventional observations producing the large scale analysis, and nudging systems used to update the high resolution moisture and surface precipitation data. The system uses a cloud nudging (CN) procedure to nudge humidity increments, whereas surface precipitation rates are assimilated via latent heat nudging (LHN). The increments produced by 3D-Var and nudging procedures are used to correct the model trajectory at each time step during the DA window. The main differences between the 1.5-km and the 4-km model DA configurations are as follows. The former uses hourly assimilation cycles, rather than the three-hourly used in the operational 4-km system, starting from fields interpolated to the southern England and Wales 1.5-km grid from the operational UK 4-km forecast. The 1.5-km DA system also uses more frequent cloud (hourly) and precipitation (every 15 minuntes) observations.

2.3 Generation of high-resolution perturbations

To obtain the initial condition ensemble at 1.5 km resolution, the following steps are performed (see figure 7 in Appendix 1). First, an ensemble of atmospheric fields is formed by adding an ensemble of operational MOGREPS NAE analysis perturbations at 24 km resolution valid at a given time (18Z), denoted as $\mathbf{X}'_{24km}$, to the operational 4D-Var

atmospheric analysis at 18Z over the NAE domain, denoted as $\overline{\mathbf{x}}_{24km}$, resampled at a resolution of 24 km from its original resolution of 12 km. This 24 km resolution ensemble, denoted as $\mathbf{X}_{24km}$, and the 4D-Var control, $\overline{\mathbf{x}}_{24km}$, are interpolated to a 1.5 km resolution grid to obtain $\mathbf{X}_{1.5/24km}$ and $\overline{\mathbf{x}}_{1.5/24km}$, respectively. The ensemble of perturbations at 1.5 km resolution, denoted as $\mathbf{X}'_{1.5/24km}$, is then obtained by subtracting the reconfigured 4D-Var analysis $\overline{\mathbf{x}}_{1.5/24km}$ from each column of the ensemble matrix $\mathbf{X}_{1.5/24km}$. Finally, to obtain the ensemble of initial conditions at 1.5 km resolution at 18Z, denoted as $\mathbf{X}_{1.5km}$, the 1.5 km ensemble perturbations $\mathbf{X}'_{1.5/24km}$ are combined with a 3D-Var high-resolution analysis – including nudging of precipitation and cloud observations – valid at 18Z from the 1.5 km model and data assimilation system over the southern UK [6]. The 18Z analysis was generated as part of a 1.5-km assimilation experiment with hourly DA cycles that were started at 2Z on 27 July 2008 and initialized with a 4-km model forecast valid at the same time and started at 0Z. To initiate the ensemble forecast at 1.5 km resolution, a 1.5 km version of UM is integrated forward in time, starting with the 18Z 3D-Var 1.5 km analysis and adding in the perturbations over 20 min using incremental analysis update [2]. Forecasts are produced at 19Z, 20Z, and 21Z on the same day. All ensemble members use the same boundary conditions taken from the operational 4 km UK forecasts.

The full ensemble matrix has dimensions of $\mathbf{X} \in \mathcal{R}^{(360\times288\times70)\times24}$. However, in this paper we have considered only the vertical part of the state by extracting vertical components from each ensemble member at a given location. For a given location and a given state variable, the ensemble matrix is then given by $\mathbf{X} \in \mathcal{R}^{70\times24}$.

3 Hydrostatic balance in the perturbations

In this section we derive the approximation of hydrostatic balance for perturbations. This allows us to compute the hydrostatically balanced potential temperature perturbations, θ'_H (where subscript H stands for hydrostatically balanced), for each of the ensemble members. In section 4 we use θ'_H to measure how close the perturbations are to hydrostatic balance.

Available fields at a given location at a given vertical level are Exner pressure Π, potential temperature θ, and specific humidity q. The ensemble mean values for each of these fields are denoted by $\overline{\Pi}$, $\overline{\theta}$, and $\overline{q}$, and Π', θ', and q' denote the perturbation from the mean of an individual ensemble member at the given location and height.

Hydrostatic balance is given by (e.g. see [9]),

$$\frac{dp}{dz} = -\frac{gp}{RT}, \tag{10}$$

where $p(z)$ is pressure, $T(z)$ is temperature, z is vertical height, and $R = 287.06$ J K^{-1}kg^{-1} is the gas constant for dry air. We can rewrite equation (10) in terms of the available variables, i.e. Π and θ, using the definition of the Exner pressure

$$\Pi = \left(\frac{p}{p_0}\right)^{R/c_p} \tag{11}$$

$$= \frac{T}{\theta^v}, \tag{12}$$

where $c_p = 1005$ J kg^{-1}K^{-1} is the specific heat at constant pressure, $p_0 = 1000$ hPa is a reference pressure at the ground level z_0, and θ^v is the virtual potential temperature, given by (e.g [9])

$$\theta^v = \theta\left(1 + (\epsilon^{-1} - 1)q\right), \tag{13}$$

with ϵ being the ratio of the mass of water liquid to dry air in the atmosphere and q being the specific humidity. Using θ^v instead of θ allows us to take the moisture of the atmosphere into account. Hence, we may write

$$p = p_0\Pi^{c_p/R} \tag{14}$$
$$T = \Pi\theta^v. \tag{15}$$

Substituting equations (14)–(15) into (10) and using (13), we find that the hydrostatic balance may be expressed in terms of the available variables Π and θ as

$$\frac{d\Pi}{dz} = -\frac{g}{c_p}\left(1 + (\epsilon^{-1} - 1)q\right)^{-1}\theta^{-1}. \tag{16}$$

This equation may be linearised, giving a hydrostatic equation for the mean variables

$$\frac{d\overline{\Pi}}{dz} = -\frac{g}{c_p}\frac{1}{(1 + (\epsilon^{-1} - 1)\overline{q})\overline{\theta}} \tag{17}$$

and a first order approximation to the hydrostatic equation for the perturbations

$$\frac{d\Pi'}{dz} = \frac{g}{c_p}\left[\frac{(\epsilon^{-1} - 1)q'}{(1 + (\epsilon^{-1} - 1)\bar{q})^2\bar{\theta}} + \frac{\theta'}{(1 + (\epsilon^{-1} - 1)\bar{q})\bar{\theta}^2}\right], \tag{18}$$

where $\Pi = \overline{\Pi} + \Pi'$, $\theta = \overline{\theta} + \theta'$ and $q = \overline{q} + q'$.

This equation may be rearranged to give

$$\theta'_H = -\frac{(\epsilon^{-1} - 1)q'\overline{\theta}}{1 + (\epsilon^{-1} - 1)\overline{q}} + \frac{c_p}{g}\frac{d\Pi'}{dz}\overline{\theta}^2(1 + (\epsilon^{-1} - 1)\overline{q}). \tag{19}$$

Hence, using the ensemble perturbations and mean, i.e. q', $\overline{q}$, Π', and $\overline{\theta}$ we can calculate the perturbation values of the potential temperature θ'_H that are in hydrostatic balance with these variables at each location and height.

The perturbation ensemble of each variable is defined as

- Potential temperature ensemble

$$\mathbf{\Theta}' = [\theta'_1, \theta'_2, ..., \theta'_{24}] \tag{20}$$

- Specific humidity ensemble

$$\mathbf{Q}' = [\mathbf{q}'_1, \mathbf{q}'_2, ..., \mathbf{q}'_{24}] \tag{21}$$

- Exner pressure ensemble

$$\mathbf{\Pi}'_e = [\mathbf{\Pi}'_1, \mathbf{\Pi}'_2, ..., \mathbf{\Pi}'_{24}] \tag{22}$$

where $.'_i$, $i = 1, ..., 24,$ denote the ensemble members – vectors containing the fields at all vertical grid points at a given location.

Then the hydrostatically balanced potential temperature perturbation, θ'_H, is computed for each element of each ensemble member, $i = 1, ..., 24$, leading to the ensemble vertical error covariance matrix

$$\mathbf{P}_e = \left\langle \mathbf{\Theta'}_H \mathbf{\Theta'}_H^T \right\rangle, \tag{23}$$

where $\mathbf{\Theta'}_H$ is the hydrostatically balanced potential temperature perturbation ensemble. The correlation matrix $\mathbf{C}_e$ is obtained by scaling matrix $\mathbf{P}_e$ by its own variance.

4 Results

This section discusses the results from applying the equation (19) to the data obtained from the ensemble of 1.5 km forecasts. The model was initialised with a set of ensemble atmospheric states valid at 18Z on 27/07/2008 and determined as explained in section 2 and Appendix 1. Ensemble forecasts were produced at each hour for the following 3 hours, 19Z, 20Z, and 21Z, on the same day, 27/07/2008. This case was selected as the convection had already occurred before 18Z and at the time of initialisation, 18Z, the system was fully convective with convection moving in the domain over the next three hours.

Although figures showing the degree of hydrostatic balance present in forecast perturbations were computed for various vertical columns, here only the "extreme" cases are investigated. The column for which the ensemble mean precipitation was zero over the entire 3 hours is labeled 'Non-Conv' and columns for which the ensemble mean precipitation was the highest for each hour [1] are labeled, 'Conv19Z', 'Conv20Z', and 'Conv21Z', respectively. Column location is indicated in the figure 1.

To find the degree to which hydrostatic balance holds in the perturbations as a function of horizontal scale the original 1.5 km resolution data were aggregated into boxes ranging from sides 3 km up to 90 km resolution around the non-convective point 'Non-Conv' and the convective column 'Conv19Z'.

The following quantities were calculated from both the 1.5 km and the coarsened resolution data:

- Correlation matrices for $\mathbf{\Theta'}$, $\mathbf{\Theta'_H}$, see figures 2, 3.
- Explained variances computed at each vertical level and 19Z, 20Z, 21Z, see figures 4, 5.
- Root mean square (RMS) errors between $\mathbf{\Theta'}$ and $\mathbf{\Theta'_H}$ computed for each of the four columns as a mean error over all members, levels and the three hour forecast window, see table 1.
- Mean ensemble time-dependent error between $\mathbf{\Theta'}$ and $\mathbf{\Theta'_H}$ computed for each of the four columns at each vertical level and for each hour as a mean error over all ensemble members, see figure 6.

[1] E.g. The column 'Conv19Z' has the highest rain rate at 19Z.

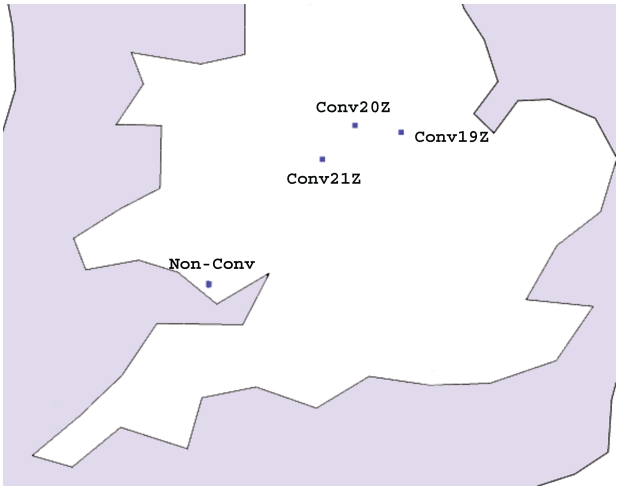


Fig. 1: 1.5 km resolution domain and the chosen vertical columns for testing

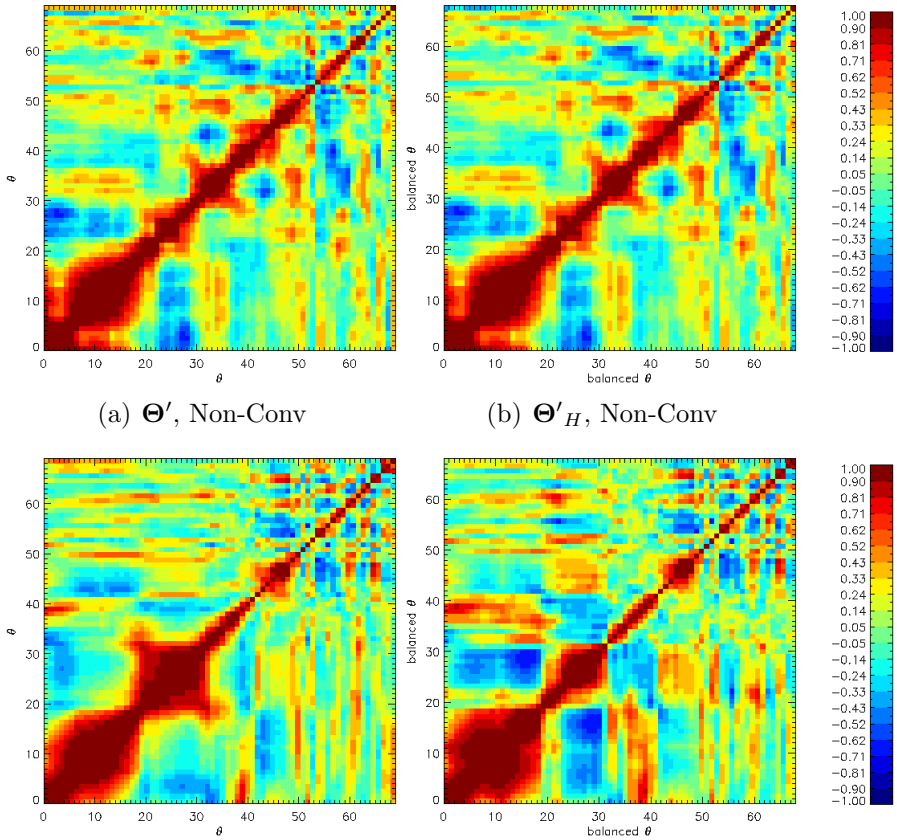
The dependence of hydrostatic balance can be ascertained in a quantitative way by means of the two error measures and the explained variance and in a more subjective way by inspection of the correlation matrices.

4.1 Correlation matrices for Θ' and Θ'_H

The balanced and "raw" data ensemble correlation matrices of $\mathbf{\Theta'}$ and $\mathbf{\Theta'}_H$ for the columns 'Non-Conv' and 'Conv19Z', at 1.5 km resolution, are shown in figure 2 for 19Z. As discussed in section 3, the balanced variable θ'_H is computed using equation (19) for each ensemble member and each vertical level. From figure 2 we can see that at 1.5 km resolution in the case of no convection (figures 2(a) and 2(b)) the ensemble correlation matrices for $\mathbf{\Theta'}$ and $\mathbf{\Theta'_H}$ are indistinguishable, meaning that hydrostatic balance holds very well in the perturbations when convection is not present. However, in the presence of convection (figures 2(c) and 2(d)) the ensemble correlation matrices for $\mathbf{\Theta'}$ and $\mathbf{\Theta'}_H$ are clearly different, especially just above the boundary layer (between vertical levels 20 and 40) where the convection is the strongest. Hence, in the cases (vertical levels) where the convection is present, the balance is no longer valid in the perturbations.

At 1.5 km resolution when convection is present, the $\mathbf{\Theta'}$ is less correlated in the boundary layer (vertical levels 0 to 20) than $\mathbf{\Theta'}_H$. However, $\mathbf{\Theta'}$ is more correlated than $\mathbf{\Theta'}_H$ right above the boundary layer at vertical levels 20 to 30.

By coarsening the grid we expect the perturbations to become more hydrostatically balanced. This is visually confirmed in figure 3, where correlation matrices of $\mathbf{\Theta'}$ and $\mathbf{\Theta'}_H$ for convective columns at 4.5 km and 12 km resolutions are shown. Notice, that even though at these resolutions the perturbations are not in hydrostatic balance in the mid-atmosphere (levels 20 - 40), they appear to be much more in the balance in the boundary layer.



(a) $\mathbf{\Theta}'$, Non-Conv (b) $\mathbf{\Theta}'_H$, Non-Conv

(c) θ', Conv19Z (d) θ'_H, Conv19Z

Fig. 2: Auto-correlations for $\mathbf{\Theta}$ and auto-correlations for the corresponding $\mathbf{\Theta}'_H$ at 19Z for non-convective column 'Non-Conv' and convective column 'Conv19Z' at 1.5 km data resolution

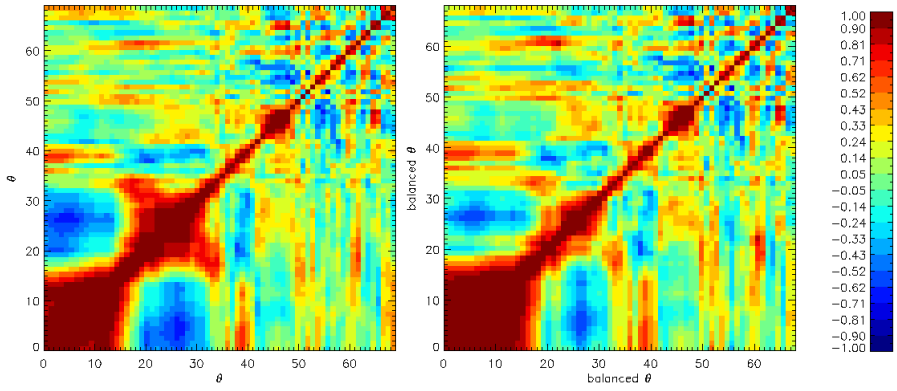
4.2 Explained variances

Plots in figure 2 show qualitatively that hydrostatic balance in perturbations is disturbed when convection is present. To determine the extent to which the hydrostatic balance holds in the perturbations quantitatively, we consider a measure known as explained variance [5]. The explained variance shows the extent to which the assumed balance relationship actually exists in the model and gives an indication of the optimality of the assumed algebraic form of the relationship.

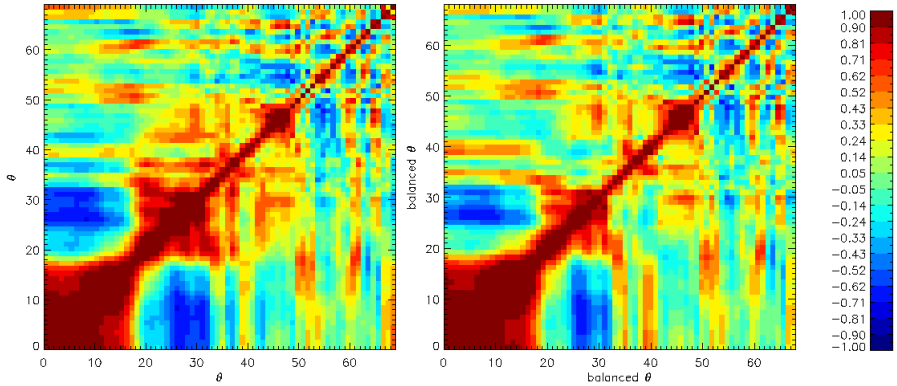
The explained variance is given by

$$E(z) = \left(1 - \frac{\sigma_U^2(z)}{\sigma^2(z)}\right), \tag{24}$$

where σ^2 is the grid-point variance of $\mathbf{\Theta}'$ and σ_U^2 is the variance of the unbalanced part of the perturbations, i.e. $\mathbf{\Theta}'_U = \mathbf{\Theta}' - \mathbf{\Theta}'_H$, and z is the vertical level. Thus, if $E \approx 1$ then perturbations are close to hydrostatic balance and if $E \approx 0$ then they are unbalanced. Figure 4 shows the explained variance as a function of height for all of the columns of the original 1.5 km data at 19Z, 20Z, and 21Z. This clearly shows that for the columns in the convective regions (i.e. 'Conv19Z', 'Conv20Z', and 'Conv21Z') at the time of convection the perturbations are not consistent with hydrostatic balance. By contrast, the perturba-



(a) $\mathbf{\Theta}'$, Conv19Z4km (b) $\mathbf{\Theta}'_H$, Conv19Z4km

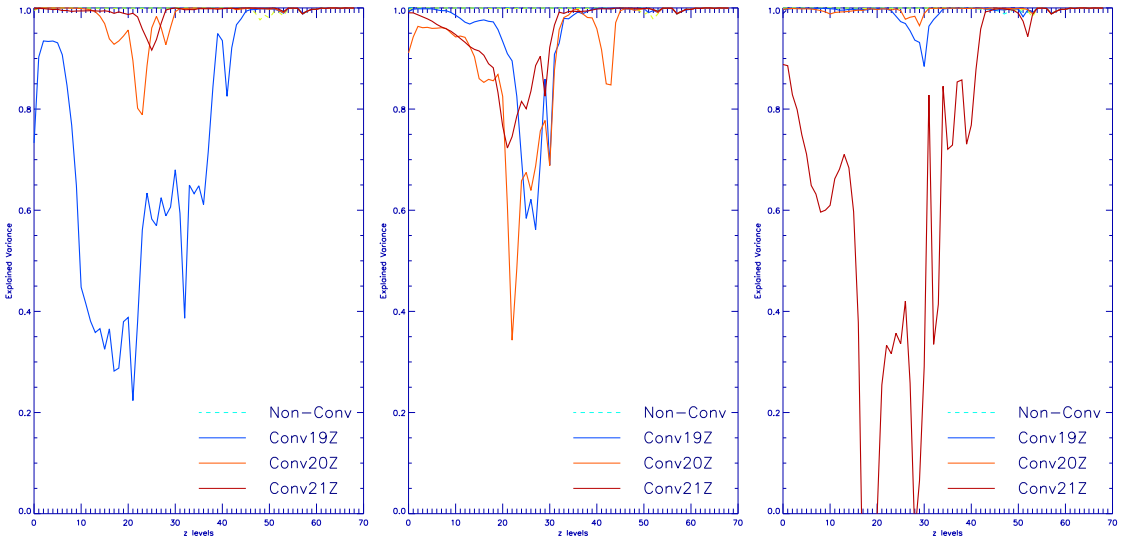


(c) $\mathbf{\Theta}'$, Conv19Z12km (d) $\mathbf{\Theta}'_H$, Conv19Z12km

Fig. 3: Auto-correlations for $\mathbf{\Theta}'$ and auto-correlations for the corresponding $\mathbf{\Theta}'_H$ at 19Z for the convective column 'Conv19Z4km' at 4.5 km resolution and the convective column 'Conv19Z12km' at 12 km resolution

tions in all of the non-convective columns are very well explained by hydrostatic balance. Figure 5 shows the explained variance as a function of height for the convective columns from 3 km up to 22.5 km coarsened data at 19Z, 20Z, and 21Z. This shows the degree to which hydrostatic balance (in convective regions) increases as a function of horizontal scale and demonstrates that at resolutions coarser than about 20 km hydrostatic balance holds very strongly. It follows that the resolution at which hydrostatic balance holds in the perturbations over the whole domain, i.e. for convective and non-convective regions, is coarser than 20 km. At 45 km and 90 km resolutions (not shown) the perturbations are consistent with hydrostatic balance, as are perturbations for the non-convective columns for all coarsened resolutions (not shown).

Notice that the convective coarsened column data was obtained by coarsening around the position of 'Conv19Z', i.e. the column with the highest rain rate in the ensemble mean at 19Z. As the convection moves out from 'Conv19Z' column over the two hours, the variances explain more of hydrostatic balance at 20Z and 21Z compared with 19Z. However, since with coarser horizontal resolution a larger area is encompassed, a coarser column may have more convection on average than the 1.5 km grid point. Hence, hydrostatic balance at some coarser resolutions may be less well explained by variances than at 1.5 km resolution. Compare the peaks of figures 4 (b), (c) with figures 5 (b), (c).



(a) 19Z (b) 20Z (c) 21Z

Fig. 4: Explained variance $E(z)$ at 1.5 km for each vertical level at a) 19Z, b) 20Z and c) 21Z

For the 1.5 km data the 'Conv20Z' and 'Conv21Z' are more in hydrostatic balance initially at 19Z, and they become more unbalanced as convection becomes stronger in these columns at 20Z and 21Z, respectively.

4.3 High resolution RMS errors between Θ' and Θ'_H

Another measure of the hydrostatic balance in the perturbations is the RMS error between $\boldsymbol{\Theta}'_H$ and $\boldsymbol{\Theta}'$, using the standard formula for the RMS,

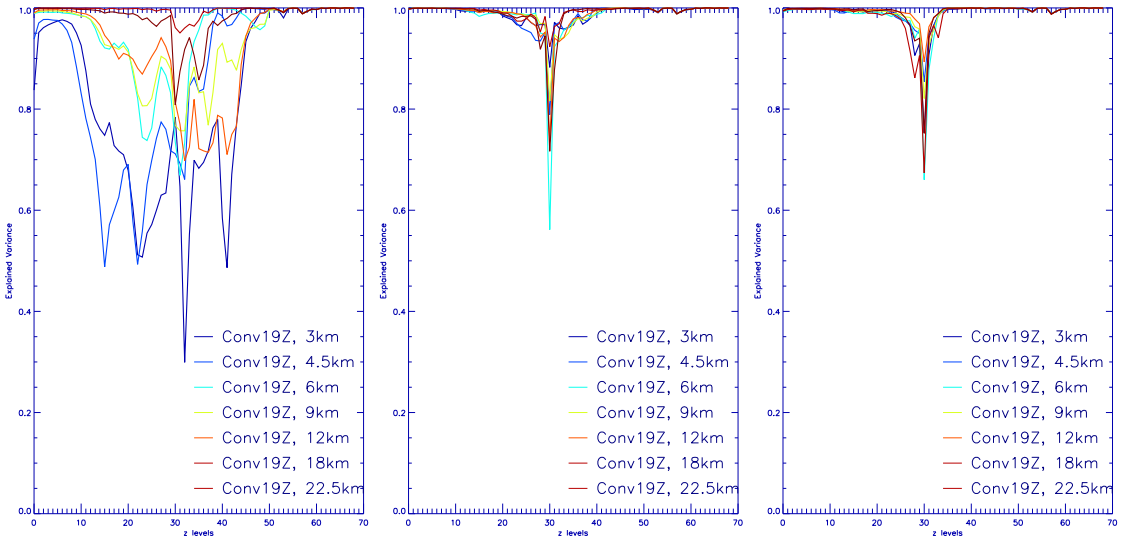
$$\sqrt{\overline{(\boldsymbol{\Theta}'_{\mathbf{H}} - \boldsymbol{\Theta}')^2}}. \tag{25}$$

This provides an average measure of how far the full perturbations are from the hydrostatic perturbations, averaged over all the ensemble members and vertical levels. Results for the 1.5 km data are displayed in table 1 and results for the coarse data are given in Appendix 2.

Column index	19Z	20Z	21Z
'Non-Conv'	0.82	0.83	1.59
'Conv19Z'	17.97	3.81	3.44
'Conv20Z'	3.26	9.41	1.89
'Conv21Z'	2.64	8.59	28.29

Tab. 1: RMS error ($\times 10^{-2}$) for the selected vertical columns at 1.5 km resolution

¿From table 1 we see that for the non-convective column the RMS error grows slightly over the three hour period, but this growth is insignificant. However, for the convective columns, hydrostatic balance holds much better in each column when convection is either



(a) 19Z (b) 20Z (c) 21Z

Fig. 5: Explained variance $E(z)$ for resolutions of 3 km to 22.5 km of convective 'Conv19Z' column only for each vertical level at a) 19Z, b) 20Z and c) 21Z

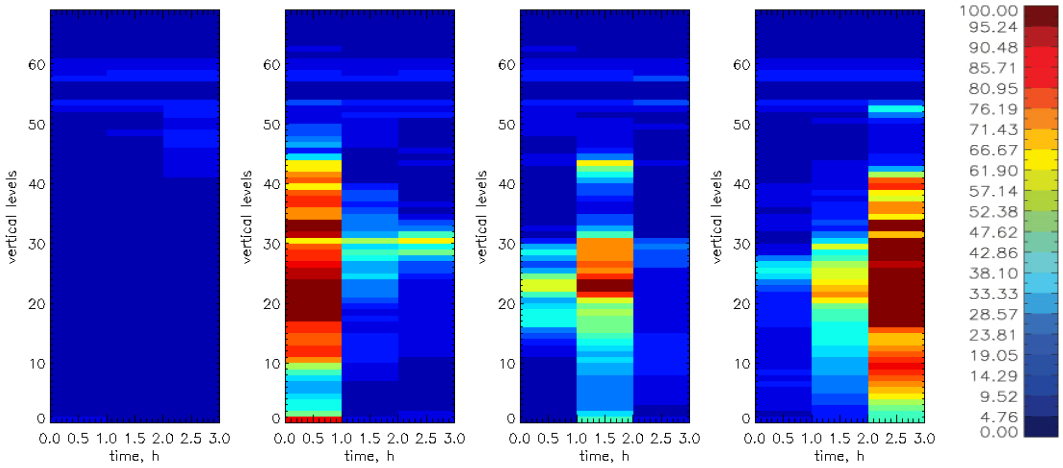
not present or weak. For example, for column 'Conv20Z' the RMS error is small at 19Z since convection is very weak at this time in the region; however at 20Z, when convection in this column is strongest in the ensmeble mean, the RMS error is three times larger. Finally, at 21Z the convection has moved out of the column and the RMS error for 'Conv20Z' has decreased again. Similarly, for columns 'Conv19Z' and 'Conv21Z'. The ensemble mean precipitation rates are shown in Appendix 3.

4.4 The time-dependence of relative error

Here the error between an average ensemble value for each vertical level and time is computed, given by

$$rel.error = \frac{\sqrt{\overline{(\Theta'_H - \Theta')^2}}}{\overline{|\Theta'_H|}} \times 100. \tag{26}$$

Figure 6 shows the relative error for all five columns at 1.5 km resolution. The perturbations from convective columns 'Conv19Z', 'Conv20Z', and 'Conv21Z' are far from being in balance at the time when the convection is the strongest in each column, whereas, for the non-convective column 'Non-Conv', the perturbations are close to being balanced at all times. From the figure it is possible to see that convection moves in space over the three hours and this is consistent with the results in sections 4.2 and 4.3 obtained from the explained variances and the RMS error.



(a) 'Non-Conv' (b) 'Conv19Z' (c) 'Conv20Z' (d) 'Conv21Z'

Fig. 6: Relative error between $\mathbf{\Theta'}_H$ and $\mathbf{\Theta'}$ averaged over all ensemble members in vertical space and time at 1.5 km resolution

5 Summary and conclusions

To investigate how well the hydrostatic balance holds for forecast errors at convective scales we used an ensemble with 24 members obtained from running the 1.5 km resolution version of the UM that was initialised according to the procedure described in section 2.3 and in Appendix 1. The ensemble was initialised at 18Z on 27/07/2008 when convection was fully developed and data for analysis were available at 19Z, 20Z, and 21Z on the same day. In this paper we focused on the vertical analysis of forecast errors. Four vertical columns from the whole domain were selected for testing purposes for each forecast hour: a column with no precipitation and three columns with highest rain rates in the ensemble mean, for each forecast hour 19Z, 20Z and 21Z, respectively. Data around two columns, one non-convective and one convective were aggregated from 3 km up to 90 km resolution. For each of these columns the hydrostatically balanced potential temperature perturbations, θ'_H, were calculated using the approximated hydrostatic equation for perturbations, expressed in the terms of the available fields - potential temperature θ, Exner pressure Π and specific humidity q.

By constructing the correlation matrices for these columns it was shown that at 1.5 km resolution the hydrostatic balance does not hold in the perturbations in the regions of convection but does hold in the regions where convection is not present. Note, that from the explained variances and mean error in the vertical we establish not only the extent to which the perturbations are not in hydrostatic balance but also, that perturbations are in balance at all resolutions in all columns in the stratosphere (above vertical level 55) where the atmosphere is dry. Also, from the explained variances we see that at the vertical levels $10 - 30$ the perturbations are very far from being balanced. This suggests that the hydrostatic balance should be relaxed around these columns and levels in the correlation matrices at 1.5 km resolution. This would require a redesign of the control variable transform in variational data assimilation system used by the Met Office.

We also showed using the explained variances that 20 km horizontal resolution is the

limit at which the hydrostatic balance becomes valid over the entire domain.

Acknowledgements

The authors would like to thank the Met Office and the Natural Environment Research Council (NERC) for financial support.

References

[1] C.H. Bishop, B.J. Etherton, S. J. Majumdar, *Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects*, Mon. Weather Rev., 2001, Vol 129, pp 420–436

[2] S. Bloom, L. Takacs, A. da Silva, D. Ledvina, *Data Assimilation Using Incremental Analysis Updates*, 1996, Mon. Wea. Rev., Vol 124, pp 1256–1271

[3] N.E. Bowler, A. Arribas, K.R. Mylne, K.B. Robertson, S.E. Beare, *The MOGREPS short-range ensemble prediction system*, QJRMS RMetS, 2008, Vol 134, pp 703–722

[4] T. Davies, M.J.P. Cullen, A.J. Malcolm, M.H. Mawson, A. Staniforth, A.A. White, N. Wood, *A new dynamical core for the Met Office's global and regional modelling of the atmosphere*, QJRMS RMetS, 2006, Vol 131, pp 1759–1782

[5] J. Derber, F. Bouttier, *A reformulation of the background error covariances in the ECMWF global data assimilation*, Tellus, 1999, Vol 51A, pp 195–221

[6] M. Dixon, L. Zhihong, H. Lean, N. Roberts, Sue Ballard, *Impact of Data Assimilation on Forecasting Convection over the United Kingdom Using a High-Resolution Version of the Met Office Unified Model*, Mon. Weather Rev., 2009, Vol 137, pp 1562–1584

[7] G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo mothods to forecast error statistics*, J. Geophys. Res., 1994, Vol 99, pp 10143–10162

[8] R.E. Kalman, *A new approach to linear filtering and prediction problems*, Transactions of the AMSE -Journal of Basic Engineering, 1960, Vol 82D, pp 35 - 45

[9] J.M. Wallace, *Atmospheric science : an introductory survey*, Elsevier, 2006, 2nd edition

[10] X.G. Wang, C.H. Bishop, S.J. Julier, *Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble?*, Mon. Weather Rev., 2004, Vol 132, pp 1590–1605

1 Appendix: Flow of the high resolution EPS

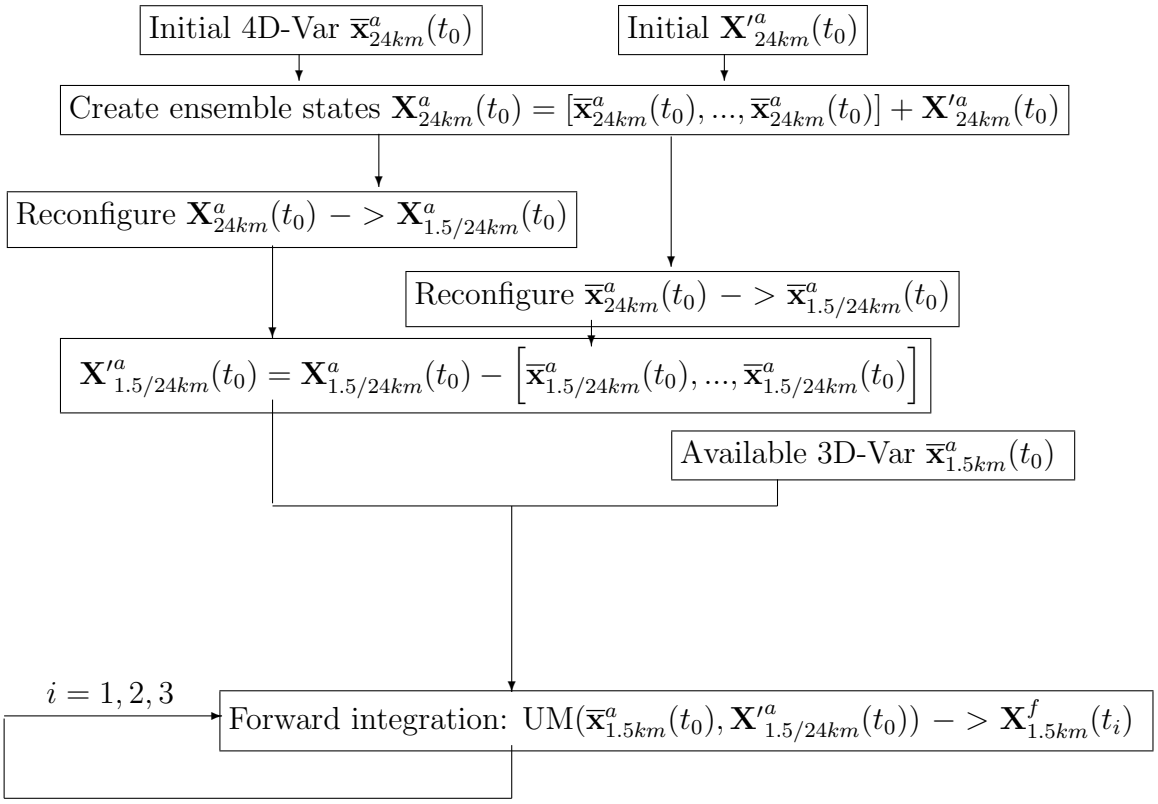


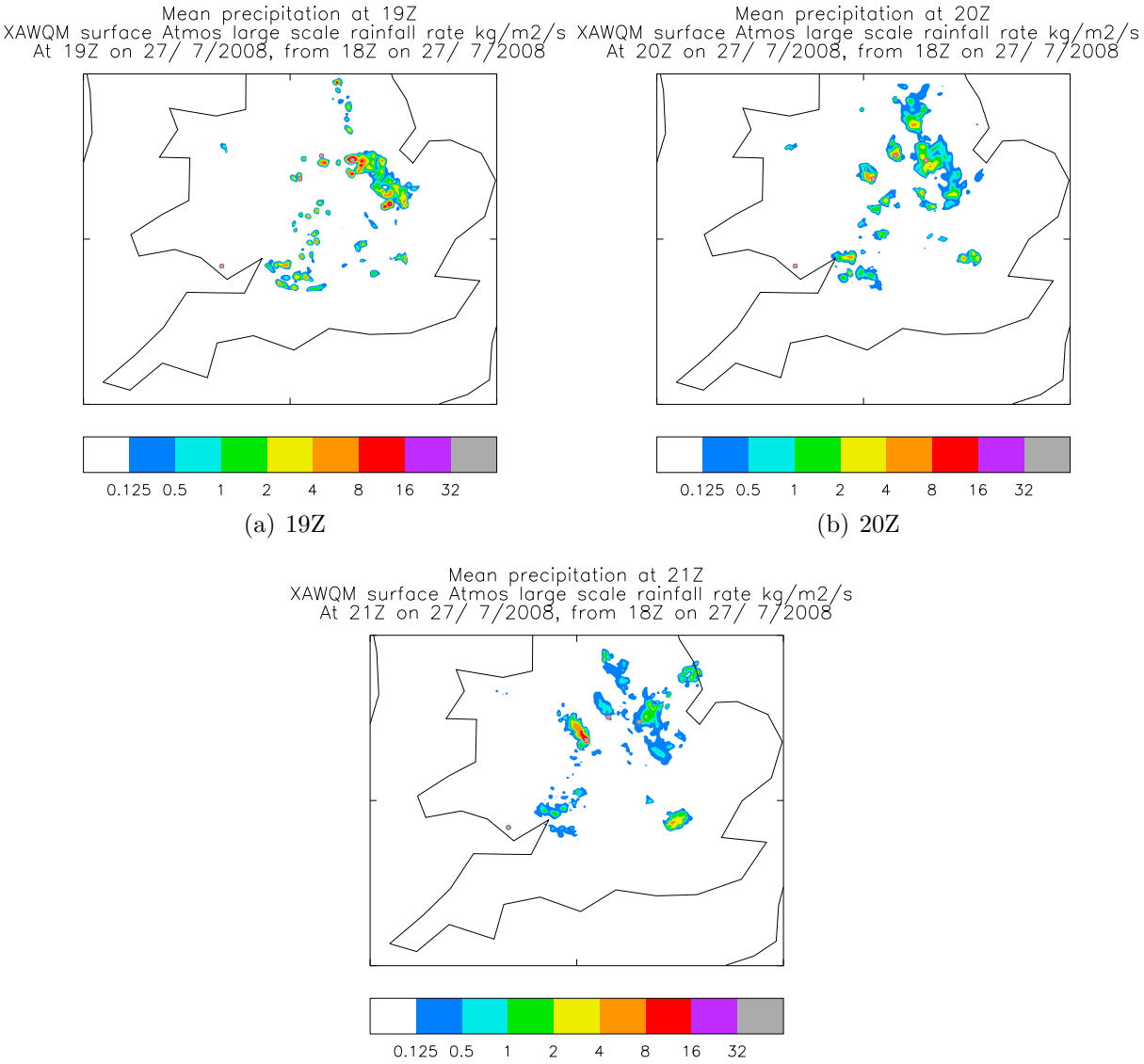
Fig. 7: Setup and flow of the reconfigured ensemble prediction system. Here t_0=18Z, t_1=19Z, t_2=20Z, and t_3=21Z all on 27/07/2008. The UM forward integration step at 1.5 km includes the addition of perturbations over 20 min via incremental analysis update [2] from 18Z onwards. All ensemble members use the same operational 4 km boundary conditions.

2 Appendix: RMS error of coarsened data

Resolution (km)	Col. type	19Z	20Z	21Z
3	conv.	12.0619	3.3082	3.2973
3	non-conv.	0.8335	0.8379	1.4559
4.5	conv.	7.8751	3.1099	2.8045
4.5	non-conv.	0.8445	0.8359	1.3021
6	conv.	5.2717	2.7469	3.0582
6	non-conv.	0.8446	0.8281	1.2450
9	conv.	4.9439	2.5643	2.4611
9	non-conv.	0.8239	0.7909	1.1596
12	conv.	4.7447	2.4807	1.9634
12	non-conv.	0.8176	0.7759	1.1191
18	conv.	1.4827	1.9645	2.4036
18	non-conv.	0.7968	0.7369	1.0726
22	non-conv.	1.7442	1.8898	2.2769
22	non-conv.	0.7794	0.8594	1.0379
45	conv.	0.9137	1.2158	1.2769
45	non-conv.	0.7640	0.7295	0.9387
90	conv.	0.7494	0.9231	0.9995
90	non-conv.	0.7189	0.7951	0.8869

Tab. 2: RMS error with coarsened resolutions around 'Conv19Z' and 'NonConv' columns, $\times 10^{-2}$.

3 Appendix: Mean precipitation rate



(a) 19Z

(b) 20Z

(c) 21Z

Fig. 8: Mean precipitation rates (mm/hr), at 19Z, 20Z, and 21Z. Black dots are coordinates of the selected vertical columns.