School of Mathematics, Meteorology
and Physics

University of Reading

Department of Mathematics

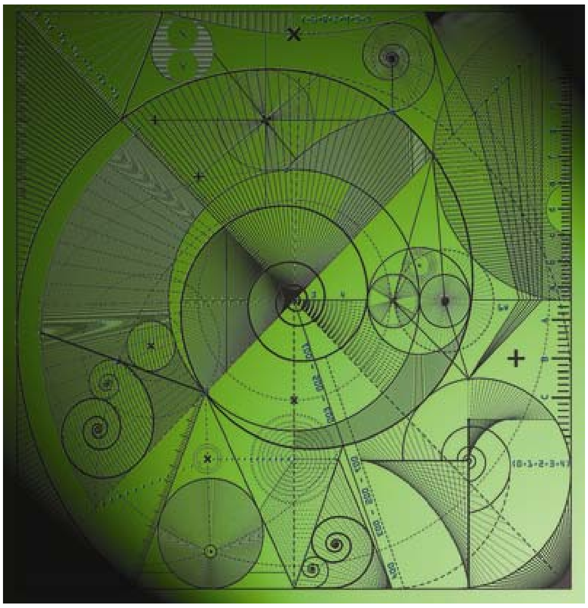
Preprint MPS_2010_13

21 April 2010

Four-dimensional variational data assimilation for high resolution nested models

by

G.M. Baxter, S.L. Dance, A.S. Lawless, N.K. Nichols



Four-dimensional variational data assimilation for high resolution nested models

G.M. Baxter, S.L. Dance, A.S. Lawless*, N.K. Nichols

Department of Mathematics, University of Reading, U.K.

Abstract

Four-dimensional variational data assimilation (4D-Var) is used in environmental prediction to estimate the state of a system from measurements. When 4D-Var is applied in the context of high resolution nested models, problems may arise in the representation of spatial scales longer than the domain of the model. In this paper we study how well 4D-Var is able to estimate the whole range of spatial scales present in nested models. Using a model of the one-dimensional advection-diffusion equation we show that small spatial scales that are observed can be captured by a 4D-Var assimilation, but that information in the larger scales may be degraded. We propose a modification to 4D-Var which allows a better representation of these larger scales.

Keywords: advection-diffusion equation, sine transform, 4D-Var

1. Introduction

In many applications of environmental forecasting, such as numerical weather prediction, it is necessary to estimate the current state of the system in order to make a forecast. Usually the number of available measurements of the system is not sufficient to define the state uniquely and so the measurements are combined with a numerical model forecast, using techniques of data assimilation, in order to provide the best estimate of the system state. In operational numerical weather prediction a common data assimilation technique is that of four-dimensional variational data assimilation

*Corresponding author, email a.s.lawless@reading.ac.uk

(4D-Var). This technique formulates the assimilation problem as an optimization problem over space and time, constrained by a numerical model of the equations describing the atmospheric flow. The solution to this optimization problem provides the state estimate from which a forecast can be produced.

An important challenge in weather prediction is the improvement of our ability to forecast small-scale localized weather systems, such as convective storms. Such systems are often associated with severe weather events, such as localized flooding, and so have a large impact on people and society. Detailed local weather forecasts are usually provided by running high resolution regional models nested within lower resolution models covering a larger domain. The lower resolution model is used to provide lateral boundary conditions for the higher resolution model throughout the forecast period. However, current regional models in numerical weather prediction are usually unable to resolve the scales needed for storm-scale weather forecasting. To improve forecasts at these scales the Met Office are developing a very high resolution forecasting model covering the U.K. with a horizontal resolution down to 1 km, nested in a lower resolution European-scale model [9], [13]. The development of these very high resolution nested models presents a challenge for current data assimilation methods such as 4D-Var [4], [12]. One such challenge is that for reasons of computational efficiency the highest resolution models usually only cover a small-sized domain. In such cases large-scale atmospheric features, which may contain scales which are larger than the model domain, cannot be correctly represented within the scales of the nested numerical model. However, an accurate representation of these long waves is important since meteorological processes are inherently multiscale and there are strong feedbacks between phenomena at large and small spatial scales [4].

In this paper we study the ability of 4D-Var to capture information on large length scales in a nested model. By considering a spectral transform of the long waves, we show how information from these waves is aliased onto other wave components in a nested model 4D-Var assimilation. We propose a new adaptation of the 4D-Var scheme which uses this aliasing property to provide a suitable constraint on the large scales. In section 2 we explain in more detail the 4D-Var assimilation method and consider the problem of representing long-wave information in a small domain. In section 3 we present some numerical experiments using the one-dimensional advection-diffusion equation. We demonstrate how a 4D-Var scheme may be adapted to capture better the long-wave information by using our knowledge of how such waves are aliased. Finally we draw conclusions in section 4.

2. Four-dimensional variational data assimilation

2.1. Formulation

The aim of four-dimensional data assimilation (4D-Var) is to estimate the system state $\mathbf{x}_0 \in \mathbb{R}^n$ at initial time t_0 given a prior estimate of the state $\mathbf{x}_0^b \in \mathbb{R}^n$ at time t_0 and observations $\mathbf{y}_j \in \mathbb{R}^{p_j}$ at times $t_j, j = 0, \ldots, N$. We assume that $\mathbf{x}_0^b$ has random, unbiased Gaussian errors with error covariance matrix $\mathbf{B}_0$ and that the observations have random, unbiased Gaussian errors with error covariance matrices $\mathbf{R}_i$, where the matrices $\mathbf{B}_0$ and $\mathbf{R}_i$ are assumed to be symmetric positive definite. Then 4D-Var defines the best estimate as the state which minimizes the objective function

$$\begin{aligned} \mathcal{J}[\mathbf{x}_0] &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) \\ &+ \frac{1}{2}\sum_{j=0}^{N}(h_j(\mathbf{x}_j) - \mathbf{y}_j)^T \mathbf{R}_j^{-1}(h_j(\mathbf{x}_j) - \mathbf{y}_j), \end{aligned} \tag{1}$$

subject to the nonlinear dynamical system equations

$$\mathbf{x}_j = m(t_j, t_0, \mathbf{x}_0), \tag{2}$$

where m is the solution operator of the nonlinear model. The operator h_j maps the state space to the space of observations. Under the statistical assumptions given, the solution to this minimization problem is equal to the maximimum posterior Bayesian estimate of the initial state [10]. We refer to this state estimate as the *analysis*. In operational weather forecasting the function (1) is minimized using a few iterations of a Gauss-Newton method [3], [8]. On each iteration (1) is linearized around the current estimate $\mathbf{x}_0^{(k)}$, beginning with $\mathbf{x}_0^{(0)} = \mathbf{x}_0^b$, and the linear cost function is solved for an increment $\delta\mathbf{x}_0$ to this estimate. Thus we obtain the linearized cost function

$$\begin{aligned} \tilde{\mathcal{J}}[\mathbf{x}_0] &= \frac{1}{2}(\delta\mathbf{x}_0 - \delta\mathbf{x}^b)^T \mathbf{B}_0^{-1}(\delta\mathbf{x}_0 - \delta\mathbf{x}^b) \\ &+ \frac{1}{2}\sum_{j=0}^{N}(\mathbf{H}_j\delta\mathbf{x}_j - \mathbf{d}_j^{(k)})^T \mathbf{R}_j^{-1}(\mathbf{H}_j\delta\mathbf{x}_j - \mathbf{d}_j^{(k)}), \end{aligned} \tag{3}$$

where $\delta\mathbf{x}^b = \mathbf{x}_0^b - \mathbf{x}_0^{(k)}$, $\mathbf{d}_j^{(k)} = \mathbf{y}_j - h_j(\mathbf{x}_j^{(k)})$ is the innovation vector, $\mathbf{H}_j$ is the Jacobian of h_j evaluated at $\mathbf{x}_j^{(k)}$ and the increment $\delta\mathbf{x}_j$ satisfies the linear model equation

$$\delta\mathbf{x}_j = \mathbf{M}(t_j, t_0)\delta\mathbf{x}_0, \tag{4}$$

where $\mathbf{M}$ is the Jacobian of the nonlinear model m with respect to the current estimate $\mathbf{x}_0^{(k)}$. In order to minimize (3) efficiently we transform to a different set of variables $\delta\mathbf{z}$ defined by the linear transform $\delta\mathbf{z} = \mathbf{U}^{-1}\delta\mathbf{x}$, where the transform $\mathbf{U}^{-1} \in \mathbb{R}^{n\times n}$ is defined to ensure that components of $\delta\mathbf{z}$ are independent. Thus we obtain the linear cost function

$$\begin{aligned} \tilde{\mathcal{J}}[\delta\mathbf{z}_0] &= \frac{1}{2}(\delta\mathbf{z}_0 - \delta\mathbf{z}^b)^T\mathbf{\Sigma}^{-1}(\delta\mathbf{z}_0 - \delta\mathbf{z}^b) \\ &+ \frac{1}{2}\sum_{j=0}^{N}(\mathbf{H}_j\mathbf{U}\delta\mathbf{z}_j - \mathbf{d}_j^{(k)})^T\mathbf{R}_j^{-1}(\mathbf{H}_j\mathbf{U}\delta\mathbf{z}_j - \mathbf{d}_j^{(k)}), \end{aligned} \tag{5}$$

where $\delta\mathbf{z}^b = \mathbf{U}^{-1}\delta\mathbf{x}^b$ and $\mathbf{\Sigma}^{-1} = \mathbf{U}^T\mathbf{B}_0^{-1}\mathbf{U}$. The transform $\mathbf{U}$ is chosen to ensure that the components of $\delta\mathbf{z}$ are independent and so the matrix $\mathbf{\Sigma}$ is diagonal. Such a transformation always exists since $\mathbf{B}_0$ is symmetric positive definite and so has a diagonal eigendecomposition.

2.2. 4D-Var in nested models

When 4D-Var is applied to nested models the lateral boundary conditions play an important role. In nested modelling for weather forecasting the lateral boundary conditions are usually provided by a larger-domain model, here referred to as the 'parent' model, which may, for example, cover the whole globe. This parent model is usually at a lower spatial and temporal resolution than the nested model and so the boundary conditions must be interpolated in time and space. Within the 4D-Var system the nested version of the nonlinear model (2) is run, in order to calculate the innovation vectors $\mathbf{d}_j$ and to provide the state values at which the Jacobians of h_j and m are calculated. This nested model run requires boundary conditions from the parent model on each Gauss-Newton iteration and at each time step during the observation time window. Since the parent model is usually run separately from the nested model, these boundary conditions are usually not updated during the assimilation and the increments $\delta\mathbf{x}$ are assumed to be zero on the lateral boundaries. One way to enforce this is to define the transform $\mathbf{U}^{-1}$ to be the sine transform and $\mathbf{U}$ to be its inverse. This transform is commonly used in data assimilation for nested models, for example at the Met Office [11], and it is the transform we will use in this paper. In this case the variables $\delta\mathbf{z}$ correspond to the different wave numbers in the discrete sine transform.

One possible problem with the use of this transform is the treatment of waves that have a wavelength longer than the domain of the nested model. In this case the true scale of the wave cannot be represented by the sine transform and so information from such long waves will be projected onto shorter scales by the transform. This is essentially a reverse of the classical aliasing problem. Whereas usually aliasing is considered as the misinterpretation of small-scale waves as larger-scale waves, here we have large-scale waves being misinterpreted as shorter-scale waves. To illustrate this effect we define a sine wave with wavenumber one over the periodic domain $[0, 1)$ and calculate its sine transform over the whole domain and then the sine transform of the part of the wave in the nested domain $[0, 0.25]$. The results are shown in Figure 1. Both transforms are calculated using 32 spatial points. We see that for the transform over the whole domain the power is all at wavenumber two. This is a property of the sine transform, in which the power of a sine wave of wavenumber k appears at wavenumber $2k$ when the transform is applied on a domain of length one. When the transform is applied on the smaller domain there is only a quarter wavelength that fits into the domain, which is not within the discrete spectrum on the nested model grid. In this case we see that most of the power is projected onto wavenumber one, with significant power also in higher wavenumbers.

Previous studies have examined different methods for treating the large spatial scales in a nested model data assimilation using a three-dimensional variational data assimilation scheme, a variation of 4D-Var in which the observations are all considered to be at the same time as the background. Within this context the authors of [1] examined the possibility of taking the large scales completely from a parent model analysis and using the nested model data assimilation to update only the small scales. An alternative method that has been proposed is to constrain the large scales to be close to those of the parent model analysis and a nested model background field by the addition of an extra term in the objective funtion [7]. Here we examine how the aliasing of the long waves affects their representation in a 4D-Var scheme and propose a new modification to the data assimilation system to allow for this.

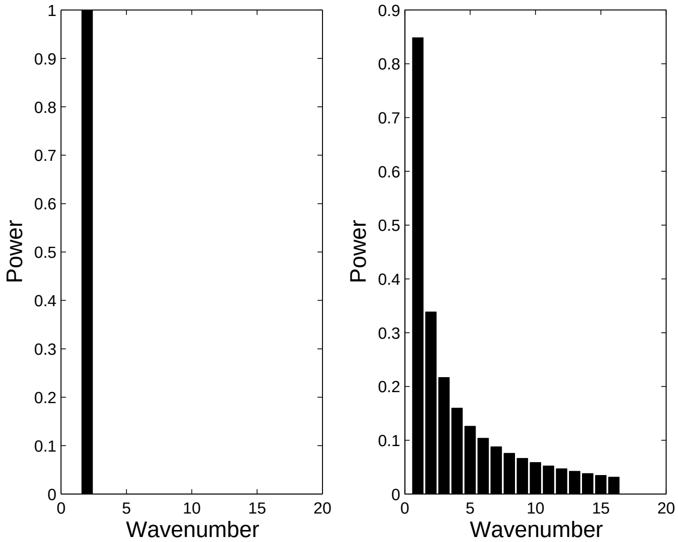


Figure 1: Sine transform of sine wave with wavenumber $k = 1$ on domain $[0, 1)$ (left) and on domain $[0, 0.25]$ (right).

3. Assimilation experiments

3.1. Model

The model we use to test the 4D-Var algorithm is the one-dimensional advection-diffusion equation,

$$u_t + cu_x = \sigma u_{xx}, \tag{6}$$

where $u(x, t)$ is the temperature, x is the spatial coordinate, t is the time, $c \geq 0$ is the constant advection velocity, $\sigma \geq 0$ is the diffusion constant and subscripts indicate derivatives. The equation for the parent model is defined on the domain $x \in (0, 1]$ with periodic boundary conditions.

The model is discretized using an explicit Euler scheme for the time derivative, centred differences for the diffusion term and upwind differences for the advection term. We define a spatial step Δx and time step Δt. Then $u(x_j, t_n)$ is approximated by $u_{j,n}$, where for each point we have the discrete update equation

$$u_{j,n+1} = (\nu + \mu)u_{j-1,n} + (1 - \nu - 2\mu)u_{j,n} + \mu u_{j+1,n}, \tag{7}$$

with $\nu = c\Delta t/\Delta x$ and $\mu = \sigma\Delta t/(\Delta x)^2$. The scheme is first-order in time and space, but is close to second-order spatial accuracy when the discrete Peclet number $c\Delta x/\sigma$ is small [6, p. 138].

For the nested model we discretize (6) on the limited domain $[0.5, 1]$, with the boundary conditions $u(0.5, t)$ and $u(1, t)$ taken from a run of the parent model at all times t. In the interior of the domain the discretization scheme is exactly as in the parent model, with a higher resolution spatial step Δx_H and time step Δt_H. Close to the boundaries the nested model solution is relaxed to the parent model solution using a Davies relaxation scheme over a small buffer zone [5].

3.2. Experimental design

Idealised 4D-Var experiments are set up by running the model from a known initial state, which we refer to as the true trajectory, and then generating observations from this true trajectory to use in the assimilation. The truth is generated at a higher resolution than either the parent or nested models, with spatial step Δx_T and time step Δt_T. For the experiments presented here we use values of $\Delta x = 0.0625, \Delta x_H = \Delta x/4, \Delta x_T = \Delta x/8$ and $\Delta t = 0.05, \Delta t_H = \Delta t/16, \Delta t_T = \Delta t/64$. This means that there are 16 spatial points in the parent model over the range $(0, 1]$ and 32 spatial points in the nested model over the range $[0.5, 1]$. The assimilation is performed over the time window $[0, 0.5]$, with observations at every spatial point of the nested model and at every time step of the nested model. The diffusion constant is set to $\sigma = 10^{-3}$ and the advection velocity $c = 1$.

To generate the true trajectory we run the model from an initial value consisting of the sum of three sine waves.

$$u_T(x, 0) = 5\sin\pi x + \sin 2\pi x + \sin 36\pi x. \quad (8)$$

The first of these waves ($\sin\pi x$) is a wave that is longer than the nested model domain. The other two waves, with wavenumbers $k = 1$ and $k = 18$, can both be resolved by the nested model, but the wave with $k = 18$ is too short to be resolved by the parent model. Synthetic observations are generated by adding a random, Gaussian error to values of the true state trajectory with variance $\sigma_o^2 = 0.25$.

In order to define the innovation vectors $\mathbf{d}_j$ in the objective function (5) we need to define a suitable background trajectory of the nested model at all observation times within the time window. We suppose that we have a perfect analysis on the parent model grid which includes the components of the truth u_T at the initial time that can be represented on this grid, that is

$$u_P(x,0) = 5\sin\pi x + \sin 2\pi x. \tag{9}$$

To generate a background field with known covariance for the nested model, we choose to add random noise to this field in spectral space at the nested model resolution. We interpolate u_P to the nested model grid and then apply the sine transform to obtain the field in spectral space. Random, Gaussian noise, with variance $\sigma_b^2 = 0.25$, is then added to the wavenumbers contained in the spectrum of the parent model (wavenumbers $1,\ldots,7$). The variance of the background error is thus equal to the level of noise we have on the observations. The higher wavenumbers of the background are set to zero. The inverse sine transform is then applied to obtain the nested model background field at grid point values. The nested model can then be run from this initial background, using boundary conditions provided from the parent model run with initial state $u_P(x,0)$, in order to provide a background trajectory throughout the assimilation window from which we can calculate the innovation vectors $\mathbf{d}_j$. It is important that this background trajectory is obtained by running the numerical model at the resolution of the assimilation model, as is done in practice, rather than just interpolating from the background state of the lower resolution parent model [2]. The synthetic observations are then assimilated using the 4D-Var formulation (5), which is minimized using a conjugate gradient algorithm in order to obtain the analysis at the initial time. The error in this analysis is calculated with respect to the truth u_T at the grid points of the nested model and compared to the error in the background calculated in a similar manner.

We note that due to the effect illustrated in section 2.2, the long wave will be projected mainly onto wavenumber one by the sine transform, with some power also in the other low wavenumbers. The true $k = 1$ wave will also be projected onto wavenumber one, and not wavenumber two as in the example of section 2.2, since the domain is of length 0.5. Hence the signal at wavenumber one will be a mixture of the true $k = 1$ wave and some of the long wave information. Our aim is to understand how the 4D-Var treats these two separate sources of information.

3.3. Results

We first aim to understand how the 4D-Var assimilation treats the long wave when the correct variance information is used in the assimilation. The error covariance matrix $\mathbf{\Sigma}$ in (5) is defined to be a diagonal matrix, in which the first seven components are set to the true variance of the background error, 0.25. For the higher wavenumbers there is no useful information coming from the background, so we assume a variance of 5.0, which ensures that the observations will be given a much greater weight than the background at these scales. In Figure 2 we show the power spectrum of the error in the background and the error in the analysis. For clarity we show only the lowest and highest wavenumbers, which are the parts of the spectrum containing the true solution.

The first thing that we notice is that the data assimilation in the nested model is able to capture the high resolution information at wavenumber $k = 18$. This wave appears in the true solution, but it cannot be resolved by the parent model and so is set to zero in the background field. The use of the 4D-Var system with the high resolution nested model enables information at this scale to be inferred from the observations. Further experiments show that a necessary condition to infer this wave is that the observations are also at the high resolution; it is not sufficient to have a high resolution data assimilation system with only low resolution observations [2]. At the low wavenumbers we see that the errors in the analysis are worse than those in the background for $k = 1$ and 2. The large-scale information coming from the background has been degraded during the assimilation process. Since these wavenumbers include information from the long wave that cannot be represented on the nested model domain, the nested assimilation does not treat this information correctly in this case.

In general we may expect the large scales provided by the parent model to be reasonably accurate and we would like to use the nested model assimilation to improve the small scales. Hence we would ideally like the 4D-Var scheme applied to the nested model to retain the large-scale information from the parent model. Since we have seen that this large-scale information is projected onto low wavenumbers by the sine transform, we may expect to improve the analysis if we constrain the solution to be closer to the background in these low wavenumbers. To test this possibility we run the assimilation experiment again, but within the matrix $\mathbf{\Sigma}$ we assume that variance on the low wavenumber components ($k = 1, \ldots, 7$) is 0.05 rather than the true variance 0.25. Thus the background field for these wavenumbers is given more weight than a simple statistical argument would justify. The spectrum of the error in the analysis from this experiment is compared to that of the background in Figure 3. We see that the error on the low wavenumbers has

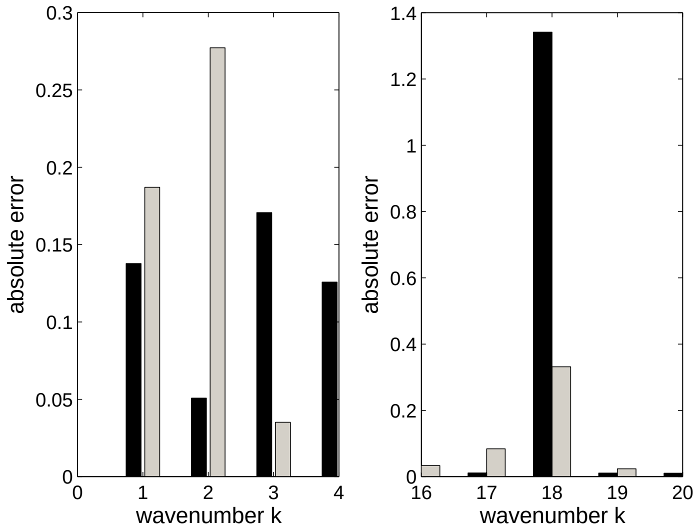


Figure 2: Power spectrum of errors in background (black) and analysis (grey) for low wavenumbers (left) and high wavenumbers (right) when the true error variances are used.

been much reduced with respect to the experiment using the true variances. The wavenumber one component of the solution is now more accurate than the background. At wavenumber $k = 2$ the analysis is still worse than the background, but it is improved with respect to the first experiment. Other choices of the variances at the low wavenumbers lead to further improvements at these scales [2]. An important aspect of this experiment is that the analysis at wavenumber $k = 18$ is still as accurate as in the first experiment. Hence, by over-weighting the low wavenumbers in the background field, we have been able to retrieve the small-scale information while retaining the accuracy of the background in the large scales.

4. Conclusions

The development of data assimilation schemes for very high resolution nested models is an important component of the development of future weather prediction systems. In this paper we have analysed one particular aspect of such schemes, namely the treatment of very long waves within a 4D-Var data assimilation system. Information on large spatial scales is provided to a nested model by a larger-domain parent model and it is important

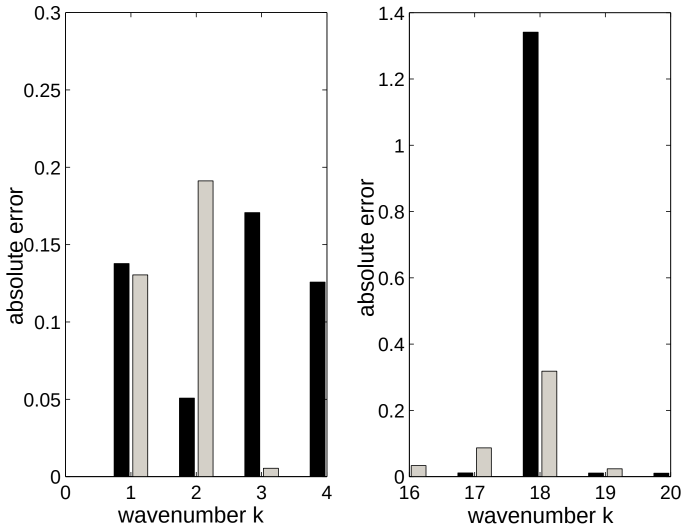


Figure 3: As Figure 2, for the experiment in which the low wavenumbers are over-weighted in the assimilation.

that the assimilation in the nested model does not degrade this information. We have shown that within a nested model domain these scales are projected onto low wavenumbers by a spectral transform. Hence the low wavenumbers in the background contain a combination of information at scales which the nested model can resolve and information at scales larger than the domain. When a standard 4D-Var assimilation is performed in spectral space it is not able to disinguish between these two sources of information. Hence the high resolution assimilation is able to improve the estimate of the state at small scales, but this occurs at the expense of a loss of information at the large scales. We have proposed a modification to 4D-Var for these cases, in which the low wavenumbers in the background are given more weight in order to allow for the fact that they contain information on larger scales than can be represented in the nested model. By performing the assimilation in spectral space and over-weighting the low wave numbers, we are able to improve the estimates of these large scales, while still keeping the same accuracy in the smaller scales.

Previous studies reported in [1] and [7] have also used the large scales of the parent model analysis as a constraint in the nested model analysis, in the context of a three-dimensional variational assimilation. Whereas the approach of [1] enforces the large scales of the nested model to be exactly equal to those of the parent analysis, that of [7] weakly constrains the large scales by a combination of these scales from the parent analysis and a previous nested model forecast. In the new approach presented here we use the large scales from only the parent analysis as a constraint, as in the work of [1], but they act as only a weak constraint and observations are allowed to alter these scales within the assimilation process. Experiments in a simple numerical model indicate that this new approach can provide benefit, but a more comprehensive comparison with the other approaches would be valuable.

Acknowledgements

This work was supported by the Natural Environment Research Council and the Met Office through a CASE studentship. We would like to thank Sue Ballard of the Met Office for useful discussions throughout this research.

References

[1] Ballard S., Z. Li, M. Dixon, S. Swarbrick, O. Stiller and H. Lean. Development of 1-4km resolution data assimilation for nowcasting at the Met Office. World Weather Research Program Symposium on Nowcasting and Very Short Range Forecasting (WSN05) 2005; paper 3.02.

[2] Baxter G.M. 4D-Var for high resolution, nested models with a range of scales. 2009. *PhD thesis*. Department of Mathematics, University of Reading.

[3] Courtier P., J.-N. Thépaut and A. Hollingsworth. A stategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society* 1994; **120**:1367–1387.

[4] Dance S.L. Issues in high resolution limited area data assimilation for quantitive precipitation forecasting. *Physica D* 2004; **196**:1–27.

[5] Davies H.C. A lateral boundary formulation for multi-level prediction models. *Quarterly Journal of the Royal Meteorological Society* 1976; **102**:405–418.

[6] Durran D.R. Numerical Methods for Wave Equations in Geophysical Fluid Dynamics. Springer, 1999.

[7] Guidard V. and C. Fischer. Introducing the coupling information in a limited area variational assimilation. *Quarterly Journal of the Royal Meteorological Society* 2008; **134**:723–735.

[8] Lawless, A.S., S. Gratton and N.K. Nichols . An investigation of incremental 4D-Var using non-tangent linear models. *Quarterly Journal of the Royal Meteorological Society* 1994; **131**:459–476.

[9] Lean H.W., P.A. Clark, M. Dixon, N.M. Roberts, A. Fitch, R. Forbes and C. Halliwell. Characteristics of high-resolution versions of the Met Office Unified Model for forecasting convection over the United Kingdom. *Monthly Weather Review* 2008; **136**:3408–3424.

[10] Lorenc A.C. Optimal nonlinear objective analysis. *Q. J. R. Met. Soc.* 1988; **114**:205–240.

[11] Lorenc A.C. Development of an operational variational assimilation scheme. *Journal of the Meteorological Society of Japan* 1997; **75**:339–346.

[12] Park S.K. and D. Zupanski. Four-dimensional variational data assimilation for mesoscale and storm-scale applications. *Meteorology and Atmospheric Physics* 2003; **82**:173–208.

[13] Roberts N.M. and H.W. Lean. Scale-selective verification of rainfall accummulation from high-resolution forecasts of convective events. *Monthly Weather Review* 2008; **136**:78–97.