The University of Reading

Towards Data Assimilation for High Resolution Nested Models

G.M. Baxter[1], S.L. Dance[1], A.S. Lawless[1], N.K. Nichols[1] and S.P. Ballard[2]

NUMERICAL ANALYSIS REPORT 2/2007

Presented at the ICFD International Conference on Numerical Methods for Fluid Dynamics, Reading, March 2007

[1]*Department of Mathematics*
The University of Reading
Whiteknights, PO Box 220
Reading
Berkshire RG6 6AX

[2]*Met Office*
FitzRoy Road
Exeter
Devon EX1 3PB

Department of Mathematics

Abstract

Accurate prediction of convective storms is important because these storms can lead to dangerous flooding events. Improvements in computer power have allowed operational forecasting centres to begin running research models at resolutions down to O(1km). There are still limits to computer power however, and if a model is to have a very high resolution it can only have a limited domain size; this creates some problems. One is there may be waves present in the atmosphere that have wavelengths longer than the domain of the model; this is what we refer to as the 'long wave' problem. Another is that if a model domain has a limited area, then it has 'edges' and the model will need boundary conditions. In order to investigate the 'long wave' problem we nest a limited area domain within a simple 1D PDE model. Fourier analysis shows that the smallest wavenumber that can be represented by the limited area model is larger than the wavenumber of the 'long wave'. This has implications for data assimilation. Large scale information from observations within the limited area domain will require special treatment.

1 Introduction

An important challenge for numerical weather prediction (NWP) is to improve our ability to forecast convective storms. Accurate prediction of convective storms is important because these storms can lead to dangerous flooding events, such as the flood in Boscastle, UK in 2004.

In order to represent convective storms accurately, spatial resolutions of order 100m are needed to resolve the dominant motions properly. However, running operational forecast models at such high resolutions requires computer power beyond that currently available. Nevertheless research has shown that very useful results can be obtained with horizontal resolutions of order 1km [6], because the mesoscale flows leading to storm initiation, organisation and propagation are generally well represented. The 1km resolution also allows us to do away with convective parameterisations and allows us to use more accurate orography. If a model is to run at a very high resolution it can only have a limited domain size. For example, the UK

Met Office High Resolution Trial Model has a horizontal domain of 300km×300km and 76 vertical levels [6].

In order to generate a weather forecast with a limited area model (LAM) we need initial conditions and lateral boundary conditions (LBCs). To generate initial conditions that accurately describe the observed reality we combine a previous model forecast (background) with observations. The tool that allows us to do this is data assimilation and the initial conditions are known as the 'analysis'. The LAM also needs LBCs because the grid is nested within a larger grid and has 'edges'. This is not necessary for the global model because the grid stretches over the whole globe.

The operational assimilation-forecast cycle has the following steps:

Step 1 Run a coarse resolution model to provide a forecast over a large domain.

Step 2 Run the LAM to provide a background forecast over the assimilation time window, using LBCs from the forecast in Step 1.

Step 3 Run the data assimilation to combine the LAM background forecast with observations.

Step 4 Run the LAM over the full forecast period, using the updated initial conditions from Step 3 and the LBCs from Step 1.

The limited area nature of storm-scale forecasting models can create some significant problems for data assimilation. Firstly, there may be waves present in the atmosphere with wavelengths that are longer than the domain of the model; for example, a Rossby wave is of the scale O(1000km). Meteorological processes are known to be multiscale phenomena and there are strong feedbacks between synoptic and convective scale behaviour [4]. Atmospheric measurements can contain a large range of scales and it is important to preserve this information in the analysis.

Secondly, the LAM LBCs come from a model with a coarser resolution in both space and time, so have to be interpolated to the LAM grid at every timestep. The LAM solution is then relaxed to these interpolated values over a 'buffer zone'; operationally this is usually done using Davies Relaxation [5]. To control reflection at the boundaries Davies Relaxation requires the buffer zone to be sufficiently wide because the reflection coefficient reduces with buffer width [3].

In order to investigate these and other possible problems that may exist, we insert a limited area domain in a simple 1D-PDE model, to mimic the operational system in a simple model. We aim to explore possible problems caused by the presence of a boundary, differences in resolution and multiscale properties. The model is described in Section 2. In Section 3 we discuss some preliminary results and we conclude in Section 4 and consider future work.

2 Method

We use the 1D heat equation. This has the advantage of having an analytic solution so we can compare our results with the truth and check the accuracy and convergence of the model. Using this simple model we explore possible problems caused by the presence of a boundary and differences in resolution.

We have the 1D heat equation

$$u_t = \sigma u_{xx}, \tag{1}$$

with homogeneous boundary conditions and $x \in [0,1]$, $t \in [0,0.5]$, where $u = u(x,t)$ is the heat, x is the spatial coordinate, t is time and $\sigma > 0$ is the diffusion constant. We call the spatial domain $[0,1]$ the global domain. We use an explicit Euler discretisation

$$u_{i,k} = u_{i,k-1} + \mu \left(u_{i-1,k-1} - 2u_{i,k-1} + u_{i+1,k-1}\right), \tag{2}$$

where $\mu = \sigma\delta t/\delta x^2$, $i = 1, 2, \cdots, N$ and $k = 1, 2, \cdots, T$. N is the number of internal spatial gridpoints and T is the number of timesteps. $\delta x = 1/(N+1)$ and $\delta t = 0.5/T$. Into the global domain we insert a smaller, limited area domain. It covers D global grid spaces where $D = b2 - b1$ and $b1, b2$ are the global gridpoints corresponding to the boundaries of the limited area domain. The spatial and temporal resolutions of the LAM are defined by their ratio to the global resolutions. If the distance between two global grid points is δx then the distance between two LAM grid points is $\delta\xi = \delta x/h$, and if the global model uses a timestep of δt then the LAM timestep is $\delta\tau = \delta t/g$. The values of h and g are chosen such as to keep the same value of μ in both models. The boundary values for the limited area model come from the global model values at $b1$ and $b2$. In order to relax the solution on the interior of the limited area domain to the values prescribed at the boundaries there is a buffer zone implemented at the boundaries covering bz LAM gridpoints. A Davies Relaxation scheme [5] is used in the buffer zone. We use Davies Relaxation because this is what is used operationally. Since our discretisation is explicit it can be shown algebraically that the Davies Relaxation is equivalent to the interpolation

$$u_i^{new} = (1 - \omega_i)u_i^L + \omega_i u_i^G, \tag{3}$$

where u_i^{new} is the value of u in the buffer zone, u_i^L is the value of u calculated by the limited area model, u_i^G is the value of u coming from the global model and ω is a interpolation function

$$\omega_i = 1 - [(i-1)/bz].$$

To investigate the outputs of our models we consider their power spectra. The absolute value of the fast Fourier transform (FFT) squared, is plotted against the wavenumber κ, where the FFT for a function f_j, defined on $[0, 2\pi)$ is

$$FFT(f_j) = \sum_{j=0}^{N-1} f_j e^{-i\zeta_j k}, \qquad \text{where} \quad \zeta_j = \frac{2\pi j}{N}.$$

Methods of using the power spectra to understand a wave solution are described in [2]. We now consider the effect the limited area domain has on the model output when the initial conditions are a sine wave with wavelength longer than the domain of the LAM.

3 Results

We take our global model to have $N = 15$ internal gridpoints. The LAM boundaries are at global gridpoints $b1 = 4$ and $b2 = 12$ so the LAM domain covers $D = 8$ global gridspaces. The spatial resolution of our LAM is $\delta\xi = \delta x/h$ where $h = 2$, and the temporal resolution of our LAM is $\delta\tau = \delta t/g$ where $g = 4$. The buffer zone covers $bz = 2$ LAM gridpoints. We begin with the initial conditions

$$u(x_i, 0) = \sin(2\pi x_i), \qquad \text{where} \quad i = 1, \cdots, N$$

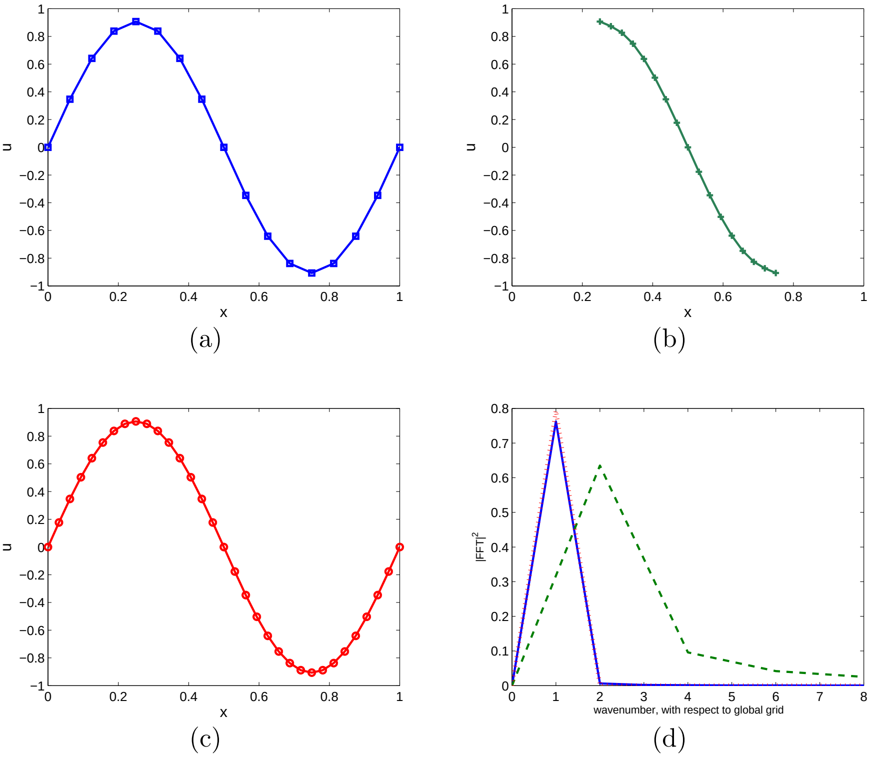
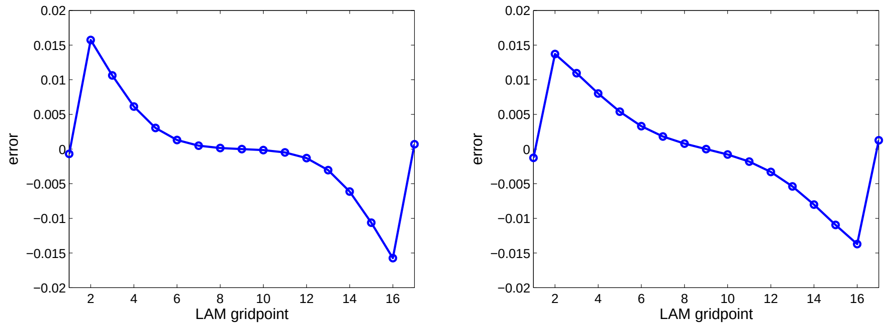


Figure 1: (a) is the output of the global model run at resolution δx and δt, (b) is the output of the LAM run at resolution $\delta\xi$ and $\delta\tau$, (c) is the output of the global model run at resolution $\delta\xi$ and $\delta\tau$ and (d) is the power spectra of all three outputs plotted against wavenumber κ with respect to the global grid with resolution δx. The power spectra of (a) is the solid line, (b) the dashed line and (c) the dotted line.



(a) (b)

Figure 2: The differences between the fine resolution global model and LAM outputs. (a) at time t=0.25, (b) at time t=0.50.

$$u(\xi_j, 0) = \sin(2\pi\xi_j), \qquad \text{where} \quad j = 1, \cdots, (D \times h) + 1$$

for the global and LAM respectively. In both models the constant of diffusivity $\sigma = 0.01$.

Figure 1 shows the model outputs for the global and LAM at time $t = 0.25$. Figure 1(a) shows the global model output, Figure 1(b) shows the LAM output and Figure 1(c) shows the output of the global model when run at the resolution of the LAM, i.e. the global model with $N = 31$ and $T = 160$. Clearly in Figure 1(b), the limited area of the model domain has meant that the model cannot resolve the entire wave. In order to inspect what has been captured by the LAM, the power spectra of all three model outputs are shown in Figure 1(d). As can be seen both global model outputs capture the correct wavenumber, $\kappa = 1$. The LAM output does not capture this wavenumber. This is because wavenumber $\kappa = 1$ cannot be represented on the limited area domain. The smallest wavenumber the LAM can resolve is $\kappa = 2$. As can be seen in Figure 1(d), the LAM has shifted most of the power in the output to wavenumber $\kappa = 2$. There are also small amounts of power in the subsequent wavenumbers.

In order to consider the effect of taking boundary conditions for the LAM from a global model with coarser resolution, we compare the output of the LAM with the output of the global model, both run at resolution $\delta\xi$, $\delta\tau$. We then extract the part of the global output which covers the domain of the LAM. This allows us to examine the effect of the boundaries as the only difference between the two models is that the LAM takes boundary conditions from a coarser resolution global model. Figure 2(a) shows the error at $t = 0.25$ and Figure 2(b) shows the error at $t = 0.50$.

As can be seen in Figure 2(a), at time $t = 0.25$ there are small discrepancies between the two models at the boundary. The error is at a maximum at the second gridpoint in; this is the edge of the buffer zone. The error then decreases as we move inwards through the domain to become zero at several of the central gridpoints. This pattern is repeated at time $t = 0.5$ in Figure 2(b), although there are some differences worth noting. The magnitude of the maximum error is smaller at time $t = 0.5$. However the number of gridpoints with zero error is fewer, with only the very central gridpoints having no error at all. This is because the errors at the boundary are diffused inwards over time. This simple experiment shows the impact that can be made on a LAM output by taking LBCs from a model with coarser resolution.

4 Summary and Future Work

We nested a LAM within a larger model of the 1D heat equation. We have shown that taking LBCs from a model with coarser resolution can introduce errors to the model and that these errors propagate inwards over time. We have also demonstrated the 'long wave' problem and shown that the LAM output does not capture the same wavenumbers as the global model output. This has implications for data assimilation. Large scale information from observations within the limited area domain will require special treatment. One method for this has been attempted by [1] where the LAM background is modified using large scale information from the coarser model. However, this method is not suitable when the LAM is run more frequently than the coarse model providing the LBCs or when extra sources of high resolution observations are used only in the LAM.

Having considered the model itself we are now implementing a 4D-Var data assimilation scheme on the limited area domain. 4D-Var data assimilation means we find an analysis which minimizes a cost function as a nonlinear least squares problem, subject to model equations, in space and time [7]. The cost function is a measure of the distance of the analysis from a first guess field (background) and the observations, weighted by the inverse of the error covariance matrices. The status and progress of 4D-Var and in particular the problems associated with applying it to the prediction of storm-scale atmospheric phenomena are reviewed by [4], [8] and [10]. With the data assimilation we will again consider problems coming from differences in resolution and the presence of a boundary. Once we have investigated these different sources of potential problems using the heat equation, we aim to extend the model to the 1D Kuramoto-Sivashinski equation. This nonlinear wave equation illustrates self-sustained chaotic behaviour of a multiscale nature [9] and will allow us to identify problems associated with nonlinear advection rather than diffusion, as well as multiscale properties.

5 Acknowledgements

This work is supported by NERC and a CASE sponsorship with the Met Office.

References

[1] Ballard S, Li Z, Dixon M, Swarbrick S, Stiller O. Lean H. 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 GM, Dance SL, Lawless AS. Nichols NK. Scale Analysis of Observational Information in Data Assimilation. Numerical Analysis Report, Department of Mathematics, University of Reading, UK. 2006; Report 11/2006.

[3] Clark TL. Lateral and upper boundary conditions. *Numerical Methods in Atmospheric Models, Seminar Procedings, ECMWF* 1991; Volume II:43–71.

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

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

[6] Lean HW, Ballard SP, Clark PA, Dixon M, Li Z, Roberts NM. Trials of a 1km version of the unified model for short range forecasting of convective events. *World Weather Research Program Symposium on Nowcasting and Very Short Range Forecasting (WSN05)* 2005; paper 6.20.

[7] Nichols NK. Data assimilation: Aims and basic concepts. In *Data Assimilation for the Earth System,* Swinbank R, Shutyaev V, Lahoz WA (eds). Kluwer Academic: 2003; 9–20.

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

[9] Protas B, Bewley TR, Hagen G. A computational framework for the regularization of adjoint analysis in multiscale PDE systems. *Journal of Computational Physics* 2004; **195**:49–89.

[10] Sun J. Convective-scale assimilation of radar data: Progress and challenges. *Quarterly Journal of the Royal Meteorological Society* 2005; **131**:3439–3463.