The University of Reading

Radial Velocity Assimilation and Experiments with a Simple Shallow Water Model

S.J. Rennie[2] and S.L. Dance[1,2]

NUMERICAL ANALYSIS REPORT 1/2008

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

[2]*Department of Meteorology*
The University of Reading
Earley Gate, PO Box 243
Reading
Berkshire, RG6 6BB

Department of Mathematics

Abstract

The assimilation of radial wind from Doppler radar into numerical forecast models can improve weather predictions. Radial wind contains only the velocity component of wind moving along a radial line from the radar. Experiments were performed with a simple model based on the shallow water equation with no rotation. 'Radial' velocity and geopotential observations were assimilated using a 4D-Var assimilation scheme. An observation operator was introduced for the observations, to accommodate radial wind observations. A variety of experiments were performed in order to examine the effect of different parameters in finding a solution close to the truth. The quantity and error of the observations affected the accuracy of the analysis.

1 Introduction

Data assimilation is a means of incorporating observations into a forecasting model, to adjust the simulation closer to reality, and create better initial conditions for subsequent forecasts. Meteorological forecasting is making more and more use of observation assimilation in numerical weather prediction. Wind velocity from Doppler radar is one type of weather observation that is not currently fully utilised. Present effort is directed towards assimilation of radar data in high resolution models, including the Met Office's new 1.5 km resolution model (Ballard et al., 2006), as radar is one of the few means of making observations at high resolution over a large spatial scale. In this study the idea of using radial wind vectors for assimilation is considered by simulating radial velocity observations in a simple shallow water model. For radial velocity observations, a new operator was created.

This shallow water equation model was used to examine different schemes to linearise the model (see below; Lawless et al., 2003; Lawless et al., 2005). The usual tangent linear model was compared with a perturbation forecast model and shown to have similar convergence rates, although there was some difference between the analyses when observation error was present.

The model used here is based on the shallow water equation in one dimension, with no rotation. The fluid flow is modelled over an obstacle. The equations describing the model are

$$\frac{Du}{Dt}+\frac{\partial \phi}{\partial x}=-g\frac{\partial \bar{h}}{\partial x} \tag{1}$$

$$\frac{D(\ln \phi)}{Dt}+\frac{\partial u}{\partial x}=0 \tag{2}$$

$$\text{where} \quad \frac{D}{Dt}=\frac{\partial}{\partial t}+u\frac{\partial}{\partial x}. \tag{3}$$

Here, $\bar{h}=\bar{h}(x)$ is the height of the bottom orography, u is the velocity of the fluid and $\phi=gh$ is the geopotential, where g is gravity and h is the depth of the fluid above the orography (Lawless et al., 2003, Lawless et al., 2005). The model uses a discrete form of these equations using a two-time-level semi-Lagrangian integration scheme. The variables are located on a staggered grid and the new values of φ and u are not determined simultaneously.

There are several schemes that can be used to assimilate observations. Here four-dimensional variational assimilation (4D-Var) was used. In this scheme, the observations are assimilated within a

space and time domain and optimized within that domain. The fit of the model to the observations incorporates the propagation of the observations through time as the model runs. As such, the best fit of the observations into the model must also be a solution to the model.

Data assimilation involves a background term, which is the initial or prior estimate of the model state, and the analysis, which is the updated model state. The 'best fit' of the model and the observations is found by minimising a cost function, which represents the difference between the model and the observations, and the background state. The cost function, *J*, is given by

$$J[\mathbf{x_0}]=\frac{1}{2}(\mathbf{x_0}-\mathbf{x}^{b})^{\mathrm{T}}\mathbf{B_0^{-1}}(\mathbf{x_0}-\mathbf{x}^{b})+\frac{1}{2}\sum_{i=0}^{n}(H_i[\mathbf{x}_i]-\mathbf{y}_i^{o})^{\mathrm{T}}\mathbf{R}_i^{-1}(H_i[\mathbf{x}_i]-\mathbf{y}_i^{o}) \qquad (4)$$

where $\mathbf{x}_0$ is the model state at the initial time, $\mathbf{x}^b$ is the background field and $\mathbf{y}_i^o$ are the observations. The values of $\mathbf{x}$ are constrained to be solutions to the model, such that $\mathbf{x}_i=M(\mathbf{x}_{i-1})$ where M is the model operator. H is the observation operator which maps from the model space to the observation space, by producing model values at the observation location and deriving the variable if the observation type is not a model variable. $\mathbf{B}$ and $\mathbf{R}$ are the background error and observation error covariance matrices. These matrices affect the relative weight of the background and the observation terms when minimising the cost function is indicated by the inverse of the covariance matrices. The values of $\mathbf{R}$ are calculated from the observation error variances and the non-diagonal elements indicate error correlations between observations. For $\mathbf{B}^{-1}$ an inverse Laplacian is used, which is calculated according to (5).

$$\mathbf{B}^{-1}=\left(\begin{bmatrix}2c^2 & c^2 & 0 & \cdots & c^2\\ c^2 & 2c^2 & c^2 & \cdots & 0\\ 0 & c^2 & 2c^2 & \cdots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ c^2 & 0 & 0 & \cdots & 2c^2\end{bmatrix}^2+I\right)\cdot\begin{bmatrix}\sigma_u^2 & 0 & \cdots & 0 & 0\\ 0 & \sigma_u^2 & & \vdots & \vdots\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ \vdots & \vdots & & \sigma_\phi^2 & 0\\ 0 & 0 & \cdots & 0 & \sigma_\phi^2\end{bmatrix} \qquad (5)$$

where c is the correlation length and σ^2 is the variance of the error for background values of u and φ.

To calculate the cost function, a discrete version of this equation is used, and solved with an iterative algorithm (described in Lawless et al., 2005). The observation operator is linearised (denoted **H**) about the model state for each iteration, to reduce computational effort. The model operator M is also linearised, which makes the cost function quadratic, which for the model used here can be achieved with either a tangent linear model or a perturbation forecast model. The incremental cost function is given by (6)

$$J^{(k)}[\delta\mathbf{x}_0^{(k)}]=\frac{1}{2}(\delta\mathbf{x}_0^{(k)}-[\mathbf{x}^{b}-\mathbf{x}_0^{(k)}])^{\mathrm{T}}\mathbf{B_0^{-1}}(\delta\mathbf{x}_0^{(k)}-[\mathrm{x}^{b}-\mathbf{x}_0^{(k)}])+\frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i\delta\mathbf{x}_i^{(k)}-\mathbf{d}_i^{(k)})^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{H}_i\delta\mathbf{x}_i^{(k)}-\mathbf{d}_i^{(k)})$$

where $\mathbf{d}_i^{(k)}=\mathbf{y}_i^{o}-H_i[\mathbf{x}_i^{(k)}]$ are the innovation vectors for each observation and the increment

$\delta \mathbf{x}_0^{(k)} = \mathbf{x}_0^{(k+1)} - \mathbf{x}_0^{(k)}$ where k is the iteration number. These iterations constitute the outer loop of the solution Each outer loop involves improving the guess for $\mathbf{x_0}$, which is used to run the nonlinear model to calculate $x_i^{(k)}$ at each timestep t_i, used in the incremental cost funtion (6).

Within each outer loop, the cost function is minimised, using an iterative procedure referred to as the inner loop. The minimisation is accomplished by finding the minimum of the gradient of the cost function. The gradient indicates the direction for each iteration of the inner loop to improve the estimate. The gradient of the cost function is determined using an adjoint model. The adjoint model consists of the derivatives of the cost function with respect to the variables. In this model, a Beale restarted memoryless quasi-Newtonian conjugate gradient method is used for the minimisation, implemented by the CONMIN routine[1].

For a solitary Doppler radar, the only radial velocity can be measured directly, i.e. the component of movement directed towards or away from the radar. For example, in the case of a monotonic wind, the velocity will be negative on one side of the radar, positive on the other, and zero at a tangent. For the 1D case here, an observation operator was created as a diagonal matrix, where the velocity was multiplied by -1 on one side of the radar, and multiplied by 1 on the other side, so the diagonal elements h_i of $\mathbf{H}$ were:

$$h_i = \begin{cases} -1 & x < x_r \\ 1 & x > x_r \end{cases} \qquad (7)$$

where the location x_r of the radar on the grid was specified, and x represents the grid locations. For simplicity, x_r was assumed to fall between grid points, as the radar does not make observations at it's own location so this value would be zero. Off-diagonal elements of $\mathbf{H}$ would also be zero. In a full 3D wind field, the observation operator would map the velocity vector at any point to the position vector of the velocity with respect to the radar. The spatial extent and spatial resolution of observations could be specified independently of the observation operator, for simulating limited range of Doppler velocities. Another consideration is that the spatial frequency of the observations may not match that of the model grid.

In this report, section 2 describes the model setup, lists the model parameters and associated values, as were used in various experiments. Section 3 describes the effect of various parameters and describes the results of several specific experiments.

1 In the ACM TOMS package available from the GAMS software library at gams.nist.gov

2 Model Parameters

The model has a one-dimensional domain of grid length 200 and grid spacing 0.01. The timestep was 9.2e-3. A run of the model providing u and φ everywhere provided 'truth' for comparison with the analysis, and was used to generate observations and the background. The background was the truth with a phase shift. Obsevations were sampled from the truth and then random noise was added according to the variance specified, unless perfect observations were being used (see section 3.3).

A range of experiments were conducted with the model, changing various model parameters and observation choices. These parameters (with range of values in brackets) include:

- Presence of u and/or φ observations to be assimilated.
- Range of observations of u and/or φ (everywhere, limited to 50 <x<150)
- Spatial frequency of observations(every gridpoint in x or every 10)
- Correlation length (used for **B**) (10 to 50)
- Variance of observation errors (used for **R** diagonal elements) (0.001-0.02)
- Number of assimilation time steps (normally 50, tried 100)
- Forecasting time steps (normally 0, tried 100)
- Background weight (additional term to weight the background term in the cost function (4)) (0.001 - 1)

Various parameters were also varied to test the effect on convergence.

- Outer loops, ideally more than required for convergence (normally 50, tried up to 150)
- Solver tolerance for convergence of the final solution (1e-8 – 1e-3)
- Number of conmin iterations (the inner loop) (200 usually, tried 300 and 400)
- Inner tolerance for convergence of cost function and its gradient in conmin (O(0.1 -0.5))
- Outer tolerance for convergence within outer loop.(O(0.0005-0.001))
- Solver max iterations for minimising the cost function (normally 200, used 50 for perfect observations)

Other options.

- Background type – how the background was calculated from the (truth + phase error)
- Covariance B matrix type (I or Laplace (see 5))
- Covariance R matrix type (I or real, i.e. using real variances of observations)

The parameters used in the various trials and experiments are tabulated in Appendix 1. For the main

set of experiments, most trials are displayed graphically in Appendix 2. Each graph has four panels, showing the background, truth, and analysis for velocity (u) and for geopotential (φ), the cost function value, and the cost function relative gradient.

3 Results

3.1 Effect of some parameters, based on experiments

This section discusses briefly the effect of various parameters which were noted in Section 2. Initiailly, various experiments were run with different choices of parameters, in order to select one suitable to act as a control experiment. The control is described as Trial 1 in the Appendices, and used 'typical' values for quantitative parameters, and observations coinciding with all grid point locations. For the subsequent trials, as listed in the Appendices, one or two parameters were varied. All were performed with the radar observation operator in place, as whether the velicity observations were 'radial' or not would not effect the analysis. The effect of various parameters is now described.

Increasing the number of assimilation time steps visibly improved the analysis, i.e. it more closely resembled the truth. Changing the number of time steps from 50 to 100 (as in trials 16-19 as shown in Appendix 2) lowered the final cost function value and its gradient by an order of magnitude, because the number of observation was doubled, thus a far more precise analysis could be obtained.

The analysis was much smoother (not shown) when using the real **R** matrix based on observation covariances, and the inverse Laplacian **B** matrix (5), compared to using the Identity matrix for the **B** and **R** matrices. (All experiments described here use the real **R** and inverse Laplacian **B** matrices.) The correlation length affected the smoothness of the analysis as it affected how the observations were spread within the spatial domain. There was an issue with the φ analysis values that occurred when there were no φ observations, which resulted in the φ analysis values offset from the truth and background by ~0.1. This offset was affected by the correlation length. For a correlation length of 20 the offset manifested (Figure 1); for a correlation length of 50 there was no offset (Figure 2; see trials 10-12). The reason for this was not determined, though one possibility is that the spread of observations was not sufficient to allow for calculation of the value of φ, even though the slope of φ was able to be calculated with some accuracy.

Typically, the cost function would reach a constant value (e.g. to four significat figures) very quickly, but the (relative) gradient, used as the stopping criteria, would decline erratically. With no background term (background weight = 0), the gradient declined without the noise. However, the analysis was not as smooth when the background was disregarded (see Trial 19 in Appendix 2).

The background term can provide the starting point for assimilation. The background variance and weight of the background term determined how much the background contributed to the analysis, and to the cost function. If the background weight was zero, the analysis converged much closer to the observations. However, the analysis was also noisy, which was not due to the observation noise (as demonstrated when the observations were perfect). Using limited-domain observations and ignoring the background term caused the model to explode (Trial 18, not shown), which implied there was insufficient information from which to produce a viable analysis.

A non-zero background weight substantially increased the cost function value (J) from $O(500)$ to $O(20000)$. However, the cost function also depended upon the number of observations, which made it harder to compare the size of J in runs with a different number of observations. Results from a study by Sun and Crook (1997), showed that for assimilation of single Doppler radar winds, including a background term improved the analysis.

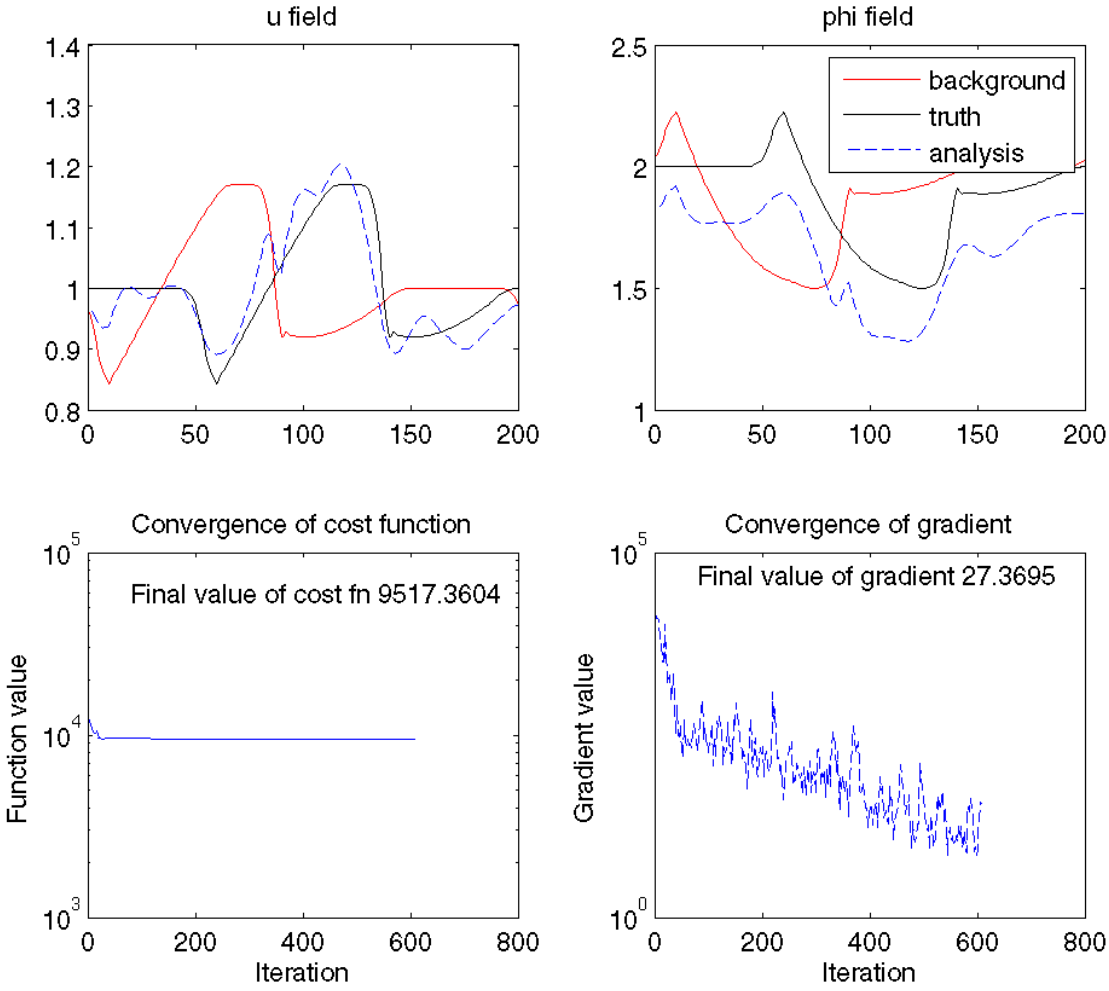


Figure 1: (Trial 3) No φ observations and correlation length was 20. The φ value was offset.

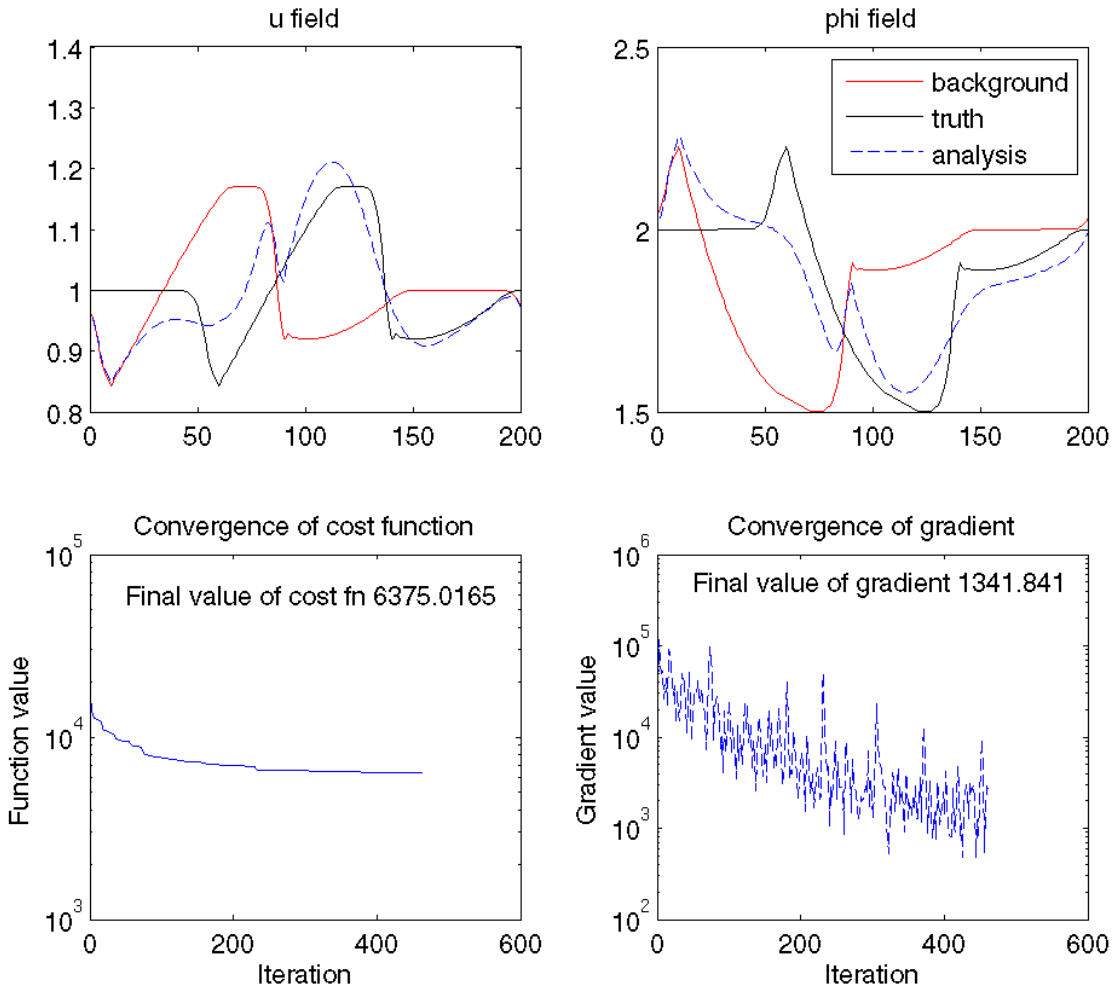


Figure 2. (Trial 9) No φ observations and u observations were limited to between 50 and 150. The correlation length was 50.

Several trials were performed varying the availability of observations. With only φ observations, the solution did not converge as well. u observations alone gave a better solution, but observations of both u and φ gave the best solution. Limiting the spatial extent of observations within the domain to close to the radar also resulted in poorer convergence. Naturally, more observations will result in a solution closer to the truth.

Reducing the tolerance of the inner and outer loops allowed the model to converge to a solution faster. The number of conmin iterations was usually (half to two thirds of loops) insufficient to allow convergence within the outer loop, using the outer tolerance of 2e-4. Convergence was improved with more perfect observations or a higher outer tolerance (around 2e-3, Trial 17). For solution convergence within the specified number of outer loops, a more relaxed solver tolerance was required. A solver tolerance of 1e-3 was found suitable for noisy observations of limited range (see

Trials 16-19). For most experiments, the solver tolerance was probably too small.

3.2 Single observation tests

Several experiments were performed using one single observation location, for both variables and at all forecast timesteps. The paramters are listed in Table 2 of Appendix 1. In this case the analysis resembled the background in that the phase offset was not altered; however the amplitude of the variables was affected throughout (Figure 3). The location of the observation did not seem to matter (not shown) and the shape of the wave was not altered at all, when the background weight was not zero As the background is the truth with a phase shift, this brought the the estimate closer to the truth in some parts, and moved further away in other parts. When the background weight was zero, there was a localised shift in the analysis closer to the truth.

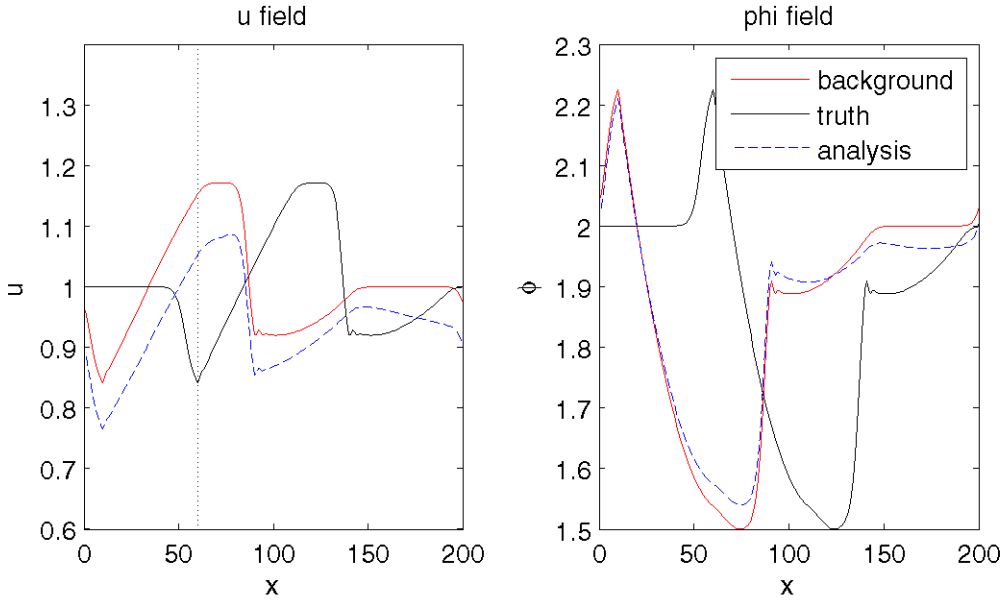


Figure 4. Analysis using one observation of *u* located at x=60, marked by the dotted line.

3.3 Perfect observation test

For this test the observations were made exactly equal to the truth (no noise was added) and were located at every grid point. The variance of the observation errors , used for the **R** matrix, was varied between 0.02 and 0.002. The number of assimilation time steps was increased from 50 to 100 and the maximum solver iterations was reduced to 50 (to match the setup described by Lawless et al., 2004).

This resulted in the analysis exactly matching the truth, unless the background weight was 0. With the background weight equal to 0 the analysis was noisy. This resulted in cost function values $O(500)$ and gradient values $O(0.002)$. These cases reached convergence within the number of outer loops specified.

Running with a limited range of data and 100 assimilation time steps reduced the gradient by a factor of 4. The cost function magnitude depended upon the number of observations. The analysis appeared very similar to the result from 50 assimilation time steps.

3.4 Radar Observation Set Up

To simulate the assimilation of radar data, an observation was introduced operator to make the model appear like the observations, as discussed in the introduction. The observation operator was stored as a vector of 1 and -1 values (7). In this case the velocity observations were negative before the radar (see Figure 4). The location of the last observation before the radar was input in vars_user_mod.f90 to specify the radar location. When the observations were generated (from the truth, plus noise), the sign was changed for u observations before the radar. (The radar does not observe φ.) For example, see Figure 1. As might be expected in a situation where radar targets were not universally distributed, the range of observations was also limited to a part of the domain. Similarly the radar observations could be of lower resolution than the model, as ground clutter interference or superobbing might produce. Thus the spatial resolution was decreased from 1 to 10 for some model runs (not shown). The result of this was to reduce the accuracy of the analysis compared to the truth, as would be expected.

Using 4D-var, the observations were advected and the analysis improved beyond the range of the observations in the direction of advection. In the opposite direction the observations had no influence and the analysis was closer to the background. Figure 2 indicates this, where for $0<x<50$ the analysis follows the background, but for the rest of the domain, the analysis is close to the truth, even where there were no observations. Trials 9-12 matched the radar setup described here, where the correlation length was varied. Trials 13-15 had velocity observations everywhere in the domain. Trials 9-15 all excluded φ observations, as a radar can not observe φ directly.

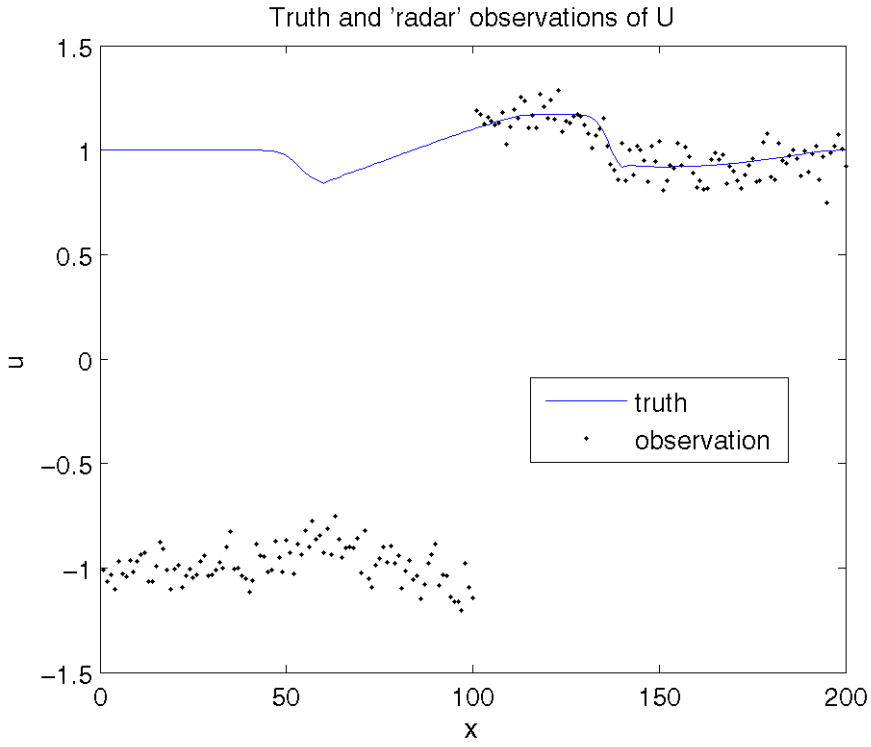


Figure 4. Truth and observations for the radar assimilation experiments.

4 Summary

The 4D-Var assimilation of velocity and geopotential observations into a simple shallow water equation model was explored here, with two objectives. The first was to become familiar with various parameters and their effect, and the second was to consider the setup for radial velocity observations, like that from a Doppler radar. The 'radar' setup involved creating an observation operator that reflected the change in velocity sign of the observation about the location of the radar.

It was clear that the number of observations had an impact upon how closely the analysis resembled the truth. The number of observations could change through the number of timesteps with observations, the spatial range and spatial frequency of observations, and the absence of u or φ observations.

The tolerance which marks when convergence is achieved (of either the inner loop, the outer loop, or the solution) must be sufficiently large for cases where the observations are noisy. A solver tolerance inthe order of 10^{-3} was found suitable for noisy observations.

The use of a background can improve the analysis by providing the field where there are no observations. With 4D-Var, the observations can also be advected through to improve the analysis where there are no observations.

Bibliography

Ballard, S., Dixon, M., Swarbrick, S., Li, Z. & Stiller,O. (2006), Assimilation of radar data for convective scale NWP at the Met Office. *Proceedings of ERAD 2006 (Extended abstract)*.

Lawless, A. S., Nichols, N. K., & Ballard, S. P. (2003) A comparison of two methods for developing the linearisation of a shallow-water model, *Quarterly Journal of the Royal Meteorological Society*, 129, 1237-1254.

Lawless, A. S., Gratton, S., & Nichols, N. K. (2005) An investigation of incremental 4D-Var using non-tangent linear models, *Quarterly Journal of the Royal Meteorological Society*, 131, 459-476.

Nichols, N. K. *Data Assimilation: Aims and Basic Concepts*, Department of Mathematics, University of Reading.

Rihan, F. A., Collier, C. G., & Roulstone, I. (2005) Four-dimensional variational data assimilation for Doppler radar wind data, *Journal of Computational and Applied Mathematics*, 176, 15-34.

Sun, J. & Crook, N. A. (1997) Dynamical and Microphysical Retrieval from Doppler Radar Observations Using a Cloud Model and Its Adjoint. Part I: Model Development and Simulated Data Experiments, J*ournal of Atmospheric Sciences*, 54, 1642-1661.

Appendix 1

Tables indicating parameters of experiments.

Table 1. Parameter variation experiments.

Description	Model 1 (M)	Obs everywhere 2	No φ obs 3	No *u* obs 4	M+corr=50 5	M+corr50 +tol 6
Number of assim timesteps	50	50	50	50	50	50
Number of forecast timesteps	0	0	0	0	0	0
Solver max iterations	200	200	200	200	200	200
Solver tolerance	1.00E-008	1.00E-008	1.00E-008	1.00E-008	1.00E-008	1.00E-008
First *u* ob timestep	0	0	0	-	0	0
Last *u* ob timestep	End	End	End	-	End	End
First φ ob timestep	0	0	-	0	0	0
Last φ ob timestep	End	End	-	End	End	End
Time frequency of *u* obs	1	1	1	-	1	1
Time frequency of φ obs	1	1	-	1	1	1
First *u* ob point	50	0	0	-	50	50
Last *u* ob point	150	End	End	-	150	150
First φ ob point	50	0	-	0	50	50
Last φ ob point	150	End	-	End	150	150
Space frequency of *u* obs	1	1	1	-	1	1
Space frequency of φ obs	1	1	-	1	1	1
Variance of *u* obs	0.02	0.02	0.02	-	0.02	0.02
Variance of φ obs	0.02	0.02	-	0.02	0.02	0.02
Variance of *u* background	1	1	1	1	1	1
Variance of φ background	2	2	2	2	2	2
Corr length	20	20	20	20	50	50
Background weight	1	1	1	1	1	1
Max number outer loops	50	50	50	50	50	50
Minim max iters	200	200	200	200	200	200
Cov B matrix*	Laplace	Laplace	Laplace	Laplace	Laplace	Laplace
Cov R matrix**	Real	Real	Real	Real	Real	Real
Stop criteria	Relative grad	Relative grad	Relative grad	Relative grad	Relative grad	Relative grad
Inner tolerance	3.00E-001	3.00E-001	3.00E-001	3.00E-001	3.00E-001	5.00E-001
Outer tolerance	2.00E-004	2.00E-004	2.00E-004	2.00E-004	2.00E-004	1.00E-003
φbar	1.5	1.5	" "			
Comments				Not converging		Converging faster

* Inverse Laplacian as shown in Equation 5.
** Real **R** uses the variance of the observations.

Continued on the next two pages.

Table 1 continued.

M+corr50+tol	M+corr50+tol 6+no φ	Corr=10	Corr=20	Corr=50	Corr=50, *u* allover
7	8	9	10	11	12	13
50	50	50	50	50	50	50
0	0	0	0	0	0	0
200	200	200	200	200	200	200
1.00E-008	1.00E-007	1.00E-008	1.00E-008	1.00E-008	1.00E-008	1.00E-008
0	0	0	0	0	0	0
End	End	End	End	End	End	End
0	0	-	-	-	-	-
End	End	-	-	-	-	-
1	1	1	1	1	1	1
1	1	-	-	-	-	-
50	50	50	50	50	50	1
150	150	150	150	150	150	End
50	50	-	-	-	-	-
150	150	-	-	-	-	-
1	1	1	1	1	1	1
1	1	-	-	-	-	-
0.02	0.02	0.02	0.02	0.02	0.02	0.02
0.02	0.02	0.02	0.02	0.02	0.02	0.02
1	1	1	1	1	1	1
2	2	2	2	2	2	2
50	50	50	10	20	50	50
1	1	1	1	1	1	1
50	50	50	50	50	50	50
200	200	200	300	300	300	300
Laplace	Laplace	Laplace	Laplace	Laplace	Laplace	Laplace
Real	Real	Real	Real	Real	Real	Real
Relative grad	Relative grad	Relative grad	Relative grad	Relative grad	Relative grad	Relative grad
1.00E-001	5.00E-001	5.00E-001	5.00E-001	5.00E-001	5.00E-001	5.00E-001
1.00E-004	1.00E-003	1.00E-003	1.00E-003	1.00E-003	1.00E-003	1.00E-003
				1.5	1.5	
Conv slower Like 6		Test for φ offset problem				

Table 1 continued.

3+300conmin	φ_bar=2	n. forecast	Long run for convergence	Tolerance for Convergence	No Bkgd	No bkgd, obs everywhere
14	--	15	16	17	18	19
50	50	50	100	100	100	100
0	0	10	0	0	0	0
200	200	200	200	200	200	200
1.00E-008	1.00E-008	1.00E-008	5.00E-004	1.00E-003	1.00E-003	1.00E-003
0	0	0	0	0	0	0
End	End	End	End	End	End	End
-	-	-	0	0	0	0
-	-	-	End	End	End	End
1	1	1	1	1	1	1
-	-	-	1	1	1	1
1	1	1	50	50	50	1
End	End	End	150	150	150	End
-	-	-	50	50	50	1
-	-	-	150	150	150	End
1	1	1	1	1	1	1
-	-	-	1	1	1	1
0.02	0.02	0.02	0.02	0.01	0.01	0.01
0.02	0.02	-	0.02	0.02	0.02	0.02
1	1	1	1	1	1	1
2	2	2	2	2	2	2
20	20	50	50	50	50	50
1	1	1	0.001	0.001	0	0
50	50	50	150	50	50	50
300	300	300	400	200	200	200
Laplace	Laplace	Laplace	Laplace	Laplace	Laplace	Laplace
Real	Real	Real	Real	Real	Real	Real
Relative grad	Relative grad	Relative grad	Relative grad	Relative grad	Relative grad	Relative grad
5.00E-001	5.00E-001	5.00E-001	5.00E-001	3.00E-001	3.00E-001	3.00E-001
1.00E-003	1.00E-003	1.00E-003	1.00E-003	2.00E-003	2.00E-003	2.00E-003
	2					Conv in 19 loops
	No effect	No effect		Converged in 49 loops	Crashed	Much higher J More obs

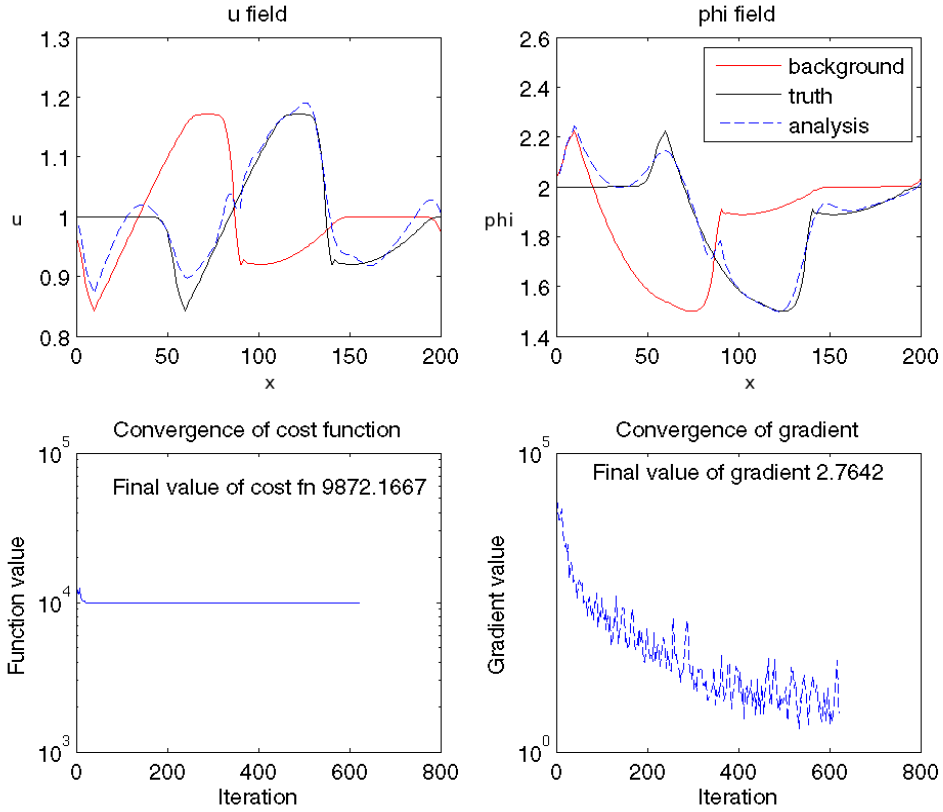
Table 2. Single observation test.

	1		
Number of assim timesteps	50	50	50
Solver max iterations	200	200	200
Solver tolerance	1.00E-008	1.00E-008	1.00E-008
First *u* ob timestep	1	1	1
Last *u* ob timestep	End	End	End
First φ ob timestep	-	-	-
Last φ ob timestep	-	-	-
Time frequency of *u* obs	1	1	1
Time frequency of φ obs	-	-	-
First *u* ob point	60	60	60
Last *u* ob point	60	60	60
First φ ob point	-	-	-
Last φ ob point	-	-	-
Space frequency of *u* obs	1	1	1
Space frequency of φ obs	-	-	-
Variance of *u* obs	0.02	0.02	0.02
Variance of φ obs	0.02	0.02	0.02
Variance of *u* background	1	1	1
Variance of φ background	2	2	2
Corr length	50	20	50
Background weight	1	1	0
Max number outer loops	50	50	50
Minim max iters	500	500	500
Cov B matrix	Laplace	Laplace	-
Cov R matrix	Real	Real	Real
Stop criteria	Relative grad	Relative grad	Relative grad
Inner tolerance	5.00E-001	5.00E-001	5.00E-001
Outer tolerance	1.00E-003	1.00E-003	1.00E-003
Comments	Shift		Has a localised change

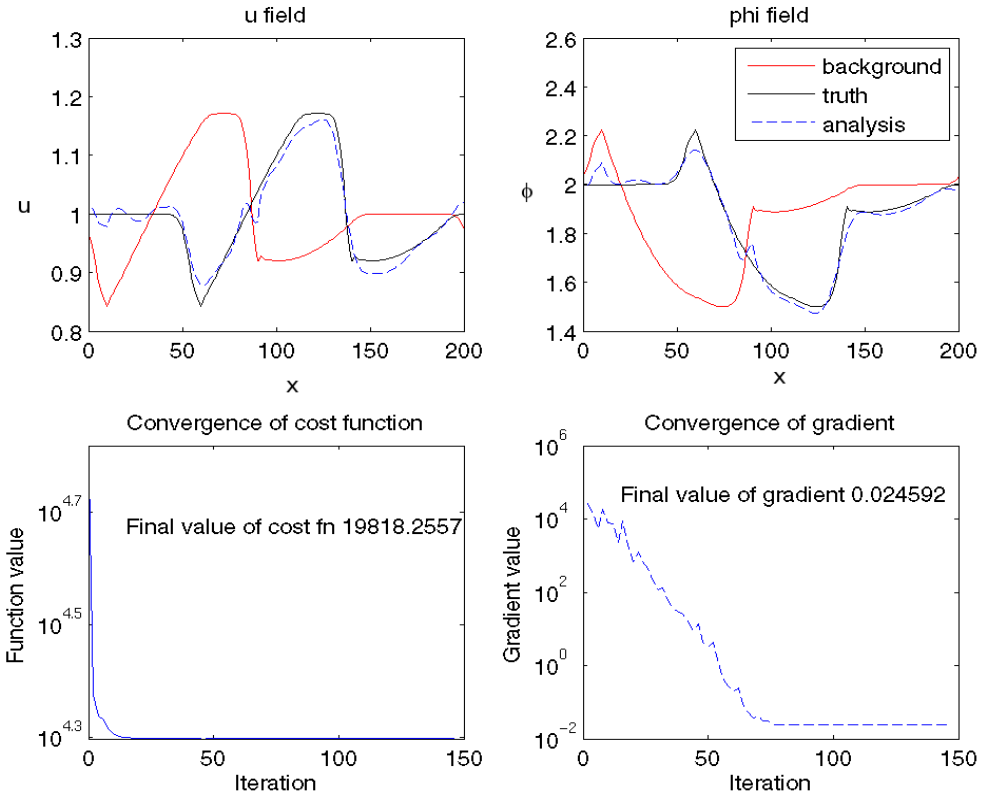
Table 3. Perfect observation tests (Observations = truth).

		Solver tol	No variance	Lower variance		
Number of assim timesteps	50	50	50	50	50	50
Solver max iterations	50	50	50	50	50	50
Solver tolerance	1.00E-008	1.00E-004	1.00E-008	1.00E-008	1.00E-008	1.00E-008
First *u* ob timestep	1	1	1	1	1	1
Last *u* ob timestep	End	End	End	End	End	End
First φ ob timestep	1	1	1	1	1	1
Last φ ob timestep	End	End	End	End	End	End
Time frequency of *u* obs	1	1	1	1	1	1
Time frequency of φ obs	1	1	1	1	1	1
First *u* ob point	1	1	1	1	1	1
Last *u* ob point	End	End	End	End	End	End
First φ ob point	1	1	1	1	1	1
Last φ ob point	End	End	End	End	End	End
Space frequency of *u* obs	1	1	1	1	1	1
Space frequency of φ obs	1	1	1	1	1	1
Variance of *u* obs	0.02	0.02	0	0.001	0.001	0.001
Variance of φ obs	0.02	0.02	0	0.001	0.001	0.001
Variance of *u* background	1	1	1	1	1	1
Variance of φ background	2	2	2	2	2	2
Corr length	50	50	50	50	50	50
Background weight	1	1	1	1	2	0.5
Max number outer loops	50	50	50	20	20	20
Minim max iters	500	500	500	500	500	500
Cov B matrix	Laplace	Laplace	Laplace	Laplace	Laplace	Laplace
Cov R matrix	Real	Real	Identity	Real	Real	Real
Stop criteria	Relative grad	Relative grad	Relative grad	Relative grad	Relative grad	Relative grad
Inner tolerance	5.00E-001	5.00E-001	5.00E-001	5.00E-001	5.00E-001	5.00E-001
Outer tolerance	1.00E-003	1.00E-003	1.00E-003	1.00E-003	1.00E-003	1.00E-003
Comments		Not much difference		Much better convergence	J increase	J decrease. Better. Still not converging.

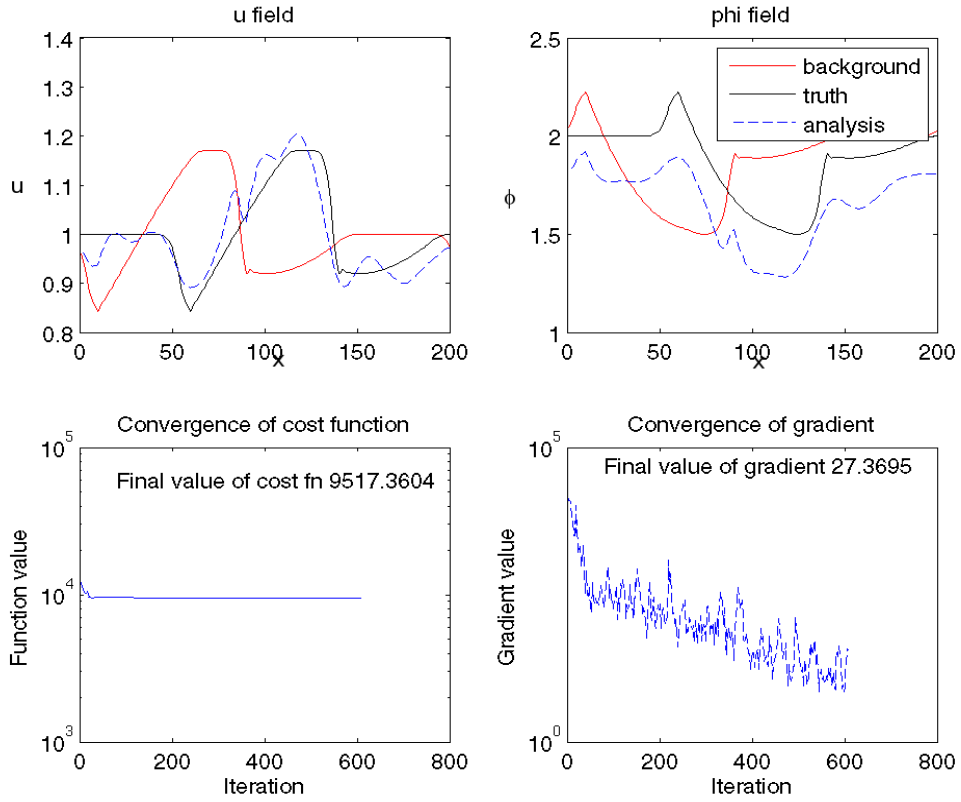
Appendix 2. Figures showing output from trials in Table 1.



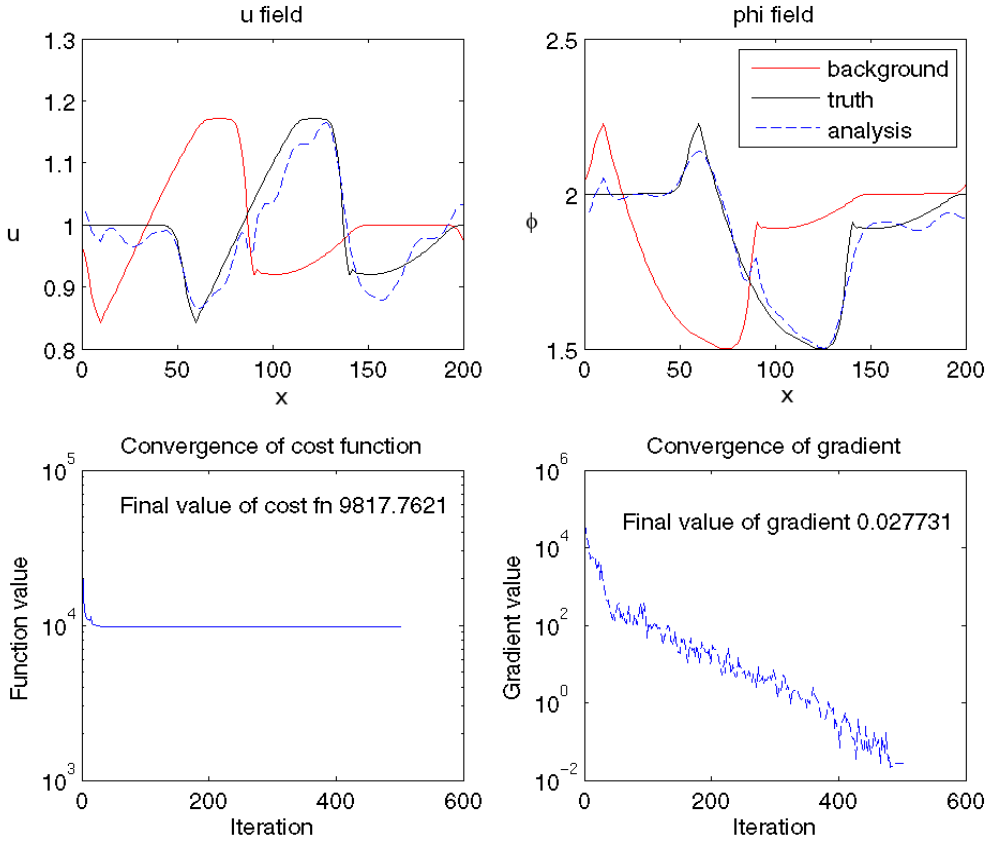
Trial 1.



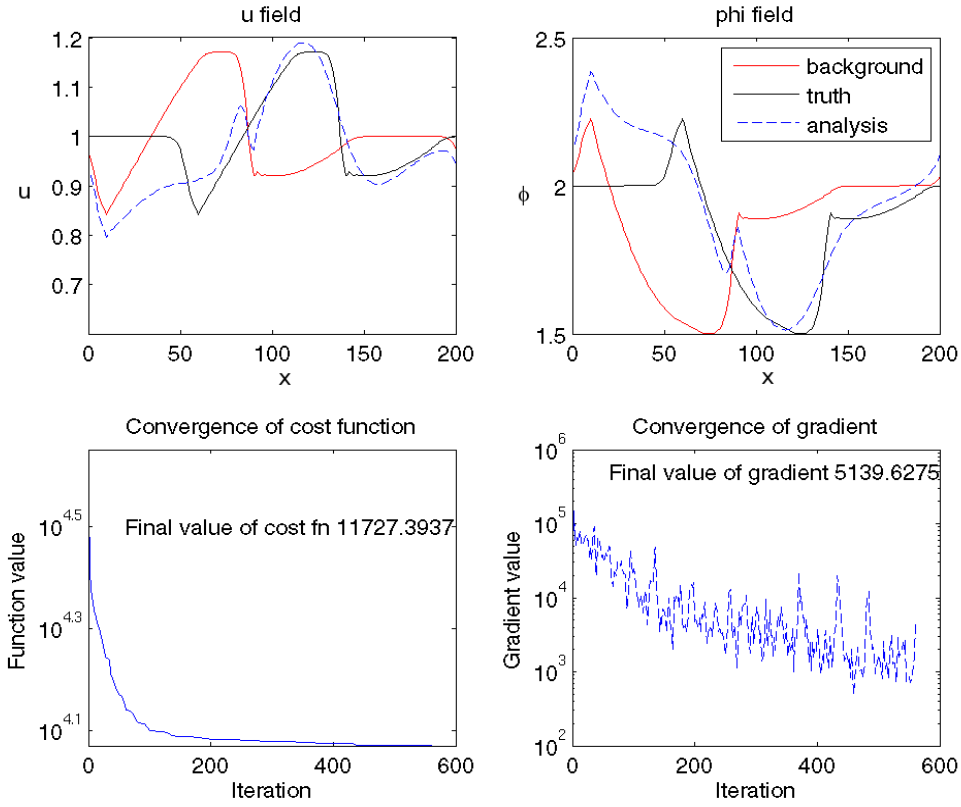
Trial 2.



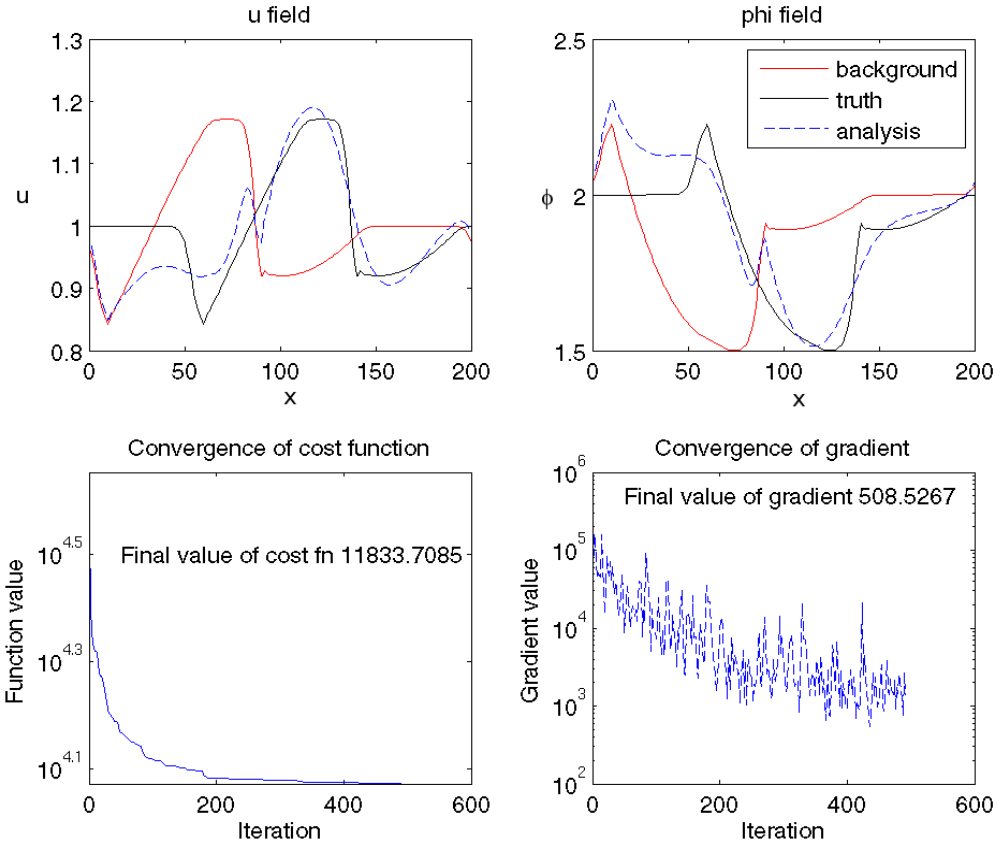
Trial 3.



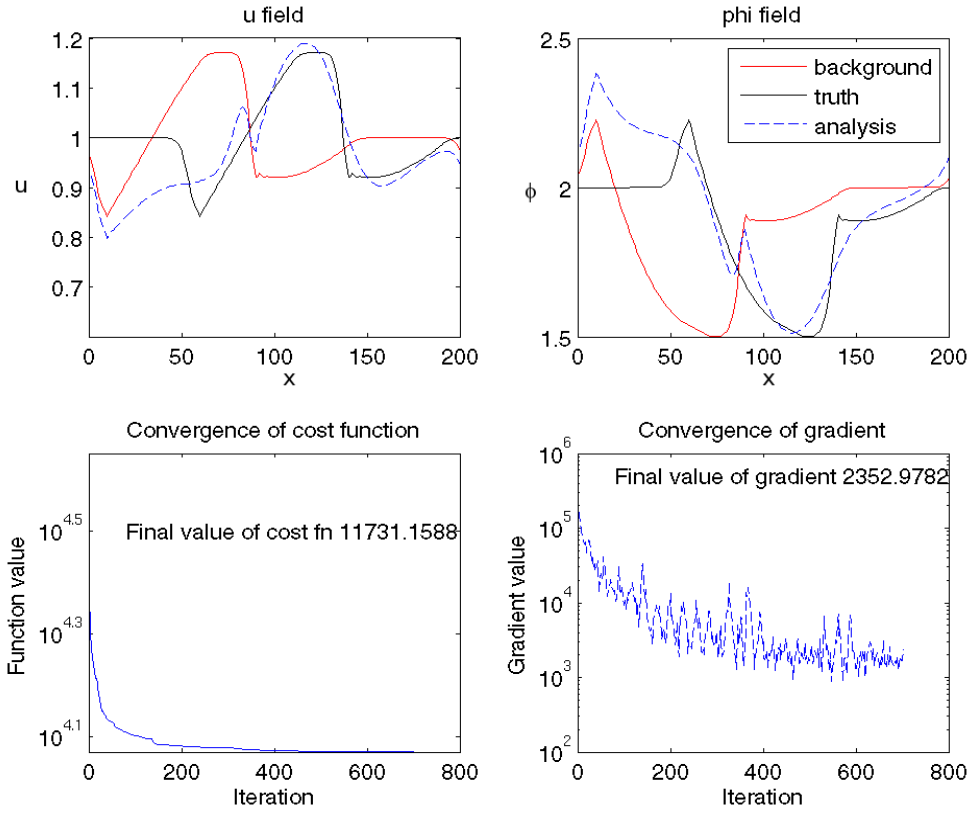
Trial 4.



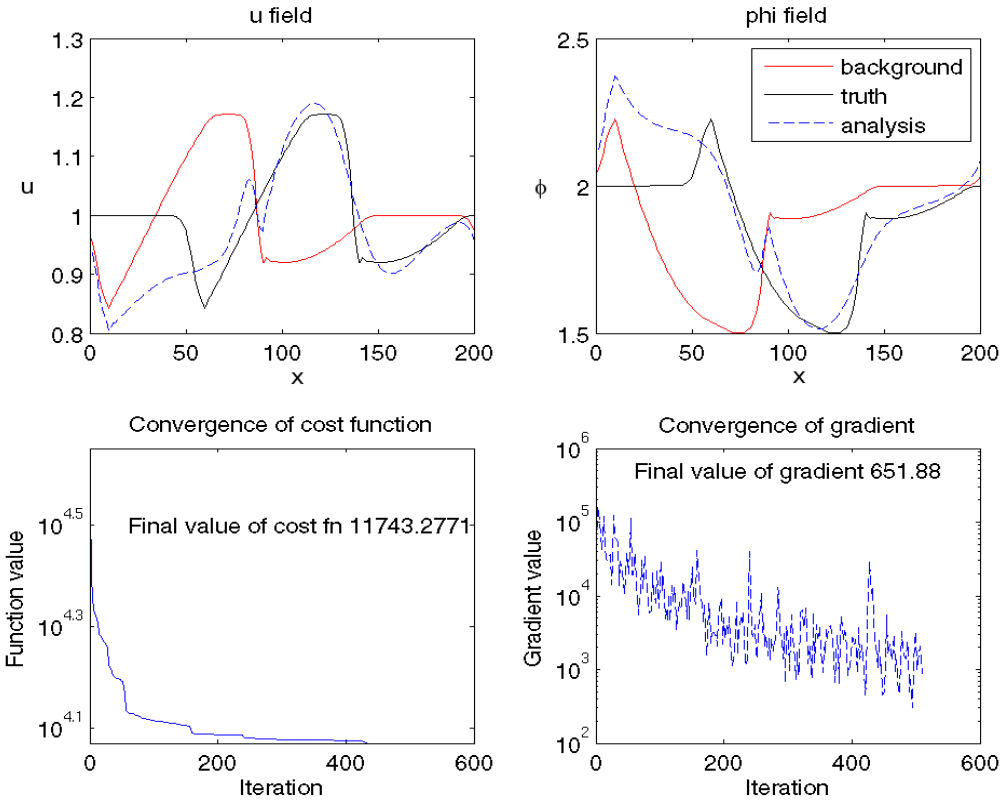
Trial 5.



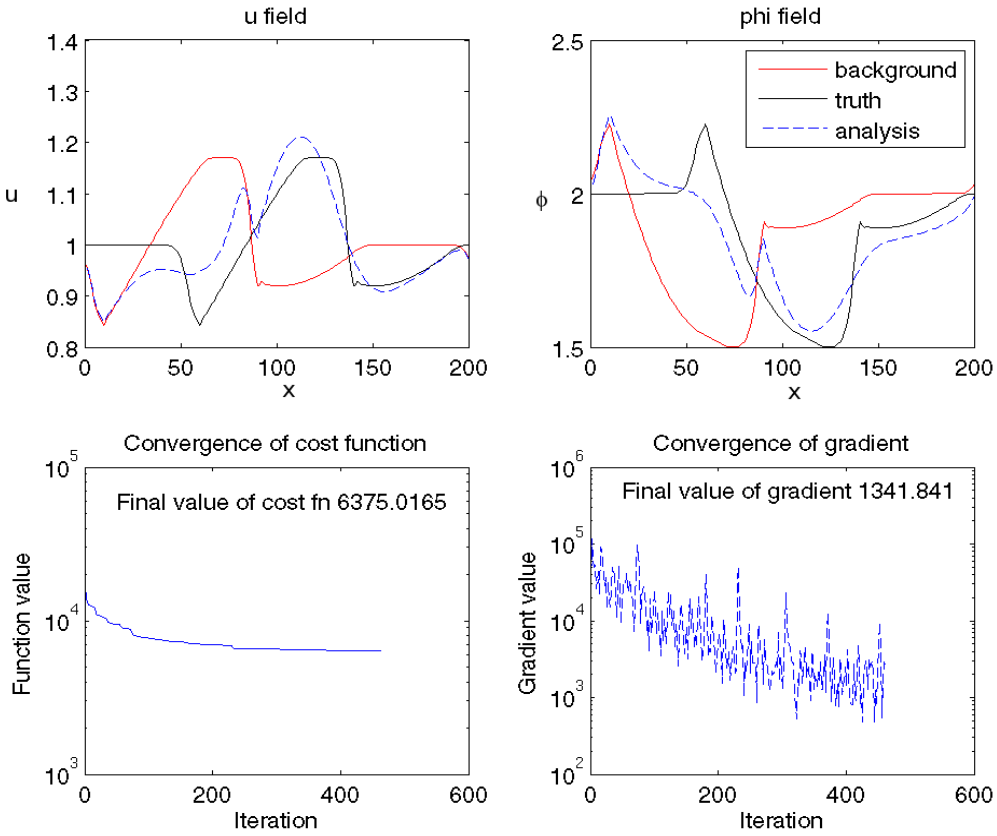
Trial 6.



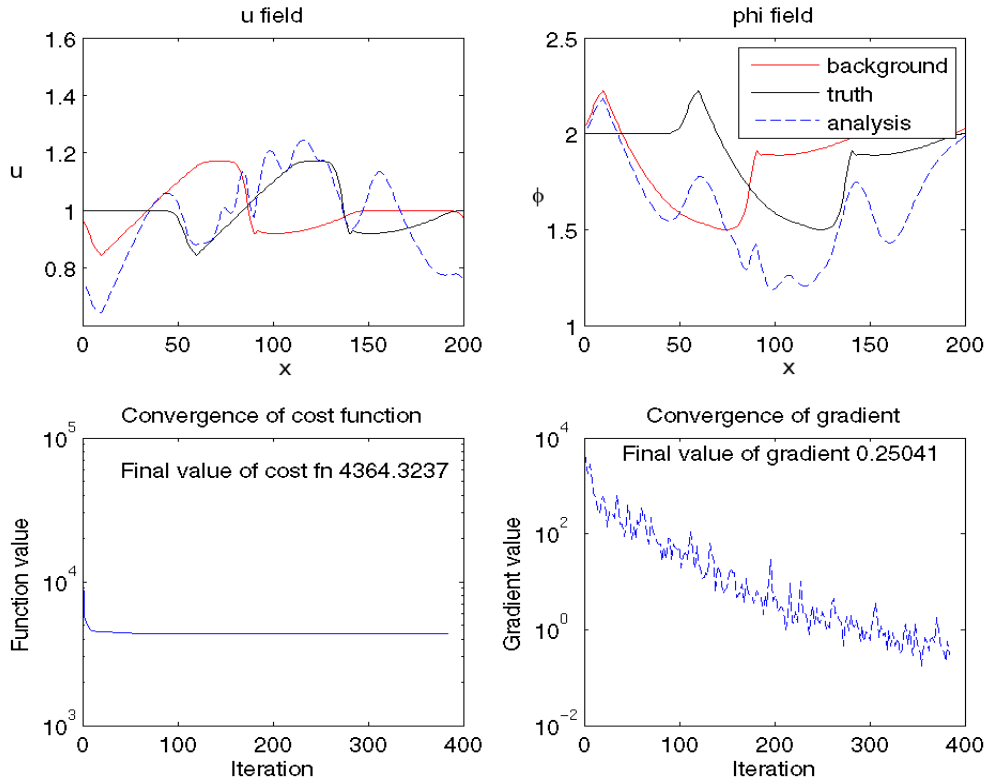
Trial 7.



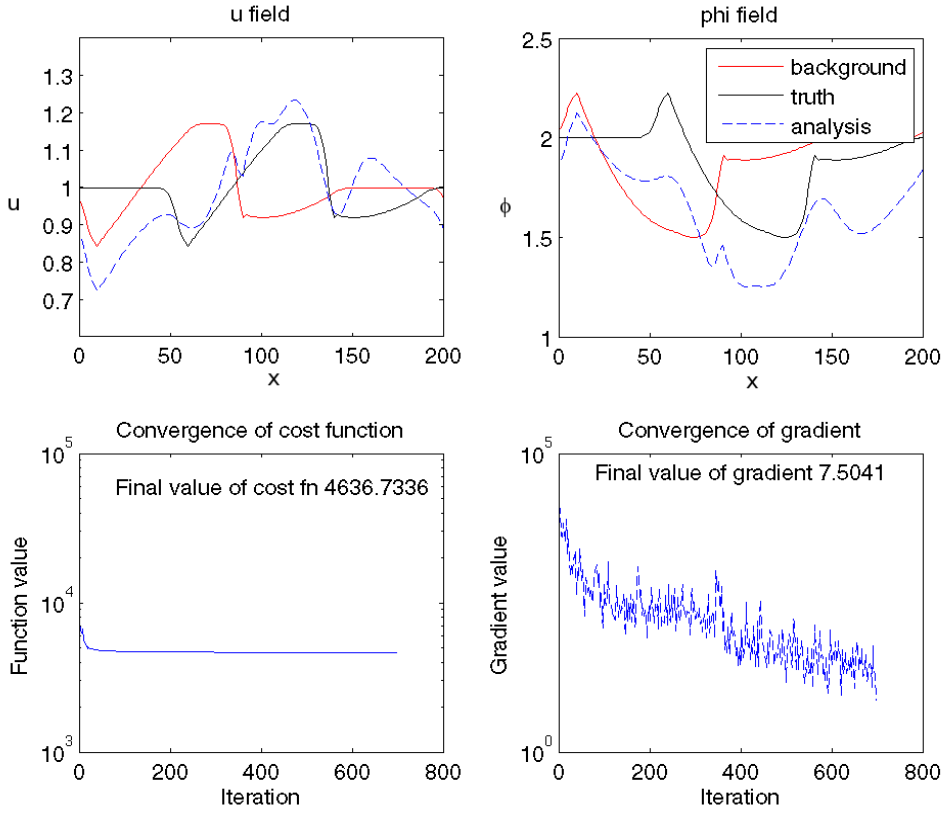
Trial 8.



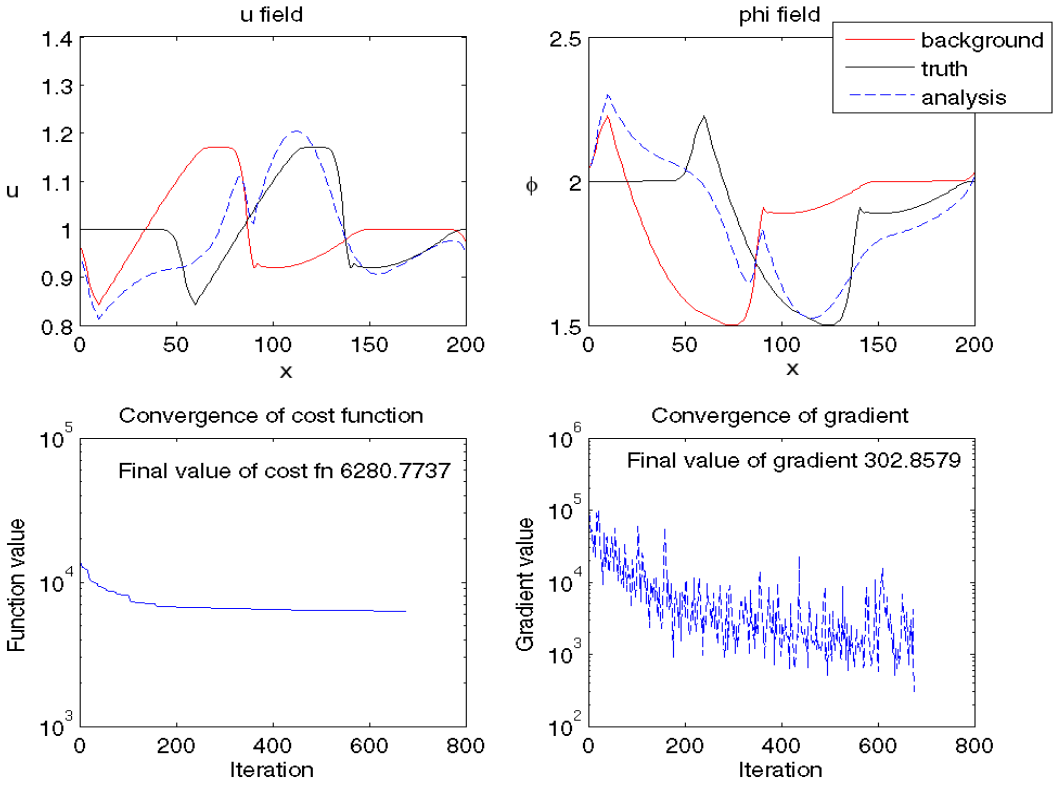
Trial 9.



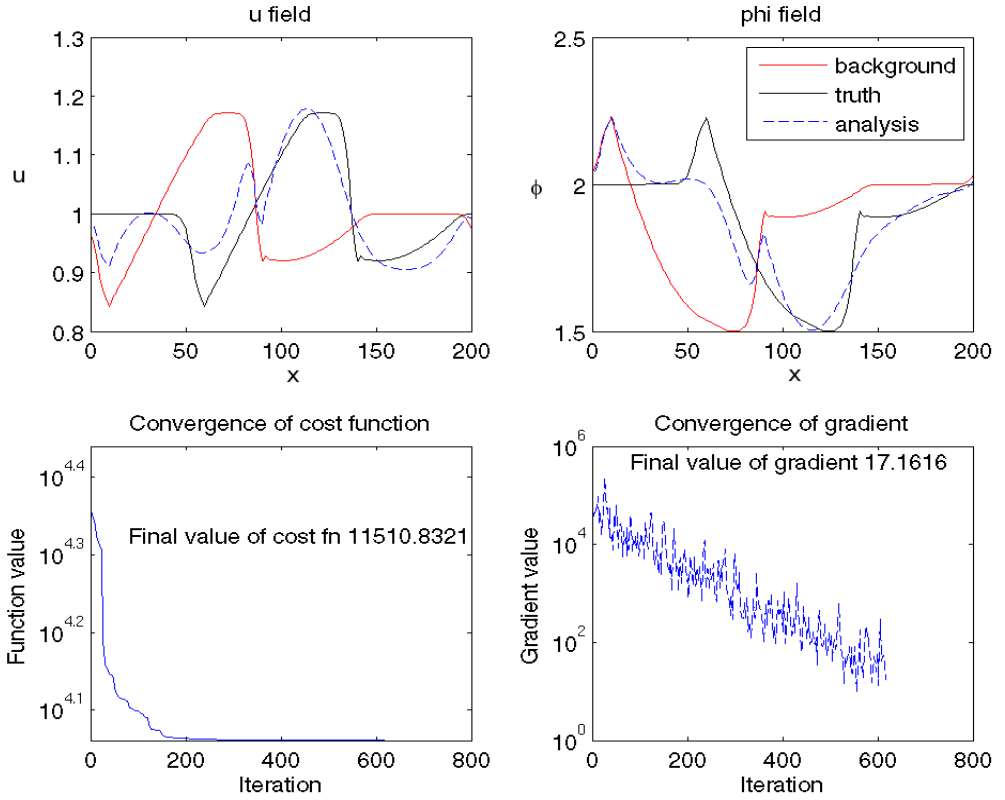
Trial 10.



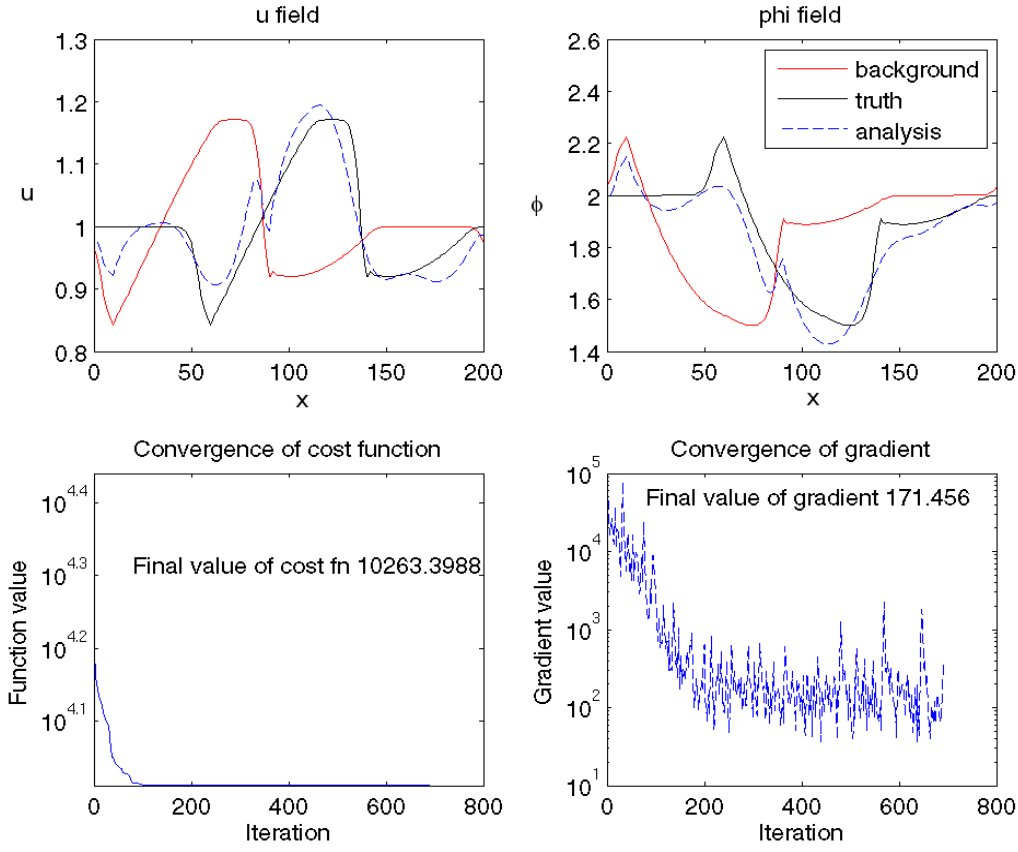
Trial 11.



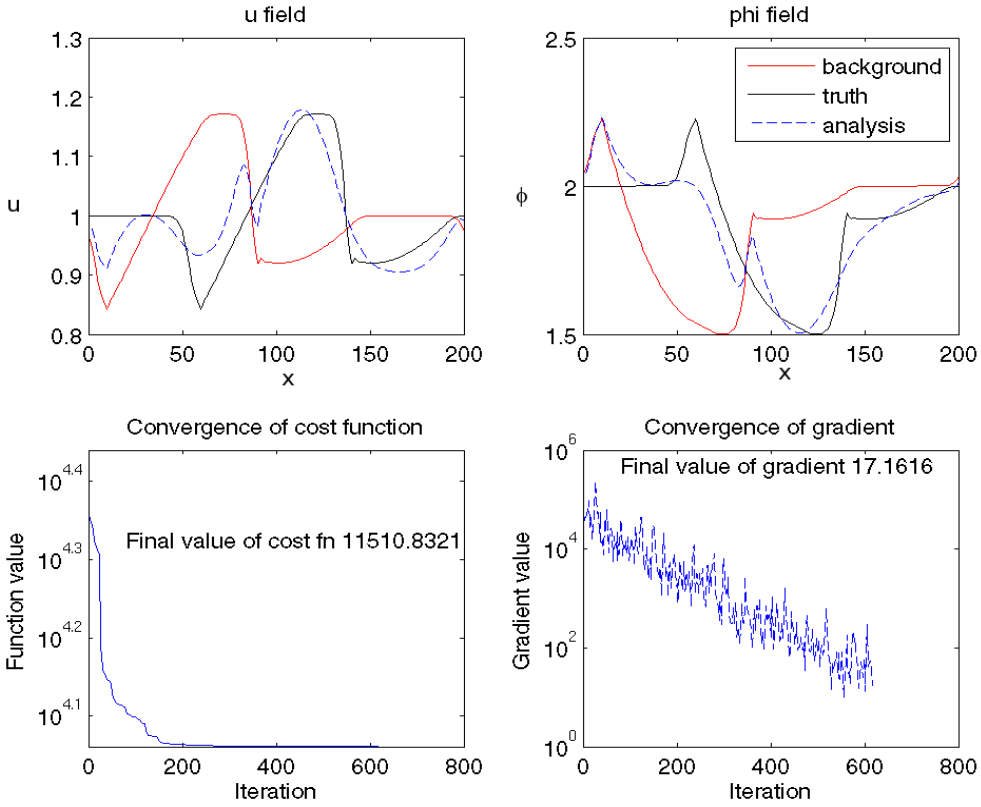
Trial 12.



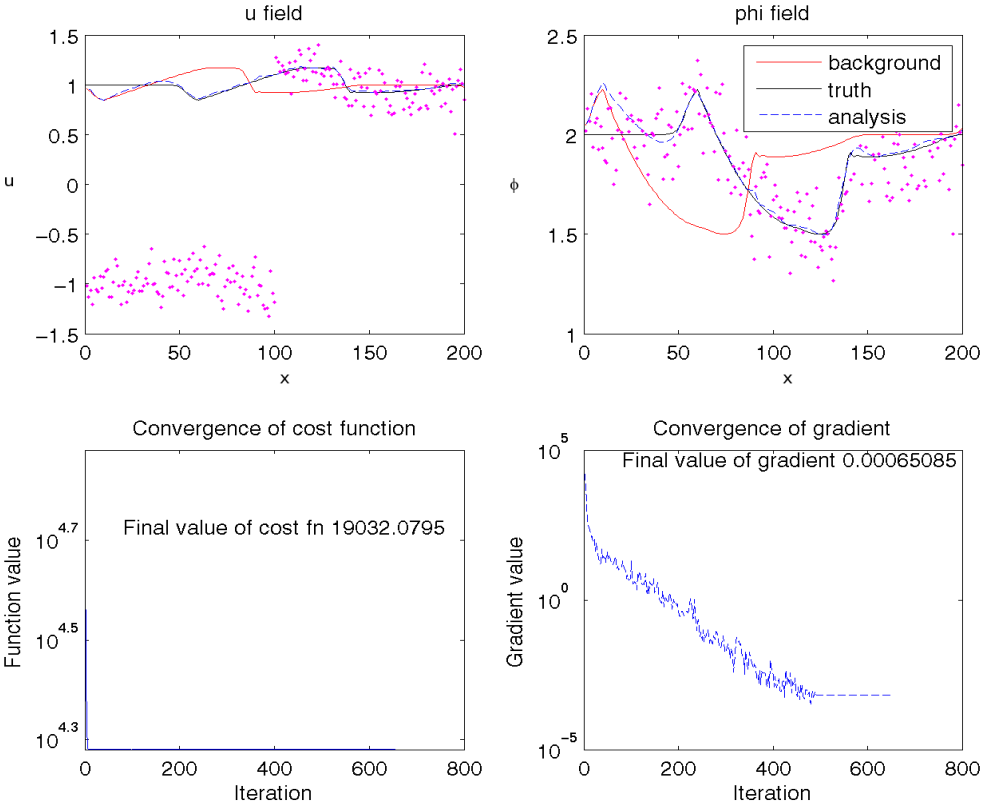
Trial 13.



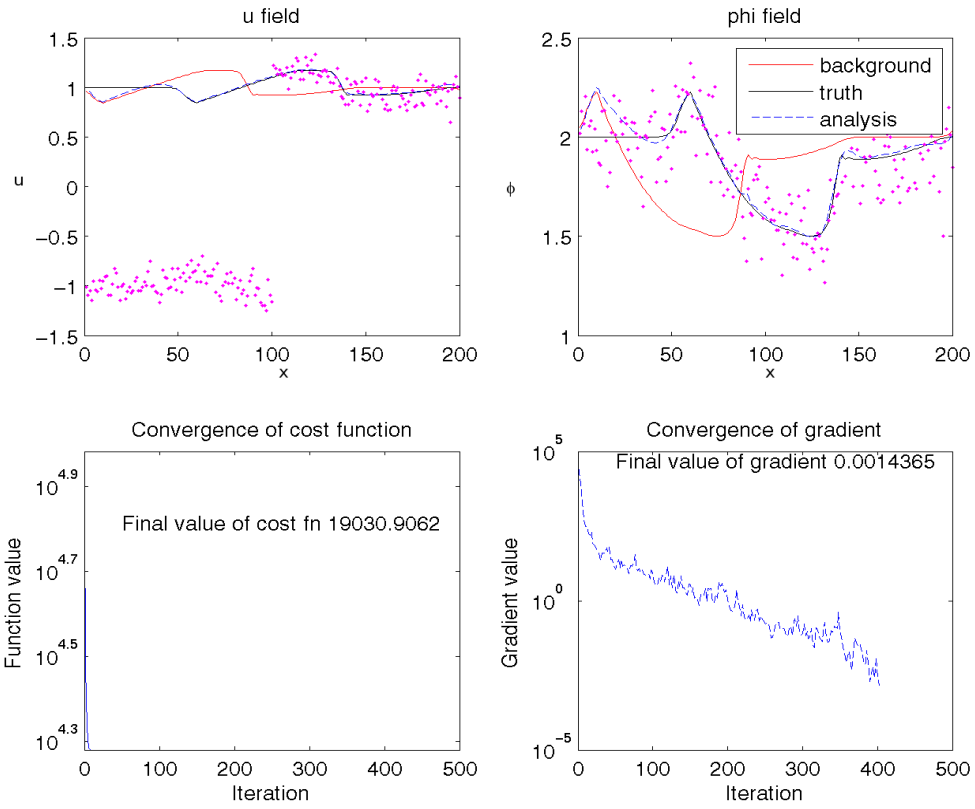
Trial 14.



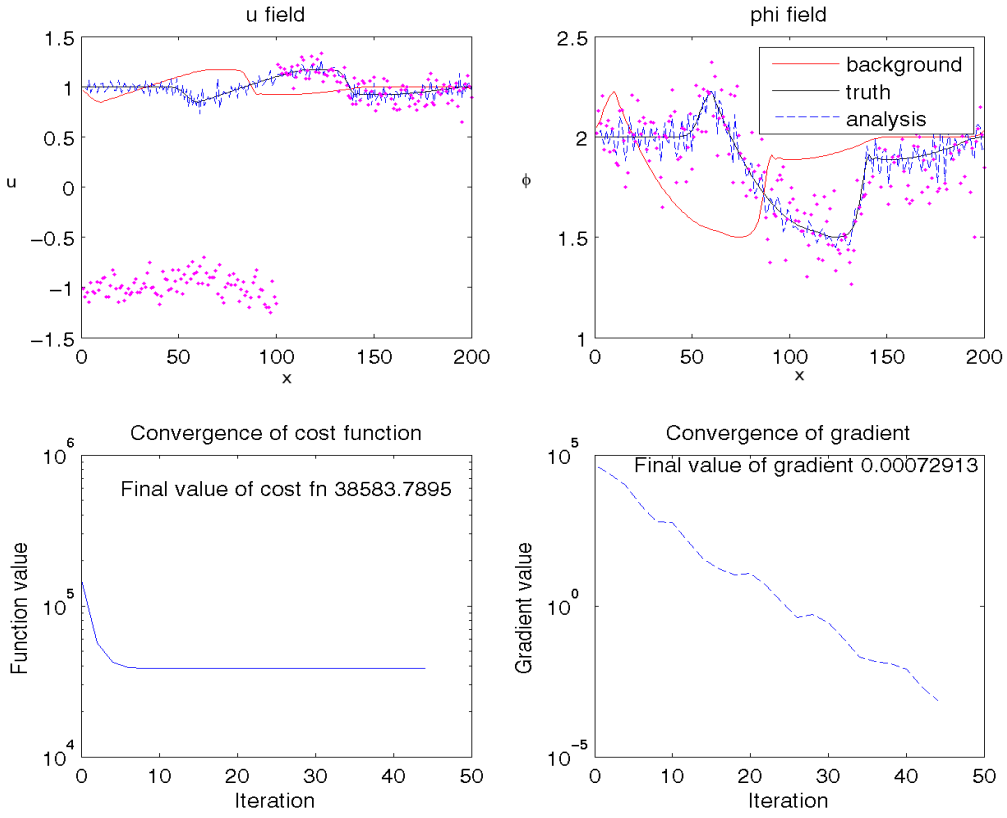
Trial 15.



Trial 16.



Trial 17.



Trial 19.