Resolution of sharp fronts in the presence of model error in variational data assimilation

QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY
00: 1–16 (2010)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/qj.000

Resolution of sharp fronts in the presence of model error in variational data assimilation

M. A. Freitag[a*], N. K. Nichols[b] and C. J. Budd[a]

[a]

[b]

Abstract: We show that the four-dimensional variational data assimilation method (4DVar) can be interpreted as a form of Tikhonov regularisation, a very familiar method for solving ill-posed inverse problems. It is known from image restoration problems that L_1-norm penalty regularisation recovers sharp edges in the image more accurately than Tikhonov, or L_2-norm, penalty regularisation. We apply this idea from stationary inverse problems to 4DVar, a dynamical inverse problem, and give examples for an L_1-norm penalty approach and a mixed Total Variation (TV) L_1-L_2-norm penalty approach. For problems with model error where sharp fronts are present and the background and observation error covariances are known, the mixed TV L_1-L_2-norm penalty performs better than either the L_1-norm method or the strong-constraint 4DVar (L_2-norm) method. A strength of the mixed TV L_1-L_2-norm regularisation is that in the case where a simplified form of the background error covariance matrix is used, it produces a much more accurate analysis than 4DVar. The method thus has the potential in numerical weather prediction to overcome operational problems with poorly tuned background error covariance matrices.

KEY WORDS Data assimilation; Tikhonov regularisation; L_1 regularisation; TV regularisation, model error; linear advection equation

1 Introduction

Data assimilation is a method for combining model forecast data with observational data in order to forecast more accurately the state of a system. One of the most popular data assimilation methods used in modern numerical weather prediction is four-dimensional variational data assimilation (4DVar) (Sasaki (1970); Talagrand (1981); Lewis . (2006)), which seeks initial conditions such that the forecast best fits both the observations and the background state (which is usually obtained from the previous forecast) within an interval called the assimilation window. Currently, in most operational weather centers, systems and states of dimension $\mathcal{O}(10\ )$ or higher are considered, whereas there are considerably fewer observations, usually $\mathcal{O}(10\ )$ (see Daley (1991); Nichols (2010)) for revie 197stio420.3e4k6ET10iI

Using these adjoint equations we avoid having to compute $M_{\ ,\ -1}(x_{\ -1})$ several times. We note that $\ $, $i = 0; : : : ; N$ are vectors whereas $\nabla\ $, $i = 0; : : : ; N$ are square matrices of the dimension of the system state.

The approximate Hessian $\nabla g \nabla \mathcal{J}(x_0)$ and $\nabla \mathcal{J}(x_0)$ are then used in (3), which is equivalent to a linearised least

and

$$v^+ = \max(v;0); \qquad v^- = \max(-v;0):$$

Problem (20) can then be written as

$$\min_{,+,-} \left(\left\| \begin{bmatrix} f \\ 0 \end{bmatrix} - \begin{bmatrix} G \\ I \end{bmatrix} z \right\|_2^2 + 1^T v^+ + 1^T v^- \right) : \tag{21}$$

subject to the constraints

$$D(C_B^{1/2} z + x_0) = v^+ - v^-; \tag{22}$$
$$v^+; v^- \geq 0: \tag{23}$$

Here 1 denotes the vector of all ones of appropriate size. This problem can then be written as

$$\min \left\{ \frac{1}{2} w^T H w + c^T w \right\} \tag{24}$$

subject to

$$Ew = g \quad \text{and} \quad Fw \geq 0; \tag{25}$$

where

$$w = \begin{bmatrix} z \\ v^+ \\ v^- \end{bmatrix}; \quad H = \begin{bmatrix} 2(G^T G + \ I) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix};$$

$$c = \begin{bmatrix} -2G^T f \\ 1 \\ 1 \end{bmatrix}; \quad E = \begin{bmatrix} DC_B^{1/2} & -I & I \end{bmatrix};$$

$$F = \begin{bmatrix} 0 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}; \quad g = - \ Dx_0;$$

and the block matrices I and 0 as well as the vectors 1 of all ones in the matrices H, E, F and c are of appropriate size. The objective function in (24) is convex as H is symmetric positive semi-definite. In order to solve the quadratic programming problem (24) with constraints (25) we use the MATLAB in-built function `quadprog.m`, which readily solves problems of the form (24),(25). For our problem we use an active-set quadratic programming strategy (also known as a projection method), which is described in Gill . (1981). For details on the implementation of the MATLAB quadratic programming tool we refer to the MATLAB Product Documentation Matlab (R2012a).

In the following sections we consider a square wave propagated by the linear advection equation as an example. We use a 'true' model (from which we take the observations) and another model, which is different from the truth and hence introduces a model error. For the different regularisation approaches we keep the regularisation parameter fixed, as we are only investigating the influence of the norm in the regularisation term, but not the size of the regularisation parameter . In all the examples we observe that the new edge-preserving mixed TV L_1-L_2-norm regularisation indeed gives better results than the standard L_2-norm approach and the simple L_1-norm regularisation.

5 Numerical experiments

We consider the linear advection equation

$$u \ + u \ = 0; \tag{26}$$

on the interval $x \in [0;1]$, with periodic boundary conditions. We discretise the equation using the upwind scheme

$$U^{\ +1} = U \ - \frac{\Delta t}{\Delta x} \left(U \ - U_{\ -1} \right); \tag{27}$$

where $j = 1; : : : ; N$, and the CFL condition $\Delta t < \Delta x$

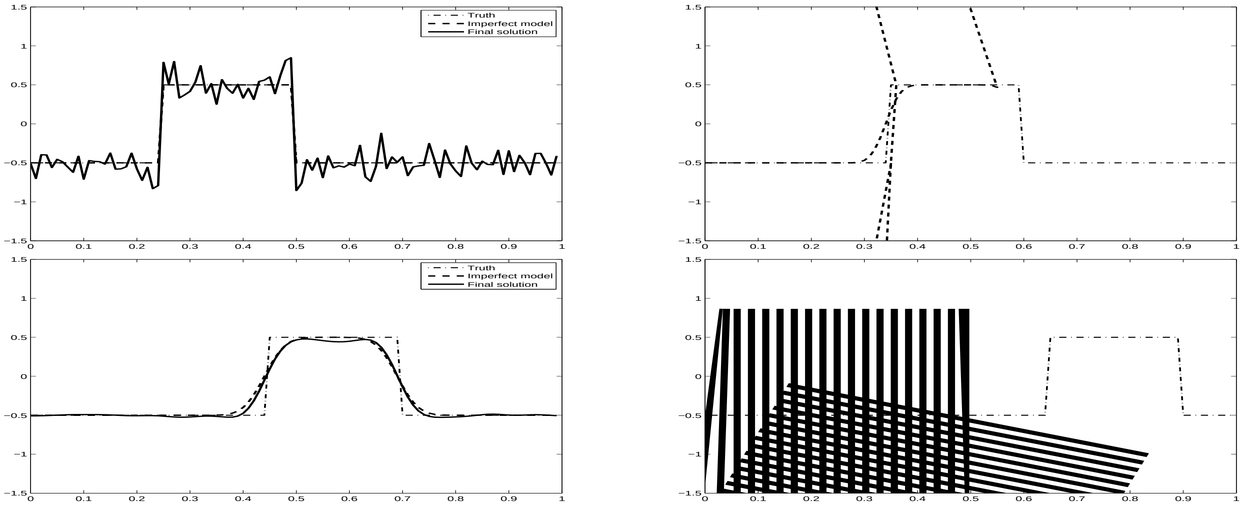


Figure 1. Results for **4DVar** applied to the linear advection equation where the initial condition is a square wave. We take **imperfect observations every** 20 **points in space and every** 2 **time steps**. 4DVar leads to bad oscillations in the initial condition and also to a phase error in the forecast.

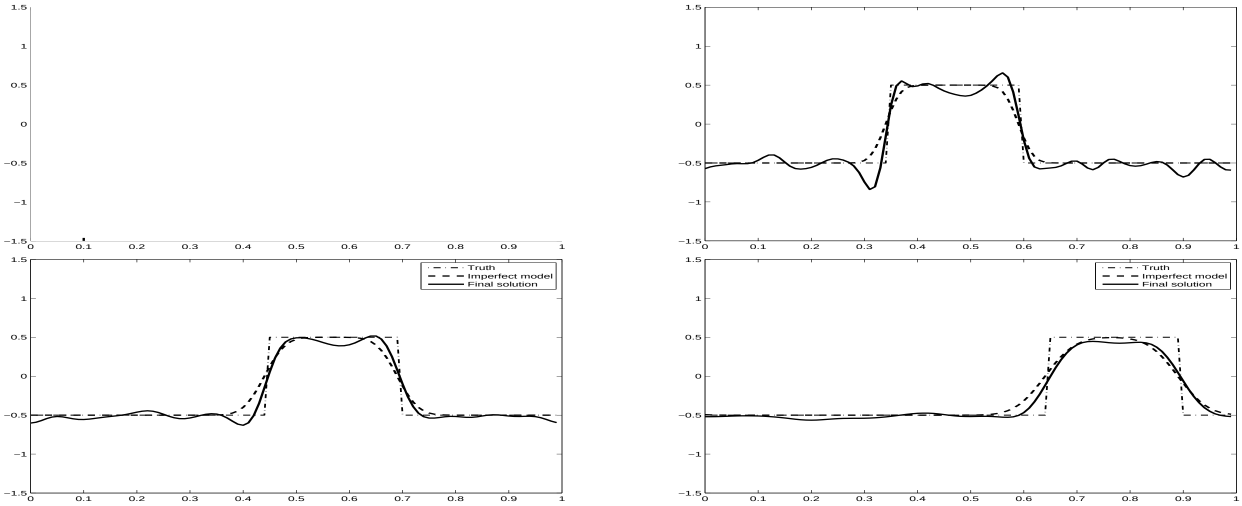
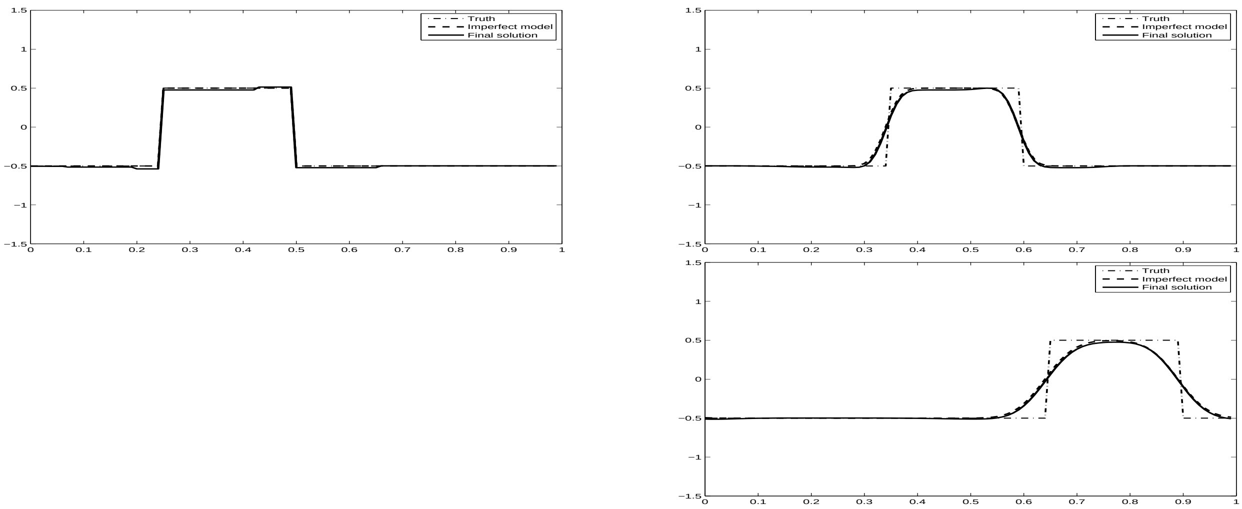


Figure 2. Results for L_1 **regularisation** for the same data as in Figure 1.



FigureL1

Figures 1 - 3 show the results for this example. We only present the case for imperfect noisy observations, as it is the most realistic one. In the next subsection we consider a non-diagonal background error covariance matrix - for which we present the results for cases (1)-(3). We also note that a summary of results is presented in Table I.

In the plots the true solution is represented by a thick dot-dashed line (called 'Truth' in the legend). This true solution is unknown in practice. We take (noisy) observations by perturbing the true trajectory using zero-mean Gaussian noise. The model solution (which is derived from the upwind method) is shown as a dashed line (called 'Imperfect model' in the legend). This solution represents the model solution, that is the solution that is obtained if we use the correct initial conditions and the (imperfect) model. It represents the best solution that we are able to achieve (if data assimilation gives us the perfect initial condition), as the model error is always present. The solution obtained from the assimilation process by incorporating the (perfect/partial/noisy) observations is given by the solid line (called 'Final solution' in the legend).

The result for 4DVar is shown in Figure 1 (minimisation problem (17)), that for L_1-regularisation in Figure 2 (minimisation problem (16)) and that for mixed TV L_1-L_2-norm regularisation in Figure 3 (minimisation problem (24)). The analysis obtained by 4DVar and L_1-regularisation is very inaccurate, with many oscillations and large over/undershoots near the discontinuities (first plots in Figures 1 and 2). When L_1-norm regularisation with the gradient (mixed TV L_1-L_2

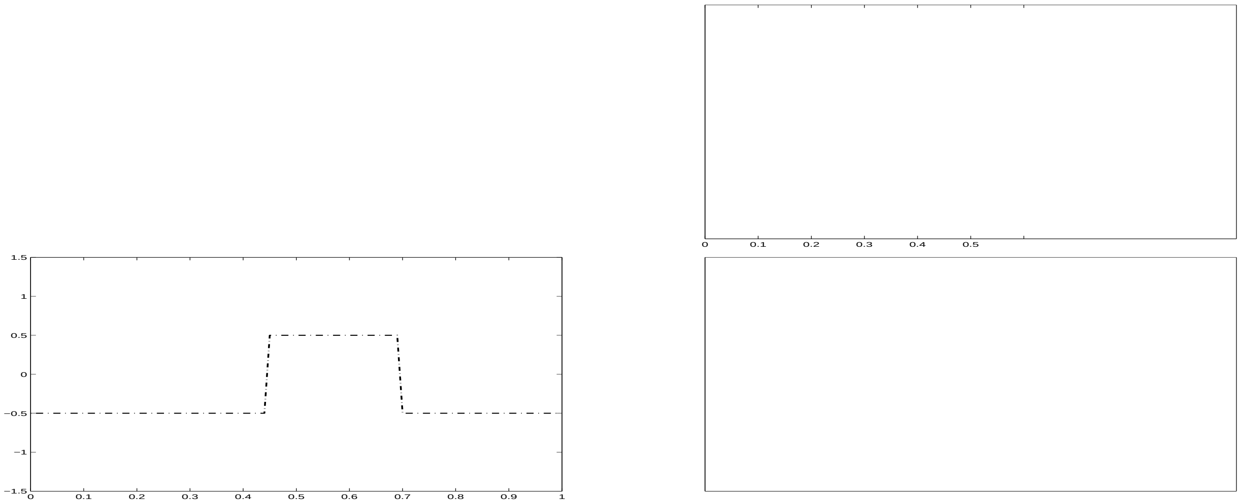


Figure 4. Results for **4DVar**. We take **perfect observations at each point in time and space** over the assimilation interval which is 40 time steps. The four plots show the initial conditions at $t = 0$ and the result after 20, 40 and 80 time steps. We choose B with $B_{ij} = 0.01\, e^{-\frac{|i-j|}{2L^2}}$, where $L = 5$.

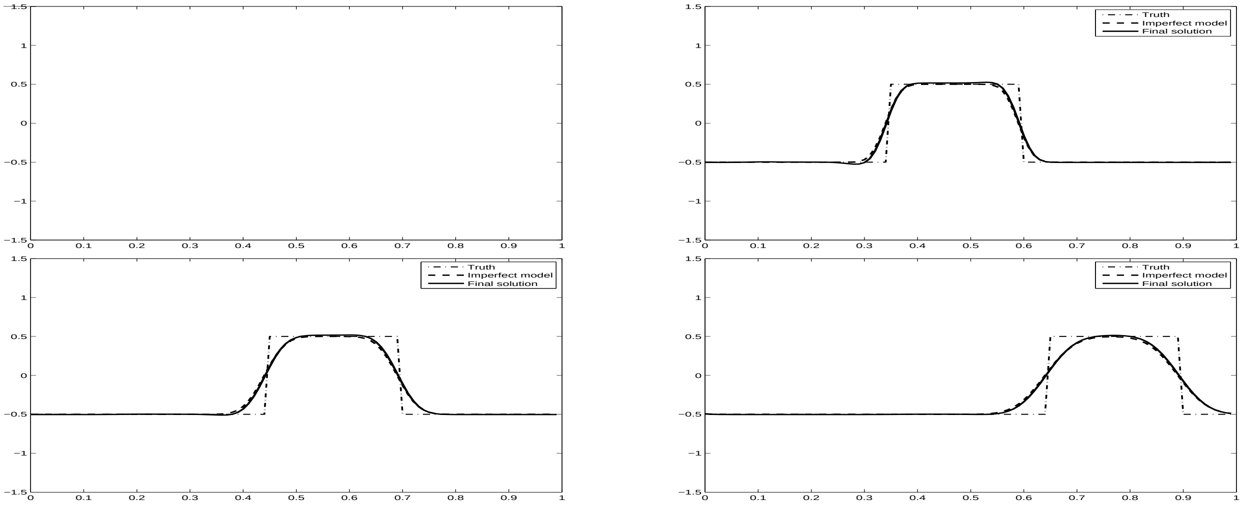


Figure 5. Results for **mixed TV L_1-L_2-norm regularisation** for the same data as in Figure 4.

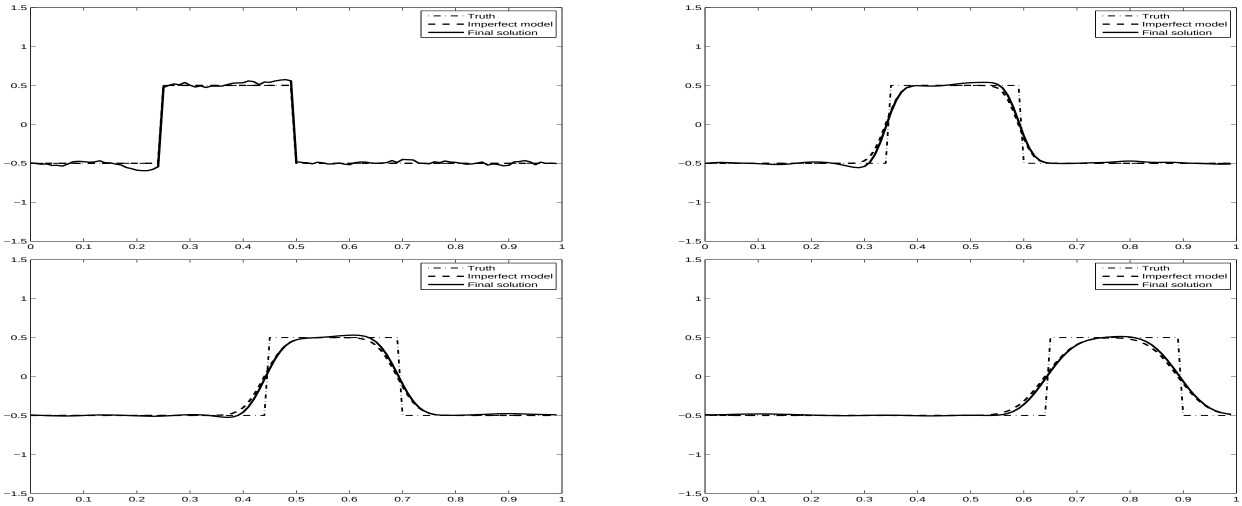


Figure 6. Results for **4DVar** for the same data as in Figure 4 but with **perfect observations every** 20 **points in space and every** 2 **time steps** for B with $B_{ij} = 0.01\, e^{-\frac{|i-j|}{2L^2}}$, where $L = 5$.

00: 1–16 (2010)
DOI: 10.1002/qj

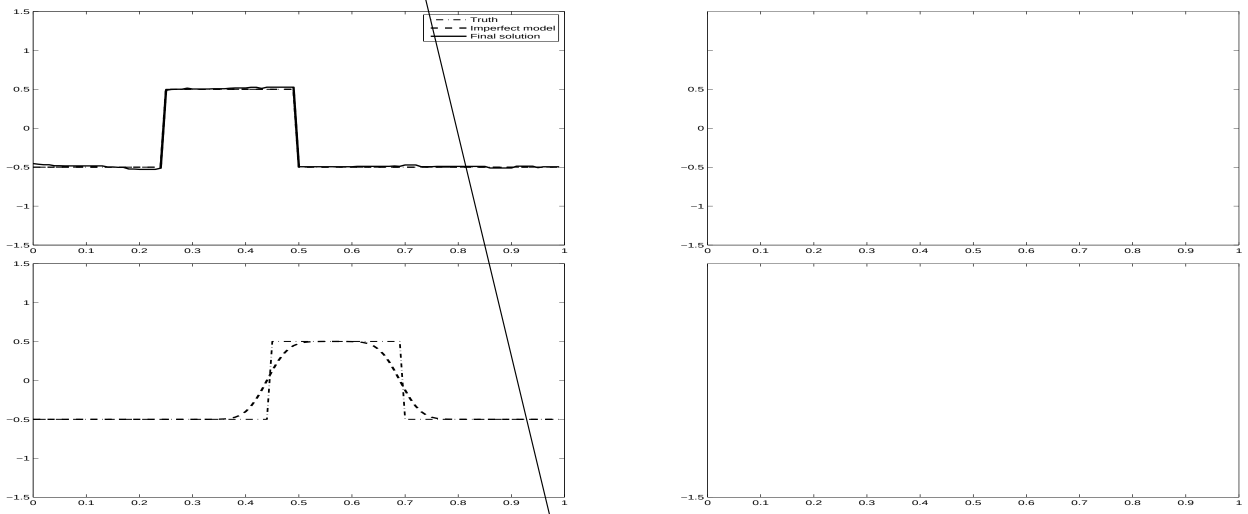


Figure 7. Results for **mixed TV L_1-L_2-norm regularisation** for the same data as in Figure 6.

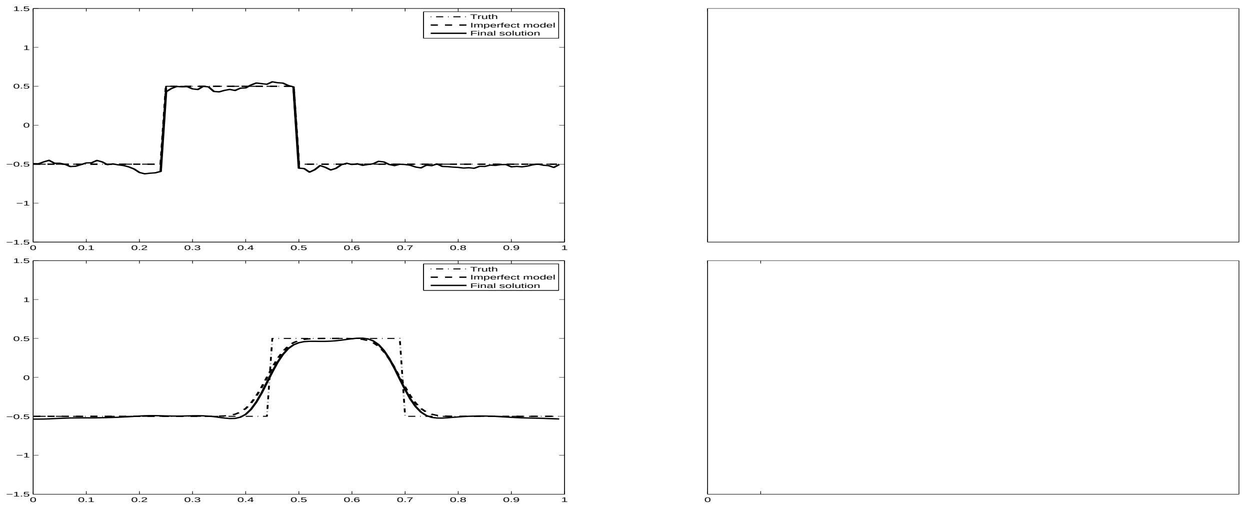


Figure 8. Results for **4DVar** for the same data as in Figure 1, but for B with $B_{ij} = 0.01\, e^{-\frac{|i-j|}{2L^2}}$, where $L = 5$.

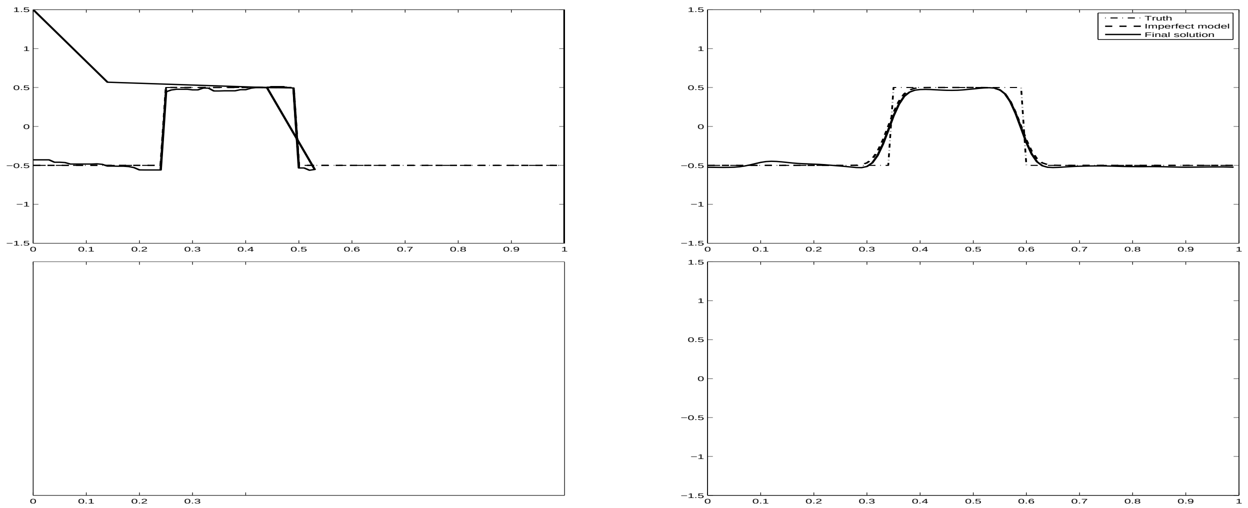


Figure 9. Results for **mixed TV L_1-L_2-norm regularisation** for the same data as in Figure 8, but for B with $B_{ij} = 0.01\, e^{-\frac{|i-j|}{2L^2}}$, where $L = 5$

DOI: 10.1002/qj

condition in standard 4DVar (first plot in Figure 4) and the forecast is slightly better than the forecast in 4DVar. For the case of partial perfect observations we obtain similar results. Mixed TV L_1-L_2-norm regularisation (Figure 7) gives better initial conditions than standard 4DVar (Figure 6).

Finally, Figures 8 and 9 show the results for partial noisy observations. Note that with this choice of B, the results for 4DVar (Figure 8) are better than the results for the diagonal matrix B (Figure 1) because information is spread via the covariance matrix B, and we see that the oscillations in the analysis are significantly reduced. It is notable, however, that the mixed TV L_1-L_2-norm regularisation (Figure 3) eliminates oscillations in the analysis even when the matrix B provides no smoothing. Moreover, where correlations are taken into account via the matrix B, then mixed L_1-L_2-norm regularisation (Figure 9) gives still better results than 4DVar (Figure 8). The quantities of the errors in the initial conditions for this particular case are summarised in the fifth row of Table I where we see that the errors using mixed TV L_1-L_2-norm regularisation are the smallest.

5.3 Changing the length of the assimilation window

Again, we take the same experimental data as in Subsection 5.1; this time, however, we reduce the size of the assimilation window from 40 time steps to 5 time steps and carry out the following test: we take imperfect observations every 5 points in space and every 2 time steps with Gaussian noise of mean zero and variance 0.01. For the background we again take the truth perturbed by Gaussian noise with covariance B taken from (29) with $^2 = 0.01$. Figures 10 and 11 show the results for a reduced size of the assimilation window. The first observation that we can make is that again the regularisation using the mixed TV L_1-L_2-norm (Figure 11) is consistently better than that using the L_2-norm (Figure 10). Standard 4DVar produces oscillations, in particular in the initial conditions, whereas the mixed TV L_1-L_2-norm regularisation does not show any oscillations. The oscillations in the initial conditions in standard 4DVar then lead to errors in the forecast (see plots for $t = 5$, $t = 20$ and $t = 45$ in Figure 10). Again, for 4DVar, the forecast of the analysis does not keep the amplitude correctly (final plot in Figure 10), whereas the mixed TV L_1-L_2-norm regularisation provides a more accurate amplitude in the forecast (final plot in Figure 11).

5.4 Summary of initial condition errors

In Table I we summarise the analysis errors (the errors between the analysis and the truth at $t = 0$, that is, the initial condition errors) measured in the L_2 vector norm for the different regularisation techniques. The results are shown for all three test cases described in Section 5.1 where either perfect observations are taken at all spatial and time points, partial perfect observations are taken less frequently in time and space, or partial imperfect (noisy) observations are taken, also with less frequency. We choose observation errors with covariance $R = 0.01I$ and assimilation windows of length 40 and length 5. We consider the two background covariance matrices $B = {}^2 I$, and the double-sided exponential covariance matrix B given by (29), with three different variances: $^2 = 1$, $^2 = 0.01$ and $^2 = 0.005$.

For the mixed TV L_1-L_2-norm regularisation method, we also give results for different values of in (19). The emphasis on the sparsity of the gradient of the initial condition depends on this regularisation parameter. We have looked at three different values for and the best of all three results (that is the smallest error in the initial condition) is underlined in the table. The regularisation depends on the regularisation parameter, but investigating the influence of this parameter and finding the optimal choice of is beyond the scope of this paper. We remark that for the plots in the previous subsections we have used the value of from the table that gives the smallest initial condition error.

We see from the entries in the table that the errors in the analysis at time $t = 0$ are consistently smaller for mixed TV L_1-L_2-norm regularisation than for standard 4DVar or L_1-norm regularisation. Mixed TV L_1-L_2-norm regularisation gives an error of about one magnitude smaller than standard 4DVar. We also observe from the table that, for standard 4DVar, L_1-norm regularisation and mixed TV L_1-L_2-norm regularisation, the errors in the initial condition (analysis) decrease as the variance in the background error is reduced, that is, as the ratio of the background to observation variance decreases. This is consistent with the results of Haben . (2010), which show that the standard 4DVar assimilation problem becomes more well-conditioned (well-posed) as this ratio decreases. These examples demonstrate that, even where the noise in the background and observations is Gaussian with known covariances, the standard 4DVar approach does not produce as accurate an analysis as mixed TV L_1-L_2-norm regularisation in the presence of sharp fronts and model error.

6 Further experiments

We now investigate how the 4DVar and mixed TV $L_1 - L_2$-norm regularisation methods perform in cases where the position of the shock in the background is displaced from the truth and where the frontal gradient of an advected wave in the background is incorrect. As discussed in the introduction, it is recognized that if a shock in the background field is displaced, then the 4DVar method may not give a good analysis. Similarly, the assimilation method may be unable to capture a sharp shock if

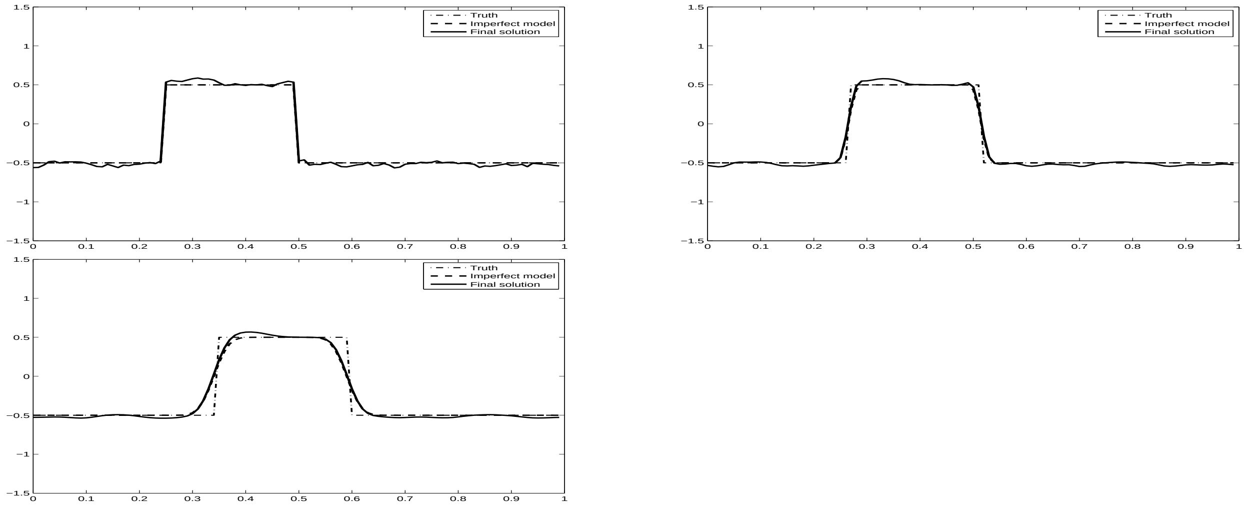


Figure 10. Results for **4DVar** applied to the linear advection equation where the initial condition is a square wave. We take **imperfect observations every** 5 **points in space and every** 2 **time steps** over the assimilation interval which is 5 time steps. The four plots show the initial conditions at $t = 0$ and the result after 5, 20 and 45 time steps. 4DVar leads to oscillations in the initial condition and a misplaced discontinuity in the forecast.

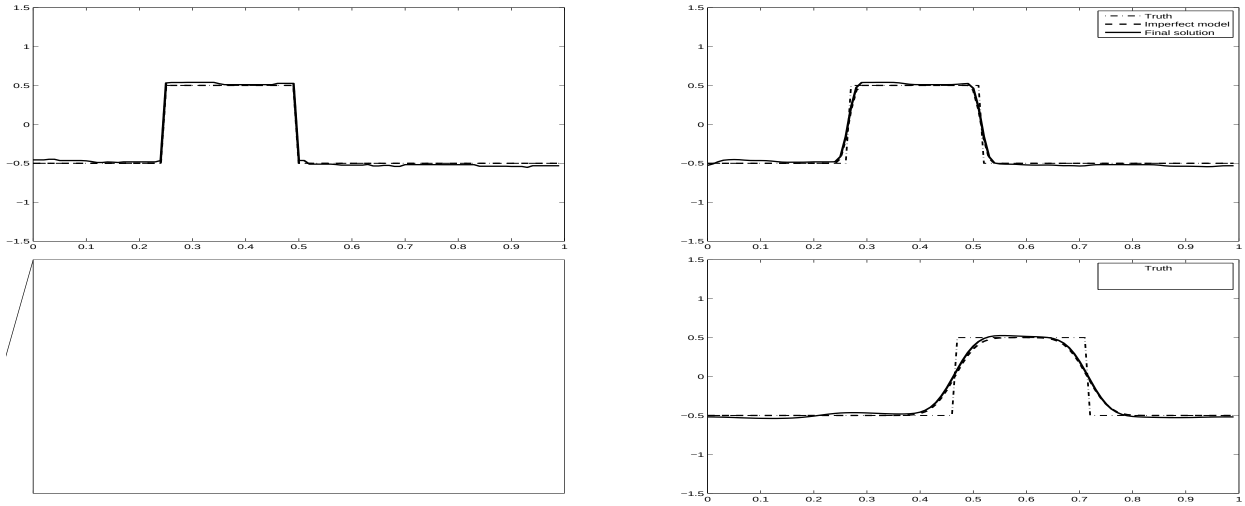


Figure 11. Results for **mixed TV** L_1-L_2**-norm regularisation** for the same data as in Figure 10. Mixed TV L_1-L_2-norm regularisation gives the best possible result for the initial condition.

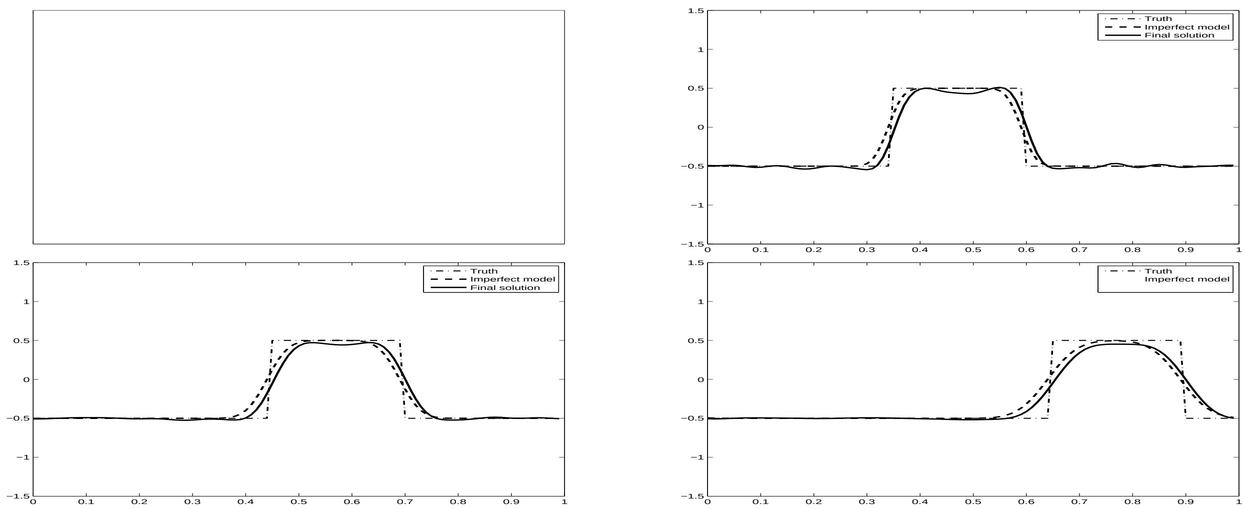


Figure 12. Results for **4DVar** for a shifted (and noisy) background and for background error covariance matrix $B = 0.01I$

00: 1–16 (2010)
DOI: 10.1002/qj

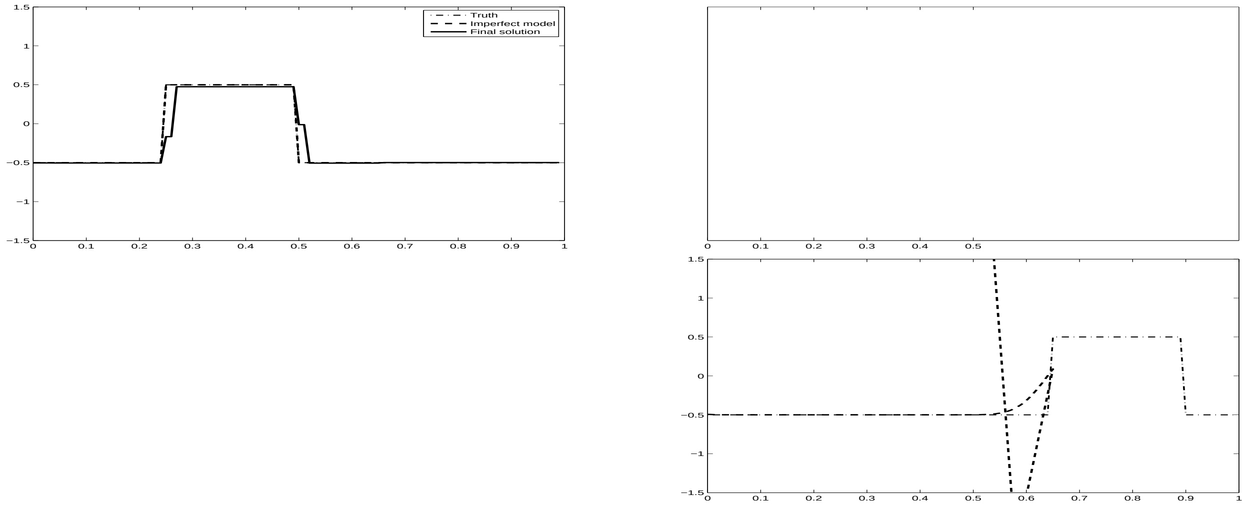


Figure 13. Results for **mixed TV L_1-L_2-norm regularisation** for the same data as in Figure 12.

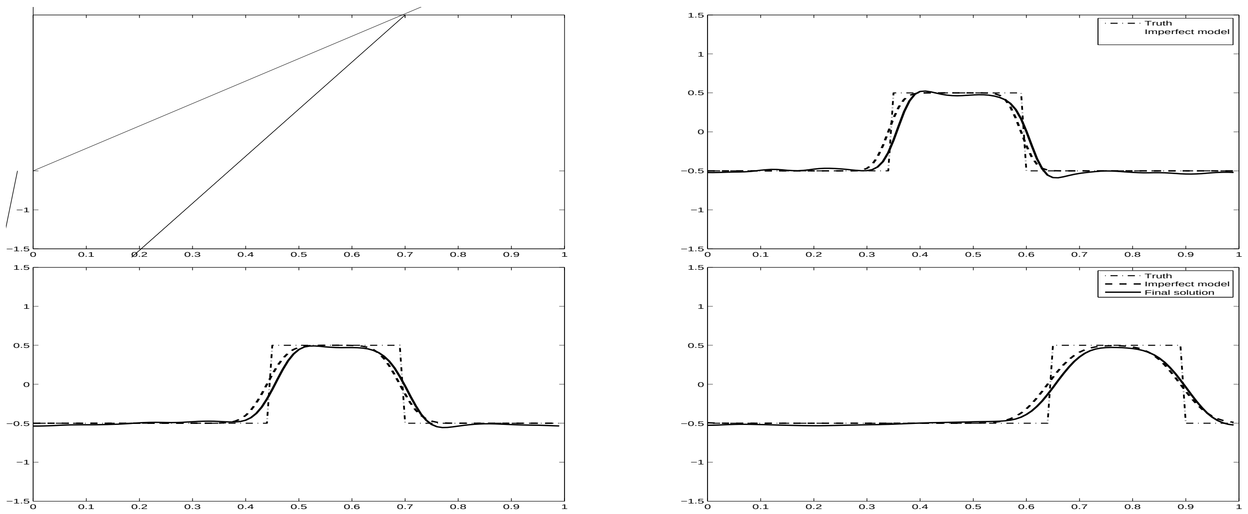


Figure 14. Results for **4DVar** for a shifted (and noisy) background and for background error covariance matrix B taken from (29) with $\sigma_b^2 = 0.01$

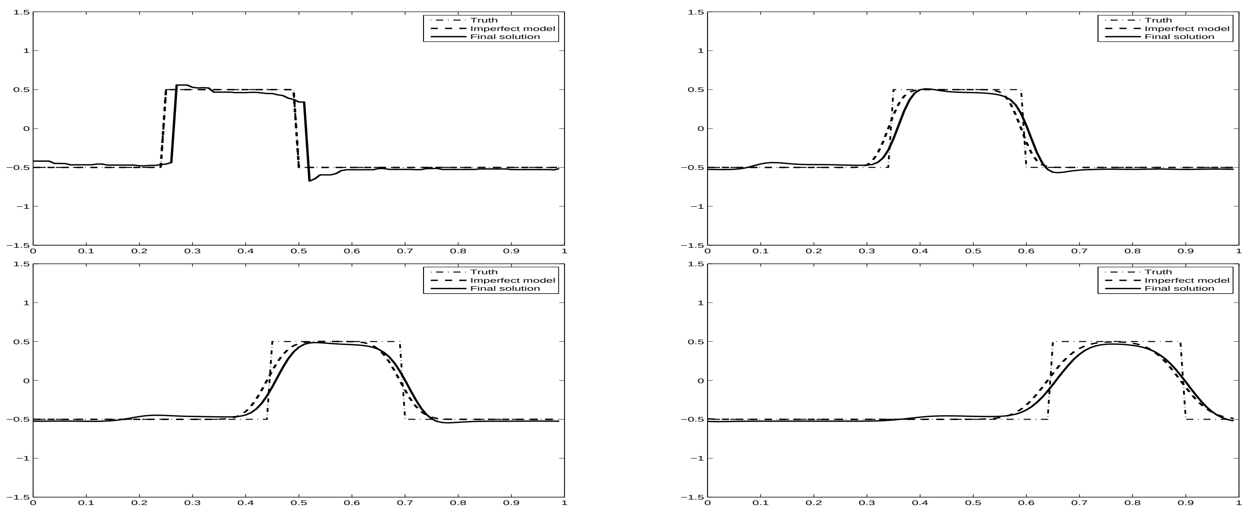


Figure 15. Results for **mixed TV L_1-L_2-norm regularisation** for the same data as in Figure 14.

and 15. The initial condition in 4DVar is clearly recovered poorly, with many oscillations (see first plot in Figure 14). Furthermore, at the end of the assimilation window the solution gives undershoots (see third plot in Figure 14) and the amplitude of the front is reduced (see second and third plot in Figure 14). The solution at the initial time provided by the mixed TV L_1-L_2-norm regularisation has less oscillation present in the shock wave (see first plot in Figure 15) and produces somewhat less distortion of the wave front over the window (see second and third plot in Figure 15). The errors in the analysis in this case for the standard 4DVar and the mixed TV L_1-L_2-norm regularisation (with $= 10$) are similar, with a value of 1.80. Both methods smear the shock front and both produce an initial phase error which is reproduced in the forecast.

The plots in Figures 14 and 15 show that choosing an exponential (non-diagonal) covariance matrix B is not necessarily advantageous when there is a sharp front with a phase error. In this case both 4DVar and mixed TV L_1-L_2-norm regularisation with a diagonal covariance matrix B capture the shock front more accurately, but the mixed TV L_1-L_2-norm technique also eliminates the oscillations in the analysis arising from the effects of the model error (see Figures 12 and 13).

In general, the mixed norm approach removes oscillations and sharpens fronts - but the position of the shock is not recovered precisely where there is a phase error in the background.

6.2 A slanted front for the background

Finally, with the same experimental data as in Subsection 5.1 we consider a slanted background given by the slanted square wave

$$u_1(x;0) = \begin{cases} -0.5 + \frac{50}{7}(x - 0.18); & 0.18 < x < 0.32 \\ 0.5; & 0.32 \le x \le 0.43 \\ 0.5 - \frac{50}{7}(x - 0.43); & 0.43 < x < 0.57 \\ -0.5 \quad 0 \le 18 x \ll x < 0.32 & \end{cases}$$

Acknowledgement

The authors thank Nathan Smith (University of Bath) for