# The Effect of Numerical Model Error on Data Assimilation

S.E.Jenkins[a,b,*], C.J.Budd[b], M.A.Freitag[b], N.D.Smith[a]

[a]*Department of Electronic and Electrical Engineering, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom.*
[b]*Department of Mathematical Sciences, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom.*

**Abstract**

Strong constraint 4D-Variational data assimilation (4D-Var) is a method used to create an initialisation for a numerical model, that best replicates subsequent observations of the system it aims to recreate. The method does not take into account the presence of errors in the model, using the model equations as a strong constraint. This paper gives a rigorous and quantitative analysis of the errors introduced into the initialisation through the use of finite difference schemes to numerically solve the model equations. The 1D linear advection equation together with circulant boundary conditions, are chosen as the model equations of interest as they are representative of the advective processes relevant to numerical weather prediction, where 4D-Var is widely used. We consider the deterministic error introduced by finite difference approximations in the form of numerical dissipation and numerical dispersion and identify the relationship between these properties and the error in the 4D-Var initialisation. In particular, we find that a solely numerically dispersive scheme has the potential to introduce destructive interference resulting in the loss of some wavenumber components in the initialisation. Bounds for the error in the initialisation due to finite difference approximations are determined with and without observation errors. The bounds are found to depend on the smoothness of the true initial condition we wish to recover and the numerically dissipative and dispersive properties of the scheme. Numerical results are presented to demonstrate the effectiveness of the bounds. These lead to the conclusion that there exists a critical number of discretisation points when considering full sets of observations, where the effects of both the considered numerical model error and observational errors on the initialisation are minimised. The numerically dissipative and dispersive properties of the finite difference schemes also have the potential to alter the properties of the noise found in observations. Correlated noise structures may be introduced into the 4D-Var initialisation as a result. We determine when this occurs for observational errors in the form of additive white noise and find that the effect is reduced through the use of numerically non-dissipative finite difference schemes.

*Keywords:* data assimilation, numerical model error, observation error, deterministic error

## 1. Introduction

### *1.1. Summary of problem and results*

This paper presents a rigorous and quantitative study of the influence of finite difference approximations on the accuracy of the initialisation produced by *strong constraint Four-Dimensional Variational data assimilation (4D-Var)*. Given a forward model for the considered model equations, 4D-Var compares the forecast from this model using an a priori initialisation, with data obtained from observing the physical system, to create an improved initialisation. This leads to an improved forecast for the system. This is accomplished through the minimisation of a cost functional with respect to the initialisation for the forward model, creating an optimal initialisation. The method is described, for example in [1, 2, 3, 4]. 4D-Var is of particular interest due to its applications in numerical weather prediction (NWP). In this instance, the model equations are typically a system of advection dominated PDEs.

*Corresponding author.
*Email addresses:* S.E.Jenkins@bath.ac.uk (S.E.Jenkins), C.J.Budd@bath.ac.uk (C.J.Budd), M.A.Freitag@bath.ac.uk (M.A.Freitag), N.Smith@bath.ac.uk (N.D.Smith)

The accuracy of the optimal initialisation (also known as the analysis vector) and its subsequent forecast are affected by many different sources of error [5]. Examples are observation errors due to systematic errors in instrumentation [4] and model errors in the forward model [3]. Model error in a deterministic forward model can be viewed in two forms; inaccurate model equations and numerical model error. The former is introduced by a failure of the model equations to capture a property of the physical system, whilst the latter is due to errors introduced when numerically solving the model equations in the forward model. Solving the model equations numerically utilising finite difference approximations, is one such source of numerical model error. These errors then enter into the 4D-Var problem, affecting the resulting initialisation and subsequent forecast.

Here we consider the 1D linear advection equation together with circulant boundary conditions, as our prototype problem. This system is representative of the advective processes relevant to NWP and can be solved using various well known finite difference schemes [6, 7]. The study of linear problems in the context of data assimilation is relevant to both linear and non-linear data assimilation problems. The adjoint method and the tangent linear model assumption in incremental 4D-Var, make use of local linearisations of non-linear problems to identify the optimal initialisation [1], making the analysis of a linear problem important. Pfeffer et al. [8] analysed the sensitivity of the non-linear NASA-GLAS forecast model to the time-differencing scheme used to solve the model equations. It was found that some aspects of the results exhibited the effects of properties indicated by their linear analysis of the considered scheme. Hence the results of a linear problem may also be relevant to the results from a non-linear problem. The analysis of the behaviour of strong constraint 4D-Var for the problem given in this paper, is quite involved despite the apparent simplicity of the equation itself. Our aim is to use the insights into the nature of the errors in 4D-Var (particularly the effects of the smoothness of the initial condition and the nature of the errors from the numerical method) given by the analysis of this equation, as a stepping stone towards understanding the effects of numerical model errors on the accuracy of the initialisation from 4D-Var for more complex advective processes.

The numerical model error introduced by the finite difference schemes used to solve the 1D linear advection equation can be completely described in terms of numerical dissipation and numerical dispersion, including aliasing errors [7]. However, it is not sufficient to study the impact of these errors alone as in practice, many different forms of error will interact to affect the accuracy of the initialisation produced through strong constraint 4D-Var. To this end, we initially analyse the effects of this form of numerical model error without any other form of error and then together with observational errors. The forecast from 4D-Var experiments have been shown to be most sensitive to observational errors [9] so it is key to understand their combined impact on the initialisation.

In this paper, Section 2 states the assumptions placed upon strong constraint 4D-Var throughout this paper and introduces the 1D linear advection equation as the forward problem. Three finite difference schemes are chosen as forward models; the Upwind, Preissman Box and Lax-Wendroff schemes. The effects of numerical dissipation and dispersion on the initialisation are then reviewed through these representative schemes. Section 3 develops a formulation for the initialisation which allows the effects of numerical dissipation and dispersion on the true initial condition to be analysed using spectral methods. Section 4 estimates bounds for the $l_2$-norm of the error in the initialisation, in terms of the numerically dissipative and dispersive properties of the scheme, along with the smoothness of the true initial condition. These are found in Lemmas 3 and 4 and form the main results of the paper. We find that in the absence of observation errors the rate of decay of the error in the initialisation, with respect to the number of spatial mesh points $N$ when considering full sets of observations, increases with the smoothness of the true initial condition. In the presence of observation errors, the same result holds until a critical value of $N$ is reached when considering full sets of observations. At this point, the error begins to increase due to observation errors. Performing strong constraint 4D-Var at this value of $N$ when considering full sets of observations, minimises the error in the initialisation due to the numerical model error and observation errors. Section 5 presents a discussion of the relevance of the results presented in this paper to non-linear systems and possible future work to extend this to more meteorologically relevant problems.

### *1.2. Background*

Data assimilation is a vibrant and active area of research. We will present the results presented in this paper in the context of research already conducted in the area. The derivation of strong constraint 4D-Var data assimilation makes the assumption that the forward model used to solve the model equations is perfect [3]. In order to account for the effects of model error on strong constraint 4D-Var, a modified formulation was proposed by Sasaki [10] that did not rely as heavily on satisfying the constraints of the forward model. This was termed weak constraint 4D-Var data assimilation leading to the original formulation being termed strong constraint 4D-Var. Strong constraint 4D-Var uses

the model as a strong constraint for the minimisation process, not altering the form of the model to account for model errors. This method is described in Section 2. Le Dimet et al. [9] performed sensitivity analyses to identify the impact of different forms of error associated with strong constraint 4D-Var, on the accuracy of the forecast produced by the analysis vector. They identified that the error in this forecast is most sensitive to observation errors and advocate the use of regularisation to ensure that the prediction error remains stable. The strong constraint 4D-Var cost function can be interpreted as a Tikhonov regularisation problem [11]. Budd et al. showed using this formulation and a mixed total variation $L_1 - L_2$-norm regularisation in the presence of model error, that sharp fronts can be recovered more accurately [12]. Zou et al. [13] also make use of penalty functions to reduce the effects of model error in the shallow water equations. They also investigate the effects of incomplete observations on the minimisation process of the cost function and how penalty functions can be used to improve results.

Weak constraint 4D-Var uses the model as a weak constraint for the minimisation process where an alternative formulation for the numerical model is chosen to account for model errors [14, 15, 16]. The model equations are augmented with a model error term, usually assumed to be a Gaussian random variable. However the Gaussian assumption is unlikely to be valid in practical situations [14]. The cost function gains an additional term for the square error of the model error, weighted by the model error covariance matrix, similar to the minimisation terms for the background and observation errors. The cost functional is then minimised with respect to the initialisation and the model errors. However the computational cost of applying this to large systems such as those used in NWP, makes this method impractical [14]. The work by Trémolet [17] investigates the formulation of the model error covariance matrix and some of the limitations of weak constraint 4D-Var.

Griffith et al. [14] propose treating model error in the form of a time evolving function together with the addition of a stochastic term so as to account for both the deterministic and stochastic properties of model error. Vidard et al. [16] also propose a similar form focusing on the deterministic errors. When only deterministic errors are considered, this approach simplifies the control of the errors so that only the initial error needs to be considered in the minimisation problem, reducing the number of variables to be controlled [14]. Results in [14, 16, 18] have shown that treating deterministic errors in this way can be effective. However, there is still the problem of gathering prior knowledge of the model error to constrain the initial error in the minimisation problem [18]. Akella et al. [15] investigate the impact of choosing growing, constant and decaying linear forms for the deterministic errors.

Numerical model errors in advection problems can lead to physically unrealistic anomalies in the solution that can have wide reaching consequences on the model results over time [19]. Vukićević et al. [20] examined the numerical values of and errors in the analysis vector achieved through strong constraint 4D-Var data assimilation experiments, using three different schemes to solve the 2D linear advection equation. The observations used were assumed to be perfect and three different initial errors were considered for the background estimate in the minimisation of the cost function. The numerical results obtained exhibited behaviours due to the effects of numerical dissipation and dispersion in the advection schemes. The accuracy of the results was also found to be positively correlated to the accuracy of the forward and adjoint models. Both Gerdes et al. [19] and Vukićević et al. [20] discuss the impact of numerically dissipative and dispersive effects from their advection schemes on the results of their 4D-Var experiments. Some of these can be desirable whilst others can lead to physically unrealistic results. Hence it is important to understand analytically the impact of numerical dissipation and dispersion on the results of 4D-Var.

The work presented in this paper extends the results of Vukićević et al. [20] by making rigorous quantitative error estimates of the initialisation, using spectral methods to investigate the effects of numerical dissipation and dispersion in finite difference schemes on the results of strong constraint 4D-Var. The aim is both to provide some prior knowledge on the deterministic error in the considered problem and to inform the choice of model error term in weak constraint 4D-Var.

## 2. Problem formulation

The method of 4D-Var is a procedure used to solve a particular inverse problem; given a set of (typically sparse) observations ($\{\mathbf{y}_l\}_{l=0}^{L}$), of a physical system taken over a period of time, a numerical model of the system ($\mathcal{M}_{l+1,l}(\cdot)$) and a priori information on the initial condition for the system ($\mathbf{x}_b$), the method estimates an initial condition for the numerical model that best replicates the true state of the system. Here we consider the effects of numerical model error on strong constraint 4D-Var. To find the best estimate for the initial condition $\mathbf{x}_a \in \mathbb{R}^N$, the following cost function is

minimised with respect to the initial condition $\mathbf{x}_0 \in \mathbb{R}^N$ of the numerical model:

$$J(\mathbf{x}_0) = (\mathbf{x}_b - \mathbf{x}_0)^T B^{-1}(\mathbf{x}_b - \mathbf{x}_0) + \sum_{l=0}^{L} \left[\mathbf{y}_l - \mathcal{H}_l(\mathbf{x}_l)\right]^T R_l^{-1} \left[\mathbf{y}_l - \mathcal{H}_l(\mathbf{x}_l)\right], \tag{1}$$

$$\mathbf{x}_{l+1} = \mathcal{M}_{l+1,l}(\mathbf{x}_l). \tag{2}$$

The minimiser $\mathbf{x}_a$, the optimal initialisation for the numerical model, will be termed the *analysis vector* (following the convention of NWP literature) and satisfies $\nabla J(\mathbf{x}_a) = 0$. The period of time the observations are taken over is known as the *assimilation window*. Here the cost function uses $L + 1$ sets of observations and subscript $l$ denotes the state of a variable at the time of the $l$th set of observations. A set of observations contains observations of all observed points in space at the considered point in time.

The variable $\mathbf{x}_b \in \mathbb{R}^N$ is the estimated initial condition termed the *background estimate*; $\mathcal{M}_{l+1,l} : \mathbb{R}^N \to \mathbb{R}^N$ is the forward model taking the $l$th state of the numerical model to its $(l + 1)$th state; $\mathbf{x}_l \in \mathbb{R}^N$ is the $l$th state of the numerical model; $\tilde{\mathbf{y}}_l \in \mathbb{R}^{m_l}$ is the $l$th set of perfect observations of the physical system (that is, no observation errors); $\boldsymbol{\epsilon}_l \in \mathbb{R}^{m_l}$ is the observation error in the $l$th set of observations; $\mathbf{y}_l \in \mathbb{R}^{m_l}$ is the $l$th set of observations of the true physical system such that $\mathbf{y}_l = \tilde{\mathbf{y}}_l + \boldsymbol{\epsilon}_l$; $\mathcal{H}_l : \mathbb{R}^N \to \mathbb{R}^{m_l}$ is the $l$th observation operator; $B \in \mathbb{R}^{N\times N}$ is the symmetric positive definite (SPD) background error covariance matrix; and $R_l \in \mathbb{R}^{m_l \times m_l}$ is the SPD $l$th observation error covariance matrix. Here $N, m_l \in \mathbb{N}$ for all $l$. More details on these variables and 4D-Var can be found in [2, 3, 4, 21, 22].

We choose the 1D linear advection equation as the physical system, which can be solved using several numerical schemes. This problem is representative of (and a prototype for) the advective processes in more complex systems of interest in NWP. Despite looking deceptively simple, this system provides a challenge numerically and has historically been an essential test equation for the development and analysis of many numerical methods, for example see [23]. Such linear problems are also directly relevant for adjoint methods and to tangent linear models used in incremental 4D-Var [22]. The schemes introduce numerical model error through the approximation of derivatives [7]. In order to fully investigate the effects of this deterministic numerical model error, all other errors present in the problem will be initially removed. Therefore, the background term of the cost function is neglected as in Griffith et al. [14] and Vukićević et al. [20] in order to allow the full impact of deterministic numerical model error to be seen. We take a set of observations at every time step of the numerical model. Hence $m_l = N$ and $\mathcal{H}_l = I_N$ for all $l$. Each set contains observations at every spatial grid point of the numerical model, which results in the set of observations taken at time $t = 0$ acting to regularise the problem so it remains well-posed. This is demonstrated in Section 3. Also, $\boldsymbol{\epsilon}_l$ is assumed to be an iid Gaussian random variable, $\mathcal{N}(\mathbf{0}, \sigma_o^2 I_N)$, $\sigma_o \in \mathbb{R}^+$, leading to $R_l = \sigma_o^2 I_N$ for all $l$. These assumptions result in the following cost function,

$$J(\mathbf{x}_0) = \frac{1}{\sigma_o^2} \sum_{l=0}^{L} \left[\mathbf{y}_l - \mathcal{M}_{l,0}(\mathbf{x}_0)\right]^T \left[\mathbf{y}_l - \mathcal{M}_{l,0}(\mathbf{x}_0)\right]. \tag{3}$$

Initially the problem will be investigated in the absence of observation errors. If $R_l^{-1}$ was chosen so as to reflect the statistical properties of numerical model error, then at this initial point we know nothing about these statistics, so we choose $R_l^{-1} = I_N$ for all $l$ by taking $\sigma_o = 1$. This choice gives the observations an equal weighting, assuming nothing about the error statistics of the numerical model. These values were also chosen by Daley [2], Griffith et al. [14] and Vukićević et al. [20]. In Section 4.2, observation errors will be reintroduced to the problem.

### *2.1. The physical system*

Consider the 1D linear advection equation for the function $u : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, (x, t) \mapsto u(x, t)$, together with circulant boundary conditions and initial condition $u_0 : \mathbb{R} \to \mathbb{R}$ given by,

$$\begin{aligned} &u_t(x,t) + \mu u_x(x,t) = 0, && x \in [0,1), \quad t > 0, \\ &u(x,t) = u(x+1,t), && x \in \mathbb{R}, \quad t \geq 0, \\ &u(x,0) = u_0(x), && x \in [0,1). \end{aligned} \tag{4}$$

Here the *wave speed* $\mu \in \mathbb{R}$ remains constant. In the context of data assimilation, this problem is also considered in Freitag et al. [12] and Griffith et al. [14]. It is important to note that the scalar $x$ is the spatial dimension whilst the vectors $\{\mathbf{x}_l\}_{l=0}^L$ denote the state of the numerical model.

**Definition 1.** *We define $u_0(x)$ to have regularity $r \in \mathbb{N}_0$ over $(0,1)$ when $r$ denotes the maximum number of times the function $u_0(x)$ can be differentiated with respect to $x$, such that $u_0^{(\alpha)}(x)$ is continuous and piecewise differentiable over $(0,1)$, for $\alpha = 0, \ldots, r-1$ and $u_0^{(r)}(x)$ is piecewise continuous over $(0,1)$.*

The solution to this problem, $u(x,t) = u(x-\mu t, 0) = u_0([x-\mu t]_1)$ [6], preserves the shape of the initial condition over time and propagates it through space with speed $\mu$. Here $[\cdot]_1$ denotes modulo one. Problem (4) can be solved numerically using a finite difference scheme as the forward model. These find a numerical approximation to the analytic solution, introducing numerical model error, which will be considered in the form of *numerical dissipation* and *numerical dispersion* [24].

### *2.2. Numerical dissipation and dispersion*

In order to introduce the concept of numerical dissipation and dispersion, it is helpful to represent the solution $u(x,t)$ as a Fourier series, with Fourier basis function $e^{2\pi ikx}$ and corresponding coefficients $c_k : \mathbb{R} \to \mathbb{C}$,

$$u(x,t) \sim \sum_{k=-\infty}^{\infty} c_k(t) e^{2\pi ikx}, \quad \text{where } c_k(t) := \int_0^1 u_0(x,t) e^{-2\pi ikx} \mathrm{d}x.$$

Also, define $c_k(0) := c_k \in \mathbb{C}$ for the Fourier series of $u_0(x)$,

$$u(x,0) \sim \sum_{k=-\infty}^{\infty} c_k e^{2\pi ikx}, \quad \text{where } c_k = \int_0^1 u_0(x) e^{-2\pi ikx} \mathrm{d}x. \tag{5}$$

We term the $k$th Fourier basis function, multiplied by its corresponding coefficient $c_k$, the $k$th wavenumber component. In a linear problem, the finite difference scheme propagates the wavenumber components of (5) through time, by multiplying each coefficient $c_k$ by the eigenvalue $d_k$ of the matrix used to implement the scheme (see below and Section 3). The variable $N$ denotes the number of discretisation points of the scheme and hence the number of eigenvalues associated with it. This creates a new Fourier series where, in an ideal situation, the $k$th wavenumber coefficient is equal to the wavenumber coefficient $c_k(\Delta t)$ ie: $c_k(\Delta t) = d_k c_k$, $d_k \in \mathbb{C}$. However, most of the time this is not the case and the coefficient $d_k c_k$ forms an approximation to $c_k(\Delta t)$. The error in $d_k c_k$ when compared to $c_k(\Delta t)$ can be described in terms of amplitude and phase errors. Amplitude errors are described as numerical dissipation and phase errors as numerical dispersion [24].

In the case of problem (4), we have $c_k(t) = c_k e^{-2\pi ik\mu t}$. Therefore the coefficient required to multiply $c_k$ to create $c_k(\Delta t)$ is $e^{-2\pi ik\mu\Delta t}$. The numerically dissipative and dispersive properties of the scheme are analysed in comparison to these coefficients. The magnitude of each coefficient is 1, so that the magnitude of $c_k(t)$ does not change with time. Its phase is $-2\pi k\mu\Delta t$, a linear function with respect to $k$. In the case of problem (4), numerical dissipation occurs when at least one eigenvalue of the scheme does not have unit magnitude ie: there exists $k$ such that $|d_k| \neq 1$. This results in the amplitude of at least one wavenumber component of (5) not being preserved over time. Numerical dispersion occurs when at least one wavenumber component is out of phase from the others [24]. This occurs when $e^{i\psi_k} \neq e^{-2\pi ik\mu\Delta t}$ for some $k$, where $\psi_k \in \mathbb{R}$ is the phase of $d_k$, that is $d_k = |d_k| e^{i\psi_k}$. This results in an out of phase wavenumber component travelling with an incorrect phase speed. Numerical dissipation and dispersion are important forms of numerical error to consider as their impact can be widespread and lead to physically unrealistic results. Limitations may sometimes be placed on model variables to avoid these effects, restricting the accuracy of the model [19].

This paper will consider three finite difference schemes which solve problem (4) numerically over the domain $[0,1]$. These schemes are chosen to be representative of solely numerically dissipative, solely numerically dispersive or both numerically dissipative and numerically dispersive schemes with respect to the resolvable wavenumber components. In order to define these scheme, we require the following assumptions.

**Assumptions 1.** *Divide the domain $[0,1]$ into $N+1$ equally spaced mesh points, $N \in \mathbb{N}$. This gives a spatial step size of $\Delta x = 1/N$ and grid points $x_j = j\Delta x$, $j = 0, \ldots, N$. Define the time step $\Delta t \in \mathbb{R}^+$ for the finite difference schemes and $t^n = n\Delta t$ for $n \in \mathbb{N}_0$. Let $U_j^n$ be the numerical solution at $(x_j, t^n)$, such that $U_j^n \approx u(x_j, t^n)$ for $j = 0, \ldots, N$ and $n \in \mathbb{N}$. When $n = 0$, $U_j^0$ is created by sampling $u(x,0)$, such that $U_j^0 := u(x_j, 0)$ for $j = 0, \ldots, N$. Define the vector $\mathbf{U}^n \in \mathbb{R}^N$ such that the $j$th element of $\mathbf{U}^n$ is defined by $\{\mathbf{U}^n\}_j := U_{j-1}^n$ for $j = 1, \ldots, N$. Also, define $h := |\mu|\Delta t/\Delta x$, the CFL number [25] and set $h \leq 1$ so all considered schemes are numerically stable.*

Each scheme is implemented by applying the matrix $M \in \mathbb{R}^{N\times N}$ to $\mathbf{U}^n$, which advances the numerical solution at $x_0, \ldots, x_{N-1}$, forward $\Delta t$ in time, ie: $\mathbf{U}^{n+1} = M\mathbf{U}^n$ for all $n$. The circulant boundary conditions mean that $M$ is a circulant matrix [26] and $u(x_N, t^n) = u(x_0, t^n)$ for all $n$, hence $U_N^n = U_0^n$ for all $n$.

As $\mathbf{U}^n$ is an $N$-dimensional vector, it can be constructed from the $N$ vectors of the *Discrete Fourier Transform* (DFT) basis [27] $\{\mathbf{v}_p\}_{p=1}^N$ such that,

$$\{\mathbf{v}_p\}_q = \frac{1}{\sqrt{N}} e^{\frac{2\pi i(p-1)(q-1)}{N}} = \frac{1}{\sqrt{N}} e^{2\pi i(p-1)x_{q-1}}, \quad p, q = 1, \ldots, N, \tag{6}$$

is the $q$th element of the $p$th vector, $\mathbf{v}_p \in \mathbb{C}^N$. This is the $(p-1)$th Fourier basis function of the Fourier series, sampled at $x_{q-1}$, with amplitude $1/\sqrt{N}$. The numerical solution is constructed from $N$ Fourier basis functions of the Fourier series, that are resolvable on the finite grid, represented by the basis vectors. The remaining Fourier basis functions of the Fourier series are indistinguishable on the discrete mesh, due to *aliasing* [28]. As a result, the coefficients of the $N$ resolvable wavenumber components are determined by the *Poisson summation* [27] when $u_0(x)$ is continuous and has a convergent Fourier series. There is no spectral leakage in this problem due to 1-periodicity, see [27]. This allows $M$ to propagate all the wavenumber components of the Fourier series, by only directly acting on $N$ of them.

Aliasing results in $M$ applying the same magnitude and phase changes to unresolvable wavenumber components, as it applies to the resolvable wavenumber component they alias to. This is not necessarily the correct magnitude or phase change for the considered unresolvable wavenumber component. The matrix $M$ can also introduce numerical dissipation and dispersion into the resolvable wavenumber components of the solution. In the case of problem (4), if $M$ is numerically non-dissipative and non-dispersive with respect to the resolvable wavenumber components, then aliasing is solely a form of numerical dispersion.

### *2.3. Forward models for the linear advection equation*

There are many numerical methods for solving the 1D linear advection equation. Rather than studying them all we consider the Upwind, Preissman Box and Lax-Wendroff schemes as three quantitatively different finite difference schemes used to solve problem (4). These are 'representative schemes' chosen as they exhibit three different types of deterministic numerical model error. These schemes are defined by the following schematics when $\mu > 0$.

- Upwind scheme (explicit scheme) [6],

$$U_j^{n+1} = hU_{j-1}^n + (1-h)U_j^n. \tag{7}$$

- Preissman Box scheme (implicit scheme) [29],

$$(1-h)U_j^{n+1} + (1+h)U_{j+1}^{n+1} = (1+h)U_j^n + (1-h)U_{j+1}^n. \tag{8}$$

- Lax-Wendroff scheme (explicit scheme) [6],

$$U_j^{n+1} = \frac{h}{2}(h+1)U_{j-1}^n + (1-h^2)U_j^n + \frac{h}{2}(h-1)U_{j+1}^n. \tag{9}$$

Selecting $h = 0.5$ allows the effects of numerical dissipation and numerical dispersion on the resolvable wavenumber components to be investigated as individual processes and in combination. For simplicity, $\mu = 1$ is chosen in the following.

Let $M$ denote the matrix used to implement either the Upwind, Preissman Box or Lax-Wendroff scheme as in Section 4. As $M$ is a circulant matrix, its eigenvalue decomposition can be constructed using the unitary DFT matrix, denoted $V \in \mathbb{C}^{N\times N}$ [26],

$$M = V\Lambda V^{-1} = V\Lambda V^*, \tag{10}$$

where $\cdot^*$ denotes Hermitian. The columns of $V$ are the DFT basis vectors in (6), ie: $\{V\}_{p,q} = \{\mathbf{v}_p\}_q$, $p, q = 1, \ldots, N$. The corresponding eigenvalues are found in $\Lambda := \mathrm{diag}(\lambda_p) \in \mathbb{C}^{N\times N}$. The eigenvalues of $M$ are scheme dependent, whereas the eigenvectors are scheme independent. In the case of a linear system, the eigenvalues affect the propagation of the

initial state of the system through time and introduce any numerical dissipation and dispersion into the state of the system [7] (this is not physical dissipation or dispersion).

The eigenvalues $\lambda_p := |\lambda_p|e^{i\theta_p}$, $\theta_p \in [-\pi, \pi)$ for $p = 1, \ldots, N$, are such that the magnitude $|\lambda_p|$ and phase $\theta_p$ affect the magnitude and phase of the corresponding wavenumber component of the DFT basis, respectively. Aliasing results in $d_k = \lambda_{[k]_N+1}$ where $[\cdot]_N$ denotes modulo $N$. The DFT basis is complex; given an eigenpair of $M$, its complex conjugate is also an eigenpair of $M$ [30]. This results in $\lambda_1 \in \mathbb{R}$ and $\lambda_p = \overline{\lambda}_{N-p+2}$ for $p = 2, \ldots, N$, hence $\theta_1 = 0$ and $\theta_p = -\theta_{N-p+2}$ for $p = 2, \ldots, N$. Summing the conjugate pair of eigenvectors, scaled by their respective complex coefficients for the state of the system, results in a real wavenumber component for the linear system. We consider these real wavenumber components in Sections 3.1-3.4 of this paper.

### *2.4. Generating perfect observations*

Consider the following finite difference scheme, the *Numerical Implementation of the Method of Characteristics* (NIMC):

$$U_j^{n+1} = \text{sgn}(\mu)hU_{j-1}^n. \tag{11}$$

This is an explicit finite difference scheme and can be implemented as described in Section 2.2, via a circulant matrix $M_{NIMC} \in \mathbb{R}^{N\times N}$. This scheme is only consistent when $h = 1$. At this value of $h$, the scheme is numerically non-dissipative and non-dispersive with respect to all wavenumber components. The scheme can be used to generate perfect observations for the physical system in this instance. This leads to $M_{NIMC}$ generating a perfect observation every $\Delta t = \Delta x/|\mu|$ in time. However, the imperfect schemes in this paper progress the forward model $\Delta t = h\Delta x/|\mu|$ in time with each application of $M$, where $h$ is not necessarily equal to one. This means that given the same $\Delta t$, $\Delta x$ and $\mu$, the NIMC cannot provide perfect observations at every time step of the imperfect forward model.

Hence another finite difference scheme is used to create perfect observations. This scheme will be known as the *Modified NIMC* (MNIMC) scheme.

**Definition 2** (The MNIMC scheme)**.** *Let Assumptions 1 hold true with* $\mathbf{U}^n$ *replaced by* $\tilde{\mathbf{U}}^n$ *to mark the difference in the schemes. Define the matrix* $\tilde{M} \in \mathbb{R}^{N\times N}$ *where N is odd, by* $\tilde{M} := V\tilde{\Lambda}V^*$*, where V is defined as in Section 2.3 and* $\tilde{\Lambda} := diag(\tilde{\lambda}_p)$ *the diagonal matrix of eigenvalues of the scheme,* $\tilde{\lambda}_p \in \mathbb{C}$ *for* $p = 1, \ldots, N$*. The eigenvalues of the MNIMC scheme are defined by* $\tilde{\lambda}_p = e^{i\tilde{\theta}_p}$ *such that,*

$$\tilde{\theta}_p = \begin{cases} \frac{-2\pi(p-1)sgn(\mu)h}{N}, & for\ p = 1, \ldots, \frac{N+1}{2}, \\ \frac{2\pi(N-p+1)sgn(\mu)h}{N}, & for\ p = \frac{N+3}{2}, \ldots, N. \end{cases} \tag{12}$$

*The scheme is implemented by multiplying* $\tilde{\mathbf{U}}^n$ *by the matrix* $\tilde{M}$ *to move the state of the system forward* $\Delta t$ *in time, ie:* $\tilde{\mathbf{U}}^{n+1} = \tilde{M}\tilde{\mathbf{U}}^n$.

This scheme is numerically non-dissipative with respect to all wavenumber components and non-dispersive with respect to the resolvable wavenumber components for any $h \in \mathbb{R}^+$, producing the state of the system every $\Delta t = h\Delta x/|\mu|$ in time. It is also numerically stable for any $h \in \mathbb{R}^+$. If the current time of the solution is divisible by $\Delta x/|\mu|$, then the solution of the system is exact. However in between these times, the state of the system at each spatial mesh point is interpolated in time, due to aliasing errors. This is due to the scheme being dispersive for unresolvable wavenumber components of the solution. Let $\tilde{\mathbf{x}}_0 \in \mathbb{R}^N$ denote the true initial condition $u_0(x)$, sampled at the spatial grid points $x_0, \ldots, x_{N-1}$ defined in Assumptions 1, such that $\{\tilde{\mathbf{x}}_0\}_j := u_0(x_{j-1})$. Now define $\tilde{\mathbf{x}}_l \in \mathbb{R}^N$ by $\tilde{\mathbf{x}}_l := \tilde{M}^l\tilde{\mathbf{x}}_0$ for all $l \in \mathbb{N}$. The global error in the MNIMC scheme is defined by,

$$\mathbf{r}_l := \tilde{\mathbf{y}}_l - \tilde{M}^l\tilde{\mathbf{x}}_0, \tag{13}$$

where $\mathbf{y}_l := \tilde{\mathbf{y}}_l$ denotes the $l$th set of perfect observations. As only aliasing errors are introduced by the MNIMC scheme, $\mathbf{r}_l$ can be viewed as an additive correction term to correct for aliasing errors in $\tilde{M}^l\tilde{\mathbf{x}}_0$ such that,

$$\tilde{\mathbf{y}}_l = \tilde{M}^l\tilde{\mathbf{x}}_0 + \mathbf{r}_l. \tag{14}$$

Choosing $h = 1$ results in $\tilde{M} = M_{NIMC}$ and consequently $\mathbf{r}_l = \mathbf{0}$ for all $l \in \mathbb{N}_0$. Lemma 1 provides some insight into the properties of the aliasing error introduced by the MNIMC scheme.

**Lemma 1.** *Let the conditions in Assumptions 1 hold true so the MNIMC scheme can be defined as in Definition 2. Also, let $u_0(x)$ be bounded and piecewise continuous on $[0, 1)$ and suppose that the left- and right-hand derivatives of $u_0(x)$ exist for all $x \in [0, 1]$.*

*Additionally, consider the CFL number to be a rational number $h \in \mathbb{Q}^+$ expressed as $h = q/b$, $q, b \in \mathbb{N}$ such that $gcd(q, b) = 1$ (greatest common divisor). Then the global error in the MNIMC scheme at time $l\Delta t$, defined by Equation (13), is such that,*

$$\mathbf{r}_l = \begin{cases} \mathbf{0}, & \text{for } [l]_b = 0, \\ \tilde{M}^{l-[l]_b}\mathbf{r}_{[l]_b}, & \text{for } [l]_b = 1, \dots, b-1, \end{cases} \tag{15}$$

*for all $l \in \mathbb{N}_0$ where $[\cdot]_b$ denotes modulo $b$.*

The proof of Lemma 1 can be found in [31]. Examining expression (15), we see that the aliasing error in $\tilde{M}$ has a shifted $b$-periodic nature. Raising the matrix $\tilde{M}$ to the power $l - [l]_b$ results in $\tilde{M}$ being raised to a power which is an integer multiple of $b$. Suppose $l - [l]_b = sb$ for some $s \in \mathbb{N}_0$, then $\tilde{M}^{l-[l]_b} = M_{NIMC}^{sq}$ (here $\tilde{M}$ is defined using any $h \in \mathbb{Q}^+$, and $M_{NIMC}$ is defined using $h = 1$). Applying this matrix to $\mathbf{r}_{[l]_b}$ shifts it $sq\Delta x$ in space introducing no numerical dissipation or dispersion. This means that $\mathbf{r}_l$ is $\mathbf{r}_{[l]_b}$ shifted an integer number of discretisation points in space.

## 3. The effect of numerical dissipation and dispersion on the analysis vector

Numerical dissipation and dispersion are introduced into the inverse problem through the forward model. This Section explores how these errors affect the analysis vector. This is achieved by formulating the analysis vector in terms of the true initial condition, allowing the direct impact of numerically dissipative and/or dispersive eigenvalues of the imperfect scheme, to be seen. Under Assumptions 1 $\mathcal{M}_{l+1,l} := M$ and $\mathbf{x}_l \equiv \mathbf{U}^l$ for all $l \in \mathbb{N}_0$, and the cost function becomes,

$$J(\mathbf{x}_0) = \sum_{l=0}^{L} [\mathbf{y}_l - M^l\mathbf{x}_0]^T [\mathbf{y}_l - M^l\mathbf{x}_0]. \tag{16}$$

Let $\mathcal{F} : \mathbb{R}^N \to \mathbb{C}^N$, $\mathbf{x} \mapsto \mathcal{F}(\mathbf{x}) := V^*\mathbf{x}$, be the DFT operator and $\mathcal{F}_p(\cdot)$ denote the $p$th element of $\mathcal{F}(\cdot)$, $p = 1, \dots, N$. The analysis vector $\mathbf{x}_a$, is the solution to the inverse problem, ie: $\nabla J(\mathbf{x}_a) = 0$. Then,

$$\mathbf{x}_a = \left[\sum_{k=0}^{L} (M^T M)^k\right]^{-1} \sum_{l=0}^{L} (M^T)^l \mathbf{y}_l = V \left[\sum_{k=0}^{L} (\Lambda^*\Lambda)^k\right]^{-1} \sum_{l=0}^{L} (\Lambda^*)^l V^* \mathbf{y}_l, \tag{17}$$

using (10). This can be re-written using the DFT,

$$\mathcal{F}(\mathbf{x}_a) = \left[\sum_{k=0}^{L} (\Lambda^*\Lambda)^k\right]^{-1} \left[\sum_{l=0}^{L} (\Lambda^*)^l \mathcal{F}(\mathbf{y}_l)\right] = \left[I_N + \sum_{k=1}^{L} (\Lambda^*\Lambda)^k\right]^{-1} \left[\sum_{l=0}^{L} (\Lambda^*)^l \mathcal{F}(\mathbf{y}_l)\right]. \tag{18}$$

Here the diagonal matrices $\Lambda$ and $\Lambda^*$ are known as the *forward* and *adjoint models* [4] respectively, in the DFT basis. In the inverse problem, each set of observations is mapped back in time to $t = 0$, by the adjoint model $M^T$. Once the observations have been mapped back to the initial time, they are then summed. This process has the potential to create interference between the corresponding wavenumber components constructing each set of observations in time. The result is then normalised with respect to the eigenvalues of the scheme. The observation at $t = 0$ acts to regularise the solution of the inverse problem so that the matrix applying the normalisation is always invertible.

Expression (18) forms the coefficients of the DFT basis in the construction of the analysis vector $\mathbf{x}_a$, ie: $\mathbf{x}_a = V\mathcal{F}(\mathbf{x}_a)$. The following Lemma provides an expression for the analysis vector in terms of the sum of a matrix operation on $\tilde{\mathbf{x}}_0$, $A_L\tilde{\mathbf{x}}_0$ and an aliasing correction term $\boldsymbol{\rho}_L$. The matrix $A_L$ is constructed from the MNIMC scheme and the matrix $M$ implementing the considered numerically dissipative and/or numerically dispersive scheme.

**Lemma 2.** *Let the assumptions of Lemma 1 hold true, allowing* $\mathbf{x}_a$ *to be stated as in (17). Consider perfect observations of the physical system ie:* $\mathbf{y}_l := \tilde{\mathbf{y}}_l$ *for all* $l = 0, \dots, L$*, where* $L \in \mathbb{N}_0$ *is finite, in the form of (14). Then the analysis vector can be expressed as,*

$$\mathbf{x}_a = A_L \tilde{\mathbf{x}}_0 + \boldsymbol{\rho}_L, \tag{19}$$

*where the* model resolution matrix $A_L \in \mathbb{R}^{N\times N}$ *is such that,*

$$A_L := V \left[ \sum_{k=0}^{L} (\Lambda^* \Lambda)^k \right]^{-1} \left[ \sum_{l=0}^{L} (\Lambda^* \tilde{\Lambda})^l \right] V^*, \tag{20}$$

*and* $\boldsymbol{\rho}_L \in \mathbb{R}^N$ *is given by,*

$$\boldsymbol{\rho}_L := V \left[ \sum_{k=0}^{L} (\Lambda^* \Lambda)^k \right]^{-1} \left[ \left\{ \sum_{l=0}^{\frac{L-[L]_b}{b}-1} (\Lambda^* \tilde{\Lambda})^{lb} \right\} \left\{ \sum_{j=1}^{b-1} (\Lambda^*)^j V^* \mathbf{r}_j \right\} + \left( \Lambda^* \tilde{\Lambda} \right)^{L-[L]_b} \left\{ \sum_{j=1}^{[L]_b} (\Lambda^*)^j V^* \mathbf{r}_j \right\} \right]. \tag{21}$$

*Here we consider* $\sum_{j=1}^{0} (\Lambda^*)^j V^* \mathbf{r}_j = \mathbf{0}$ *and* $\sum_{l=0}^{-1} \left( \Lambda^* \tilde{\Lambda} \right)^{lb} = 0_N \in \mathbb{R}^{N\times N}$ *as we assume* $\mathbf{r}_0 = \mathbf{0}$*.*

The proof of Lemma 2 can be found in [31]. Expression (19) can be viewed as the sum of two analysis vectors created when solving the same problem but with two different sets of observations; $\mathbf{y}_l = \tilde{\mathbf{x}}_l$ and $\mathbf{y}_l = \mathbf{r}_l$. As a result the aliasing error in $\mathbf{y}_l := \tilde{\mathbf{x}}_l$ does not play a part in $A_L$, and is solely found in $\boldsymbol{\rho}_L$. Consequently, $\boldsymbol{\rho}_L$ acts as a correction term for the aliasing errors introduced into the analysis vector by the MNIMC scheme.

The eigenvalues of $A_L$ in (20) determine the magnitude and phase change applied to each wavenumber component of $\tilde{\mathbf{x}}_0$, in the construction of $\mathbf{x}_a$. In this way, they can be described as *amplification factors* for the wavenumber components of $\tilde{\mathbf{x}}_0$. Let $\nu_p$ be an eigenvalue of $A_L$ such that $\nu_p = |\nu_p| e^{i\kappa_p}$, $\kappa_p \in [-\pi, \pi)$ for $p = 1, \dots, N$. Due to the diagonal structures of $\Lambda$ and $\tilde{\Lambda}$, $\nu_p$ is constructed solely from $\lambda_p$ and $\tilde{\lambda}_p$, the $p$th eigenvalues of $M$ and $\tilde{M}$ respectively,

$$\nu_p = \frac{\sum_{l=0}^{L} \overline{\lambda}_p^l \tilde{\lambda}_p^l}{\sum_{k=0}^{L} |\lambda_p|^{2k}}. \tag{22}$$

Numerical model error can enter into $\nu_p$ via both $\lambda_p$ and $\tilde{\lambda}_p$. In the case of $\tilde{\lambda}_p$, any error introduced is due to aliasing. As $\overline{\lambda}_p = \lambda_{N-p+2}$ and $\overline{\tilde{\lambda}}_p = \tilde{\lambda}_{N-p+2}$ for $p = 2, \dots, N$, $\overline{\nu}_p = \nu_{N-p+2}$ and $\kappa_p = -\kappa_{N-p+2}$ for $p = 2, \dots, N$. Define $\phi_p := \tilde{\theta}_p - \theta_p$, for $p = 1, \dots, N$ as the error in the phase shift applied by $\lambda_p$ with respect to the corresponding resolvable wavenumber component of the DFT basis. The complex conjugate property of the eigenvalues results in $-\phi_p = \phi_{N-p+2}$ for $p = 2, \dots, N$. Then,

$$\nu_p = \begin{cases} 1, & \text{for } |\lambda_p| = 1 \text{ and } \phi_p = 2\pi s, \\ \frac{1+|\lambda_p|}{1+|\lambda_p|^{L+1}}, & \text{for } |\lambda_p| < 1 \text{ and } \phi_p = 2\pi s, \\ \frac{1}{L+1} \left| \frac{\sin\left[(L+1)\frac{\phi_p}{2}\right]}{\sin\left[\frac{\phi_p}{2}\right]} \right| e^{i\kappa_p}, & \text{for } |\lambda_p| = 1 \text{ and } \phi_p \neq 2\pi s, \\ \frac{[1-|\lambda_p|^{L+1} e^{i(L+1)\phi_p}][1-|\lambda_p|^2][1-|\lambda_p| e^{-i\phi_p}]}{[1-|\lambda_p|^{2(L+1)}][1+|\lambda_p|^2-2|\lambda_p|\cos(\phi_p)]}, & \text{for } |\lambda_p| < 1 \text{ and } \phi_p \neq 2\pi s, \end{cases} \tag{23}$$

where $s \in \mathbb{Z}$, by the sum of a geometric progression. When $|\lambda_p| = 1$ and $\phi_p \neq 2\pi s$ for some $s \in \mathbb{Z}$,

$$\tan(\kappa_p) = \tan\left(\frac{L\phi_p}{2}\right), \quad \kappa_p \in [-\pi, \pi),$$

for $p = 1, \dots, N$.

When $\lambda_p$ does not introduce numerical model error into the corresponding resolvable wavenumber component, $\nu_p = 1$, so the corresponding resolvable wavenumber component of $\tilde{\mathbf{x}}_0$ is preserved in $\mathbf{x}_a$. A solely numerically dissipative $\lambda_p$ with respect to the corresponding resolvable wavenumber component, creates an amplification factor that affects the amplitude and not the phase of the corresponding resolvable wavenumber component.

In the case of a solely numerically dispersive eigenvalue of $M$, the amplification factor affects both the phase and amplitude of the corresponding resolvable wavenumber component of $\tilde{\mathbf{x}}_0$. The affect on the magnitude is due to interference between the corresponding resolvable wavenumber components making up each set of observations $\mathbf{y}_l$ in the construction of $\mathbf{x}_a$, as discussed in Section 3 after Equation (18).

A numerically dissipative and dispersive eigenvalue of $M$, creates an amplification factor that appears to combine the solely numerically dissipative and solely numerically dispersive amplification factors. However, it is not possible to isolate the dissipative and dispersive effects from one another. The magnitude and phase of the spectra of the model resolution matrix for each scheme are plotted in Figures 1, 3 and 5.

The contribution of $\boldsymbol{\rho}_L$ to the analysis vector is not as easy to analyse, but can be reduced by choosing an $\tilde{\mathbf{x}}_0$ that is minimally constructed from unresolvable wavenumber components or by increasing $N$. As a result, a higher regularity initial condition will reduce $\boldsymbol{\rho}_L$. Choosing $h = 1$ leads to $\boldsymbol{\rho}_L = \mathbf{0}$.

We begin our analysis by examining the effects of the model resolution matrix on $\tilde{\mathbf{x}}_0$ and the contribution of $\boldsymbol{\rho}_L$ to the analysis vector, by considering a low regularity $u_0(x)$ in the form of a square function defined by,

$$u_0(x) = \begin{cases} 0.5, & \text{for } x \in [0.25, 0.5], \\ -0.5, & \text{for } x \in [0, 0.25) \cup (0.5, 1). \end{cases} \tag{24}$$

This square function has regularity zero, requiring many high wavenumber components to resolve the edges of the function. The vector $\tilde{\mathbf{x}}_0$ is then a discrete sample of the square function. The square function allows us to analyse the ability of strong constraint 4D-Var data assimilation, to reconstruct initial conditions that contain unresolvable wavenumber components, in the presence of numerical dissipation and/or dispersion. This tests the effects of numerical dissipation and/or dispersion on strong constraint 4D-Var, in the same way as Durran's "spike test" [24].

In Sections 3.1-3.4, the magnitude and phase of the spectra of $A_L$ are analysed for the three schemes, in terms of the real wavenumber components of the solution, together with the result of applying $A_L$ to $\tilde{\mathbf{x}}_0$ for the square function when using $L = 4$. Here $L = 4$ is chosen so as to be consistent with our results in Section 4. Here we remind the reader that $L$ is the number of sets of observations in time. The corresponding $\boldsymbol{\rho}_L$ and $\mathbf{x}_a$ for the square function are also shown for $L = 4$. The reader is reminded that $A_L$ acts upon all wavenumber components of $\tilde{\mathbf{x}}_0$ through the effects of aliasing. The eigenpair property of $\nu_p$ can be seen through the line of symmetry in the centre of the plots for the magnitude of $\nu_p$ and the rotational symmetry in the plots for the phase of $\nu_p$ ie: $\kappa_p$.

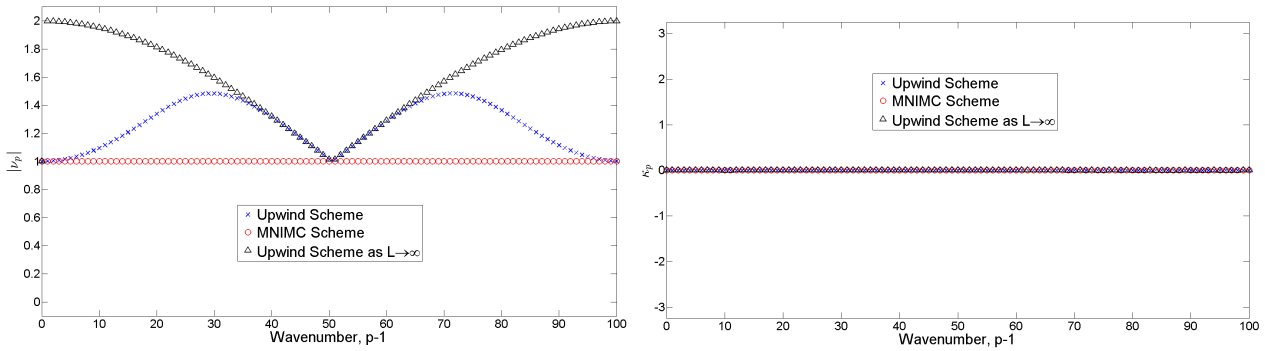
We remind the reader here the effects of $A_L$ on the real wavenumber components of $\tilde{\mathbf{x}}_0$, can be seen in the first $(N + 1)/2$ ($N$ odd) values of $p$, due to the complex conjugate properties of the eigenvalues of the schemes. This property results in the discontinuity seen in Figure 5(b). The magnitude and phase of $\nu_p$ are plotted against $(p - 1)$ as these are the wavenumbers of the resolvable wavenumber components of the Fourier series for the numerical solution. Increasing $p$ over $p = 1, \ldots, \frac{N+1}{2}$, represents increasing the wavenumber of the resolvable real wavenumber component from low to high. The discussions below will make use of this terminology.

### *3.1. The Upwind scheme*

When $h = 0.5$, the Upwind scheme is a numerically dissipative and non-dispersive scheme with respect to the resolvable wavenumber components of the numerical solution. This results in the aliasing error introduced by the scheme, being both numerically dissipative and dispersive [31]. These properties of the Upwind scheme and the numerically dispersive aliasing errors introduced by the MNIMC scheme, dictate the oscillations in $A_L\tilde{\mathbf{x}}_0$ and $\boldsymbol{\rho}_L$, compared to $\tilde{\mathbf{x}}_0$ and $\mathbf{0} \in \mathbb{R}^N$ respectively.

Examining the phase of the eigenvalues of $A_L$ in Figure 1(b), we see that all the resolvable wavenumber components of $\tilde{\mathbf{x}}_0$ are propagated with the correct phase speed. As a result, there are no destructive or constructive interference effects affecting the magnitude of the eigenvalues of $A_L$ in Figure 1(a). Examining the magnitude of the eigenvalues of $A_L$, we see that all but the lowest real resolvable wavenumber components (ie: $p = 1$) of $\tilde{\mathbf{x}}_0$ are amplified by $A_L$. The greatest amplification effects are experienced by the medium real resolvable wavenumber components. As $L$ increases, the amplification of the lower real resolvable wavenumber components of $\tilde{\mathbf{x}}_0$ increases.

The plots of $A_L\tilde{\mathbf{x}}_0$ and $\boldsymbol{\rho}_L$ in Figure (2) demonstrate oscillations at the locations of the discontinuities making up the square function in $\tilde{\mathbf{x}}_0$. The discontinuities are formed from the unresolvable wavenumber components of the square function, so these oscillations represent a failure to propagate the unresolvable wavenumber components of $\tilde{\mathbf{x}}_0$. As $\boldsymbol{\rho}_L$ corrects for the aliasing errors introduced by the MNIMC scheme, this verifies that the oscillations are due to errors in the propagation of the unresolvable wavenumber components of $\tilde{\mathbf{x}}_0$. Adding $\boldsymbol{\rho}_L$ to $A_L\tilde{\mathbf{x}}_0$ removes the effects of aliasing introduced by the MNIMC scheme into $A_L\tilde{\mathbf{x}}_0$, in order to construct $\mathbf{x}_a$ in Figure 2(c). This visibly improves the width of the oscillations in $\mathbf{x}_a$ in Figure 2(c) when compared to $A_L\tilde{\mathbf{x}}_0$ in Figure 2(a). This indicates how important accounting for the effects of aliasing can be. The error in $\mathbf{x}_a$ is solely due to numerical model error introduced by using the Upwind scheme as the forward model. Similar results follow for the remaining schemes in Figures (4) and (6), with regard to the effects of aliasing errors.




(a) Magnitude of the amplification factors.

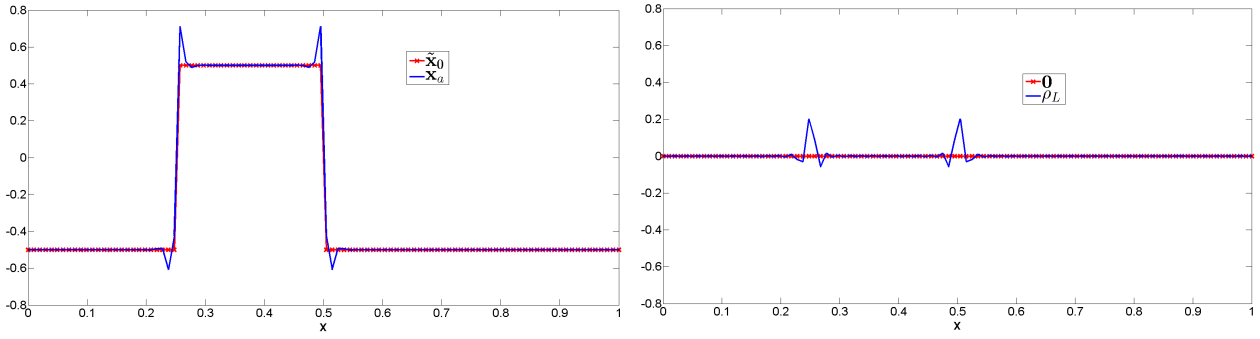
(b) Phase of the amplification factors.

Figure 1: The magnitude and phase of the spectrum of the model resolution matrix, $A_L$ for $L = 4$, together with their limit as $L \to \infty$, for the Upwind scheme when $h = 0.5, \mu = 1$ and $N = 101$ ($\Delta t = \frac{1}{202}$). The magnitude and phase of the spectrum of $A_L$ for the MNIMC scheme is included for comparison, using the same variables.

### *3.2. The Preissman Box scheme*

The Preissman Box scheme is always numerically non-dissipative with respect to all wavenumber components of the numerical solution, when solving the 1D linear advection problem. When $h = 0.5$ the scheme is numerically dispersive with respect to the resolvable wavenumber components except when $p = 1$. Aliasing is also introduced in the form of numerical dispersion [31]. These properties of the Preissman Box scheme and the numerically dispersive aliasing errors introduced by the MNIMC scheme, determine the oscillations in $A_L\tilde{\mathbf{x}}_0$ and $\boldsymbol{\rho}_L$ compared to $\tilde{\mathbf{x}}_0$ and $\mathbf{0} \in \mathbb{R}^N$ respectively, in Figure 4. This means that only numerically dispersive effects introduce errors into $A_L\tilde{\mathbf{x}}_0$ and $\boldsymbol{\rho}_L$.

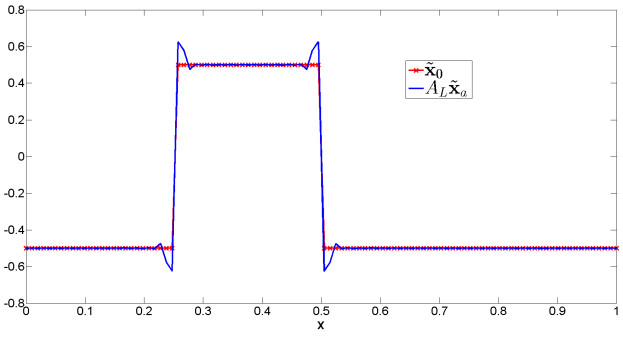
Examining the eigenvalues of $A_L$ in Figure 3, we see that the numerically dispersive effects of the schemes affect both the magnitude (Figure 3(a)) and phase (Figure 3(b)) of the eigenvalues. As there is no numerical dissipation taking place, it is solely the affects of destructive interference between the wavenumber components of the sets of observations in time, that is causing the attenuation of the resolvable wavenumber components of $\tilde{\mathbf{x}}_0$. This was discussed in Section 3 after Equation (18). The amplitude of the lowest resolvable real wavenumber component is the only one not affected ($p = 1$) by destructive interference, as this wavenumber is always correctly propagated by the Preissman Box and MNIMC schemes. In this instance, the low to medium resolvable real wavenumber components experience a small attenuation effect, whilst the medium to high resolvable real wavenumber components experience a much larger attenuation. The highest resolvable real wavenumber components are almost attenuated to zero. As the number of observations is increased, it is not possible to define a limit for the phase of the eigenvalues of $A_L$ as $L \to \infty$. However Figure 3(a) shows that as $L \to \infty$, the magnitude of all eigenvalues of $A_L$ except $\nu_1$, decay to zero. This will be discussed in Section 3.4. The effects of destructive interference on the square function initial condition in $\tilde{\mathbf{x}}_0$, can be seen in Figure 4. The discussion of the effects of adding $A_L\tilde{\mathbf{x}}_0$ and $\boldsymbol{\rho}_L$ in Figures 4(a) and 4(b) respectively, to create $\mathbf{x}_a$ in Figure 4(c), is similar to that in Section 3.1 for the Upwind scheme.




(a) $A_L\tilde{\mathbf{x}}_0$

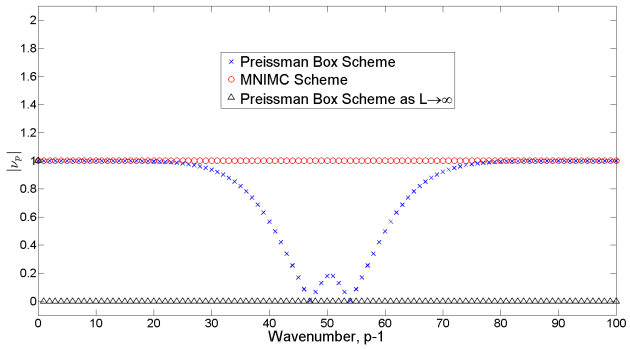



(b) $\boldsymbol{\rho}_L$

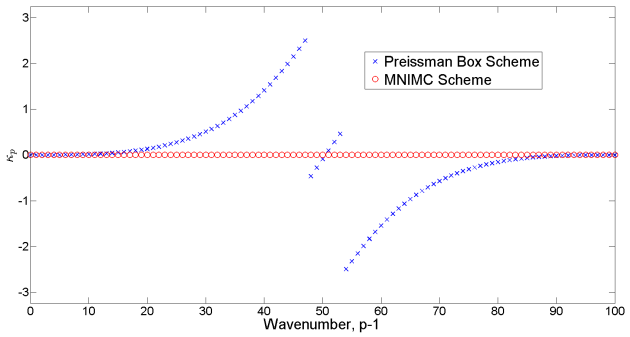



(c) $\mathbf{x}_a = A_L\tilde{\mathbf{x}}_0 + \boldsymbol{\rho}_L$

Figure 2: The analysis vector, $\mathbf{x}_a = A_L\tilde{\mathbf{x}}_0 + \boldsymbol{\rho}_L$, for the square function initial condition, when using the Upwind scheme and perfect observations, $\mathbf{y}_l = \tilde{\mathbf{y}}_l = \tilde{\mathbf{x}}_l + \mathbf{r}_l$, for $h = 0.5$, $\mu = 1$, $N = 101$ and $L = 4$ ($\Delta t = \frac{1}{202}$).




(a) Magnitude of the amplification factors



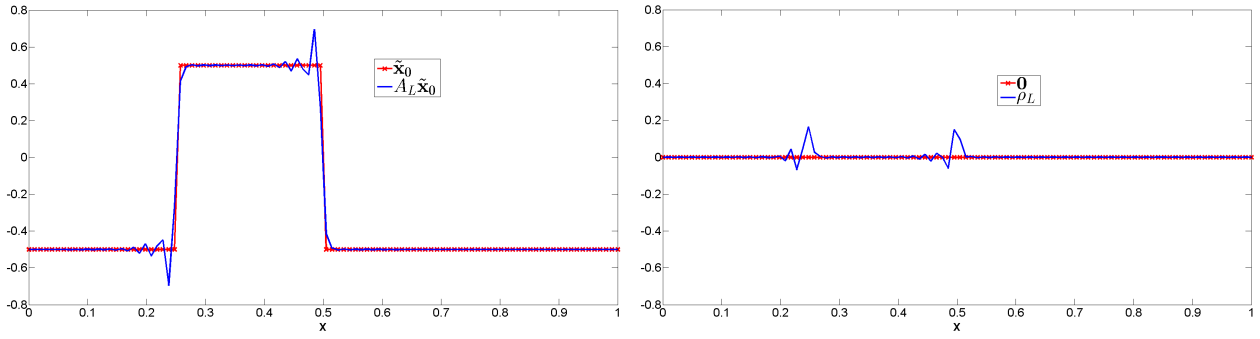

(b) Phase of the amplification factors

Figure 3: The magnitude and phase of the spectrum of the model resolution matrix, $A_L$ for $L = 4$, together with the limit as $L \to \infty$ for the magnitudes, for the Preissman Box scheme when $h = 0.5$, $\mu = 1$ and $N = 101$ ($\Delta t = \frac{1}{202}$). The magnitude and phase of the spectrum of $A_L$ for the MNIMC scheme is included for comparison, using the same variables.

### *3.3. The Lax-Wendroff scheme*

When $h = 0.5$, the Lax-Wendroff scheme is both numerically dissipative and dispersive with respect to the resolvable wavenumber components of the numerical solution. This results in the aliasing error introduced by the scheme being both numerically dissipative and dispersive [31]. These properties of the scheme, along with the numerically dispersive aliasing errors introduced by the MNIMC scheme, dictate the oscillations present in $A_L\tilde{\mathbf{x}}_0$ and $\boldsymbol{\rho}_L$ compared to $\tilde{\mathbf{x}}_0$ and $\mathbf{0} \in \mathbb{R}^N$, respectively.

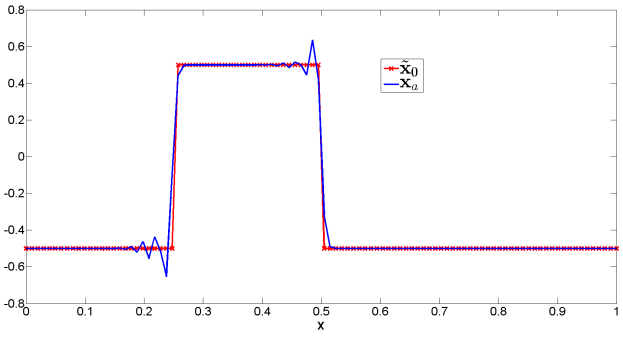
Examining the eigenvalues of $A_L$ in Figure 5, we see that the amplitude of the eigenvalues in 5(a) appear to




(a) $A_L\tilde{\mathbf{x}}_0$



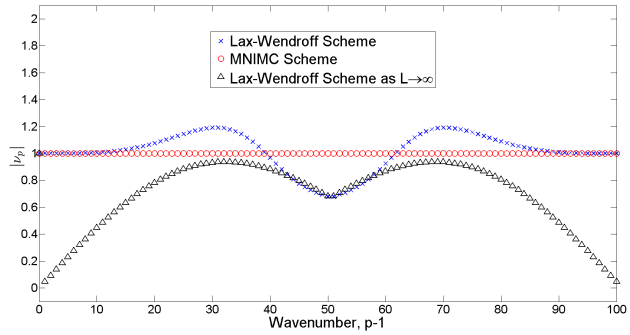

(b) $\boldsymbol{\rho}_L$



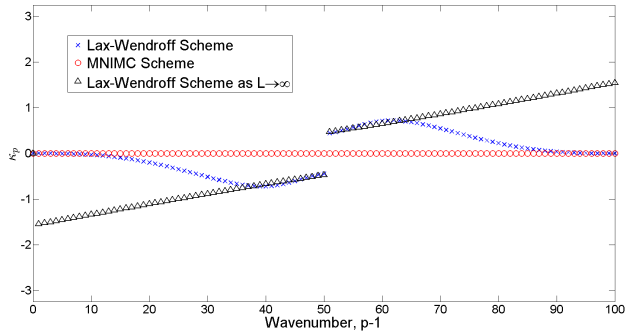

(c) $\mathbf{x}_a = A_L\tilde{\mathbf{x}}_0 + \boldsymbol{\rho}_L$

Figure 4: The analysis vector, $\mathbf{x}_a = A_L\tilde{\mathbf{x}}_0 + \boldsymbol{\rho}_L$, for the square function initial condition, when using the Preissman Box scheme and perfect observations, $\mathbf{y}_l = \tilde{\mathbf{y}}_l = \tilde{\mathbf{x}}_l + \mathbf{r}_l$, for $h = 0.5$, $\mu = 1$, $N = 101$ and $L = 4$ ($\Delta t = \frac{1}{202}$).

experience a combination of the amplification affects seen in Figure 1(a) for the Upwind scheme and the attenuation affects seen in Figure 3(a) for the Preissman Box scheme. This was also observed in the formulation of $\nu_p$ for this type of scheme, in the text between Equations (23) and (24). The combination of effects sees the medium and the highest real resolvable wavenumber components of $\tilde{\mathbf{x}}_0$, amplified and attenuated respectively for the Lax-Wendroff scheme, when $L = 4$. The amplification effects seem to balance the attenuation effects so no real resolvable wavenumber components are attenuated to zero. The discussion of the effects of adding $A_L\tilde{\mathbf{x}}_0$ and $\boldsymbol{\rho}_L$ in Figures 6(a) and 6(b) respectively, to create $\mathbf{x}_a$ in Figure 6(c), is similar to that in Section 3.1 for the Upwind scheme.

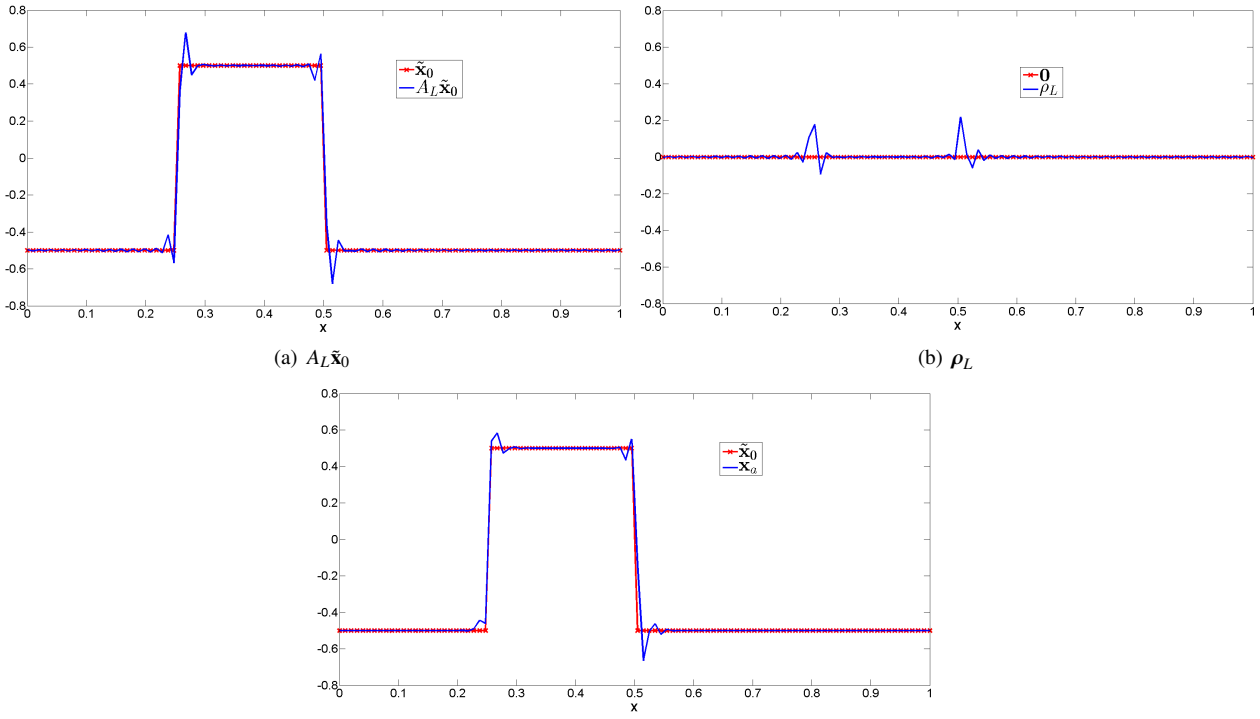



(a) Magnitude of the amplification factors




(b) Phase of the amplification factors

Figure 5: The magnitude and phase of the spectrum of the model resolution matrix, $A_L$ for $L = 4$, together with their limit as $L \to \infty$, for the Lax Wendroff scheme when $h = 0.5$, $\mu = 1$ and $N = 101$ ($\Delta t = \frac{1}{202}$). The magnitude and phase of the spectrum of $A_L$ for the MNIMC scheme is included for comparison, using the same variables.




(a) $A_L\tilde{\mathbf{x}}_0$ (b) $\boldsymbol{\rho}_L$

(c) $\mathbf{x}_a = A_L\tilde{\mathbf{x}}_0 + \boldsymbol{\rho}_L$

Figure 6: The analysis vector, $\mathbf{x}_a = A_L\tilde{\mathbf{x}}_0 + \boldsymbol{\rho}_L$, for the square function initial condition, when using the Lax-Wendroff scheme and perfect observations, $\mathbf{y}_l = \tilde{\mathbf{y}}_l = \tilde{\mathbf{x}}_l + \mathbf{r}_l$, for $h = 0.5$, $\mu = 1$, $N = 101$ and $L = 4$ ($\Delta t = \frac{1}{202}$).

### *3.4. The length of the assimilation window*

Another factor that affects the behaviour of numerical model error is the length of the assimilation window. It is important to understand whether the extra time and processing power required to include more observations will yield an improvement in the solution. To understand the behaviour of $\nu_p$ for large $L$, we consider $\nu_p$ as $L \to \infty$. As $L \to \infty$,

$$\nu_p \to \begin{cases} 1, & \text{for } |\lambda_p| = 1 \text{ and } \phi_p = 2\pi s,\ s \in \mathbb{Z}, \\ 1 + |\lambda_p|, & \text{for } |\lambda_p| < 1 \text{ and } \phi_p = 2\pi s,\ s \in \mathbb{Z}, \\ 0, & \text{for } |\lambda_p| = 1 \text{ and } \phi_p \neq 2\pi s,\ s \in \mathbb{Z}, \\ \frac{(1-|\lambda_p|^2)(1-|\lambda_p|e^{-i\phi_p})}{1+|\lambda_p|^2-2|\lambda_p|\cos(\phi_p)}, & \text{for } |\lambda_p| < 1 \text{ and } \phi_p \neq 2\pi s,\ s \in \mathbb{Z}. \end{cases} \tag{25}$$

When $|\lambda_p| \ll 1$, $\nu_p$ is very close to its limit for $L \to \infty$, for a relatively small value of $L$ when considering numerically dissipative eigenvalues. This can be seen in Figures 1(a) and 5(a) where the amplification factors for the highest real resolvable wavenumber components are approaching their limit for $L \to \infty$, when $L = 4$. Hence increasing the length of the assimilation window for the Upwind and Lax-Wendroff schemes, will not affect the contribution of the high resolvable real wavenumber components to the analysis vector and its forecast. The amplification factor for the lower resolvable real wavenumber components can be altered by increasing the length of the assimilation window.

In the case of a numerically non-dissipative and dispersive eigenvalue $\lambda_p$, such as those found in the Preissman Box scheme, $\nu_p \to 0$ as $L \to \infty$. This can be seen in Figure 3(a). This leads to $A_L\tilde{\mathbf{x}}_0 \to \mathbf{0}$ as $L \to \infty$. Therefore as the length of the assimilation window is increased, by adding extra sets of observations in time, the contribution of $A_L\tilde{\mathbf{x}}_0$ to $\mathbf{x}_a$ decreases. This shows that as more observations are included, destructive interference increases between the corresponding wavenumber components of each set of observations in time, leading to a loss of information in $\mathbf{x}_a$ and its subsequent forecast. Hence for a solely numerically dispersive scheme, increasing the number of observations does not necessarily improve the accuracy of the analysis vector and its forecast.

## 4. Error analysis

Numerical model error can be measured through the direct error on the analysis vector or by its effect on the subsequent forecast. Both quantities are important in different applications of the inverse problem. These errors can be shown to converge to zero for sufficiently smooth initial conditions, when measured in the $l_2$-norm, by considering the global and truncation errors associated with the forward model of each scheme [31]. To investigate these errors for any regularity initial condition, a spectral approach is taken, using the formulation for the analysis vector found in Lemma 2.

### *4.1. Spectral approach in the absence of observation errors*

A spectral approach can be used to provide a bound for the $l_2$-norm of the error in the analysis vector, for any regularity initial condition. Lemma 3 derives such a bound making use of the results from Lemmas 1 and 2. Here we remind the reader that the *regularity* of $u_0(x)$ over $(0, 1)$ is defined in Definition 1.

**Lemma 3.** *Let the assumptions of Lemma 2 hold true. Also let $u_0(x)$ have regularity $r \in \mathbb{N}_0$ over $(0, 1)$, be piecewise monotone over $(0, 1)$ and $u_0^{(r)}(x)$ be bounded and piecewise monotone over $(0, 1)$. Then,*

$$\|\tilde{\mathbf{x}}_0 - \mathbf{x}_a\|_2^2 \leq N\left\{|1-\nu_1|D_1 + (|1-\nu_1| + 2\xi_1)\frac{D_3}{N^{r+1}}\right\}^2 + 2N\sum_{p=2}^{\frac{N+1}{2}}\left\{|1-\nu_p|\frac{D_2}{(p-1)^{r+1}} + (|1-\nu_p| + 2\xi_p)\frac{D_3}{N^{r+1}}\right\}^2, \quad (26)$$

*where $D_1 := \nu_1 \in \mathbb{R}$ is the bound on $u_0(x)$ over $(0, 1)$, $D_2 := \frac{4\nu_2 sT^r}{(2\pi)^{r+1}}$ where $\nu_2 \in \mathbb{R}$ is the bound on $u_0^{(r)}(x)$ over $(0, 1)$ and $s \in \mathbb{N}$ is the number of monotone pieces $u_0^{(r)}(x)$ can be broken up into over $(0, 1)$ and*

$$D_3 := \begin{cases} D_2\left[4 + 2\zeta(2)\right] + 2\nu_1 w, & \text{for } r = 0, \\ D_2\left[2^{r+1} + 2\zeta(r+1)\right], & \text{for } r \in \mathbb{N}, \end{cases} \quad (27)$$

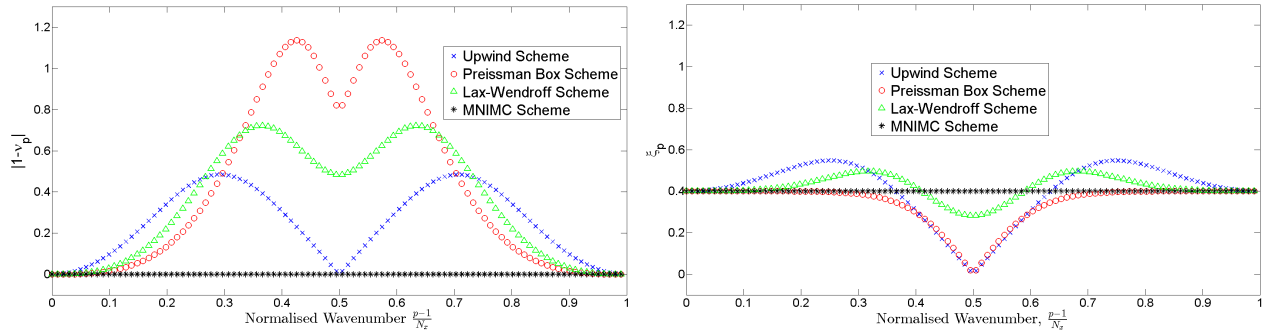
*such that $w \in \mathbb{N}$ is the number of sub-domains $[x_j, x_{j+1}]$ for $j = 0, \ldots, N-1$, where $u_0(x)$ contains a discontinuity. Here $\zeta(\cdot)$ denotes the* Riemann Zeta function*. Also define,*

$$\xi_p := \frac{\left|\sum_{l=0}^{\frac{L-[L]_b}{b}-1}\left[|\lambda_p|^b e^{ib\phi_p}\right]^l\right|\left\{\sum_{j=1}^{b-1}|\lambda_p|^j\right\} + |\lambda_p|^{L-[L]_b}\sum_{j=1}^{[L]_b}|\lambda_p|^j}{\sum_{k=0}^{L}|\lambda_p|^{2k}}. \quad (28)$$

The proof of Lemma 3 can be found in [31]. This bound can be used to analyse the order of convergence of the error to zero, with respect to either $N$ the number of discretisation points, or $L$ the number of sets of observations. Examining (26) indicates that the order of convergence of the bound will be explicitly dependent on the regularity of the initial condition, given by $r$. It should be noted here that in the following experiments, the order of convergence with respect to either $N$ or $L$ is found for constant $h$, resulting in $\Delta t = h/(|\mu|N)$ varying with $N$. Consequently, the length of the assimilation window is altered by varying either $N$ or $L$, $T := L\Delta t = Lh/(|\mu|N)$. As each set of observations in time contains observations at every spatial location, increasing $N$ increases the density of observations in space and time. Increasing $L$ lengthens the assimilation window by adding more sets of observations in time, keeping their density constant.

In the case of the Upwind, Preissman Box and Lax-Wendroff schemes, $\nu_1 = 1$. Hence the terms relating to $\nu_1$ in the bound in Equation (26), are zero. In the case of the MNIMC scheme, $A_L = I$, so the only contribution to the error is from aliasing errors. The bound is consistent with this as only the bound on the aliasing error remains. Choosing $h = 1$ results in both the error and its bound becoming zero.

The order of convergence of the bound with respect to either $N$ or $L$, is in part determined by the order of convergence of the $|1 - \nu_p|$ and $\xi_p$ terms. The term $|1 - \nu_p|$ has a direct impact on the analysis vector, whilst $\xi_p$ is a consequence of applying a bound to the error in the analysis vector. The terms $|1 - \nu_p|$ and $\xi_p$ are both dependent on $N$ and $L$; $N$ determines the number of points, whilst $L$ determines the shape, of the plots in Figure 7.




(a) $|1 - \nu_p|$ (b) $\xi_p$

Figure 7: The values of $|1 - \nu_p|$ and $\xi_p$ plotted against the corresponding normalised wavenumber ie: $(p-1)/N$, of the corresponding eigenvalue, for the Upwind, Preissman Box, Lax-Wendroff and MNIMC schemes, for $N = 101$ and $L = 4$ ($\Delta t = 1/202$).

The number of mesh points and observations used in NWP are typically $O(10^7)$ [32] and $O(10^5 - 10^6)$ [21] respectively. As a result, it is realistic to consider the order of convergence of the bound in (26) when $L$ is small in comparison to $N$ (ie: a small assimilation window).

The order of convergence of $|1 - \nu_p|$ to zero with respect to $N$, for fixed $p$, was found numerically using fixed $L = 4$. The order was found to be less than or equal to zero for all $p$ and the order remained constant for small $p$ ie: $(p-1)/N \ll 1$, where $|1 - \nu_p| = O(N^{-2})$ for the Upwind scheme and $|1 - \nu_p| = O(N^{-3})$ for the Preissman Box and Lax-Wendroff schemes, for such $p$, $p \neq 1$. As a result, $|1 - \nu_p|$ is either decaying to zero or remaining constant as $N$ increases.

Similarly, the order of convergence of $|1 - \nu_p|$ to zero with respect to $L$, for fixed $p$, was found numerically using fixed $N = 3^7$. This was found to be positive for all $p$ and at most $|1 - \nu_p| = O(L)$ for all three schemes, for each $p$, $p \neq 1$. These results show that increasing the value of $L$ causes the value of $|1 - \nu_p|$ to diverge from zero for $p \neq 1$. The identical order of convergence with respect to $L$ for each scheme ($p \neq 1$) is not surprising as the dependence of $|1 - \nu_p|$ on $L$ is similar for each scheme, unlike the dependence on $N$.

For the Upwind scheme, the numerical order of convergence for $|1 - \nu_p|$ to zero, with respect to both $N$ and $L$, can be explained through its asymptotic expansion as $N \to \infty$, for fixed $p$. Assuming $L > 0$, let $z := (p-1)/N$. Then, for fixed $p$, as $N \to \infty$, $z \to 0$, so we consider $z$ as a continuous variable and Taylor expand $|1 - \nu(z)|$ about $z = 0$, resulting in,

$$|1 - \nu(z)| = \frac{\pi^2 L}{4} z^2 + O(z^4), \quad \text{for } 0 < z < \frac{1}{2},$$

as the 4th derivative of $|1 - \nu(z)|$ with respect to $z$ is bounded over the interval $(0, 0.5)$. Considering $z = (p-1)/N$ for $p = 2, \ldots, (N-1)/2$ as $N \to \infty$, we obtain,

$$|1 - \nu_p| \sim \frac{\pi^2 L}{4} \left( \frac{p-1}{N} \right)^2. \tag{29}$$

Therefore we trial the use of Equation (29) as an approximation for $|1 - \nu_p|$. This expansion indicates that $|1 - \nu_p|$ has orders of convergence $O(N^{-2})$ and $O(L)$ for the Upwind scheme. These match the numerical orders of convergence found for $|1 - \nu_p|$ to zero with respect to $N$ when $p$ is small ($p \neq 1$) and the maximum order of convergence with respect to $L$, for the Upwind scheme.

As $|1 - \nu_1| = 0$ for the Upwind, Preissman Box and Lax-Wendroff schemes, the bound in (26) can be considered as the sum of six distinct summations. Each summation has an order of convergence to zero with respect to $N$ and $L$, which influences the overall order of convergence for the bound. The order of convergence of each summation was identified numerically, in order to identify the dominant summation in the bound.

The summation with the dominant order of convergence for each considered scheme and regularity initial condi-

| $r$ | $\alpha$ | | | $\beta$ | | |
|---|---|---|---|---|---|---|
| | Upwind | Preissman Box | Lax-Wendroff | Upwind | Preissman Box | Lax-Wendroff |
| 0 | $-6.7708 \times 10^{-15}$ | $-2.6329 \times 10^{-10}$ | $-3.7260 \times 10^{-9}$ | $5.7945 \times 10^{-1}$ | $3.6188 \times 10^{-1}$ | $4.0297 \times 10^{-1}$ |
| 1 | −2.0000 | −2.0000 | −2.0000 | 1.5053 | $9.394 \times 10^{-1}$ | 1.0230 |
| 2 | −3.0000 | −4.0000 | −4.0232 | 1.9957 | 1.6588 | 1.6731 |
| 3 | −3.0000 | −4.9565 | −5.0600 | 2.0000 | 1.9990 | 1.9955 |
| 4 | −3.0000 | −4.9377 | −5.0000 | 2.0000 | 2.0000 | 1.9999 |
| 5 | −3.0000 | −4.9345 | −5.0000 | 2.0000 | 2.0000 | 2.0000 |
| 6 | −3.0000 | −4.9338 | −5.0000 | 2.0000 | 2.0000 | 2.0000 |
| 7 | −3.0000 | −4.9336 | −5.0000 | 2.0000 | 2.0000 | 2.0000 |
| $r \gg 1$ | −3.0000 | −5.0000 | −5.0000 | 2.0000 | 2.0000 | 2.0000 |

Table 1: Numerical orders of convergence to zero, with respect to $N$ and $L$, for (30), $O(N^\alpha L^\beta)$, using the Upwind, Preissman Box and Lax-Wendroff schemes, given to 5sf (significant figures), with $h = 0.5$ a constant. The results for $N$ were identified using fixed $L = 4$ ($\Delta t = 1/2N$) and larger values of $N$ than those used to produce the results for $\alpha$ in Table 2. As a result, the values of $\alpha$ displayed here are likely to have a greater accuracy than those displayed in Table 2. The results for $L$ were identified using fixed $N = 3^7$ ($\Delta t = 1/(2 \cdot 3^7)$). The results for $r \gg 1$ were identified using (31).

tion, with respect to both $N$ and $L$ was found to be,

$$N \sum_{p=2}^{\frac{N+1}{2}} \frac{|1 - \nu_p|^2}{(p-1)^{2(r+1)}}. \tag{30}$$

This summation is composed from the amplification factors and a bound on the continuous Fourier coefficients of $u_0(x)$ ie: $c_p$ [27, 31, 33], giving it an explicit dependence on the regularity of $u_0(x)$.

In the case of an initial condition where $r$ is infinite, such as for a Gaussian function initial condition, the bound on the coefficients $c_p$ decays faster than any finite power of $p$ as $p \to \pm\infty$ [34]. Boyd [34] states that it is not appropriate to consider the bound on $c_p$ whilst taking the limit as $r \to \infty$, as the bound is designed to consider $c_p$ as $p \to \pm\infty$ for fixed $r$. If we were to consider the limit $r \to \infty$, then for any $p \neq 0$, $\frac{D_2}{|p|^{r+1}} \to 0$ as $\frac{1}{2\pi} < |p|$. This implies that $c_p = 0$ for all $p \neq 0$, which is not true. Since this bound is used to construct the bound in (26), it is not appropriate to consider this bound in the infinite limit of $r \to \infty$. Instead we will consider large $r$ for this case.

**Corollary 1.** *In the case of large r,*

$$N \sum_{p=2}^{\frac{N+1}{2}} \frac{|1 - \nu_p|^2}{(p-1)^{2(r+1)}} \sim N|1 - \nu_2|^2, \quad as\ N \to \infty. \tag{31}$$

This is due to the rapid decay of $(p-1)^{-2(r+1)}$ to zero for large $r$, when $p = 3, \ldots, \frac{N+1}{2}$, as $N \to \infty$. When $p = 2$, $(p-1)^{-2(r+1)}$ remains constant for any value of $r$, hence the term $|1 - \nu_2|^2$ determines the behaviour of (30) for large $r$. In this instance,

$$N|1 - \nu_2|^2 = O(N^{1+2\gamma}), \tag{32}$$

where $\gamma$ is the numerical order of convergence for $|1 - \nu_p|$ to zero with respect to $N$, for small $p$, $p \neq 1$, eg. for the Upwind scheme $\gamma = -2$. The orders of convergence to zero for the summation in (30), with respect to both $N$ and $L$, are given in Table 1 for varying $r$. When $r$ is large, the order of convergence of (31) to zero is considered.

Table 1 shows that the order of convergence of (30) to zero with respect to $N$, for an initial condition such that $r = 0$, is $O(N^0)$. This indicates that the error in the analysis vector does not decay as $N$ is increased. This is due to the error that always exists when a Fourier series is used to approximate a discontinuous function.

Expression (30) decays to zero for initial conditions where $r > 0$, as $N$ increases. As the regularity is increased, the order of convergence to zero with respect to $N$ is initially $O(N^{-2r})$. However, once a critical regularity is achieved the order of convergence saturates; $O(N^{-3})$ for the Upwind scheme when $r \geq 2$ and $O(N^{-5})$ for the Preissman Box and Lax-Wendroff schemes when $r \geq 3$. The orders of convergence at saturation point match the orders of convergence given by (32), when considering large $r$.

| Variable | Upwind Scheme | | |
|---|---|---|---|
| | Square Function ($r = 0$) | Triangular Function ($r = 1$) | Gaussian Function ($r \gg 1$) |
| $\alpha$ | $1.1838 \times 10^{-12}$ | $-2.2612$ | $-3.0000$ |
| $\beta$ | $5.6939 \times 10^{-1}$ | 1.5096 | 2.0000 |
| Variable | Preissman Box Scheme | | |
| | Square Function ($r = 0$) | Triangular Function ($r = 1$) | Gaussian Function ($r \gg 1$) |
| $\alpha$ | $-6.5427 \times 10^{-1}$ | $-1.2809$ | $-4.9178$ |
| $\beta$ | $3.7952 \times 10^{-1}$ | $9.8836 \times 10^{-1}$ | 2.0662 |
| Variable | Lax-Wendroff Scheme | | |
| | Square Function ($r = 0$) | Triangular Function ($r = 1$) | Gaussian Function ($r \gg 1$) |
| $\alpha$ | $5.5724 \times 10^{-1}$ | $-2.0836$ | $-4.9947$ |
| $\beta$ | $3.1248 \times 10^{-1}$ | 1.0187 | 2.0194 |

Table 2: Numerical orders of convergence to zero, with respect to $N$ and $L$, for the error in the analysis vector from strong constraint 4D-Var numerical experiments, given to 5sf, $\|\tilde{\mathbf{x}}_0 - \mathbf{x}_a\|_2^2 = O(N^\alpha L^\beta)$, with $h = 0.5$ a constant. The results for $N$ and $L$ were identified using fixed $L = 4$ ($\Delta t = 1/2N$) and fixed $N = 3^7$ ($\Delta t = 1/(2 \cdot 3^7)$), respectively.

The numerical results for $N$, for the Upwind scheme in Table 1, are seen when the right-hand side of Equation (29) is substituted into (30) to replace $|1 - \nu_p|$,

$$N \sum_{p=2}^{\frac{N+1}{2}} \frac{|1 - \nu_p|^2}{(p-1)^{2(r+1)}} = \begin{cases} O(N^{-2r}), & \text{for } r = 0, 1, \\ O(N^{-3}), & \text{for } r \geq 2. \end{cases} \tag{33}$$

Table 1 shows that the order of convergence of (30) to zero, with respect to $L$, is positive for all values of $r$. This indicates that the error in the analysis vector may increase as the length of the assimilation window is increased. The order of convergence also increases with the value of $r$ associated with the initial condition, until a critical value is reached, where the order of convergence saturates at $O(L^2)$.

Numerical experiments were performed to investigate whether (26) was a good indicator for the growth and decay of the error in the analysis vector. Strong constraint 4D-Var numerical experiments were performed using the same conditions as in the above analysis, for the same finite difference schemes. The error $\|\tilde{\mathbf{x}}_0 - \mathbf{x}_a\|_2^2$ was then determined numerically and its order of convergence to zero with respect to $N$ and $L$ was found. The functions and their value of $r$ chosen for $\tilde{\mathbf{x}}_0$ were the square function ($r = 0$), a triangular function ($r = 1$) and a Gaussian function, $\mathcal{N}(0.5, 0.01)$ ($r \gg 1$).

The numerical experiments were executed using the built-in PCG method in MATLAB®[35], a zero first guess and a tolerance of $10^{-10}$ on the relative residual, to minimise the cost function. For comparison with the results in Table 1, the same fixed values of $N$ and $L$ were chosen.

Initially consider the results for the order of convergence with respect to $N$ in Table 2. The results are close to those in Table 1 for initial conditions with the same value of $r$. Figure 8(a) plots the numerical order of convergence with respect to $N$, as $N$ is increased in powers of three. It shows that the order of convergence fluctuates about the order of convergence shown in Table 1, for each value of $r$. This probably explains why the results in Tables 1 and 2 do not match exactly for $N$. Table 2 also shows that the order of convergence with respect to $L$ is a good match to those found in Table 1. These results indicate that (26) is an appropriate bound for the considered error in the analysis vector, with respect to $N$ and $L$. Taking observations every $\Delta x$ in space, every $\Delta t$ in time, combined with the initial conditions chosen for the numerical experiments in this paper mean that the full set of properties of each considered $u_0(x)$ can be observed through its discrete sample in $\tilde{\mathbf{x}}_0$. When this is not the case, it has been shown that the bound in Equation (26) gives the worst case behaviour of the error in the analysis vector [31].

The next step is to re-introduce observation errors and understand how numerical model error and observation errors interact.

*4.2. Spectral approach with observation errors*

Section 4.1 provides a bound for the numerical model error in the analysis vector, for differing regularity initial conditions $u_0(x)$, in problem (4). It is possible to apply a similar bound to the error in the analysis vector when observation errors are also included.

Consider the case where each observation contains observation errors, $\mathbf{y}_l = \tilde{\mathbf{y}}_l + \boldsymbol{\epsilon}_l$, as described in Section 2. Specifically, let us consider the random error known as *white noise* [30] where $\boldsymbol{\epsilon}_l \sim \mathcal{N}(\mathbf{0}, I_N)$. The cost function becomes,

$$J(\mathbf{x}_0) = \frac{1}{\sigma_o^2} \sum_{l=0}^{L} \left[\mathbf{y}_l - M^l \mathbf{x}_0\right]^T \left[\mathbf{y}_l - M^l \mathbf{x}_0\right]. \tag{34}$$

Minimising (34) with respect to $\mathbf{x}_0$,

$$\mathbf{x}_a = \left[\sum_{k=0}^{L} (M^T M)^k\right]^{-1} \sum_{l=0}^{L} (M^T)^l \left[\tilde{\mathbf{y}}_l + \boldsymbol{\epsilon}_l\right]. \tag{35}$$

Using the eigenvalue decomposition of $M$ and $\tilde{M}$ as well as (19),

$$\mathbf{x}_a = A_L \tilde{\mathbf{x}}_0 + \boldsymbol{\rho}_L + V \left[\sum_{k=0}^{L} (\Lambda^* \Lambda)^k\right]^{-1} \left[\sum_{l=0}^{L} (\Lambda^*)^l V^* \boldsymbol{\epsilon}_l\right]. \tag{36}$$

The analysis vector in the presence of observation errors, is expressed in part by $\mathbf{x}_a$ without observation errors, as in (19). The observation errors form a separate term. If the errors did not possess the same variance, (36) would not have this property.

The term containing the observation errors in (36) would be the analysis vector when considering observations of the form $\mathbf{y}_l = \boldsymbol{\epsilon}_l$. The effect of numerical model error on the white noise may lead to correlations within the observation noise component of (36). If correlations have been introduced, then this will create artifacts in the analysis vector which will be propagated into its forecast. The autocorrelation function is used to determine if the observation noise contribution to $\mathbf{x}_a$ is still white noise.

The autocorrelation function is defined as in Mitra [30]. The autocorrelation of an $N$-periodic sample $\mathbf{x} \in \mathbb{R}^N$, at lag $j-1$ for $j = 1, \ldots, N$, is defined as $z_{j-1} : \mathbb{R}^N \to \mathbb{R}$, such that $\mathbf{x} \mapsto z_{j-1}(\mathbf{x})$ where,

$$z_{j-1}(\mathbf{x}) = \frac{1}{N} \sum_{p=1}^{N} \{\mathbf{x}\}_p \{\mathbf{x}\}_{[p-j+1]_N}, \tag{37}$$

where $\{\mathbf{x}\}_p$ denotes the $p$th element of $\mathbf{x}$ and $[\cdot]_N$ denotes modulo $N$. Also, define $\mathbf{z} \in \mathbb{R}^N$ such that the $j$th element of $\mathbf{z}(\cdot)$ is $z_{j-1}(\cdot)$. Then by the Wiener-Khintchine Theorem [30], the DFT of the autocorrelation of $\mathbf{x}$ is defined as,

$$\mathcal{F}[\mathbf{y}(\mathbf{x})] = \frac{1}{\sqrt{N}} \left[|\mathcal{F}_1(\mathbf{x})|^2, |\mathcal{F}_2(\mathbf{x})|^2, \ldots, |\mathcal{F}_N(\mathbf{x})|^2\right]^T.$$

Using (36), the autocorrelation of the noise component of the analysis vector is given by,

$$z_{j-1}\left(V \left[\sum_{k=0}^{L} (\Lambda^* \Lambda)^k\right]^{-1} \sum_{l=0}^{L} (\Lambda^*)^l V^* \boldsymbol{\epsilon}_l\right) = \frac{1}{N} \sum_{p=1}^{N} \left| \frac{\sum_{l=0}^{L} \overline{\lambda}_p^l \mathcal{F}_p(\boldsymbol{\epsilon}_l)}{\sum_{k=0}^{L} |\lambda_p|^{2k}} \right|^2 e^{\frac{2\pi i (j-1)(p-1)}{N}}, \tag{38}$$

for $j = 1, \ldots, N$. Hence [31],

$$\mathbb{E}\left[z_{j-1}\left(V \left[\sum_{k=0}^{L} (\Lambda^* \Lambda)^k\right]^{-1} \sum_{l=0}^{L} (\Lambda^*)^l V^* \boldsymbol{\epsilon}_l\right)\right] = \frac{\sigma_o^2}{N} \sum_{p=1}^{N} \frac{e^{\frac{2\pi i (j-1)(p-1)}{N}}}{\sum_{k=0}^{L} |\lambda_p|^{2k}}, \tag{39}$$

for all $j = 1, \ldots, N$, which relies upon the values of $j$, $N$, $L$ and $\sigma_o^2$, together with the dissipative properties of the considered finite difference scheme. It does not utilise the dispersive properties of the scheme. Expression (39) is potentially non-zero for all $j$, for a numerically dissipative finite difference scheme, indicating that the noise component of the analysis vector may no longer be random white noise. However, in the case of a non-dissipative scheme ie: $|\lambda_p| = 1$ for all $p$, only $j = 1$ is non-zero. Using a non-dissipative scheme such as the Preissman Box scheme, means that the noise component of the analysis vector will retain the white noise structure implicit in the observations.

A spectral approach as in Section 4.1 is now used to provide a bound for the $l_2$-norm of the error in the analysis vector, for any regularity initial condition, in the presence of observation errors.

**Lemma 4.** *Let the assumptions of Lemma 3 hold true, but consider observations of the form* $\mathbf{y}_l := \tilde{\mathbf{y}}_l + \boldsymbol{\epsilon}_l$*, allowing* $\mathbf{x}_a$ *to be stated as in (36). Then,*

$$\|\tilde{\mathbf{x}}_0 - \mathbf{x}_a\|_2^2 \leq E_M + E_O + E_C, \tag{40}$$

*where*

$$E_M = N\left\{|1-\nu_1|D_1 + (|1-\nu_1| + 2\xi_1)\frac{D_3}{N^{r+1}}\right\}^2 + 2N\sum_{p=2}^{\frac{N+1}{2}}\left\{|1-\nu_p|\frac{D_2}{(p-1)^{r+1}} + (|1-\nu_p| + 2\xi_p)\frac{D_3}{N^{r+1}}\right\}^2, \tag{41}$$

$$E_O = Nz_0\left(\left[\sum_{k=0}^{L}(\Lambda^*\Lambda)^k\right]^{-1}\sum_{l=0}^{L}(\Lambda^*)^l V^*\boldsymbol{\epsilon}_l\right), \tag{42}$$

$$\begin{aligned} E_C = {} & 2\sqrt{N}\left\{|1-\nu_1|D_1 + (|1-\nu_1| + 2\xi_1)\frac{D_3}{N^{r+1}}\right\}\left|\frac{\sum_{l=0}^{L}\overline{\lambda}_1^l\mathcal{F}_1(\boldsymbol{\epsilon}_l)}{\sum_{k=0}^{L}|\lambda_1|^{2k}}\right| \\ & +4\sqrt{N}\sum_{p=2}^{\frac{N+1}{2}}\left\{|1-\nu_p|\frac{D_2}{(p-1)^{r+1}} + (|1-\nu_p| + 2\xi_p)\frac{D_3}{N^{r+1}}\right\}\left|\frac{\sum_{l=0}^{L}\overline{\lambda}_p^l\mathcal{F}_p(\boldsymbol{\epsilon}_l)}{\sum_{k=0}^{L}|\lambda_p|^{2k}}\right|, \end{aligned} \tag{43}$$

*where* $D_1$ *is a constant independent of p, N and r and* $D_2$ *and* $D_3$ *are constants independent of p and N but dependent on r. Constants* $D_1$*,* $D_2$ *and* $D_3$ *are defined in Lemma 3 and* $\xi_p$ *is defined as in Equation (28).*

The proof of Lemma 4 can be found in [31]. The bound is formed from the equivalent bound in the absence of observation errors ($E_M$), together with the autocorrelation at lag 0 of the noise component in the analysis vector ($E_O$) and cross terms ($E_C$). It can be used to analyse the order of convergence of the error to zero, with respect to either $N$ or $L$.

The variables $E_O$ and $E_C$ are dependent on the random variables $\{\boldsymbol{\epsilon}_l\}_{l=0}^{L}$. However by the strong law of large numbers [36], if the experiments could be repeated, then as the number of experiments is increased, the sample means of $E_O$ and $E_C$ would tend toward their *expected values*. As a consequence, we consider the expected values of $E_O$ and $E_C$,

$$\mathbb{E}[E_O] = \sigma_o^2\sum_{p=1}^{N}\frac{1}{\sum_{k=0}^{L}|\lambda_p|^{2k}}, \quad \text{and} \quad \mathbb{E}[E_C] = 0, \tag{44}$$

in order to identify the orders of convergence of the bound with respect to both $N$ and $L$.

The expected value of $E_C$ is zero whilst the expected value of $E_O$ is dependent upon $N$, $L$, $\sigma_o^2$ and the dissipative properties of the scheme. Hence the expected value is independent of both the regularity of the initial condition and the dispersive properties of the finite difference scheme. A non-dissipative scheme leads to $\mathbb{E}[E_O] = \sigma_o^2 N/(L+1)$, so that the order of convergence for $\mathbb{E}[E_O]$ to zero with respect to $N$ and $L$, is $O(N)$ and $O(L^{-1})$ respectively.

The order of convergence of the bound in (40) to zero, with respect to $N$ or $L$, is determined by the dominant order of convergence possessed by either $E_M$ or $\mathbb{E}[E_0]$. The orders of convergence to zero for $E_M$ were analysed in Section 4.1. Table 3 displays the numerical orders of convergence to zero, with respect to $N$ and $L$, for $\mathbb{E}[E_O]$.

| Variable | Upwind | Preissman Box | Lax-Wendroff |
|---|---|---|---|
| $\alpha$ | 1.0000 | 1.0000 | 1.0000 |
| $\beta$ | $-3.3207 \times 10^{-4}$ | $-9.9719 \times 10^{-1}$ | $-2.0866 \times 10^{-3}$ |

Table 3: Numerical orders of convergence to zero, with respect to $N$ and $L$, for $\mathbb{E}[E_O]$ in (42), given to 5sf, $\mathbb{E}[E_0] = O(N^\alpha L^\beta)$, with $h = 0.5$ a constant. The results for $N$ and $L$ were identified using fixed $L = 4$ ($\Delta t = 1/2N$) and fixed $N = 3^7$ ($\Delta t = 1/(2 \cdot 3^7)$), respectively.

Initially consider the order of convergence of the bound in (40) with respect to $N$. The results of Section 4.1 show that $E_M$ remains constant or decays to zero, whilst Table 3 shows that $\mathbb{E}[E_O]$ increases, as $N$ is increased. Next we consider the order of convergence of each variable to zero, with respect to $L$. $E_M$ increases, and $\mathbb{E}[E_O]$ decreases, as $L$ is increased. Subsequently, the dominant order of convergence of (40) to zero, for both $N$ and $L$, will be determined by the order of magnitude of the coefficients of each term.

Strong constraint 4D-Var experiments similar to those in Section 4.1 were run to investigate the appropriateness of (40) as a bound for the error in the analysis vector in the presence of observation errors. The results can be seen in Figures 9(a) and 9(b), for $N$ and $L$ respectively. Numerical results were generated using the same setup as for the strong constraint 4D-Var experiments in Section 4.1.

Figure 9(a) shows that initially the error in the analysis vector decreases as $N$ increases, according to the order of convergence seen for $E_M$. Once a critical value of $N$ has been reached, the order of convergence then begins to increase with the order of convergence demonstrated by $\mathbb{E}[E_O]$. This provides a critical value for $N$ at which the effect of both numerical model error and observation errors on the accuracy of the analysis vector is minimised. $L$ and $\sigma_o^2$ form part of the coefficient for $N$ in $\mathbb{E}[E_O]$. Increasing $L$ or decreasing $\sigma_o^2$ will result in the critical value of $N$ increasing, whilst decreasing $L$ or increasing $\sigma_o^2$ will result in the critical value of $N$ decreasing. The critical value for $N$ shown in Figure 9(a) is between $3^4$ and $3^5$, depending on the chosen finite difference scheme.

Figure 9(b) shows a similar picture to that seen in Figure 9(a). However in this instance, the initial decrease in the error in the analysis vector corresponds to the order of convergence of $\mathbb{E}[E_O]$. As $L$ is increased further, a critical value is reached where $E_M$ becomes dominant over $\mathbb{E}[E_O]$, and the error begins to increase with $L$. As with $N$, this critical value of $L$ is determined by the coefficients of $\mathbb{E}[E_O]$. Decreasing either $N$ or $\sigma_o^2$ will result in the critical value of $L$ decreasing, whilst increasing either $N$ or $\sigma_o^2$ will result in the critical value of $L$ increasing.

When considering the orders of convergence with respect to either $N$ or $L$, reducing $\sigma_o^2$ corresponds to reducing the error in the observations. As a result, it is not surprising that reducing $\sigma_o^2$ results in $E_M$ becoming the dominant order of convergence, for a given value of $N$ or $L$.

This analysis suggests that (40) is an appropriate bound to demonstrate the order of convergence for the error in the analysis vector. As a result, given a fixed value for $\sigma_o^2$ and either $N$ or $L$, it is possible to choose a value for $L$ and $N$ respectively, that minimises the error in the analysis vector due to numerical model error and observation errors. The result for $N$ in some way works towards answering the question posed by Akella et al. [15] as to whether increasing the number of discretisation points would continue to decrease the effects of discretisation errors on the results of 4D-Var. In this instance, we have shown that when considering numerical model and observation errors in strong constraint 4D-Var, increasing the number of discretisation points past the optimal value of $N$ when considering a full sets of observations, would result in an increase in the error in the analysis vector.

## 5. Discussion

The results in this paper are for the chosen representative finite difference schemes; the Upwind, Preissman Box and Lax-Wendroff schemes. However the theory in the paper is applicable to any finite difference scheme that can be implemented to solve the 1D linear advection problem in a similar way. Lemma 3 can be used to identify the theoretical order of convergence for the numerical model error in the analysis vector, once an asymptotic expansion of $|1 - \nu_p|$ has been found for the new scheme. The orders of convergence for $\mathbb{E}(E_O)$ and $\mathbb{E}(E_C)$ are discussed in Section 4.2 in terms of the numerically dissipative and dispersive properties of the finite difference scheme and hence are applicable to an alternative scheme with these properties. If an optimal value of $N$ at which to perform strong constraint 4D-Var exists to minimise the effects of both numerical model and observations errors, this would require further numerical experiments to identify.

The analysis conducted in Section 3 into the numerically dissipative and dispersive properties of the considered finite difference schemes is important for aiding our understanding into the impact of different forms of numerical error on the quality of the analysis vector seen in Section 4. Williams [37] conducted a similar analysis for the Leapfrog scheme to understand the "spurious computational modes" which occur in it. Through this analysis, he was able to design a modification for the Robert-Asselin time filter to improve the results from meteorological models [38]. It is envisioned that the results of Sections 3 and 4 can be used in a similar way to design modifications to numerically dissipative and/or dispersive schemes, to improve the accuracy of the analysis vector. Williams [37] investigates the Leapfrog scheme due to its wide-scale use in weather and climate models eg. [8], where it is chosen for being a simple and computationally inexpensive scheme [37].

The results presented in this paper are for a linear problem. Most practical applications of 4D-Var are for non-linear problems, so it is important to address the question of how these results relate to non-linear problems in 4D-Var. The analysis conducted in this paper can be applied to quasi-linear PDEs such as Burgers' equation. Pfeffer et al. [8] performed a similar linear analysis to that of this paper, on the Matsuno and Leapfrog schemes, before using them to solve the non-linear model equations of the NASA-GLAS climate model. The aim was to investigate the sensitivity of the models' forecast to the choice of time-differencing scheme used to produce the model results. Their results showed that some of the properties of the schemes found through this analysis could be seen in the results of the non-linear problem. This leads us to believe that the results of the linear analysis presented here, are to some extent, directly relevant to non-linear problems.

Linear problems such as the one considered here, are of immediate relevance to applications such as incremental 4D-Var data assimilation. This method was developed by Courtier and Hollingsworth [39] to provide a computationally viable implementation of strong constraint 4D-Var for non-linear systems. This iterative method updates the analysis vector through the minimisation of a linearised version of the 4D-Var cost function [40]. This involves linearising $\mathcal{H}_l$ and $\mathcal{M}_{l+1,l}$ about the current iterative state of the numerical model $\mathbf{x}_l^{(k)}$, where $k \in \mathbb{N}_0$ is the iteration number. The linearised numerical model and its transpose are referred to as the *tangent linear model* (TLM) and the *adjoint model* [1, 22]. The resulting cost function to be minimised is then constructed in part from these linearised models. As a result, they have a direct impact on the accuracy of the updates to the analysis vector. Therefore analysing the numerical model error introduced by a linear numerical model in the context of 4D-Var data assimilation, is of direct relevance to the numerical model error introduced by non-linear models in practical applications of strong constraint 4D-Var data assimilation ie: incremental 4D-Var. The 1D linear advection equation is of importance in this context as it can result from the linearisation of the non-linear shallow water equations under certain assumptions [41].

Future work would be to compare the results of Section 3 with those of a non-linear problem solved with these schemes, as in Pfeffer et al. [8]. Current research is being conducted to build upon the work in this paper to investigate if a similar analysis can be applied to a higher dimensional, linear problem in the form of the linearised shallow water equations, a meteorologically relevant problem. This is the next stepping stone in this line of research with the final goal to achieve results for a non-linear problem. The work on the linear problem can also be extended by re-introducing errors associated with 4D-Var such as errors in the model physics and representative errors.

## 6. Conclusion

This paper has detailed the results of a rigorous and quantitative analysis of the errors introduced into the initialisation produced by strong constraint 4D-Var due to finite difference approximations in the numerical model solving the forward problem. This error was initially investigated alone in the absence of all other forms of error. It was found that when using a purely numerically dispersive finite difference scheme, wavenumber components of the analysis vector could be completely lost due to destructive interference. As a consequence these components would not be present in the forecast.

The order of convergence of the error in the analysis vector to zero with respect to the number of discretisation points when considering full sets of observations, was found to depend upon the numerically dissipative and dispersive properties of the numerical scheme, together with the smoothness of the true initial condition we wish to recover. The error remained constant for discontinuous true initial conditions, whilst smoother initial conditions allowed the error to decay. A bound for the error was determined and numerical experiments on this bound, shown in this paper, support these outcomes.

Including observation errors resulted in a bound for the error in the analysis vector. As the number of discretisation points was increased, when considering full sets of observations, the error in the analysis vector decayed due to the decay of the numerical model error. However, past a critical number of grid points, the error began to increase due to observational errors. This suggests there is a critical value of $N$ when considering full sets of observations, at which the effects of numerical model error and observation errors are both minimised. The same trend is seen as the length of the assimilation window is increased, however the initial decay is due to the impact of observation errors decreasing, whilst the subsequent increase is due to the impact of numerical model error increasing.

Observation errors in the form of white noise were found to have the potential to introduce correlated noise structures into the analysis vector, possibly leading to artifacts in the analysis vector and its subsequent forecast. Using a non-dissipative scheme would reduce the presence of these artifacts.

## Acknowledgements

We would like to thank the reviewer for their feedback and comments on the paper. This research was supported by a University of Bath research studentship.

## References

[1] N. K. Nichols, Treating model error in 3-D and 4-D data assimilation, in: R. Swinbank, V. Shutyaev, W. A. Lahoz (Eds.), Data Assimilation for the Earth System, volume 26 of *NATO Science Series*, Springer Netherlands, 2003, pp. 127–135.

[2] R. Daley, Atmospheric Data Analysis, Cambridge Atmospheric and Space Science Series, Cambridge: Cambridge University Press, 1999.

[3] A. C. Lorenc, Analysis methods for numerical weather prediction, Quarterly Journal of the Royal Meteorological Society 112 (1986) 1177–1194.

[4] F. Bouttier, P. Courtier, Data assimilation concepts and methods, Meteorological Training Course Lecture Series (1999). Available from: "`http://msi.ttu.ee/~elken/Assim_concepts.pdf`" [accessed: 24/07/2010].

[5] M. A. Freitag, R. W. Potthast, Synergy of inverse problems and data assimilation techniques, in: M. Cullen, M. A. Freitag, S. Kindermann, R. Scheichl (Eds.), Large Scale Inverse Problems. Computational Methods and Applications in the Earth Sciences, volume 13 of *Radon Series on Computational and Applied Mathematics*, Walter de Gruyter, Berlin, 2013, pp. 1–54.

[6] R. J. Le Veque, Numerical Methods for Conservation Laws, Lectures in Mathematics, Basel: Birkhäuser Verlag, 1999.

[7] K. W. Morton, D. F. Mayers, Numerical solution of partial differential equations, Cambridge University Press, Cambridge, UK, 2nd edition, 2005.

[8] R. L. Pfeffer, I. M. Navon, X. Zou, A comparison of the impact of two time-differencing schemes on the NASA-GLAS climate model, Monthly Weather Review 120 (1992) 1381–1393.

[9] F. X. Le Dimet, V. Shutyaev, On deterministic error analysis in variational data assimilation, Nonlinear Processes in Geophysics 12 (2005) 481–490.

[10] Y. Sasaki, Some basic formalisms in numerical variational analysis, Monthly Weather Review 98 (1970) 875–883.

[11] C. Johnson, N. K. Nichols, B. J. Hoskins, Very large inverse problems in atmosphere and ocean modelling, International Journal for Numerical Methods in Fluids 47 (2005) 759–771.

[12] M. A. Freitag, N. K. Nichols, C. J. Budd, Resolution of sharp fronts in the presence of model error in variational data assimilation, Quarterly Journal of the Royal Meteorological Society 139 (2013) 742–757.

[13] X. Zou, I. M. Navon, F. X. Le Dimet, Incomplete observations and control of gravity waves in variational data assimilation, Tellus A 44 (1992) 273–296.

[14] A. K. Griffith, N. K. Nichols, Adjoint methods in data assimilation for estimating model error, Flow, Turbulence and Combustion 65 (2000) 469–488.

[15] S. Akella, I. M. Navon, Different approaches to model error formulation in 4D-Var: a study with high-resolution advection schemes, Tellus A 61 (2009) 112–128.

[16] P. A. Vidard, E. Blayo, F. X. Le Dimet, A. Piacentini, 4D variational data analysis with imperfect model, Flow, Turbulence and Combustion 65 (2000) 489–504.

[17] Y. Trémolet, Model-error estimation in 4D-Var, Quarterly Journal of the Royal Meteorological Society 133 (2007) 1267–1280.

[18] P. A. Vidard, A. Piacentini, F. X. Le Dimet, Variational data analysis with control of the forecast bias, Tellus A 56 (2004) 177–188.

[19] R. Gerdes, C. Köberle, J. Willebrand, The influence of numerical advection schemes on the results of ocean general circulation models, Climate Dynamics 5 (1991) 211–226.

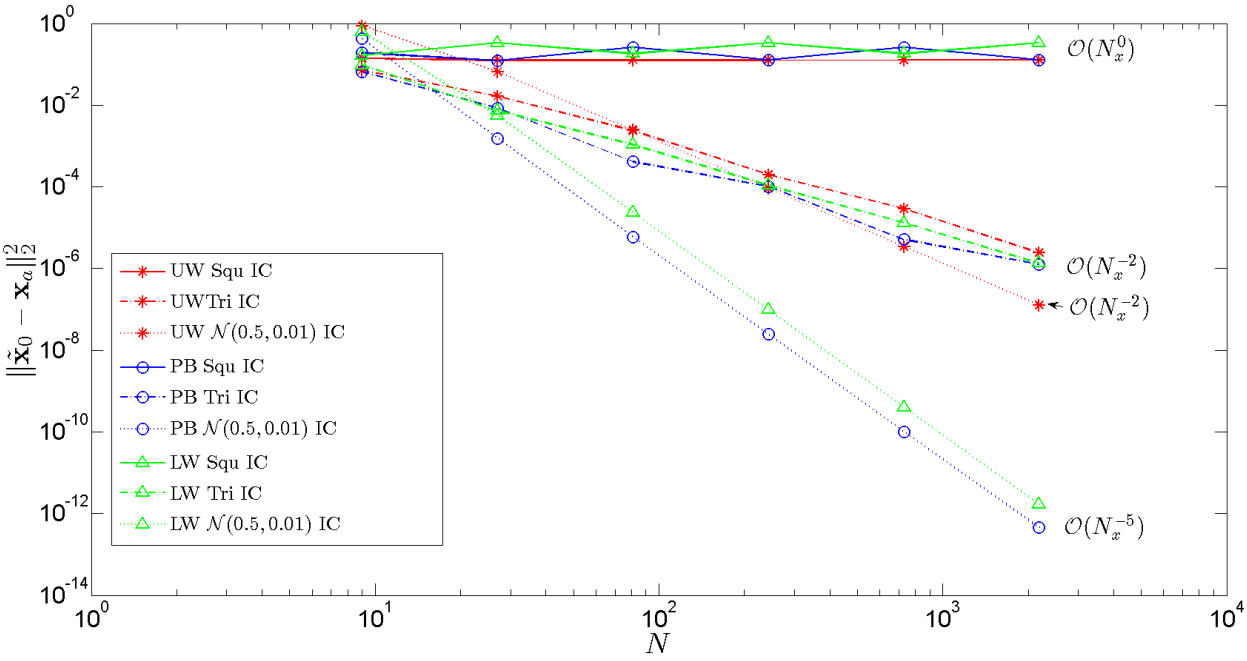
[20] T. Vukićević, M. Steyskal, M. Hecht, Properties of advection algorithms in the context of variational data assimilation, Monthly Weather Review 129 (2001) 1221–1231.

[21] N. K. Nichols, Mathematical concepts of data assimilation, in: W. Lahoz, B. Khattatov, B. Menard (Eds.), Data Assimilation, Making Sense of Observations, Springer Berlin Heidelberg, 2010, pp. 13–39.

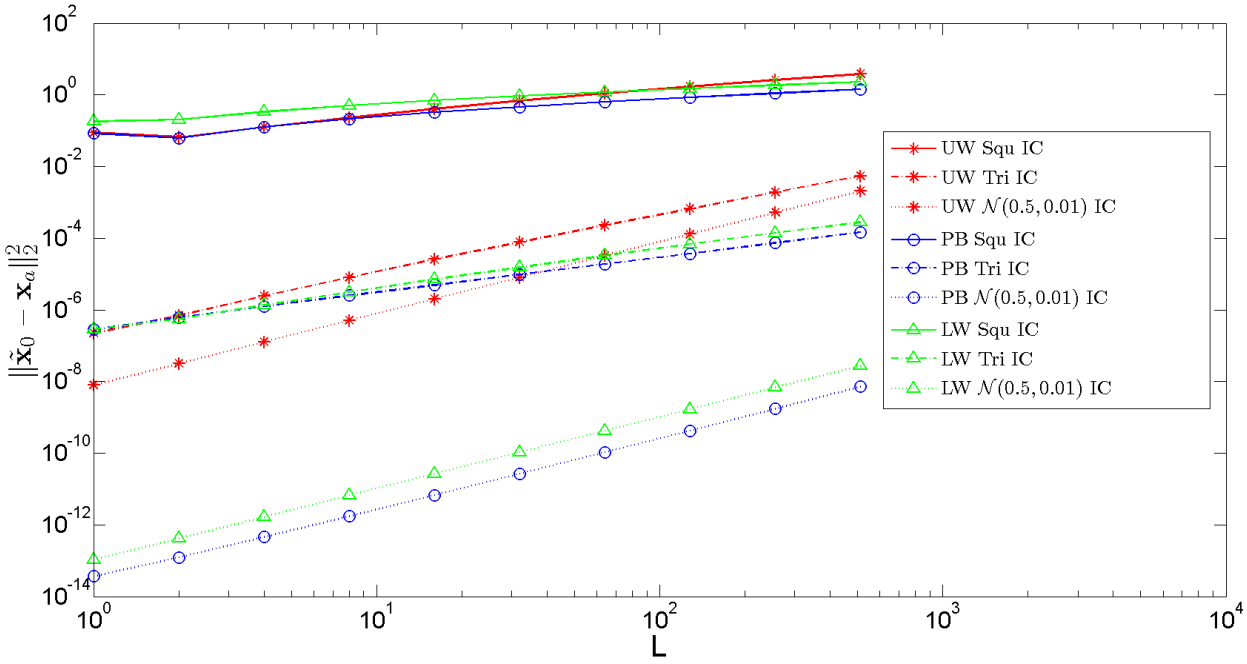
[22] A. S. Lawless, Variational data assimilation for very large environmental problems, in: M. Cullen, M. A. Freitag, S. Kindermann, R. Scheichl (Eds.), Large Scale Inverse Problems. Computational Methods and Applications in the Earth Sciences, volume 13 of *Radon Series on Computional and Applied Mathematics*, Walter de Gruyter, Berlin, 2013, pp. 55–90.

[23] H. O. Kreiss, Initial boundary value problems for hyperbolic systems, Communications on Pure and Applied Mathematics 23 (1970) 277–298.

[24] D. R. Durran, Numerical Methods for Wave Equations in Geophysical Fluid Dynamics, volume 32 of *Texts in Applied Mathematics Series*, New York: Springer New York, 1999.
[25] R. Courant, K. Freidrichs, H. Lewy, On the partial difference equations of mathematical physics, IBM Journal of Research and Development 11 (1967) 215–234.
[26] P. C. Hansen, J. G. Nagy, D. P. O'Leary, Deblurring Images: Matrices, Spectra, and Filtering, Fundamentals of Algorithms, SIAM, 2006.
[27] W. L. Briggs, V. E. Henson, The DFT : an owner's manual for the discrete Fourier transform, SIAM, 1995.
[28] M. L. Boas, Mathematical Methods in the Physical Sciences, John Wiley & Sons Inc., 3rd edition, 2006.
[29] M. A. Freitag, Transcritical Flow Modelling with the Box Scheme, Dissertation, MSc in modern applications of mathematics, University of Bath, 2003.
[30] S. K. Mitra, Digital Signal Processing: A Computer Based Approach, McGraw Hill Higher Education, 3rd edition, 2006.
[31] S. E. Jenkins, Numerical Model Error in Data Assimilation, Thesis, PhD in Engineering and Mathematics, University of Bath, 2014.
[32] E. V. Hólm, Lecture notes on assimilation algorithms, Meteorological Training Course Lecture Series (2003). Available from: "`http://old.ecmwf.int/newsevents/training/lecture_notes/pdf_files/ASSIM/Ass_algs.pdf`" [accessed: 24/07/2010].
[33] H. S. Carslaw, Introduction to the Theory of Fourier's Series and Integrals, New York: Dover Publications, Inc., 3rd edition, 1950.
[34] J. P. Boyd, Chebyshev and Fourier Spectral Methods, Dover Publications, Inc., 2nd (revised) edition, 2001.
[35] Mathworks, Inc. 1994-2013, MATLAB, 2012b edition, 2012.
[36] G. R. Grimmett, D. R. Stirzaker, Probability and random processes, Oxford: Clarendon Press, 1982.
[37] P. D. Williams, A proposed modification to the robert-asselin time filter, Monthly Weather Review 137 (2009) 2538–2546.
[38] P. D. Williams, An improvement to the robert-asselin filter in semi-implicit integrations, Monthly Weather Review 139 (2011) 1996–2007.
[39] P. Courtier, J. N. Thépaut, A. Hollingsworth, A strategy for operational implementation of 4d-var, using an incremental approach, Quarterly Journal of the Royal Meteorological Society 120B (1994) 1367–1387.
[40] S. Haben, Conditioning and Preconditioning of the Minimisation Problem in Variational Data Assimilation, Thesis, PhD in Mathematics, University of Reading, 2011.
[41] E. F. Toro, Shock-Capturing Methods for Free-Surface Shallow Flows, John Wiley & Sons, Ltd., 2001.

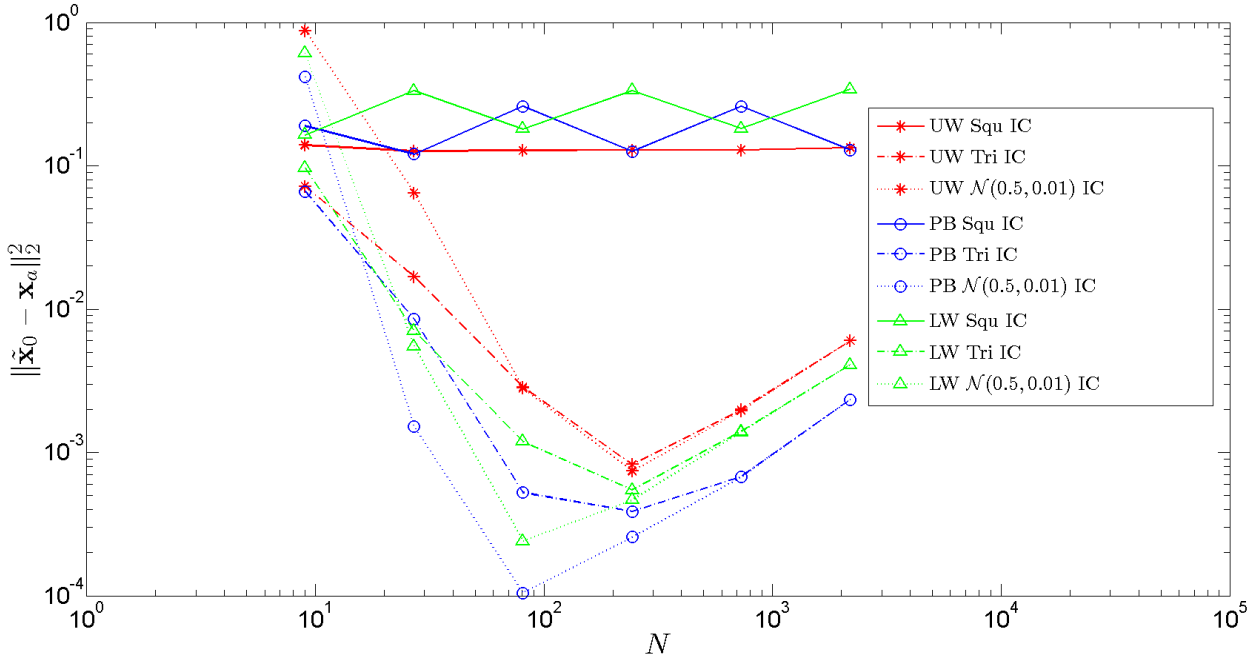



(a) The order of convergence to zero, with respect to $N$, using fixed $L = 4$ ($\Delta t = 1/2N$).

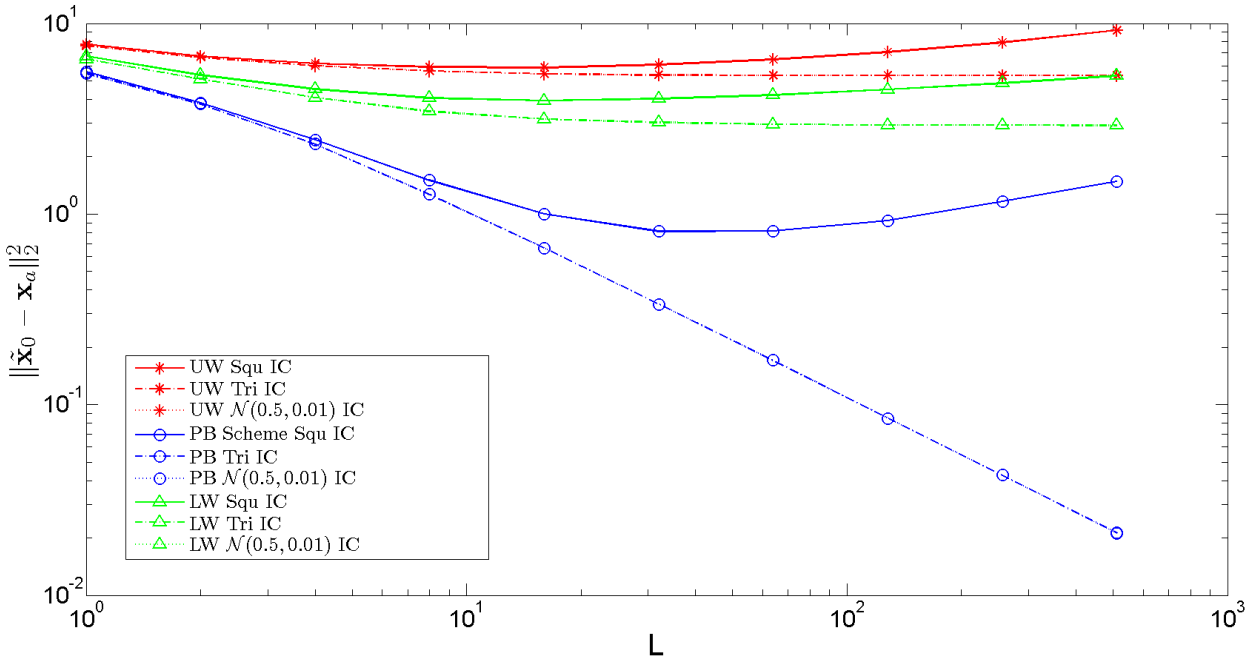



(b) The order of convergence to zero, with respect to $L$ using fixed $N = 3^7$ ($\Delta t = 1/(2 \cdot 3^7)$).

Figure 8: The square of the $l_2$-norm of the error in the analysis vector, calculated through strong constraint 4D-Var data assimilation numerical experiments, solely under the influence of errors introduced by finite difference approximations in the forward model. The results were generated using the Upwind (UW), Preissman Box (PB) and Lax-Wendroff (LW) schemes as the forward models for solving the 1D linear advection problem in (4), using $h = 0.5$ and $\mu = 1$. The functions considered for $u_0(x)$ in these experiments are the square function, a triangular function and a Gaussian function, denoted by 'squ IC', 'tri IC' and '$\mathcal{N}(0.5, 0.01)$ IC' respectively. The results are plotted using logarithmic scales to demonstrate the order of convergence of the error to zero.




(a) The order of convergence to zero, with respect to $N$, using fixed $L = 4$ ($\Delta t = 1/2N$) and $\sigma_o^2 = 5 \times 10^{-6}$.




(b) The Order of convergence to zero, with respect to $L$, using fixed $N = 3^7$ ($\Delta t = 1/(2 \cdot 3^7)$) and $\sigma_o^2 = 5 \times 10^{-3}$.

Figure 9: The square of the $l_2$-norm of the error in the analysis vector, calculated through strong constraint 4D-Var data assimilation numerical experiments, under the influence of errors introduced by finite difference approximations in the forward model and observation errors. The observations are Gaussian random variables with mean zero and variance $\sigma_o^2$. The results were generated using the Upwind (UW), Preissman Box (PB) and Lax-Wendroff (LW) schemes as the forward models for solving the 1D linear advection problem in (4), using $h = 0.5$ and $\mu = 1$. The functions considered for $u_0(x)$ in these experiments are the square function, a triangular function and a Gaussian function denoted by 'squ IC', 'tri IC' and '$\mathcal{N}(0.5, 0.01)$ IC' respectively. The results are plotted using logarithmic scales to demonstrate the order of convergence of the error to zero.