Journal of Computational and Applied Mathematics 290 (2015) 567–588

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

# 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

ARTICLE INFO

*Article history:*
Received 18 March 2014
Received in revised form 17 April 2015

*Keywords:*
Data assimilation
Numerical model error
Observation error
Deterministic error

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.



## 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

* Corresponding author at: Department of Electronic and Electrical Engineering, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom.
*E-mail 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).

http://dx.doi.org/10.1016/j.cam.2015.05.020

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 [1–4]. 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.

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