School of Mathematical
and Physical Sciences

University of Reading

Department of Mathematics and Statistics

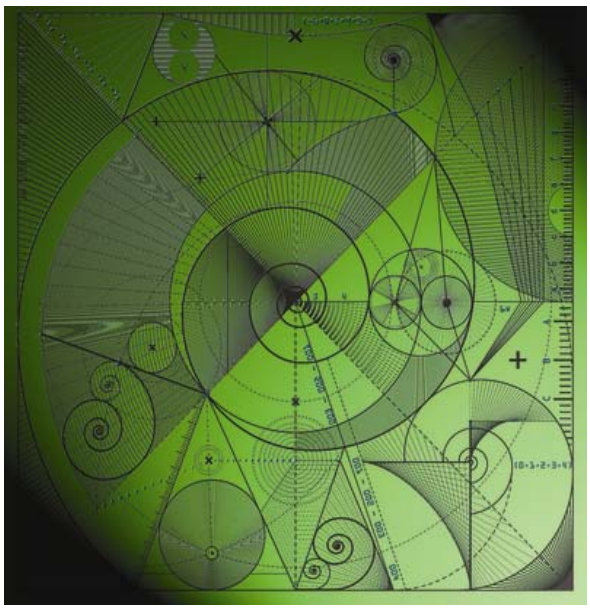
Preprint MPS-2012-13

12 June 2012

Nonlinear error dynamics for cycled data assimilation methods

by

Alexander J.F. Moodey, Amos S. Lawless, Roland W.E. Potthast, Peter Jan van Leeuwen



Nonlinear error dynamics for cycled data assimilation methods

Alexander J F Moodey[1], Amos S Lawless[1], Roland W E Potthast[1,2], Peter Jan van Leeuwen[1]

[1] University of Reading, School of Mathematical and Physical Sciences, Whiteknights, PO Box 220, Reading RG6 6AX, United Kingdom, [2] German Meteorological Service - Deutscher Wetterdienst Research and Development, Frankfurter Strasse 135, 63067 Offenbach, Germany

E-mail: A.J.F.Moodey@pgr.reading.ac.uk

Abstract. We investigate the error dynamics of the analysis error for cycled data assimilation systems. In particular, for large-scale *nonlinear dynamical systems* $\mathcal{M}$ which are Lipschitz continuous with respect to their initial states, we provide deterministic estimates for the development of the analysis error $\|\mathbf{e}_k\| := \|\mathbf{x}_k^{(a)} - \mathbf{x}_k^{(t)}\|$ between the analysis $\mathbf{x}^{(a)}$ and the true state $\mathbf{x}^{(t)}$ over time.

Large-scale dynamical systems are adequately described by an infinite-dimensional setting. Here, we investigate observations $\mathbf{y}$ arising from a linear observation operator H, such that the observation equation $H\mathbf{x} = \mathbf{y}$ is ill-posed, i.e. its solution is either non-unique or non-stable or both. In this case we need to stabilise H^{-1} by replacing it by a bounded reconstruction operator $R_\alpha = (\alpha I + H^*H)^{-1}H^*$, where $\alpha > 0$ is a regularization parameter. In the case of typical assimilation methods such as *three dimensional variational data assimilation*, α reflects the relationship between the size of the observation error covariance matrix and the size of the background error covariance matrix.

Observation error of size $\delta > 0$ leads to an analysis error in every analysis step. These errors can accumulate, if they are not (a) controlled in the reconstruction and (b) damped by the dynamical system $\mathcal{M}$ under consideration. We call a data assimilation method stable, if the analysis error is bounded in time by some constant C. For nonlinear systems we employ Lipschitz constants $K^{(1)}$ and $K^{(2)}$ on the lower and higher modes of $\mathcal{M}$ to control the damping behaviour of the dynamics. The key task of this work is to provide estimates for the analysis error $\|\mathbf{e}_k\|$, depending on the size δ of the observation error, the reconstruction error operator $I - R_\alpha H$ and the Lipschitz constants $K^{(1)}$ and $K^{(2)}$. We show that systems can be stabilised by choosing α sufficiently small, but the bound C will then depend on the data error δ in the form $c\|R_\alpha\|\delta$ with some constant c. Since $\|R_\alpha\| \to \infty$ for $\alpha \to 0$, the constant might be large. Numerical examples for this behaviour in the nonlinear case are provided using a (low-dimensional) Lorenz '63 system.

1. Introduction

Combining observational data with a dynamical system presents many challenges to different areas of science. Usually, both observations and model are uncertain, which leads to formulating the problem in a statistical framework [1][2][3]. However, this problem can also considered in a deterministic setting [1][2]. In many areas of climate and forecasting science, the dimension of the system dictates the method which is used to combine observations and dynamics. Data assimilation algorithms need to be applicable within situations where the dimension of the state space, typically in numerical weather prediction, ranges between orders of $\mathcal{O}(10^7 - 10^8)$ and observational data is of order of $\mathcal{O}(10^6)$. Since these algorithms deal with such high dimension, it is natural to extend analysis into infinite dimension to capture the key features of large-scale systems. Using an infinite dimensional approach, we are able to work within a framework that is best suited to analyse directly, challenges that exist in high dimensional data assimilation algorithms. In this work we restrict ourselves to analysing three dimensional variational data assimilation (3DVar) type methods in a large-scale or infinite dimensional setting.

This work is primarily interested in an estimate for the analysis error. We use the term *analysis* in this context to be the best estimate to the state of the system given by the data assimilation algorithm. Throughout this work, $\mathbf{x}_k^{(a)}$ will denote the analysis at a particular time t_k for $k \in \mathbb{N}_0$ which is obtained by the various data assimilation algorithms that we consider. The analysis error is defined as the difference between the analysis and the true state of the system such that

$$\mathbf{e}_k := \mathbf{x}_k^{(a)} - \mathbf{x}_k^{(t)}, \tag{1}$$

where $\mathbf{x}_k^{(t)}$ represents to true state of the system also at time t_k for $k \in \mathbb{N}_0$. We will call a data assimilation scheme *stable*, if $\|\mathbf{e}_k\|_{\mathbf{X}}$ remains bounded by some constant $C > 0$ for $k \to \infty$ with some appropriate norm $\|\cdot\|_{\mathbf{X}}$ relating to the state space $\mathbf{X}$.

In this work we will consider estimates for analysis error using weighted norms. This approach was also used in previous work [4] where the authors expressed their analysis as a cycled Tikhonov regularization. In that work, only linear and constant dynamical systems were analysed, here we now move a significant step ahead and investigate *nonlinear* model dynamics. For this work we concentrate on analysing the situation where the covariance operators are static in time and apply our theory to 3DVar-type methods with a linear observation operator H. We will show that using weighted norms enables us to directly analyse the evolution of the analysis error and demonstrate conditions whereby the analysis error remains stable for all time.

Note that the case of climatic or static covariances is a template for any cycled assimilation scheme where a reset of the background covariance matrix is carried out either regularly or where the covariance matrix is set up of climatic and dynamic components. The case of completely evolving covariances by the Kalman filter is more complicated and is treated in future work. We also note that the work of [5] considers this aspect of evolving covariances for the Navier-Stokes equations. However their analysis is restricted to the case of static covariances, where they derive similar asymptotic

behaviour of the analysis error to our work. Their work assumes that the observation operator and model dynamics commute, whereby here we study general linear ill-posed observation operators.

The paper will be split into five sections. First, we introduce our notation and the data assimilation algorithm we consider. Then, in Section 3 we work out estimates for the analysis error for a weighed norm state space and study the how the analysis error evolves. We demonstrate the theory using a simple numerical example in Section 4 using the Lorenz '63 system. Finally, we bring this work to a conclusion in Section 5.

2. Cycled data assimilation algorithms

We consider a dynamical system that we define as a nonlinear dynamical flow and a linear measurement equation. We set up our work in the framework of Hilbert spaces. Let $\mathcal{M}_k^{(t)} : \mathbf{X} \to \mathbf{X}$ be a true nonlinear model operator which maps a state $\mathbf{x}_k^{(t)} \in \mathbf{X}$ onto the state $\mathbf{x}_{k+1}^{(t)}$ for $k \in \mathbb{N}_0$ where $\mathbf{X}$ represents an infinite Hilbert space. The mapping is a discrete-time evolution where

$$\mathbf{x}_{k+1}^{(t)} = \mathcal{M}_k^{(t)} \left(\mathbf{x}_k^{(t)} \right), \tag{2}$$

for $\mathbf{x}_k^{(t)} \in \mathbf{X}$ for $k \in \mathbb{N}_0$. Here we have used the notation $\mathcal{M}_k^{(t)}$ and $\mathbf{x}_k^{(t)}$ whereby we remark that t here refers to the true nonlinear system operator and true state respectively. Furthermore, we assume that we have a modelled nonlinear model operator $\mathcal{M}_k : \mathbf{X} \to \mathbf{X}$ such that,

$$\mathcal{M}_k \left(\mathbf{x}_k^{(t)} \right) = \mathcal{M}_k^{(t)} \left(\mathbf{x}_k^{(t)} \right) + \boldsymbol{\zeta}_k, \tag{3}$$

where $\boldsymbol{\zeta}_k$ is some additive noise which we call *model error* for $k \in \mathbb{N}_0$. To keep this work as general as possible we consider all forms of noise either deterministic or stochastic. Although in this work, we require that the noise $\boldsymbol{\zeta}_k$ be bounded by some constant $\upsilon > 0$ for all time t_k for $k \in \mathbb{N}_0$, using an appropriate norm associated to the state space $\mathbf{X}$. This assumption is necessary to keep our analysis deterministic, we postpone stochastic derivations of our theory to future work. Readers interested in stochastical derivations in this field are pointed towards [6].

Furthermore we denote a constant true linear observation operator $H^{(t)}$, such that $H^{(t)} : \mathbf{X} \to \mathbf{Y}$ maps elements in the Hilbert space $\mathbf{X}$ into a Hilbert space $\mathbf{Y}$. The norms on $\mathbf{X}$ and $\mathbf{Y}$ are induced by their inner products $\langle \cdot, \cdot \rangle_{\mathbf{X}}$ and $\langle \cdot, \cdot \rangle_{\mathbf{Y}}$ respectively. In this work we will drop the the indices $\mathbf{X}$ and $\mathbf{Y}$, since it is usually clear which inner product is used. Here we state that the true observations (measurements) $\mathbf{y}_k^{(t)} \in \mathbf{Y}$ are located at discrete times, t_k, linearly such that

$$\mathbf{y}_k^{(t)} = H^{(t)} \mathbf{x}_k^{(t)}, \tag{4}$$

for $k \in \mathbb{N}_0$ where $\mathbf{x}_k^{(t)}$ represents the true state of the system and similarly $H^{(t)}$ is the true observation operator. We assume that the modelled observation operator $H : \mathbf{X} \to \mathbf{Y}$

is equal to the true observation operator and that the given observations $\mathbf{y}_k \in \mathbf{Y}$ are also taken at discrete times, t_k, linearly such that

$$\mathbf{y}_k^{(t)} = H^{(t)}\mathbf{x}_k^{(t)} = H\mathbf{x}_k^{(t)} = \mathbf{y}_k - \boldsymbol{\eta}_k, \tag{5}$$

where $\boldsymbol{\eta}_k$ is some additive noise which we call the *observation error* for $k \in \mathbb{N}_0$. Similarly to the model error, we require that the noise $\boldsymbol{\eta}_k$, either deterministic or stochastic is bounded by some constant $\delta > 0$ for all time t_k for $k \in \mathbb{N}_0$, again using an appropriate norm associated to the observation space $\mathbf{Y}$. In reality, the true or modelled observation operator might not be time-invariant or linear, however we make these assumptions for this work and postpone the time-variant and nonlinear case to future work.

The data assimilation task can be interpreted as seeking an *analysis* $\mathbf{x}_k^{(a)}$ at every assimilation step t_k, $k \in \mathbb{N}_0$ to solve the equation of the first kind

$$H\mathbf{x}_k^{(a)} = \mathbf{y}_k \tag{6}$$

given some prior or *background* knowledge $\mathbf{x}_k^{(b)}$. Since we know that the observations have errors we cannot expect to solve (6) exactly in each step and in general it would not make sense to fit exactly to noisy observations. It is well-known that with a *compact* linear operator H, (6) is ill-posed in the classical sense, whereby $H^{-1} : \mathbf{Y} \to \mathbf{X}$ is not continuous. For many modern remote sensing techniques used for example for atmospheric applications such as radiance measurements by polar satellites [7][8], for GPS/GNSS tomography [9] or even for full radar including nonlinear effects [10], this is the adequate mathematical framework to analyse the measurements.

A technique in inverse problems to resolve this issue, is to use *regularization methods*, such that we replace the unbounded operator H^{-1} with a family of bounded operators R_α. Arguably the most famous regularization method is *Tikhonov regularization* where the eigenvalues of the operator H^*H are shifted by a *regularization parameter* α, where H^* denotes the adjoint to the operator H.

Typically we have a first guess or *background*, $\mathbf{x}_k^{(b)}$, for $k \in \mathbb{N}_0$ which is an *a priori* estimate calculated from earlier analysis, subject to the nonlinear discrete-time evolution,

$$\mathbf{x}_{k+1}^{(b)} = \mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right), \tag{7}$$

for $k \in \mathbb{N}_0$. For example, in numerical weather prediction, this background state operationally comes from a previous short-term forecast [11][12]. Using this *a priori* knowledge we can formulate the following *Tikhonov functional*

$$\mathcal{J}^{(Tik)}\left(\mathbf{x}_k^{(a)}\right) := \alpha \left\|\mathbf{x}_k^{(a)} - \mathbf{x}_k^{(b)}\right\|_{\mathbf{X}}^2 + \left\|\mathbf{y}_k - H\mathbf{x}_k^{(a)}\right\|_{\mathbf{Y}}^2, \tag{8}$$

subject to (7) for $k \in \mathbb{N}_0$. For a linear operator H, the minimum is given by the solution to the *normal equations*, which can be reformulated into the update formula

$$\mathbf{x}_k^{(a)} = \mathbf{x}_k^{(b)} + R_\alpha\left(\mathbf{y}_k - H\mathbf{x}_k^{(b)}\right) \tag{9}$$

where

$$R_\alpha := (\alpha I + H^*H)^{-1}H^*, \tag{10}$$

is called the *Tikhonov inverse*, with regularization parameter $\alpha > 0$. In other disciplines, the operator R_α is known as the *Kalman gain matrix*, see [13][14][15], for an appropriate definition of H^*. Using (7) we can write (9) purely in terms of the analysis field $\mathbf{x}_k^{(a)}$,

$$\mathbf{x}_{k+1}^{(a)} = \mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right) + R_\alpha\left(\mathbf{y}_{k+1} - H\mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right)\right) \tag{11}$$

for $k \in \mathbb{N}_0$.

In the finite dimensional setting, such that $\mathbf{X} = \mathbb{R}^n$ and $\mathbf{Y} = \mathbb{R}^m$, where typically in numerical weather prediction $m \ll n$, the norms on $\mathbf{X}$ and $\mathbf{Y}$ can be defined using invertible covariance matrices $B \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{m\times m}$. These covariance matrices, B and R, statistically model the errors in the background state and the observations respectively. In the infinite dimensional setting, we call these weights, covariance operators, such that $B : \mathbf{X} \to \mathbf{X}$ and $R : \mathbf{Y} \to \mathbf{Y}$. For the purpose of this work we will assume that the operators B and R are positive definite, with bounded inverses. For convenience we will divide both covariances operators through by weightings $w_{(b)}$ and $w_{(o)}$ to obtain operators C and D. Therefore, $B := w_{(b)}C$ and $R := w_{(o)}D$ where $w_{(b)}$ and $w_{(o)}$ are the background and observation weightings with covariance operators, $C : \mathbf{X} \to \mathbf{X}$ and $D : \mathbf{Y} \to \mathbf{Y}$ respectively. By setting up the covariances in this way we will directly see how inflating these weights will correspond to how much we trust the background or the observations. We align this interpretation with variance inflation, which is popular in the data assimilation community.

With the choice of regularization parameter $\alpha = w_{(o)}/w_{(b)}$, we can connect the popular data assimilation technique, 3DVar with cycled Tikhonov regularization. The 3DVar functional takes the following form

$$\begin{aligned}\mathcal{J}^{(3D)}\left(\mathbf{x}_k^{(a)}\right) := &\left\langle B^{-1}\left(\mathbf{x}_k^{(a)} - \mathbf{x}_k^{(b)}\right), \mathbf{x}_k^{(a)} - \mathbf{x}_k^{(b)}\right\rangle_{L^2} \\ &+ \left\langle R^{-1}\left(\mathbf{y}_k - H\mathbf{x}_k^{(a)}\right), \mathbf{y}_k - H\mathbf{x}_k^{(a)}\right\rangle_{L^2},\end{aligned} \tag{12}$$

with respect to an L^2 inner product, subject to (7) for $k \in \mathbb{N}_0$. For a linear operator H, the minimum of the 3DVar functional is given by the solution to the normal equations, which can be reformulated into the update formula

$$\mathbf{x}_k^{(a)} = \mathbf{x}_k^{(b)} + K\left(\mathbf{y}_k - H\mathbf{x}_k^{(b)}\right), \tag{13}$$

where

$$K = BH'(HBH' + R)^{-1} \tag{14}$$

is again the Kalman gain in an equivalent form, where H' denotes the adjoint operator with respect to the L^2 inner product. Here (14) does not change over time since we concentrate our analysis on 3DVar-type methods. It is important to distinguish this type of algorithm from other algorithms which evolve the covariances over time, such as the Kalman filter. Of course the Kalman filter will change (14) in each assimilation step. With the appropriate definition of the adjoint H^* such that,

$$H^* := CH'D^{-1}, \tag{15}$$

the Tikhonov inverse, R_α is equivalent to the Kalman gain, K. In the next section we will analyse the asymptotic behaviour of this algorithm in a nonlinear framework, formulating the problem using weighted norms.

3. Error evolution of data assimilation systems

Here we study the error evolution of the cycled data assimilation systems which we introduced in Section 2 for weighted norms. In previous work [4], the authors connected 3DVar with cycled Tikhonov regularization using weighted norms, where

$$\left\langle \mathbf{x}_k^{(b)}, H^*\mathbf{y}_k \right\rangle_{B^{-1}} := \left\langle \mathbf{x}_k^{(b)}, B^{-1}H^*\mathbf{y}_k \right\rangle_{L^2}, \tag{16}$$

$$\left\langle H\mathbf{x}_k^{(b)}, \mathbf{y}_k \right\rangle_{R^{-1}} := \left\langle H\mathbf{x}_k^{(b)}, R^{-1}\mathbf{y}_k \right\rangle_{L^2}, \tag{17}$$

such that the index L^2 indicates a standard L^2 inner product. In this section we extend the work of [4] to nonlinear systems with respect to weighted norms. Since the algorithm we consider uses static covariances we are able to define the metric associated to the state and measurement space according to the errors in the background and observation respectively. By including the covariance information in the norms we are able to directly work with the forward nonlinear map $\mathcal{M}_k$ and the linear observation operator, H. We write (9) at time t_{k+1} in terms of the analysis field using (7), to give

$$\mathbf{x}_{k+1}^{(a)} = \mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right) + \left(\alpha I + H^*H\right)^{-1} H^*\left(\mathbf{y}_{k+1} - H\mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right)\right) \tag{18}$$

$$= \left(I - \left(\alpha I + H^*H\right)^{-1} H^*H\right)\mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right) + R_\alpha \mathbf{y}_{k+1} \tag{19}$$

$$= \left(I + \alpha^{-1}H^*H\right)^{-1}\mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right) + R_\alpha\left(\mathbf{y}_{k+1}^{(t)} + \boldsymbol{\eta}_{k+1}\right) \tag{20}$$

$$= \left(I + \alpha^{-1}H^*H\right)^{-1}\mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right) + R_\alpha\left(H\mathcal{M}_k^{(t)}\left(\mathbf{x}_k^{(t)}\right) + \boldsymbol{\eta}_{k+1}\right), \tag{21}$$

using (4) and (2). Now we subtract the true state from both sides and define $\mathbf{e}_{k+1} := \mathbf{x}_{k+1}^{(a)} - \mathbf{x}_{k+1}^{(t)}$, to obtain

$$\mathbf{x}_{k+1}^{(a)} - \mathbf{x}_{k+1}^{(t)} = \left(I + \alpha^{-1}H^*H\right)^{-1}\mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right) + \left(\left(\alpha I + H^*H\right)^{-1}H^*H - I\right)\mathcal{M}_k^{(t)}\left(\mathbf{x}_k^{(t)}\right) + R_\alpha\boldsymbol{\eta}_{k+1} \tag{22}$$

$$\mathbf{e}_{k+1} = \left(I + \alpha^{-1}H^*H\right)^{-1}\left(\mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right) - \mathcal{M}_k^{(t)}\left(\mathbf{x}_k^{(t)}\right)\right) + R_\alpha\boldsymbol{\eta}_{k+1} \tag{23}$$

$$= \left(I + \alpha^{-1}H^*H\right)^{-1}\left(\mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right) - \mathcal{M}_k\left(\mathbf{x}_k^{(t)}\right)\right) + \left(I + \alpha^{-1}H^*H\right)^{-1}\left(\mathcal{M}_k\left(\mathbf{x}_k^{(t)}\right) - \mathcal{M}_k^{(t)}\left(\mathbf{x}_k^{(t)}\right)\right) + R_\alpha\boldsymbol{\eta}_{k+1} \tag{24}$$

$$= \left(I + \alpha^{-1}H^*H\right)^{-1}\left(\mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right) - \mathcal{M}_k\left(\mathbf{x}_k^{(t)}\right)\right) + \left(I + \alpha^{-1}H^*H\right)^{-1}\boldsymbol{\zeta}_k + R_\alpha\boldsymbol{\eta}_{k+1} \tag{25}$$

$$= N\left(\mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right) - \mathcal{M}_k\left(\mathbf{x}_k^{(t)}\right)\right) + N\boldsymbol{\zeta}_k + R_\alpha\boldsymbol{\eta}_{k+1}, \tag{26}$$

where $N := (I + \alpha^{-1}H^*H)^{-1} = I - R_\alpha H$. In this work (26) will form the basis for our analysis and asymptotic behaviour. This form explicitly represents how various terms in the algorithm behave in each assimilation step. Here, N represents the reconstruction error operator which affects the analysis error, $\mathbf{e}_k$ propagated forward in time through the modelled nonlinear model operator $\mathcal{M}_k$ for $k \in \mathbb{N}_0$. Throughout this work, we will refer to the operator N as the *regularized reconstruction error operator*. Furthermore, the regularized reconstruction error operator N also has an impact on the model error term, $\boldsymbol{\zeta}_k$ for $k \in \mathbb{N}_0$. Therefore, controlling this reconstruction error is necessary in each step to keep the data assimilation scheme stable.

As a first attempt to examine the error behaviour, we take the norm on both sides, use the triangle inequality and the sub-multiplicative property as follows,

$$\|\mathbf{e}_{k+1}\| \leq \|N\| \cdot \left\|\mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right) - \mathcal{M}_k\left(\mathbf{x}_k^{(t)}\right)\right\| + \|N\|\,\upsilon + \|R_\alpha\|\,\delta, \tag{27}$$

where constants υ and δ bound the noise on the model error and observations respectively. For the Hilbert space $(\mathbf{X}, \|\cdot\|_{B^{-1}})$ this bound can be seen as a nonlinear extension to [4].

We now need some assumptions on the nonlinear map $\mathcal{M}_k$. It is natural to assume that the map is Lipschitz continuous, which does not restrict the generality of this work. However, we will require that the map has a global Lipschitz constant for all time.

Assumption 3.1. *The nonlinear mapping $\mathcal{M}_k : \mathbf{X} \to \mathbf{X}$ is Lipschitz continuous with a global Lipschitz constant such that given any $\mathbf{a}, \mathbf{b} \in \mathbf{X}$,*

$$\|\mathcal{M}_k(\mathbf{a}) - \mathcal{M}_k(\mathbf{b})\| \leq K_k \cdot \|\mathbf{a} - \mathbf{b}\| \tag{28}$$

where $K_k \leq K$, the global Lipschitz constant for all time t_k, for $k \in \mathbb{N}_0$. If the system is time-invariant then of course, $K_k = K$ for all time. However it is not necessary for the nonlinear system to be time-invariant. In this work we will refer to K without subscript as the global Lipschitz constant.

Now, applying Assumption 3.1 we obtain,

$$\|\mathbf{e}_{k+1}\| \leq K\,\|N\| \cdot \|\mathbf{e}_k\| + \|N\|\,\upsilon + \|R_\alpha\|\,\delta \tag{29}$$

where K is the global Lipschitz constant. Since everything other than the analysis error is independent of time, we can explicitly define $\Lambda := K\|N\|$ and formulate the following result.

Theorem 3.2. *If the model error $\boldsymbol{\zeta}_k$, $k \in \mathbb{N}_0$, is bounded by $\upsilon > 0$, the observation error $\boldsymbol{\eta}_k$, $k \in \mathbb{N}_0$, is bounded by $\delta > 0$ and the model operator, $\mathcal{M}$ is Lipschitz continuous, then the analysis error $\mathbf{e}_{k+1} := \mathbf{x}_{k+1}^{(a)} - \mathbf{x}_{k+1}^{(t)}$, for $k \in \mathbb{N}_0$ is estimated by*

$$\|\mathbf{e}_{k+1}\| \leq \Lambda^{k+1}\,\|\mathbf{e}_0\| + \left(\sum_{l=0}^{k} \Lambda^l\right)(\sigma + \tau), \tag{30}$$

where $\Lambda := K\|N\|$, with global Lipschitz constant K and $\sigma := \|N\|\upsilon$ and $\tau := \|R_\alpha\|\delta$, where $N := I - R_\alpha H$ and $R_\alpha := (\alpha I + H^*H)^{-1}H^*$. If $\Lambda < 1$ then,

$$\|\mathbf{e}_{k+1}\| \leq \Lambda^{k+1}\|\mathbf{e}_0\| + \frac{1 - \Lambda^{k+1}}{1 - \Lambda}(\sigma + \tau), \tag{31}$$

such that

$$\limsup_{k\to\infty} \|\mathbf{e}_{k+1}\| \leq \frac{\|N\|\,\upsilon + \|R_\alpha\|\,\delta}{1-\Lambda}. \tag{32}$$

Proof. We use induction as follows. For the base case we set $k = 0$, and from (29) we obtain

$$\|\mathbf{e}_1\| \leq \Lambda\|\mathbf{e}_0\| + \|N\|\,\upsilon + \|R_\alpha\|\,\delta, \tag{33}$$

which is equivalent to (30) substituting for σ and τ. Now we continue with the inductive step,

$$\begin{aligned}
\|\mathbf{e}_{k+1}\| &\leq \Lambda\,\|\mathbf{e}_k\| + \sigma + \tau && (34)\\
&\leq \Lambda\left(\Lambda^k\,\|\mathbf{e}_0\| + \left(\sum_{l=0}^{k-1}\Lambda^l\right)(\sigma+\tau)\right) + \sigma + \tau && (35)\\
&\leq \Lambda^{k+1}\,\|\mathbf{e}_0\| + \left(\sum_{l=0}^{k}\Lambda^l\right)(\sigma+\tau) && (36)
\end{aligned}$$

which is equal to (30), hence completing the proof by induction. Assuming $\Lambda < 1$ and using the geometric series we obtain (31), which completes the proof. □

We have described error estimates for the analysis error of cycled data assimilation scheme in Theorem 3.2. For the Hilbert space $(\mathbf{X}, \|\cdot\|_{B^{-1}})$, the sufficient condition to keep the analysis error bounded is that $\Lambda = K\|N\| < 1$. Here the regularized reconstruction error has to be strong enough, so that multiplied with the global Lipschitz constant K, Λ is kept less than one. Now we explore how we can make $\|N\|$ small enough to ensure $\Lambda < 1$. We first consider the finite dimensional case.

Lemma 3.3. *For a finite dimensional state space $\mathbf{X} = \mathbb{R}^n$, an injective operator H and a parameter $0 < \rho < 1$, by choosing the regularization parameter $\alpha > 0$ we can always achieve $\|N\| \leq \rho < 1$.*

Proof. By construction $G := H^*H$ is self-adjoint and therefore has a complete orthonormal basis $\varphi_{(1)}, \ldots, \varphi_{(n)}$ of eigenvectors with eigenvalues $\lambda_{(1)}, \ldots, \lambda_{(n)}$. We choose α and ρ such that

$$\alpha\left(\frac{1}{\rho} - 1\right) = \min_{j=1,\ldots,n} \left|\lambda_{(j)}\right| > 0. \tag{37}$$

Now,

$$\begin{aligned}
\|N\| = \left\|(I + \alpha^{-1}G)^{-1}\right\| &= \sup_{j\in\{1,\ldots,n\}} \left|\frac{1}{1+\frac{\lambda_{(j)}}{\alpha}}\right| \\
&\leq \frac{\alpha}{\alpha + \alpha\left(\frac{1}{\rho}-1\right)} = \rho < 1, && (38)
\end{aligned}$$

which completes the proof. □

In this finite dimensional setting, it is clear that from Lemma 3.3 that given any global Lipschitz constant $K > 0$, i.e. any model dynamics, we can always choose an $\alpha > 0$, sufficiently small so that $\Lambda \leq K\rho < 1$. This is an interesting conclusion to draw, since the regularization parameter controls how much we trust the background term in (8). Of course reducing α means that we solve the problem in (6) more accurately which implies that solving the problem accurately keeps the data assimilation scheme stable for all time. However, from the theory of Tikhonov regularization, we know that α must be kept large enough to shift the spectrum of H^*H to combat ill-posedness in the observation operator. Representing the problem in an infinite dimension allows us to directly represent this effect of ill-posedness into the problem. We will see that for an ill-posed observation operator, significant damping must be present on higher spectral modes for us to control the behaviour of the analysis error over time. Firstly we explore the less interesting situation of well-posed observation operators H to highlight the difficulty in treating ill-posed operators in an infinite dimension.

Lemma 3.4. *For an infinite dimensional state space* $\mathbf{X}$*, an injective well-posed operator* H *and a parameter* $0 < \rho < 1$*, by choosing the regularization parameter* $\alpha > 0$ *sufficiently small we can always achieve* $\Lambda \leq \rho < 1$*.*

Proof. As the operator H is well-posed, $G := H^*H$ has a complete orthonormal basis $\varphi_{(1)}, \ldots, \varphi_{(\infty)}$ of eigenvectors with eigenvalues $\lambda_{(1)}, \ldots, \lambda_{(\infty)} > 0$. If $K > 1$ we choose ρ such that $\rho < 1/K$, otherwise we choose any $\rho < 1$. Then we choose an α such that

$$\alpha \left(\frac{1}{\rho} - 1 \right) = \min_{j=1,\ldots,\infty} \left| \lambda_{(j)} \right| > 0. \tag{39}$$

Then,

$$\begin{aligned} \Lambda = K \left\| N \right\| = K \left\| (I + \alpha^{-1} G)^{-1} \right\| &= K \sup_{j \in \{1,\ldots,\infty\}} \left| \frac{1}{1 + \frac{\lambda_{(j)}}{\alpha}} \right| \\ &\leq \frac{K\alpha}{\alpha + \alpha \left(\frac{1}{\rho} - 1 \right)} = K\rho < 1, \end{aligned} \tag{40}$$

which completes the proof. □

From Lemma 3.4, for a well-posed observation operator H we can control the error behaviour using α. The estimates in (40) are based on lower bounds for the spectrum of H^*H. In the large-scale or infinite-dimensional case, our interest is with compact observation operators H, where the spectrum decays to zero in the infinite dimensional setting. Here, to achieve a stable cycled scheme with an ill-posed observation operator, we now show that the Lipschitz constant has to be contractive with respect to higher spectral modes of H. We remark,

Remark 3.5. For the Hilbert space $(\mathbf{X}, \| \cdot \|_{B^{-1}})$ and a compact observation operator H, (29) implies that the model operator $\mathcal{M}_k$ must be strictly damping, i.e. the global Lipschitz constant $K < 1$. This is apparent since for an infinite dimensional state space space the norm of the regularized reconstruction error, $\|N\| = 1$.

Further details of the spectrum and norm estimates of the regularized reconstruction error operator can be found in the literature [16],[4]. However, we will see that by splitting the state space $\mathbf{X}$ we are able to use the regularized reconstruction error operator N to control the Lipschitz constant K over lower spectral modes. In considering lower and higher spectral modes separately we are able to obtain a stable cycled scheme for a wider class of systems. As we have mentioned, the infinite dimensional setting here highlights the practical problem in keeping the analysis error bounded for all time. We will see that damping with respect to the observation operator H is required in the nonlinear map M_k.

We begin defining orthogonal projection operators, P_1 and P_2 such that,

$$P_1 : \mathbf{X} \to \text{span}\{\boldsymbol{\varphi}_i, i \leq n\} \tag{41}$$

$$P_2 : \mathbf{X} \to \text{span}\{\boldsymbol{\varphi}_i, i > n\} \tag{42}$$

with respect to the singular system $(\mu_n, \boldsymbol{\varphi}_n, \mathbf{g}_n)$ of H, with $n \in \mathbb{N}$. We abbreviate

$$\mathbf{X}_1 := \text{span}\{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_n\}, \quad \mathbf{X}_2 := \text{span}\{\boldsymbol{\varphi}_{n+1}, \boldsymbol{\varphi}_{n+2}, \ldots\}. \tag{43}$$

More details on the singular system of a compact linear operator can be found in [17]. Returning to (26) we have,

$$\mathbf{e}_{k+1} = N(P_1 + P_2)\left(\mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right) - \mathcal{M}_k\left(\mathbf{x}_k^{(t)}\right)\right) + N\boldsymbol{\zeta}_k + R_\alpha \boldsymbol{\eta}_{k+1} \tag{44}$$

$$\begin{aligned} &= N|_{\mathbf{X}_1} P_1 \left(\mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right) - \mathcal{M}_k\left(\mathbf{x}_k^{(t)}\right)\right) + N\boldsymbol{\zeta}_k \\ &\quad + N|_{\mathbf{X}_2} P_2 \left(\mathcal{M}_k\left(\mathbf{x}_k^{(a)}\right) - \mathcal{M}_k\left(\mathbf{x}_k^{(t)}\right)\right) + R_\alpha \boldsymbol{\eta}_{k+1}. \end{aligned} \tag{45}$$

Now defining

$$\mathcal{M}_k^{(1)}(\cdot) := P_1 \circ \mathcal{M}_k(\cdot), \qquad \mathcal{M}_k^{(2)}(\cdot) := P_2 \circ \mathcal{M}_k(\cdot), \tag{46}$$

we have

$$\begin{aligned} \mathbf{e}_{k+1} = {} & N|_{\mathbf{X}_1}\left(\mathcal{M}_k^{(1)}\left(\mathbf{x}_k^{(a)}\right) - \mathcal{M}_k^{(1)}\left(\mathbf{x}_k^{(t)}\right)\right) + N\boldsymbol{\zeta}_k \\ & + N|_{\mathbf{X}_2}\left(\mathcal{M}_k^{(2)}\left(\mathbf{x}_k^{(a)}\right) - \mathcal{M}_k^{(2)}\left(\mathbf{x}_k^{(t)}\right)\right) + R_\alpha \boldsymbol{\eta}_{k+1}. \end{aligned} \tag{47}$$

Using the triangle inequality with the sub-multiplicative property we can obtain a bound on this error as follows,

$$\begin{aligned} \|\mathbf{e}_{k+1}\| = {} & \left\| N|_{\mathbf{X}_1}\left(\mathcal{M}_k^{(1)}\left(\mathbf{x}_k^{(a)}\right) - \mathcal{M}_k^{(1)}\left(\mathbf{x}_k^{(t)}\right)\right) + N\boldsymbol{\zeta}_k \right. \\ & \left. + N|_{\mathbf{X}_2}\left(\mathcal{M}_k^{(2)}\left(\mathbf{x}_k^{(a)}\right) - \mathcal{M}_k^{(2)}\left(\mathbf{x}_k^{(t)}\right)\right) + R_\alpha \boldsymbol{\eta}_{k+1} \right\| \end{aligned} \tag{48}$$

$$\begin{aligned} \leq {} & \|N|_{\mathbf{X}_1}\| \cdot \left\| \mathcal{M}_k^{(1)}\left(\mathbf{x}_k^{(a)}\right) - \mathcal{M}_k^{(1)}\left(\mathbf{x}_k^{(t)}\right) \right\| \\ & + \|N|_{\mathbf{X}_2}\| \cdot \left\| \mathcal{M}_k^{(2)}\left(\mathbf{x}_k^{(a)}\right) - \mathcal{M}_k^{(2)}\left(\mathbf{x}_k^{(t)}\right) \right\| \\ & + \|N\boldsymbol{\zeta}_k\| + \left\| R_\alpha \boldsymbol{\eta}_{k+1} \right\| \end{aligned} \tag{49}$$

$$\begin{aligned} \leq {} & K_k^{(1)} \cdot \|N|_{\mathbf{X}_1}\| \cdot \left\| \mathbf{x}_k^{(a)} - \mathbf{x}_k^{(t)} \right\| \\ & + K_k^{(2)} \cdot \|N|_{\mathbf{X}_2}\| \cdot \left\| \mathbf{x}_k^{(a)} - \mathbf{x}_k^{(t)} \right\| \end{aligned}$$

$$+ \|N\boldsymbol{\zeta}_k\| + \left\|R_\alpha \boldsymbol{\eta}_{k+1}\right\| \tag{50}$$

$$\leq \left(\Lambda_k^{(1)} + \Lambda_k^{(2)}\right) \|\mathbf{e}_k\| + \|N\|\, \upsilon + \|R_\alpha\|\, \delta, \tag{51}$$

where we have assumed Lipschitz continuity,

$$\left\|\mathcal{M}_k^{(j)}(\mathbf{x}_k^{(a)}) - \mathcal{M}_k^{(j)}(\mathbf{x}_k^{(t)})\right\| \leq K_k^{(j)} \left\|\mathbf{x}_k^{(a)} - \mathbf{x}_k^{(t)}\right\| \tag{52}$$

for $j = 1, 2$, defining $\Lambda_k^{(1)} := K_k^{(1)} \cdot \|N|_{\mathbf{X}_1}\|$ and $\Lambda_k^{(2)} := K_k^{(2)} \cdot \|N|_{\mathbf{X}_2}\|$ with restrictions according to the singular system of H. Again, we now assume that the modelled nonlinear operator $\mathcal{M}_k$ is globally Lipschitz continuous in accordance with Assumption 3.1, where $K_k^{(1)} \leq K^{(1)}$ and $K_k^{(2)} \leq K^{(2)}$ for all time t_k. Similar to Theorem 3.2 we can form the following result.

Theorem 3.6. *For the Hilbert space $(\mathbf{X}, \|\cdot\|_{B^{-1}})$, if the model error $\boldsymbol{\zeta}_k$, $k \in \mathbb{N}_0$, is bounded by $\upsilon > 0$, the observation error $\boldsymbol{\eta}_k$, $k \in \mathbb{N}_0$, is bounded by $\delta > 0$ and the model operator, $\mathcal{M}_k$ is Lipschitz continuous, then the analysis error $\mathbf{e}_{k+1} := \mathbf{x}_{k+1}^{(a)} - \mathbf{x}_{k+1}^{(t)}$, for $k \in \mathbb{N}_0$ is estimated by*

$$\|\mathbf{e}_{k+1}\| \leq \Lambda^{k+1} \|\mathbf{e}_0\| + \left(\sum_{l=0}^{k} \Lambda^l\right)(\sigma + \tau), \tag{53}$$

where $\Lambda := \Lambda^{(1)} + \Lambda^{(2)}$, for $\Lambda^{(1)} := K^{(1)} \cdot \|N|_{\mathbf{X}_1}\|$ and $\Lambda^{(2)} := K^{(2)} \cdot \|N|_{\mathbf{X}_2}\|$ with restrictions according to the singular system of H. Furthermore $\sigma := \|N\|\upsilon$ and $\tau := \|R_\alpha\|\delta$ where $N := I - R_\alpha H$ and $R_\alpha := (\alpha I + H^*H)^{-1}H^*$. If $\Lambda < 1$ then,

$$\|\mathbf{e}_{k+1}\| \leq \Lambda^{k+1} \|\mathbf{e}_0\| + \frac{1 - \Lambda^{k+1}}{1 - \Lambda}(\sigma + \tau), \tag{54}$$

such that

$$\limsup_{k \to \infty} \|\mathbf{e}_k\| \leq \frac{\|N\|\, \upsilon + \|R_\alpha\|\, \delta}{1 - \Lambda}. \tag{55}$$

Proof. The proof is the same as that of Theorem 3.2 for a different constant $\Lambda := \Lambda^{(1)} + \Lambda^{(2)}$, with restrictions according to the singular system of H. □

Remark 3.7. Since H is compact, for an infinite dimensional state space $\mathbf{X}$, we have that $\|N\| = 1$ and by using the arithmetic-geometric mean inequality we are able to bound the Tikhonov inverse by

$$\|R_\alpha\| \leq \frac{1}{2\sqrt{\alpha}}. \tag{56}$$

Therefore from (55) we have,

$$\limsup_{k \to \infty} \|\mathbf{e}_k\| \leq \frac{\upsilon + \frac{\delta}{2\sqrt{\alpha}}}{1 - \Lambda}. \tag{57}$$

We now present two results which are obtained directly from spectral theory and can also be seen in [4] in the linear setting. These results will demonstrate how we require that the nonlinear map $\mathcal{M}_k$ be dissipative with respect to higher spectral modes of the observation operator H.

Lemma 3.8. *For the Hilbert space* $(\mathbf{X}, \|\cdot\|_{B^{-1}})$*, on the subspace* $\mathbf{X}_1$ *and for a parameter* $0 < \rho < 1$*, by choosing the regularization parameter* $\alpha > 0$ *sufficiently small, we can always achieve* $\|N|_{\mathbf{X}_1}\| \leq \rho < 1$*.*

Proof. As the space $\mathbf{X}_1$ is spanned by a finite number of spectral modes of the operator H we use Lemma 3.3 to complete the proof. □

Lemma 3.9. *For the Hilbert space* $(\mathbf{X}, \|\cdot\|_{B^{-1}})$*, the norm of the operator* $N|_{\mathbf{X}_2}$ *is given by*

$$\|N|_{\mathbf{X}_2}\| = 1. \tag{58}$$

Proof. Since H is compact, the singular values $\mu_n \to 0$ for $n \to \infty$. This means that

$$\|N|_{\mathbf{X}_2}\| = \sup_{i=n+1,\ldots,\infty} \left| \frac{\alpha}{\alpha + \mu_n^2} \right| = 1 \tag{59}$$

for all $\alpha > 0$, which completes the proof. □

As we have seen from Theorem 3.6, the sufficient condition for stability of the cycled data assimilation scheme requires that $\Lambda < 1$. Applying our norm estimates in (49) and (50) with both Lemma 3.8 and Lemma 3.9, we obtain,

$$\Lambda = K^{(1)} \cdot \|N|_{\mathbf{X}_1}\| + K^{(2)} \cdot \|N|_{\mathbf{X}_2}\| \tag{60}$$

$$\leq K^{(1)} \cdot \rho + K^{(2)}. \tag{61}$$

Here we directly see how the nonlinear growth in $\mathbf{X}_1$ can be controlled by the regularization parameter $\alpha > 0$. Furthermore, it is seen that the nonlinear system $\mathcal{M}_k$ has to be damping in $\mathbf{X}_2$ for all time. Therefore only if $\mathcal{M}_k$ is a sufficiently damping on higher spectral modes of H will we be able to stabilise the cycled data assimilation scheme. We call this type of system dissipative with respect to H as summed up in the following definition.

Definition 3.10. A nonlinear system $\mathcal{M}_k$, $k \in \mathbb{N}$, is *dissipative with respect to* H, if it is Lipschitz continuous and damping with respect to higher spectral modes of H in the sense that $\mathcal{M}_k^{(2)}$ defined by (46) satisfies

$$\left\| \mathcal{M}_k^{(2)}(\mathbf{a}) - \mathcal{M}_k^{(2)}(\mathbf{b}) \right\| \leq K_k^{(2)} \cdot \|\mathbf{a} - \mathbf{b}\| \tag{62}$$

$\forall\, \mathbf{a}, \mathbf{b} \in \mathbf{X}$, where $K_k^{(2)} \leq K^{(2)} < 1$ uniformly for $k \in \mathbb{N}_0$.

Under this assumption that $\mathcal{M}_k$ is dissipative with respect to H, we can choose the regularization parameter $\alpha > 0$ small enough such that,

$$\rho < \frac{1 - K^{(2)}}{K^{(1)}}, \tag{63}$$

to achieve a stable cycled scheme. We summarise in the following theorem.

Theorem 3.11. *For the Hilbert space* $(\mathbf{X}, \|\cdot\|_{B^{-1}})$*, assume the system* $\mathcal{M}_k$ *is Lipschitz continuous and dissipative with respect to higher spectral modes of* H*. Then, for regularization parameter,* $\alpha > 0$ *sufficiently small i.e. there exists an* $\bar{\alpha}$*, such that for* $\alpha < \bar{\alpha}$*, we have* $\Lambda := K^{(1)}\|N|_{\mathbf{X}_1}\| + K^{(2)}\|N|_{\mathbf{X}_2}\| < 1$*. Under the conditions of Theorem 3.6 the analysis error is bounded over time by*

$$\limsup_{k\to\infty} \|\mathbf{e}_k\| \leq \frac{\|N\|\,\upsilon + \|R_\alpha\|\,\delta}{1-\Lambda} \leq \frac{\upsilon + \frac{\delta}{2\sqrt{\alpha}}}{1-\Lambda}. \tag{64}$$

Proof. If $\mathcal{M}_k$ is Lipschitz continuous and dissipative with respect to H, we first show that we can achieve $\Lambda < 1$. From (61), the Lipschitz constants $K^{(1)}$ and $K^{(2)}$ determine our ability to achieve $\Lambda < 1$. Using Lemma 3.9 we have $\|N|_{\mathbf{X}_2}\| = 1$ therefore under the assumption that $\mathcal{M}_k$ is dissipative, then for the subspace $\mathbf{X}_2$, we have that $K^{(2)} < 1$. Now from Lemma 3.8 the norm $\|N|_{\mathbf{X}_1}\|$ can be made arbitrarily small such that $\|N|_{\mathbf{X}_1}\| \leq \rho < 1$ where ρ is some positive constant chosen such (63) holds. Then we obtain $\Lambda < 1$ from (61). The bound for the analysis error is then given by Theorem 3.6, which also provides the estimate in (64). The inequality completes the proof in accordance with Remark 3.7. □

As we have seen, the analysis error can be kept stable for all time with respect to the weighted norms defined in (17). These results depend on the weighted norm $\|\cdot\|_{B^{-1}}$ which is constant for all time. However, in practice other more advanced data assimilation schemes are used, where the covariances evolves in time. Notably the Kalman filter evolves the covariance in each assimilation step, though usually some additive *climatological* component of B or multiplicative *variance inflation* takes care of model errors, in which case the above analysis should carry over to more general cycled assimilation schemes.

Working in the infinite dimension demonstrates how the ill-posedness in the observation operator can lead to an unstable data assimilation scheme. However, using the regularization parameter, we were able to keep the analysis error bounded for all time. This of course has repercussions which we will explore further in Section 4 and Section 5.

4. Numerical examples

This work has focused on the analysis of the error evolution of data assimilation algorithms. To complement the theoretical results, here we present some simple numerics to demonstrate the behaviour of the analysis error in (26). Our interest lies with the choice of the regularization parameter α, as discussed in Section 3. Here we consider the famous Lorenz '63 system [18], which presents chaotic behaviour under the classical parameters. Lorenz derived these equations from a Fourier truncation of the flow equations that govern thermal convection. Despite their limited practical use, these equations present an excellent foundation to demonstrate the behaviour of the analysis

error. The Lorenz '63 equations are as follows,

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\sigma(x-y), \tag{65}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \rho x - y - xz, \tag{66}$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z, \tag{67}$$

where typically σ, ρ and β are known as the Prandtl number, the Rayleigh number and a non-dimensional wave number respectively. Throughout these experiments we will use the classical parameters, $\sigma = 10$, $\rho = 28$ and $\beta = 8/3$. We discretize the system using a fourth order Runge-Kutta approximation with a step-size $h = 0.01$. For these experiments we will omit model error from the model equations to concentrate on the behaviour of the analysis error compared with the error in the observations in each assimilation step. Since the same regularized reconstruction error operator is applied to the analysis error and the model error, there is limited inconsistency in assuming accurate model dynamics.

It is natural to set up the noise in these experiments to be Gaussian, since the 3DVar functional can be viewed as a maximum *a posteriori* probability Bayesian estimator of the state of the system under Gaussian statistics. However, as we have discussed in previous sections, we consider the 3DVar scheme purely as an algorithm and require for our analysis that the noise be bounded for time. Therefore we shall choose normally distributed noise and note that for a finite sample the noise is bounded and therefore is consistent with the theory developed in Section 3.

We set up a twin experiment, whereby we begin with an initial condition,

$$\left(x_0^{(t)}, y_0^{(t)}, z_0^{(t)}\right) = (-5.8696, -6.7824, 22.3356),$$

which were obtained using an initial reference point, $(0.001, 0.001, 2.001)$ that was spun-up 1000 time-steps to obtain the initial condition $(x_0^{(t)}, y_0^{(t)}, z_0^{(t)})$, which lies on the attractor. We produce a run of the system until time $t = 100$ with a step-size $h = 0.01$, which we call a *truth run*. Now using this truth we create observations at every tenth time-step adding random normally distributed noise with zero mean and standard deviation $\sigma_{(o)} = \sqrt{2}/40$. The background state is calculated in the same way at initial time t_0 with zero mean and standard deviation $\sigma_{(b)} = 1/400$ such that,

$$\left(x_0^{(b)}, y_0^{(b)}, z_0^{(b)}\right) = (-5.8674, -6.7860, 22.3338).$$

Now, ignoring the model error term in (26), we calculate the analysis error $\mathbf{e}_k$ for $k = 1, \ldots, 1000$. Assimilating only at every tenth time-step, we allow for the nonlinear model dynamics to play a role before we apply the data assimilation scheme. We calculate a sampled background error covariance between the background state and true state over the whole trajectory such that,

$$B = \begin{pmatrix} 117.6325 & 117.6374 & -2.3513 \\ 117.6374 & 152.6906 & -2.0838 \\ -2.3513 & -2.0838 & 110.8491 \end{pmatrix}. \tag{68}$$

We set the background weight equal to the background variance such that $w_{(b)} = \sigma^2_{(b)}$ and divide B through by $w_{(b)}$ to obtain the background error covariance matrix C. We simulate the consequence of an ill-posed observation operator H with a random 3×3 matrix with its last singular value $\mu_3 = 10^{-8}$ such that,

$$H = \begin{pmatrix} 0.4267 & 0.5220 & 0.5059 \\ 0.8384 & -0.7453 & 1.6690 \\ 0.4105 & 1.6187 & 0.0610 \end{pmatrix}. \tag{69}$$

Therefore, H is severely ill-conditioned with a condition number, $\kappa = 2.1051 \times 10^8$.

Varying the regularization parameter α we can study the averaged error, i.e. the integrated normalised total analysis error calculated by the sum of $\|\mathbf{e}_k\|_{L^2}$ from $k = 1, \ldots, 1000$, renormalised with division by 1000. In Figure 1 we plot the error integral against the regularization parameter. Here we observe that α needs to be chosen small enough to keep the analysis error bounded, although if it is chosen too small it will lead to a large analysis error bound. We select the largest regularization parameter from the error integral plot in Figure 1, which corresponds to 3DVar for $\alpha = w_{(o)}/w_{(b)}$ with weights according to the variances of the observations and background state respectively. With this value of $\alpha = 200$, in Figure 2(a) we observe that the analysis error, $\|\mathbf{e}_k^{(a)}\|_{L^2}$ fluctuates around the value of 20. The analysis is not able to track the truth and this can seen in Figure 2(b), where we plot the truth and analysis in state space from assimilation time t_{200} until t_{220}. We observe over this time interval that the analysis does not stay on the same wing of the attractor as the truth. This is evident across the whole time interval, which we confirm in Figure 2(c) and Figure 2(d) where the truth and analysis are plotted in state space from assimilation time t_{774} until t_{805} and t_{858} until t_{895} respectively. In Figure 2(c) the analysis and truth are close together at assimilation time t_{774}, however over time the analysis fails to follow the truth onto the same wing of the attractor. This is also reflected in Figure 2(d) where the analysis spends most time on the wrong wing of the attractor to the truth.

Now we inflate the background weighting $w_{(b)}$, i.e. background variance inflation, from $1/400^2$ to $1/40^2$ therefore choosing a smaller regularization parameter $\alpha = 2$, which corresponds to the parameter α where the error integral is smallest in Figure 1. Subsequently repeating with the same data, we observe in Figure 3(a) that the analysis error fluctuates around a much smaller value compared with Figure 2(a). This is reflected in Figure 3(b) where the analysis is now able to follow the trajectory of the truth better and remains for the assimilation time (t_{200} until t_{220}) on the same wing of the attractor to the truth.

In Figure 4(a) we reduce the regularization parameter even further so that $\alpha = 10^{-10}$ and we see that the analysis error becomes large again. From regularization theory we know that for a very small α the Tikhonov inverse can become very large in norm, which will affect the observation error in each assimilation step and this is confirmed in Figure 1. For this choice of α we also plot the trajectories of the analysis and the truth in Figure 4(b). Here we observe that the analysis attempts to track the

truth, however the consequence of an ill-posed observation operator, leading to a large Tikhonov inverse (in norm), is forcing the analysis off the attractor. Since the scheme trusts the observations much more than the background, the analysis is forced towards the observations, hence the analysis error is smaller for all time in Figure 4(a) compared with Figure 2(a).

Here the numerical results support the theory developed that by choosing α small enough, we can reduce the analysis error of the cycled data assimilation scheme for all time. However if it is chosen too small the analysis error will be amplified for an ill-posed observation operator.

In order to demonstrate the bounds developed in Section 3, we repeat the experiment using the same random 3×3 matrix for the observation operator H. We increase μ_3 from 10^{-8} to 10^{-3}, which is necessary since we want to achieve $\|N\| < 1/K$. If μ_3 is chosen too small, i.e. H is very ill-conditioned, then we would need to choose α sufficiently small to obtain $\|N\| < 1/K$. Theoretically this is possible, however numerically, with such a small α, the matrix N then becomes close to singular. Therefore we choose $\mu_3 = 10^{-3}$ so that we have a different observation operator,

$$H = \begin{pmatrix} 0.4275 & 0.5218 & 0.5055 \\ 0.8381 & -0.7453 & 1.6692 \\ 0.4102 & 1.6188 & 0.0612 \end{pmatrix}, \tag{70}$$

with condition number, $\kappa = 2.1051 \times 10^3$. We use the same initial condition, background state and error covariance matrix as before.

Since we are interested in our bounds on the analysis error in (29) we need to compute the global Lipschitz constant for this experiment. Using the truth and background runs we can calculate a Lipschitz constant from (28) every 10 time-steps. Choosing the largest value over all time we obtain a global Lipschitz constant $K = 1.9837$ for this experiment. This of course is only an estimate to the actual Lipschitz constant, however it will be sufficient for these experiments. Therefore we choose a regularization parameter, $\alpha = 10^{-6}$ which is small enough so that $\|N\| < 1/K$, leading to $\Lambda \approx 0.9918$. We calculate a bound on the observational noise with respect to the L^2 norm so that $\delta = 0.1571$.

In Figure 5 we plot the nonlinear analysis error from (26) and the bound on the analysis error in (29). Since we bound the analysis error by a linear update equation we observe the linear growth in our bound. For this choice of regularization parameter, we observe large fluctuations in the nonlinear analysis error arising from a Tikhonov inverse that is large in norm, which was discussed in the first experiment. Also in Figure 5 we plot the asymptotic limit of the analysis error from (64). From these numerical experiments we see that the numerical bound is not a very tight bound on the actual analysis error. This is of course expected since our approach is to use norm estimates, therefore we obtain a sufficient condition for a stable cycled scheme. Future work could address this aspect and seek a necessary condition for a stable cycled data assimilation scheme.

The theory in Section 3 was for weighted norms with respect to the error covariances. As previously discussed, we consider 3DVar-type methods which involves a static covariance matrix for all time. Therefore the inverse background covariance matrix C^{-1} is acting as a scaling on the analysis error. We calculate the analysis error $\|\mathbf{e}_k\|_{C^{-1}}$ and plot its evolution over time for regularization parameters, $\alpha = 200, 2$ and 10^{-10}. The figures are identical to Figure 2(a), Figure 3(a) and Figure 4(a) with a rescaled vertical axis. For $\alpha = 200$ the analysis error fluctuate around 0.005, then for $\alpha = 2$ the analysis error reduces to 10^{-5}. With a further inflation such that $\alpha = 10^{-10}$ the analysis error increases and fluctuates around 0.002.

We have repeated these experiments for a variety of observation operators H, different initial conditions and observation error drawn from different distributions. We obtain similar results, however we omit these experiment to keep this paper concise.

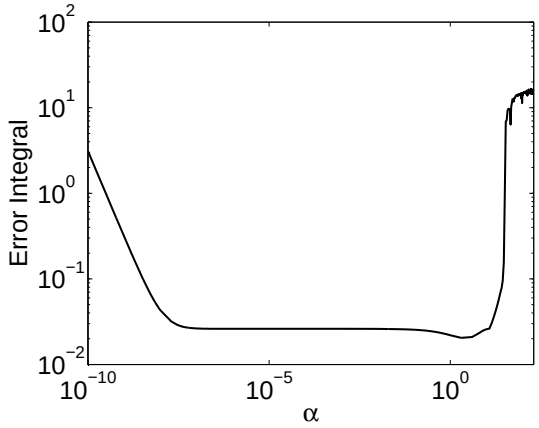


Figure 1. L^2 norm of the analysis error, $\|\mathbf{e}_k\|_{L^2}$ integrated for all assimilation time t_k for $k = 1, \ldots, 1000$, varying the regularization parameter, α.

5. Conclusions

In Section 3, we saw under weighted norms how the choice of the regularization parameter α was crucial in keeping the data assimilation scheme in a stable range, that is the analysis error remains bounded over time. Also in Section 4 we observed numerically that α has to be sufficiently small to keep the analysis error controlled over time. However, care must be taken in treating the regularization parameter, due to other instabilities arising from reducing the regularization. Moreover, in reducing α the inverse problem in (6) becomes computationally harder to solve. As we explore compact operators in a high dimensional setting we see that significant regularization is needed to shift the eigenvalues of the operator H^*H away from zero. If this is not done then the conditioning of the problem will have serious numerical implications. Therefore reducing α (which can be seen as inflating the weight $w_{(b)}$) can affect the ability with which we are able to calculate the Tikhonov inverse, when the observation operator is

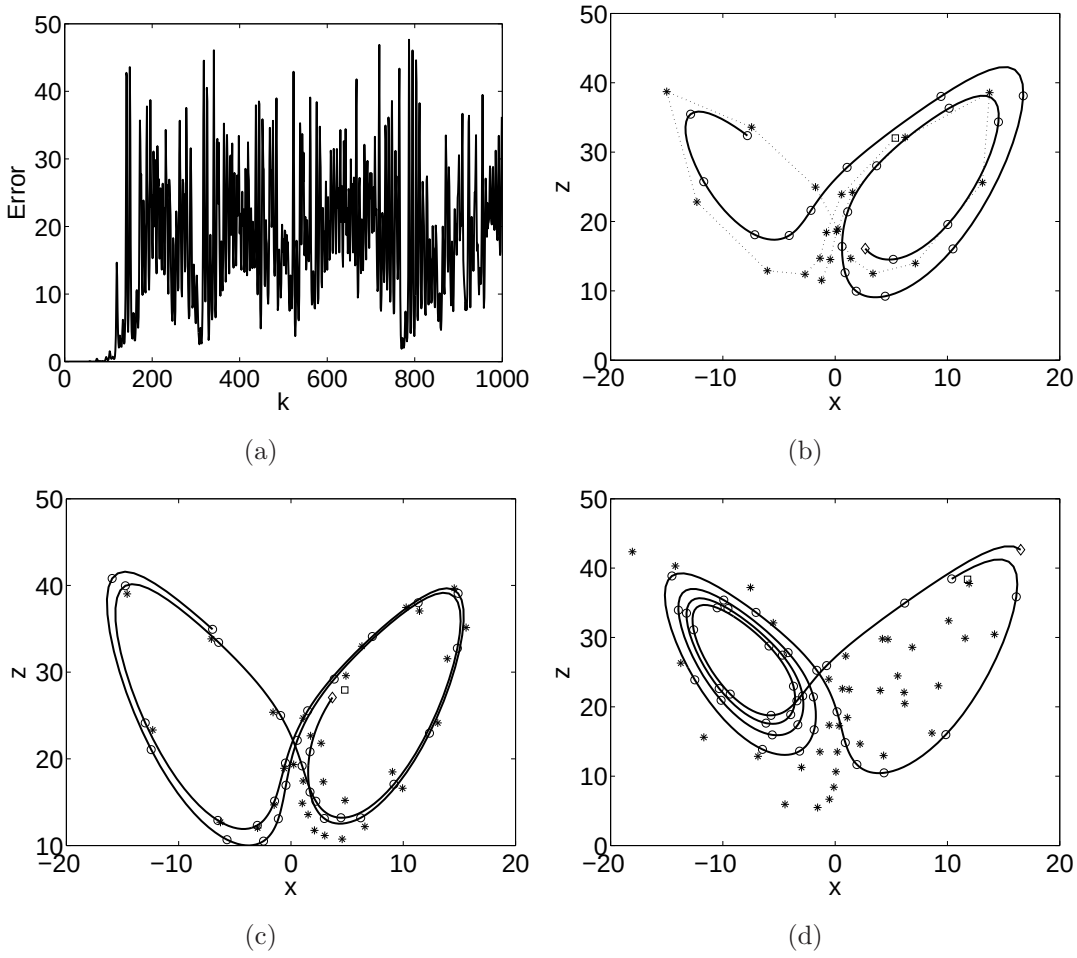
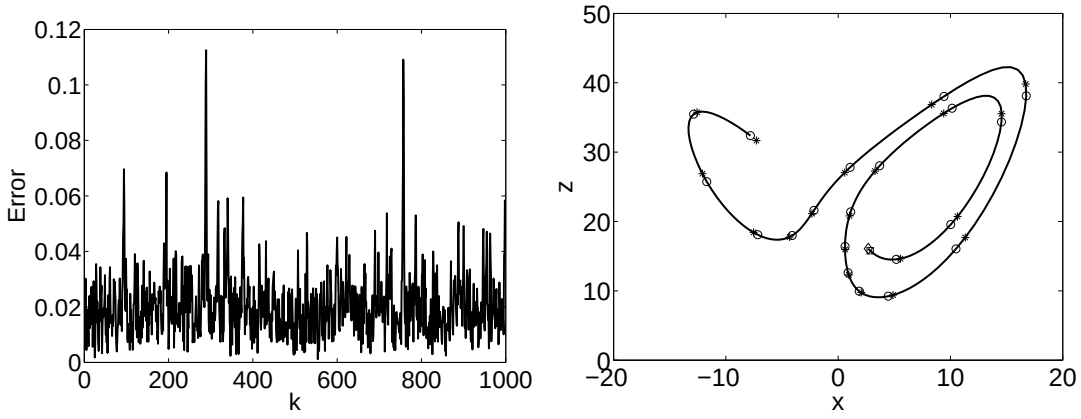


Figure 2. (a): L^2 norm of the analysis error, $\|\mathbf{e}_k\|_{L^2}$ as the scheme is cycled for index k with regularization parameter, $\alpha = 200$, which corresponds to 3DVar. Trajectories. Solid line: Truth, $\mathbf{x}^{(t)}$ for time-steps (b): $2000, \ldots, 2200$, (c): $7740, \ldots, 8050$ and (d): $8580, \ldots, 8950$. Diamond: Truth, (b): $\mathbf{x}^{(t)}_{200}$, (c): $\mathbf{x}^{(t)}_{774}$ and (d): $\mathbf{x}^{(t)}_{858}$. Circle: Truth, $\mathbf{x}^{(t)}_k$ at assimilation time t_k for (b): $k = 201, \ldots, 220$, (c): $k = 775, \ldots, 805$ and (d): $k = 859, \ldots, 895$. Square: Analysis, (b): $\mathbf{x}^{(a)}_{200}$, (c): $\mathbf{x}^{(a)}_{774}$ and (d): $\mathbf{x}^{(a)}_{858}$. Star: Analysis, $\mathbf{x}^{(a)}_k$ at assimilation time t_k for (b): $k = 201, \ldots, 220$, (c): $k = 775, \ldots, 805$ and (d): $k = 859, \ldots, 895$. Dotted line: Analysis $\mathbf{x}^{(a)}_k$ at assimilation time t_k for (b): $k = 200, \ldots, 220$.

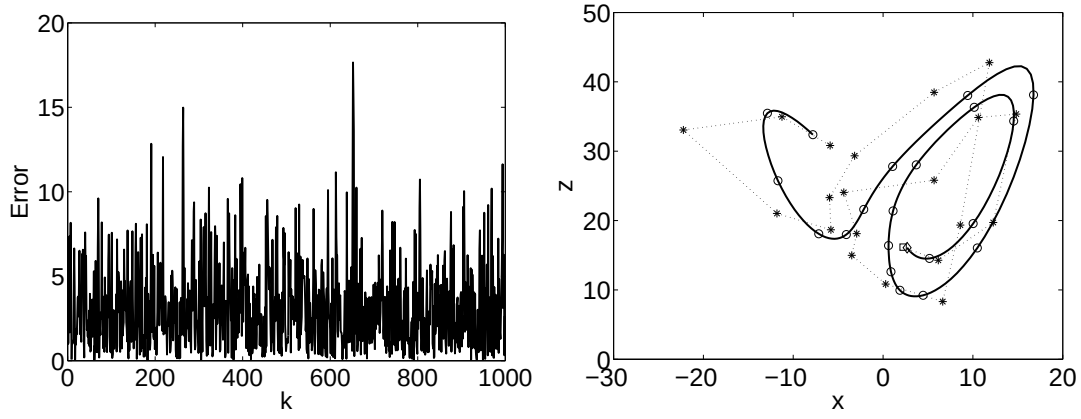
ill-conditioned. Of course the Tikhonov inverse is required in each assimilation step to solve (6).

Another difficulty in reducing the regularization parameter is evident from the bound we obtain on the analysis error. As we have seen, variance inflation enables us to control the nonlinear growth in both the analysis error and the model error. However the additional term, R_α involving the observation error is also affected by regularization parameter. As we know from regularization theory, (56) can become relatively large. This was directly seen in Section 4 where reducing the regularization too much meant the error in the observations became dominant in the analysis error. Therefore reducing



(a) (b)

Figure 3. (a): L^2 norm of the analysis error, $\|\mathbf{e}_k\|_{L^2}$ as the scheme is cycled for the index k with regularization parameter, $\alpha = 2$, an inflation in the background variance of 100%. (b): Trajectories. Solid line: Truth, $\mathbf{x}^{(t)}$ for time-steps $2000, \ldots, 2200$. Diamond: Truth, $\mathbf{x}^{(t)}_{200}$. Circle: Truth, $\mathbf{x}^{(t)}_k$ at assimilation time t_k for $k = 201, \ldots, 220$. Square: Analysis, $\mathbf{x}^{(a)}_{200}$. Star: Analysis, $\mathbf{x}^{(a)}_k$ at assimilation time t_k for $k = 201, \ldots, 220$.



(a) (b)

Figure 4. (a):L^2 norm of the analysis error, $\|\mathbf{e}_k\|_{L^2}$ as the scheme is cycled for the index k with regularization parameter, $\alpha = 10^{-10}$, an inflation in the background variance of 2×10^{12}%. (b): Trajectories. Solid line: Truth, $\mathbf{x}^{(t)}$ for time-steps $2000, \ldots, 2200$. Diamond: Truth, $\mathbf{x}^{(t)}_{200}$. Circle: Truth, $\mathbf{x}^{(t)}_k$ at assimilation time t_k for $k = 201, \ldots, 220$. Square: Analysis, $\mathbf{x}^{(a)}_{200}$. Star: Analysis, $\mathbf{x}^{(a)}_k$ at assimilation time t_k for $k = 201, \ldots, 220$. Dotted line: Analysis $\mathbf{x}^{(a)}_k$ at assimilation time t_k for $k = 200, \ldots, 220$.

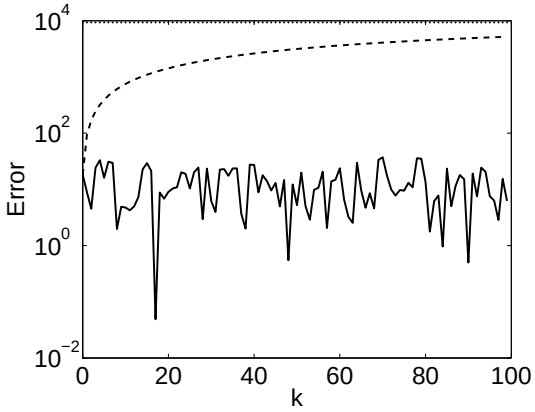


Figure 5. L^2 norm of the analysis error, $\|\mathbf{e}_k\|_{L^2}$ as the scheme is cycled for the index k with regularization parameter, $\alpha = 10^{-6}$. Solid line: Nonlinear analysis error. Dashed line: Linear bound. Dotted line: Asymptotic limit.

the regularization parameter will reduce the analysis error growth from the previous step (and the model error), although this could amplify the error in the observations. This of course depends on how ill-posed the problem is, as we have seen if the problem is very ill-posed, i.e. many singular values of the observation operator H lie very close to zero then the regularization parameter, α will need to be balanced between stability in the inverse problem (α large) and stability in the cycled data assimilation scheme (α small).

We showed in Section 3 that it is possible to keep the analysis error bounded for all time with an ill-posed observation operator in an infinite dimensional setting. This was due to the model dynamics being dissipative with respect to higher spectral modes of the observation operator. In the case of time-invariant model dynamics the constant $K^{(2)}$ must be shown to be damping with respect to the subspace $\mathbf{X}_2$. This is analogous to the previous work in [4] where the authors were able to force a contraction for linear time-invariant model dynamics. For nonlinear model dynamics, the property exploited in the linear setting does not hold, therefore it is necessary to show this dissipative property for the model dynamics under consideration. The work of [5] was able to show this property for the incompressible two-dimensional Navier-Stokes equations where the model dynamics and the observation operator commute. In the case of time-variant model dynamics then it must be shown that for all time t_k, the constants $K_k^{(2)} \leq K^{(2)} < 1$, therefore obeying the global property in Assumption 3.1, and are damping. This of course is a particular situation, however it is required for this theory to hold and must be shown for the model dynamics considered. An extension which would relax this assumption, would be for the subspace $\mathbf{X}_2$ to be also dependent on time. Then in each assimilation step, $\mathbf{X}_2$ could be chosen so that there is a contraction in the model dynamics with respect to the observation operator. In this case the regularization parameter would need to be chosen also in each assimilation step to allow for an expansion in the lower spectral modes in $\mathbf{X}_1$. This approach is consider

in future work.

Acknowledgments

The authors would like to thank the following institutions for their financial support: Engineering and Physical Sciences Research Council (EPSRC), National Centre for Earth Observation (NCEO, NERC) and Deutscher Wetterdienst (DWD). Furthermore the authors thank A. Stuart from the University of Warwick for useful discussions during this work.

References

[1] H W Engl, M Hanke, and A Neubauer. *Regularization of inverse problems*. Mathematics and its applications. Kluwer Academic Publishers, 2000.

[2] J M Lewis, S Lakshmivarahan, and S Dhall. *Dynamic Data Assimilation: A Least Squares Approach*. Cambridge University Press, 2006.

[3] A M Stuart. Inverse problems: a Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[4] R W E Potthast, A J F Moodey, A S Lawless, and P J van Leeuwen. On error dynamics and instability in data assimilation. 2012. Preprint available: MPS-2012-05 http://www.reading.ac.uk/maths-and-stats/research/maths-preprints.aspx.

[5] C E A Brett, K F Lam, K J H Law, D S McCormick, M R Scott, and A M Stuart. Accuracy and stability of filters for dissipative pdes. Submitted, 2012.

[6] C E A Brett, K F Lam, K J H Law, D S McCormick, M R Scott, and A M Stuart. Stability of filters for the Navier-Stokes equation. Submitted, 2012.

[7] C D Rodgers. *Inverse Methods for Atmospheric Sounding: Theory and Practice*. Series on Atmospheric, Oceanic and Planetary Physics. World Scientific, 2000.

[8] K N Liou. *An Introduction to Atmospheric Radiation*. International Geophysics Series. Academic Press, 2002.

[9] M Bender, G Dick, J Wickert, T Schmidt, S Song, G Gendt, M Ge, and M Rothacher. Validation of gps slant delays using water vapour radiometers and weather models. *Meteorologische Zeitschrift*, 17(6):807–812, 2008.

[10] U Blahak. *Analyse des Extinktionseffektes bei Niederschlagsmessungen mit einem C-Band Radar anhand von Simulation und Messung*. PhD thesis, Universitätsbibliothek Karlsruhe, 2004.

[11] P Courtier, E Andersson, W Heckley, D Vasiljevic, M Hamrud, A Hollingsworth, F Rabier, M Fisher, and J Pailleux. The ecmwf implementation of three-dimensional variational assimilation (3d-var). i: Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124(550):1783–1807, 1998.

[12] A C Lorenc, S P Ballard, R S Bell, N B Ingleby, P L F Andrews, D M Barker, J R Bray, A M Clayton, T Dalby, D Li, et al. The met. office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 126(570):2991–3012, 2000.

[13] R E Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME - Journal of Basic Engineering*, 82(1):35–45, 1960.

[14] A H Jazwinski. *Stochastic processes and filtering theory*. Academic Press, 1970.

[15] B D O Anderson and J B Moore. *Optimal Filtering*. Prentice Hall, 1979.

[16] B A Marx and R W E Potthast. On instabilities in data assimilation algorithms. *GEM-International Journal on Geomathematics*, pages 1–26, 2012.

[17] R Kress. *Linear Integral Equations*. Springer, 1999.

[18] E N Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.