Information Theory in Ensemble Data Assimilation: Application to the Kuramoto-Sivashinsky & Shallow-water equation

M. Jardak & M. Navon

July 5, 2011

1 Introduction

Uncertainty of various real phenomena has been for a long time a problem of scientific endeavors. However, the notion of uncertainty has a broad meaning and mathematically it can be performed in various ways. In general, we can say that uncertainty in some situations occurs whenever information pertaining to this situation is deficient. If uncertainty is interpreted as randomness then a natural way of quantifying it is probability theory; the information pertaining to random phenomena is then defined in terms of probability.

Information theory is relevant to data processing and statistical inference. Fisher [27] was the first to give a technical definition to information in his work on theory of estimation. He introducted then the concept of a measure of the amount of information supplied by data about unknown parameters. Hence, a relevant question is how information can we infer from a particular set of observations about sampled phenomenon, in particular the information theory is a tool for estimation of an unobserved quantity X through observations on another quantity Y. Information theory or communication theory is concerned with what is known as the information content of a message, which is the amount of useful information contained within a message. Information theory was first used by electrical engineers to design better telecommunication systems, but now has a wide variety of applications. In such diverse areas as complexity theory, networking analysis, financial mathematics and mathematical statistics.

It is useful to be able to quantify the amount of information provided by an observation or by an observing system. In the development of remote-sounding instruments, two popular measures of information content are entropy-reduction and degrees of freedom for signal (see, for example Rodgers [33] Rabier et al. [37]). The information theory has come to the notice of the data assimilation community, where it has been used to calculate information content of various observations. Information content of observations can potentially have many applications, including planning measurement missions, designing observational systems and defining targeted observations and data selection strategies. These applications have been underutilized so far, and were mainly oriented towards defining data selection strategies (e.g., Rabier et al. [37] and references therein). Nevertheless, progress in data assimilation methods should foster applications of information theory in many different areas.

2 The models

2.1 The Kuramoto-Sivashinsky Equation

The Kuramoto-Sivashinsky (K-S) equation is an evolution equation in one space dimension, with a Burgers nonlinearity, a fourth order dissipation term and a second anti-dissipative term. It assumes the form

$$u_t + \mu u_{xxxx} + \nu u_{xx} + u u_x = 0,\ (x,t) \in \mathbb{R} \times \mathbb{R}^+ \quad u(x,0) = u_0(x), x \in \mathbb{R} \qquad (1)$$

The K-S equation models pattern formations in different physical contexts and is a paradigm of low-dimensional behavior in solutions to partial differential equations. It arises as a model amplitude equation for inter-facial instabilities in many physical contexts. It was originally derived by Kuramoto and Tsuzuki ([1], [2]) to model small thermal diffusive instabilities in laminar flame fronts in two space dimensions. It has also been derived in the context of angular-phase turbulence for a system of reaction-diffusion modeling the Belouzov-Zabotinskii reaction in three space dimensions. Sivashinsky([3],[4])derived it independently to model small thermal diffusive instabilities in laminar flame fronts. The equation also arises in modeling small perturbations from a reference Poiseuille flow of a film layer on an inclined plane [5], while Babchin et al.[6] derived (1) as a general mechanism modeling the nonlinear saturation of instabilities in flow films as in the Rayleigh-Taylor-type instability.

The K-S equation is non-integrable, and no explicit solutions exist. It is characterized by a second-order unstable diffusion term, responsible for an instability at large scales,a fourth-order stabilizing viscosity term, which provides damping at small scales; and a quadratic nonlinear coupling term which stabilizes by transferring energy between large and small scales. This is readily apparent in Fourier space, where one may write(1)with periodic boundary condition as

$$\frac{d\hat{u}_k}{dt} = (\nu k^2 - \mu k^4)\hat{u}_k + \frac{i}{2}\sum_{k' \in \mathbb{Z}} k' \hat{u}_{k'} \hat{u}_{k-k'} \qquad (2)$$

where

$$u(x,t) = \sum_{k \in \mathbb{Z}} \hat{u}_k(t) \exp(ikx),\ k = n\frac{2\pi}{L},\ k' = m\frac{2\pi}{L},\ m,n \in \mathbb{Z},\ i = \sqrt{-1},$$

2.2 Shallow-Water equations in spherical geometry

The shallow water equations are a set of hyperbolic partial differential equations that describe the flow below a pressure surface in a fluid.
The equations are derived from depth-integrating the Navier-Stokes equations, in the case where the horizontal length scale is much greater than the vertical length scale. The shallow-water equations in spherical geometry are given by

$$\begin{cases} \dfrac{\partial u}{\partial t} + \dfrac{u}{a\cos\theta}\dfrac{\partial u}{\partial \lambda} + \dfrac{v}{a}\dfrac{\partial u}{\partial \theta} - \dfrac{\tan\theta}{a}vu - fv = -\dfrac{g}{a\cos\theta}\dfrac{\partial h}{\partial \lambda} \\ \dfrac{\partial v}{\partial t} + \dfrac{u}{a\cos\theta}\dfrac{\partial v}{\partial \lambda} + \dfrac{v}{a}\dfrac{\partial v}{\partial \theta} + \dfrac{\tan\theta}{a}u^2 + fu = -\dfrac{g}{a}\dfrac{\partial h}{\partial \theta} \\ \dfrac{\partial h}{\partial t} + \dfrac{u}{a\cos\theta}\dfrac{\partial h}{\partial \lambda} + \dfrac{v}{a}\dfrac{\partial h}{\partial \theta} + \dfrac{h}{a\cos\theta}\left[\dfrac{\partial u}{\partial \lambda} + \dfrac{\partial(\cos\theta)}{\partial \theta}\right] = 0 \end{cases} \tag{3}$$

where $V = u\vec{i} + v\vec{j}$ is the horizontal velocity vector (with respect to the surface of the sphere), gh is the free surface geopotential, h is the free surface height, g is the gravity acceleration. $f = 2\Omega\sin\theta$ is the Coriolis parameter, Ω is the angular velocity of the earth. θ denotes the angle of latitude, $\mu = \sin\theta$ is the longitude. λ the longitude,and a is the radius of the earth.

One of the major advances in meteorology was the use by [Rossby 1939] of the barotropic vorticity equation with the β -plane approximation to the sphericity of the Earth, and the deduction of solutions reminiscent of some large scale waves in the atmosphere. These solutions have become known as Rossby waves. Haurwitz (1940) then produced the equivalent solution for the sphere, now known as Rossby-Haurwitz waves (R-H waves). Rossby-Haurwitz waves are steadily propagating solutions of the fully nondivergent barotropic vorticity equation on a sphere first put forward by [Rossby 1939] and [Haurwitz 1940].

While the shallow water do not have corresponding analytic solutions they are expected to evolve in a similar way as the above R-H equations which explains why they have been widely used to test shallow water numerical models since the seminal paper of [Phillips 1959].

Following the research work of [Hoskins 1973] Rossby -Haurwitz waves with zonal wave numbers less or equal to 5 are believed to be stable. This makes the R-H zonal wave no 4 a suitable candidate for assessing accuracy of numerical schemes as was evident from its being chosen as a test case by [Williamson et al. 1992] and by a multitude of other authors.

It has been numerically shown that the R-H wave no 4 breaks down into more turbulent behaviour after long term numerical integration as recently discovered by [Thuburn and Li 2000].

The Rossby-Haurwitz waves are analytic solutions of the nonlinear barotropic vorticity equation on the sphere [Haurwitz 1940] and R. K. Smith and D. G. Dritschel [Smith 2006]. Although they are not analytic solutions of the shallow water equations they have been used so frequently for meteorological tests.

The initial velocity field for the Rossby-Haurwitz wave is defined as

$$\begin{cases} u = a\omega\cos\phi + aK\cos^{r-1}\phi(r\sin^2\phi - \cos^2\phi)\cos(r\lambda) \\ v = -aKr\cos^{r-1}\phi\sin\phi\sin(r\lambda) \end{cases} \tag{4}$$

The initial height field is defined as,

$$h = h_0 + \frac{a^2}{g}[A(\phi) + B(\phi)\cos(r\lambda) + C(\phi)\cos(2r\lambda)] \tag{5}$$

where the variables $A(\phi), B(\phi), C(\phi)$ are given by

$$\begin{cases} A(\phi) = \dfrac{\omega}{2}(2\Omega+\omega)\cos^2\phi + \dfrac{1}{4}k^2\cos^{2r}\phi[(r+1)\cos^2\phi + (2r^2-2r-2) - 2r^2\cos^2\phi] \\ \\ B(\phi) = \dfrac{2(\Omega+\omega)k}{(r+1)(r+2)}\cos^r\phi[(r^2+2r+2) - (r+1)^2\cos^2\phi] \\ \\ C(\phi) = \dfrac{1}{4}k^2\cos^{2r}\phi[(r+1)\cos^2\phi - (r+2)] \end{cases} \tag{6}$$

In here, r represents the wave number,h_0 is the height at the poles. The strength of the underlying zonal wind from west to east is given by ω and k controls the amplitude of the wave.

3 Data Assimilation

Data assimilation is the process by which observational data distributed in space and time are fused with mathematical model forecast information aimed at obtaining the best initial conditions that are as near as possible to observations while satisfying model forecast as a strong constraint. The probabilistic state space formulation and the requirement for the updating of information when new observations are encountered are ideally suited for the Bayesian approach, and thus constitute an appropriate framework for data assimilation. The Bayesian approach and in particular ensemble or particle filtering methods are a set of efficient and flexible Monte-Carlo methods to solve the optimal filtering problem. Here one attempts to construct the posterior probability density function (PDF) of the state based on all available information, including the set of received observations. Since this PDF embodies all available statistical information, it may be considered to be a complete solution to the estimation problem.

3.1 The Ensemble Kalman Filter

In order to analyze and make inference about the dynamic system at least a model equation along with an observation operator are required. First, a model describing the evolution of the state with time, and an observation operator for noisy observations of the state. Generically, stochastic filtering problem is a dynamic system that assumes the form

$$\dot{\mathbf{x}}_t = f(t, \mathbf{x}_t, \mathbf{u}_t, \mathbf{v}_t) \tag{7}$$

$$\mathbf{z}_t = h(\mathbf{x}_t, \mathbf{n}_t) \tag{8}$$

The equation (7) is the state equation or the system model, (8) is the observation operator equation,

The ensemble Kalman filter (EnKF) was first proposed by Evensen [12] and further developed by Burgers et al. [13] and Evensen ([14],[15]). It is related to particle filters in the context that a particle is identical to an ensemble member. EnKF is a sequential filter method, which means that the model is integrated forward in time and, whenever observations are available, these are used to reinitialize the model before the integration continues. The EnKF

originated as a version of the Extended Kalman Filter (EKF) ([18],[19]) for large problems. The classical KF [11] method is optimal in the sense of minimizing the variance only for linear systems and Gaussian statistics. Similar to the particle filter method, the EnKF stems from a Monte Carlo integration of the Fokker-Planck equation governing the evolution of the PDF that describes the prior, forecast, and error statistics. In the analysis step, each ensemble member is updated according to the KF scheme and replaces the covariance matrix by the sample covariance computed from the ensemble. However, the EnKF presents two potential problems namely: 1) Even though the EnKF uses full non-linear dynamics to propagate the forecast error statistics, the EnKF assumes that all probability distributions involved are Gaussian.
2) The updated ensemble preserves only the first two moments of the posterior. Let $p(\mathbf{x})$ denote the Gaussian prior probability density distribution of the state vector $\mathbf{x}$ with mean μ and covariance $\mathcal{Q}$

$$p(\mathbf{x}) \propto exp\left(\frac{-1}{2}(\mathbf{x}-\mu)^T\mathcal{Q}^{-1}(\mathbf{x}-\mu)\right)$$

We assume the data $\mathbf{z}$ to have a Gaussian PDF with covariance $\mathcal{R}$ and mean $\mathcal{H}\mathbf{x}$, where $\mathcal{H}$ is the so-called the observation matrix, is related to h of equation (8), and where the value $\mathcal{H}\mathbf{x}$ assumes the value of the data $\mathbf{z}$ would be for the state $\mathbf{x}$ in absence of observation errors. Then the conditional probability or likelihood $p(\mathbf{z}|\mathbf{x})$ assumes the form

$$p(\mathbf{z}|\mathbf{x}) \propto exp\left(\frac{-1}{2}(\mathbf{z}-\mathcal{H}\mathbf{x})^T\mathcal{R}^{-1}(\mathbf{z}-\mathcal{H}\mathbf{x})\right).$$

According to the Bayes theorem the posterior probability density follows from the relation

$$p(\mathbf{x}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{x})p(\mathbf{x}). \tag{9}$$

There are many variants of implementing the EnKF of various computational efficiency and in what follow we employ standard formulation of the EnKF for linear and nonlinear observation operators with covariance localization. See ([12], [13], [17] [16], [26]). The implementation of the standard EnKF may be divided into three steps, as follows:

- **Setting and matching**

  ■ Define the ensemble

  $$\mathcal{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N] \tag{10}$$

  be an $n_x \times N$ matrix whose columns are a sample from the prior distribution. N being the number of the ensemble members.

  ■ Form the ensemble mean

  $$\bar{\mathcal{X}} = \mathcal{X} \cdot \mathbf{1}_N, \tag{11}$$

  where $\mathbf{1}_N \in \mathbb{R}^{N\times N}$ is the matrix where each element is equal to $\frac{1}{N}$.

■ Define the ensemble perturbation matrix $\mathcal{X}'$ and set the $\mathbb{R}^{n_x \times n_x}$ ensemble covariance matrix $\mathcal{C}$

$$\mathcal{X}' = \mathcal{X} - \frac{1}{N}\bar{\mathcal{X}}, \tag{12}$$

$$\mathcal{C} = \frac{\mathcal{X}'\mathcal{X}'^T}{N-1}, \tag{13}$$

- **Sampling**

  ■ Generate

  $$\mathcal{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_N] \tag{14}$$

  be an $n_z \times N$ matrix whose columns are a replicate of the measurement vector $\mathbf{z}$ plus a random vector from the normal distribution $\mathcal{N}(0, \mathcal{R})$.

  ■ Form the $\mathbb{R}^{n_z \times n_z}$ measurement error covariance

  $$\mathcal{R} = \frac{\mathcal{Z}\mathcal{Z}^T}{N-1}, \tag{15}$$

- **Updating** Obtain the posterior $\mathcal{X}^p$ by the linear combinations of members of the prior ensemble

$$\mathcal{X}^p = \mathcal{X} + \mathcal{C}\mathcal{H}^T(\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R})^{-1}(\mathcal{Z} - \mathcal{H}\mathcal{X}) \tag{16}$$

The matrix

$$\mathcal{K} = \mathcal{C}\mathcal{H}^T(\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R})^{-1} \tag{17}$$

is the Kalman gain matrix. Since $\mathcal{R}$ is always positive definite(i.e. covariance matrix), the inverse $(\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R})^{-1}$ exists. An easy computation shows that the mean and covariance of the posterior or updated ensemble are given by

$$\bar{\mathcal{X}^p} = \mathcal{X}^p + \mathcal{K}\left[\mathbf{z} - \mathcal{H}\mathcal{X}^p\right], \tag{18}$$

and

$$\mathcal{C}^p = \mathcal{C} - \mathcal{K}\left[\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R}\right]\mathcal{K}^T, \tag{19}$$

To prevent the occurrence of filter divergence usually due to the background-error covariance estimates from small number of ensemble members as pointed out in Houtekamer and Mitchell [21], the use of covariance localization was suggested. Mathematically, the covariance localization increases the effective rank of the background error covariances. See the work of Gaspari and Cohn [24] also Hamill and Snyder [22],Hamill [23] and Ehrendorfer [25] . The covariance localization consists of multiplying point by point the covariance estimate from the ensemble with a correlation function that is 1.0 at the observation location and zero beyond some prescribed distance. Mathematically, to apply covariance localization, the Kalman gain

$$\mathcal{K} = \mathcal{C}\mathcal{H}^T(\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R})^{-1}$$

is replaced by a modified gain

$$\hat{\mathcal{K}} = [\rho \circ \mathcal{C}]\,\mathcal{H}^T(\mathcal{H}\,[\rho \circ \mathcal{C}]\,\mathcal{H}^T + \mathcal{R})^{-1} \tag{20}$$

where $\rho\circ$ denotes the Schur product (The Schur product of matrices $\mathcal{A}$ and $\mathcal{B}$ is a matrix $\mathcal{D}$ of the same dimension, where $d_{ij} = a_{ij}b_{ij}$) of a matrix $\mathcal{S}$ with local support with the covariance model generated by the ensemble. Various correlation matrices have been used, for horizontal localization Gaspari and Cohn [24] constructed a Gaussian-shaped function that is actually a fourth-order piece-wise polynomial. Houtekamer and Mitchell [21] and Evensen [14] used a cut-off radius so that observations are not assimilated beyond a certain distance from the grid points.

4 Information theory

The Bayes rule with state x and observations y states that:

$$\mathscr{P}_A(x) = \mathscr{P}(x|y) = \frac{\mathscr{P}_B(x) \cdot \mathscr{P}(y|x)}{\mathscr{P}(y)} \tag{21}$$

The information gain measured by the relative entropy is:

$$\begin{aligned} \text{info} &= \int \ln \frac{\mathscr{P}(x|y)}{\mathscr{P}(x)} \mathscr{P}(x|y)dx \\ &= \int \left[\ln \mathscr{P}(x) + \ln \mathscr{P}(y|x) - \ln \mathscr{P}(y) - \ln \mathscr{P}(x)\right] \mathscr{P}(x|y)dx \\ &= \int \left[\ln \mathscr{P}(y|x) - \ln \mathscr{P}(y)\right] \mathscr{P}(x|y)dx = \mathbb{E}_A\left[\ln \mathscr{P}(y|x)\right] - \ln \mathscr{P}(y), \end{aligned} \tag{22}$$

where $\mathbb{E}_A$ represents the expected value with respect to the posterior distribution. The marginal distribution of y does not depend on x and its expected value is a constant.

4.1 Information metrics and Gaussian probabilities

Consider a normal background distribution $\mathscr{P}_B(x_0) = \mathcal{N}(x_0^B, B)$ and a normal analysis distribution $\mathscr{P}_A(x_0) = \mathcal{N}(x_0^A, A)$. More precisely,

$$\mathscr{P}_B(x) = \frac{1}{\left((2\pi)^{n/2}\sqrt{\det B}\right)} \exp(-\frac{1}{2}(x - x_0^B)^T B^{-1}(x - x_0^B)) \tag{23}$$

and

$$-\ln \mathscr{P}_B(x) = \frac{n}{2}\ln(2\pi) + \frac{1}{2}\ln\det B + \frac{1}{2}\left(x - x_0^B\right)^T B^{-1}\left(x - x_0^B\right) \tag{24}$$

■ **Fisher information matrix**

The Fisher information matrix for a probability density $\mathscr{P}(x)$ is given by

$$\mathcal{F}(\mathscr{P}) = \int_{\mathbb{R}^n} \left[\frac{\partial(-\ln \mathscr{P}(x))}{\partial x}\right] \left[\frac{\partial(-\ln \mathscr{P}(x))}{\partial x})\right]^T \mathscr{P}(x)dx\ . \tag{25}$$

For Gaussian probabilities, the Fisher information matrix is the inverse of the

covariance,

$$\begin{cases} \left[\dfrac{\partial(-\ln \mathscr{P}_B(x))}{\partial x}\right] = B^{-1}(x - x_0^B) \\ \left[\dfrac{\partial(-\ln \mathscr{P}_B(x))}{\partial x}\right]\left[\dfrac{\partial(-\ln \mathscr{P}_B(x))}{\partial x}\right]^T = B^{-1}(x - x_0^B)(x - x_0^B)^T B^{-1} \\ \mathbb{E}_B\left[B^{-1}(x - x_0^B)(x - x_0^B)^T B^{-1}\right] = B^{-1}\mathbb{E}_B\left[(x - x_0^B)(x - x_0^B)^T\right] B^{-1} = \\ = B^{-1}BB^{-1} = B^{-1}. \end{cases} \tag{26}$$

The information content of observations can be measured by the trace of the difference of the Fisher matrices corresponding to analysis and background

$$\text{info} = \text{trace}\left(\mathcal{F}\left(\mathscr{P}_A\right)\right) - \text{trace}\left(\mathcal{F}\left(\mathscr{P}_B\right)\right) = \text{trace}\left(\mathcal{F}\left(\mathscr{P}_A\right) - \mathcal{F}\left(\mathscr{P}_B\right)\right) \tag{27}$$

which in the Gaussian case becomes

$$\text{info} = \text{trace}\left(A^{-1} - B^{-1}\right) \tag{28}$$

■ **Shannon information**
The Shannon entropy of the background pdf is

$$\begin{cases} S(\mathscr{P}_B) = \displaystyle\int_{\mathbb{R}^n} (-\ln \mathscr{P}_B(x))\mathscr{P}_B(x)dx = \\ \dfrac{n}{2}\ln(2\pi) + \dfrac{1}{2}\ln\det B + \dfrac{1}{2}\displaystyle\int_{\mathbb{R}^n}(x - x_0^B)^T B^{-1}(x - x_0^B)\mathscr{P}_B(x)dx = \\ n\ln\sqrt{2\pi} + \dfrac{1}{2}\ln\det B + \dfrac{n}{2} = n\ln\sqrt{2\pi e} + \dfrac{1}{2}\ln\det B \end{cases} \tag{29}$$

and the Shannon information content of observations reduces to

$$S(\mathscr{P}_B) - S(\mathscr{P}_A) = \frac{1}{2}\ln\det B - \frac{1}{2}\ln\det A = \frac{1}{2}\ln\det BA^{-1} \tag{30}$$

5 Numerical Experiments

The novelty of this work consists in comparing for the first time the above described and commonly used filters in the framework of the K-S equation model focusing on performance in the presence of nonlinear observation operators.

5.1 Numerical Solution of the K-S Equation

We consider the one-dimensional PDE with initial data as used in [7] and [8]

$$\begin{cases} u_t = -\mu u_{xxxx} - \nu u_{xx} - uu_x, \quad x \in [0, 32\pi] \\ u(x + L, t) = u(x, t), \quad L = 32\pi, \ \forall t > 0 \\ u(x, 0) = \cos(\dfrac{x}{16})\left(1 + \sin(\dfrac{x}{16})\right) \end{cases} \tag{31}$$

The system (31) is known to be stiff. In fact, the stiffness is due to rapid exponential decay of some modes (the dissipative part), the stiffness is also due to rapid oscillations of some modes (the dispersive part).

As the equation is periodic, a Fourier spectral method is used for spatial discretization. Despite the remarkable success of the spectral and pseudo-spectral methods for a wide range of applications [9], the set of ODEs for the mode amplitudes is stiff, due to the time scale associated with the nth mode scales as $O(n^{-m})$ for large n, where m is the order of the highest spatial derivative, so that the highest modes evolve on short time scales.

In order to carry out numerical solution of K-S, a modification of the exponential time-differencing fourth-order Runge-Kutta method (ETDRK4) has been used. This method has been proposed by Cox and Matthews [10] and further modified by Kassam and Trefethen [7]. A short review of the ETDRK4 is as follows:

First we transform (31) to Fourier space

$$\hat{u}_t = -\frac{ik}{2}\hat{u}^2 + (\nu k^2 - \mu k^4)\hat{u} \quad \text{with} \hat{u}^2 = \sum_{k' \in \mathbb{Z}} \hat{u}_{k'}\hat{u}_{k-k'}, \tag{32}$$

set

$$\mathscr{L}\hat{u}(k) = (\nu k^2 - \mu k^4)\hat{u}(k), \qquad \mathscr{N}(\hat{u}, t) = \mathscr{N}(\hat{u}) = -\frac{ik}{2}(\mathcal{F}(\mathcal{F}^{-1}(\hat{u}^2))), \tag{33}$$

$\mathscr{L}$ and $\mathscr{N}$ stand for linear and nonlinear operators, respectively. $\mathcal{F}$ denotes the discrete Fourier transform. Write (32) in an operational form

$$\hat{u}_t = \mathscr{L}\hat{u} + \mathscr{N}(\hat{u}, t). \tag{34}$$

Define $v = e^{-\mathscr{L}t}u$ where $e^{-\mathscr{L}t}$ the integrating factor to obtain

$$v_t = e^{-\mathscr{L}t}\mathscr{N}(e^{\mathscr{L}t}v). \tag{35}$$

Let h denote the time step length, then integrating (35) we obtain

$$u_{n+1} = e^{\mathscr{L}h}u_n + e^{\mathscr{L}h}\int_0^h e^{-\mathscr{L}\tau}\mathscr{N}(u(t_n + \tau), t_n + \tau)d\tau, \tag{36}$$

where u_n is the solution at the time $t = nh$ and $0 < \tau < h$.
The equation (36) is exact, and the various order EDT schemes differ only on the way one approximates the integral in (36). Cox and Matthews [10] proposed the generating formula

$$u_{n+1} = e^{\mathscr{L}h}u_n + h\sum_{m=0}^{s-1} g_m \sum_{k=0}^{m}(-1)^k \begin{pmatrix} m \\ k \end{pmatrix} \mathscr{N}_{n-k} \tag{37}$$

where s is the order of the scheme. The coefficients g_m are provided by the recurrence relation

$$\begin{cases} \mathscr{L}(hg_0) = e^{\mathscr{L}h} - I, \\ \mathscr{L}(hg_{m+1}) + \mathbf{I} = g_m + \frac{1}{2}g_{m-1} + \frac{1}{3}g_{m-2} + \cdots + \frac{1}{m+1}g_0, \quad m \geq 0. \end{cases} \tag{38}$$

We solve the K-S equation employing 64 Fourier spectral modes and integrate from $t = 0$ to $t = 150$ (nondimensional time units) using the EDTRK4 time stepping.

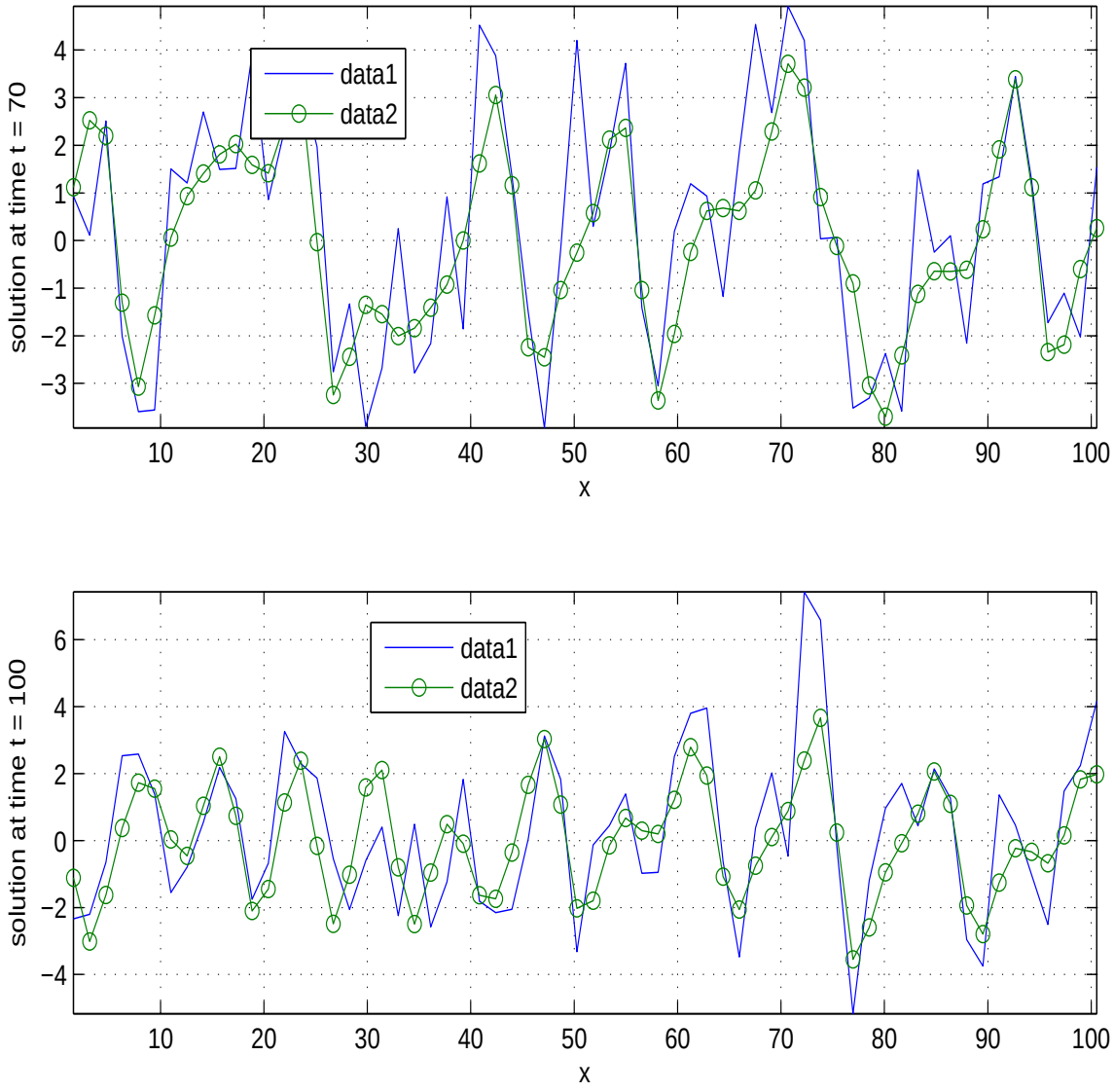

Figure 1: Background and EnKF analysis solutions at different locations and different times

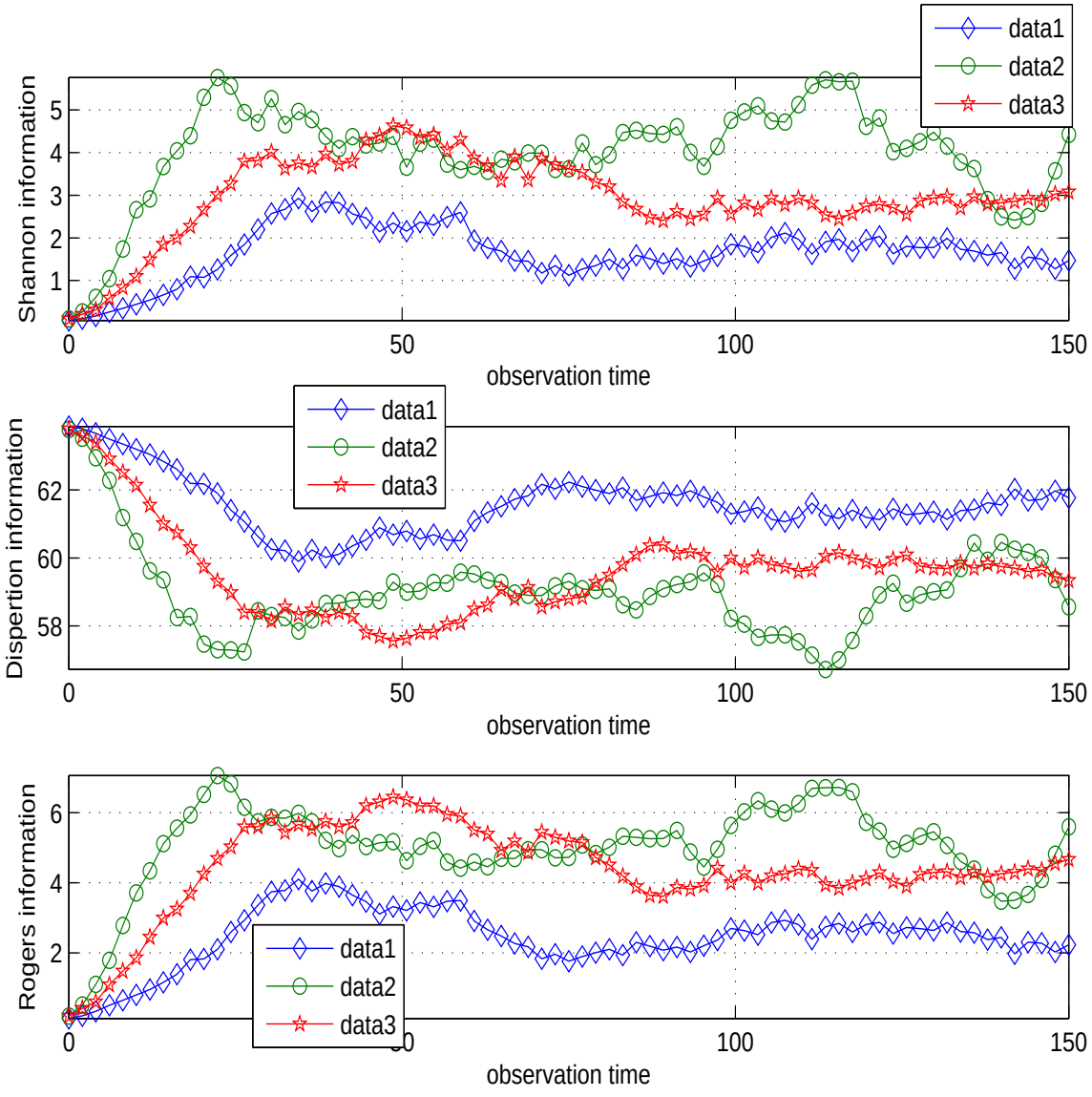

Figure 2: Model parameters effect

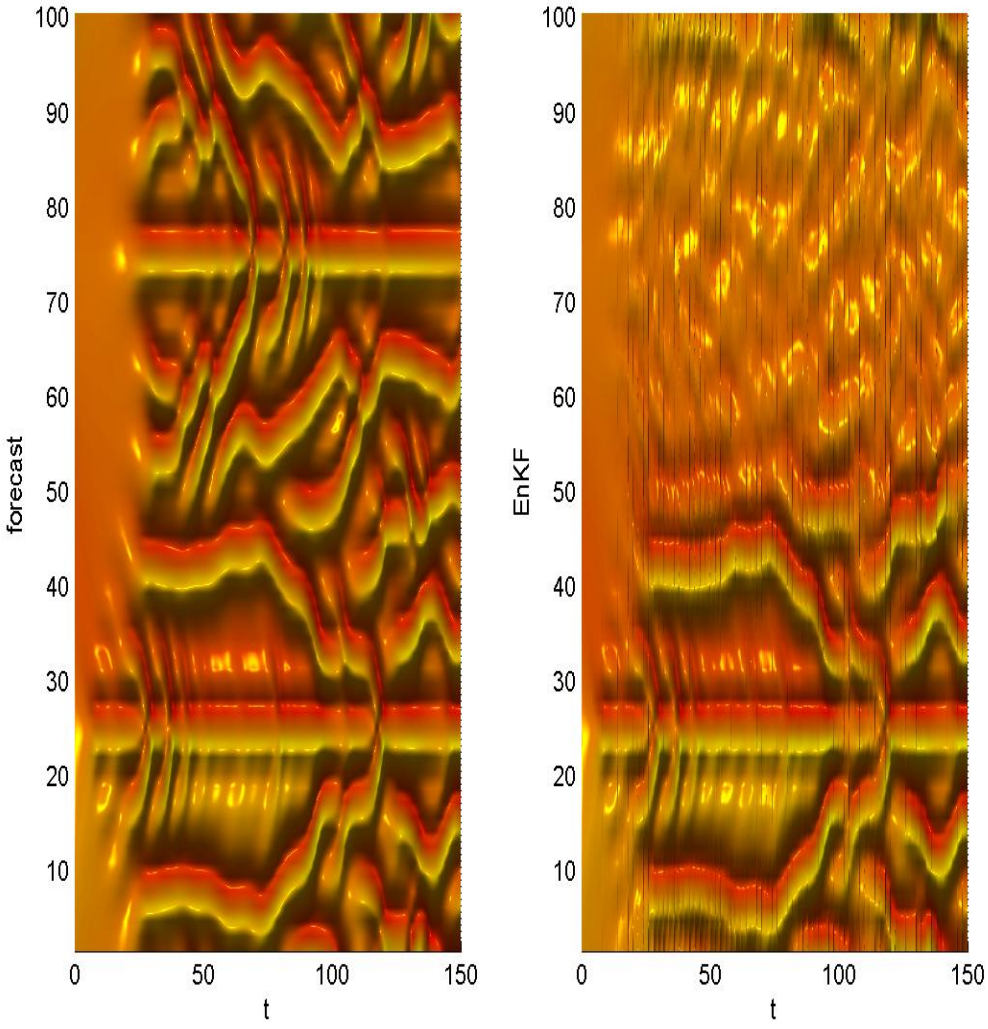

Figure 3:

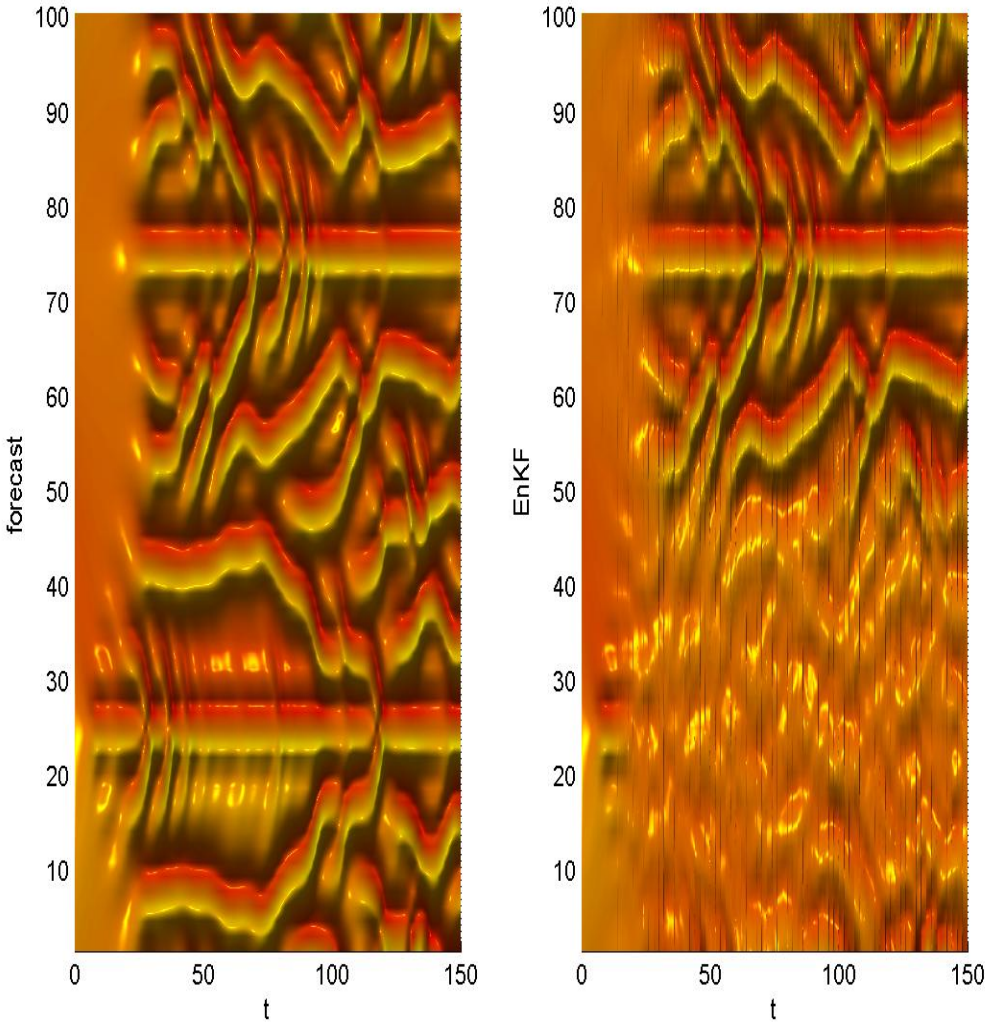

Figure 4:

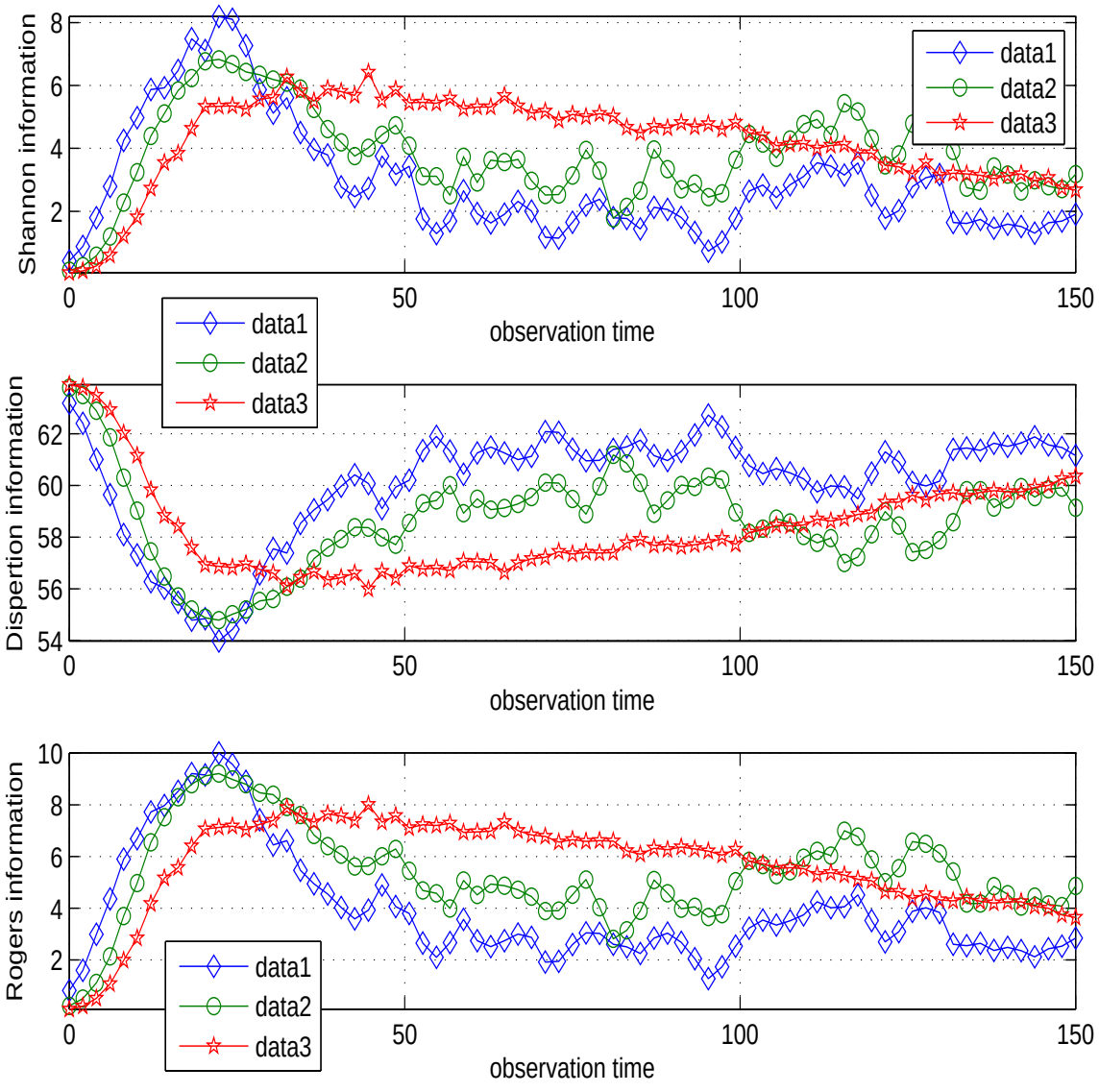

Figure 5: Effect of the number of observation

References

[1] Kuramuto Y, Tsuzuki T. On the formation of dissipative structures in reaction-diffusion systems. *Prog. Theor. Phys.* 1975; **54**:687–699

[2] Kuramuto Y, Tsuzuki T. Persistent propagation of concentration waves in dissipative media far from thermal equilibrium. *Prog. Theor. Phys.* 1976; **55**:356–369

[3] Sivashinsky GI. Nonlinear analysis of hydrodynamic instability in laminar flames I. Derivation of basic equations. *Acta Astrn.* 1977; **4**:1177–1206

[4] Sivashinsky GI. On irregular wavy flow of a liquid film down a vertical plane. *Progr. Theoret. Phys.* 1980; **63**:2112–2114

[5] Benny DJ. Long waves in liquid film. *J. Math and Phys.* 1966; **45**:150–155

[6] Babchin AJ., Frenkel AL, Levich BG, Sivashinsky GI. Nonlinear saturation of Rayleigh-Taylor instability in thin films. *Phys. Fluids.* 1983; **26**:3159–3161

[7] Kassam AK, Trefethen LN. Fourth-order time-stepping for stiff PDS's. *SIAM J.Sci.Compt.* 2005; **26**(4): 1214–1233

[8] Krogstad S. Generalized integrating factor methods for stiff PDE's. *J.Comput.Phys.* 2005; **203**: 72–88

[9] Trefethen NL. *Spectral methods in Matlab*, Software. Environments. Tools , SIAM, Philadelphia, 2000

[10] Cox SM, Matthews PC. Exponential time differencing for stiff systems. *J. Comput. Phys.* 2002; **176**: 430–455

[11] Kalman RE. A new approach to linear filtering and prediction problems *Transaction of the ASME - Journal of basic Engineering* 1960; **Series D**(82):35–45

[12] Evensen G. Sequential data assimilation in non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics *J. Geophys. Res.* 1994; **99**(C5 10):108–129

[13] Burgers G, van Leeuwen PJ, Evensen G. Analysis scheme in the ensemble Kalman Filter *Mon. Weather Rev.* 1998; **126**:1719–1724

[14] Evensen G. The Ensemble Kalman Filter: theoretical formulation and practical implementation *Ocean Dynamics* 2003; **53**:343–367

[15] Evensen G. *Data Assimilation: The ensemble Kalman filter* Springer: Berlin, 2007.

[16] Mandel J. A brief Tutorial on the ensemble Kalman filter manuscript,2007.

[17] Mandel J. Efficient Implementation of the Ensemble Kalman Filter *CCM Report* 2006; **231**

[18] Jazwinski A H. *Stochastic processes and filtering theory* Academic Press: New York, 1970.

[19] Bucy RS. Recurrent sets *Ann. Math. Statist.* 1965;**36**: 535–545.

[20] Anderson JL, Anderson SL. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts *Mon. Wea. Rev.* 1999;**127**:2741–2785.

[21] Houtekamer P, Mitchell M. Data assimilation using an ensemble Kalman filter technique *Mon. Wea. Rev.* 1998; **126**: 796–811

[22] Hamill TM, Snyder C. A Hybrid Ensemble Kalman Filter-3D Variational Analysis Scheme," *Mon.Wea.Rev.* 2000; **128**: 2905–2919

[23] Hamill TM. Ensemble-based atmospheric data assimilation *Predictability of Weather and Climate,* Cambridge Press: 2006; 124–156

[24] Gaspari G, Cohn SE. Construction of correlation functions in two and three dimensions *Quart. J. Roy.Meteor.Soc.*1999;**125**: 723–757

[25] Ehrendorfer E. A review of issues in ensemble-based Kalman filtering *Meteorologische Zeitschrift* 2007;**16**: 795-818

[26] Lewis JL, Lakshivarahan S, Dhall S K. *Dynamic Data Assimilation: A least squares approach* in the series on Encyclopedia of Mathematics and its Applications **Volume 104** Cambridge University Press:2006

[27] Fisher R. A. On the Mathematical Foundations of Theoretical Statistics *Philosophical Transactions of the Royal Society of London. Series A* 1922, **222**, 309–368 URL: http://www.jstor.org/stable/91208

[28] Abramov R. V. and Majda A. J. quantifying uncertainty for non-Gaussian ensemble in Complex Systems *SIAM J. SCI. COMPUT.* 2004, **26**, No. 2, 411–447

[29] Majda A. J. and Wang X. *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows.* Cambridge University Press, 2006

[30] Lorenz E. Predictability: A problem partly solved, *Proceedings of the Seminar on Predictability*, 1996Shinfield Park, Reading, UK, ECMWF

[31] Lorenc A. C. Analysis methods for numerical weather prediction, *Q. J. Roy. Meteor. Soc.* 1986; **112**: 1177-1194

[32] Rodgers C. D. Retrieval of atmospheric temperature and composition from remote measurements of thermal radiation. *Rev. Geophys. and Space Phys.* 1976; **14**, 609–624.

[33] Rodgers C. D. Information content and optimisation of high spectral resolution measurements. *Advances in Space Research* 1998; **21**, 361–367

[34] Rodgers C. D. *Inverse Methods for Atmospheric Sounding. Theory and Practice* World Scientic: Singapore, 2000.

[35] Shannon C. E. and Weaver W *The Mathematical Theory of Communication.* University of Illinois Press, Urbana, IL. 1949

[36] Kullback S. *Information Theory and Statistics*. Wiley, New York, New York, 1968.

[37] Rabier F., Fourrie N., Djalil C., and Prunet P. Channel selection methods for Infrared Atmospheric Sounding Interferometer radiances. *Quart. J. Roy. Meteor. Soc.* 2002; **128**, 1011–1027.