Estimating the Impact of Observational Data on Forecasts Made by Coupled Nonlinear Dynamical System

SERGEI SOLDATENKO[1] and DENIS CHICHKINE[2]
[1]Center for Australian Climate and Weather Research
700 Collins Street, Melbourne, VIC 3008
AUSTRALIA
s.soldatenko@bom.gov.au
[2]University of Waterloo
200 University Avenue West, Waterloo, ON N2L 3G1
CANADA

Abstract: In this paper, the adjoint-base technique is applied to estimate the observational impact on the forecast quality using the coupled nonlinear chaotic dynamical system obtained by combining two (fast and slow) versions of the Lorenz system. The exploration of forecast sensitivity with respect to observations and the estimation of the observational impact on the forecast quality require considerable computational resources. For simple-enough low-order models, the computational cost is minor and therefore models of this class can be used as simple test instruments to emulate more complex systems. The method is incorporated into the existing four-dimensional variational data assimilation system. The objective function represents the difference between total energy norms of the background and calculated forecast error. Since data assimilation procedure shadows the "true" state of a system given by the observational data, the shadowing property of coupled nonlinear chaotic dynamical system is considered for weak and strong coupling between fast and slow subsystems. The discussed approach can be used in designing the optimal observing network and estimating the value of various observations.

Key-Words: Dynamical system, chaotic dynamics, data assimilation, adjoint model, observation impact, sensitivity analysis

1 Introduction

Dynamical system theory provides a very powerful and comprehensive foundation for exploring, predicting, explaining and understanding numerous processes and phenomena occurring in nature and society [1]. The application areas of the dynamical systems theory are multidisciplinary and very diverse, and cover geosciences, biology, physics, chemistry, engineering, finance and other areas of human intellectual activities. Generally, any abstract dynamical system can be considered as a pair $\left(X,\varphi^t\right)$, where X is the system phase (or state) space, and is $\varphi^t : X \to X$ is a family of evolution operators parameterized by a real variable $t \in T$, where T is a time set. It is usually assumed that the phase space is a complete metric or Banach space, which can be either finite- or infinite-dimensional. A family of operators forms a semigroup, therefore:

$$\varphi^{t+\tau} = \varphi^t \circ \varphi^\tau,\ \varphi^0 \equiv I,\ \ \forall t,\tau \in T,$$

where I is the identity operator.

Suppose that $t \in \mathbb{R}_+$, then a continuous time dynamical system can be generated by the following set of autonomous ordinary differential equations

$$\dot{x} = f\left(x\right) \tag{1}$$

with given initial conditions

$$x\left(0\right) = x_0, \tag{2}$$

where $x \in X$, f is the (nonlinear) vector-valued function, and x_0 is a given initial state of the system. Since we are usually not able to solve equations (1) analytically, then the set of infinite-dimensional equations (1) has to be truncated by some means to finite-dimensional approximate model, for which a solution can be sought numerically. Applying either a projection onto a finite set of basic functions or a discretization in time and space, one can derive a discrete model, which approximates the system (1) and can be solved numerically if the initial conditions are specified.

It is known that computer simulation has various sources of errors including initial condition errors. During the numerical integration of the model

equations, errors present in the initial conditions usually tend to grow at rates that depend on the system instability. Hence, even small errors in the initial conditions can produce enormous inaccuracy in the model output results. This is especially true for nonlinear dynamical systems, which under certain conditions can exhibit the chaotic behavior. In many applications initial conditions are specified based on observations obtained from various devices including remote sensors that measure not only quantities of interest directly represented in the model, but also some quantities, which then can be used to retrieve required physical variables. To combine observations coming from various sources and incorporate them into a dynamical model, using the model's equations, procedures known as data assimilation can be applied [2, 4]. One of the most advanced and effective data assimilation techniques is the four-dimensional variational data assimilation (4D-var) [5-7]. The 4D-var is mathematically formulated as an optimization problem, in which the model equations are considered to be constraints, and initial conditions represent control variables [8].

The accuracy of computed initial conditions strongly depends, at least in an statistical terms, on the volumes and quality of observational data. The information used in data assimilation comes from different sources of observations, the reliability of which varies. However, the value added by different types of observations can vary significantly. Therefore, it is important to estimate the contribution of various types of observations to the accuracy of initial conditions and, furthermore, to the forecast quality of dynamical system evolution. This is very important not only for assessing the impact of each subset of observations on the forecast quality, but also for designing and developing optimal and effective observing networks.

The obvious approach to solving this problem is the so-called observing system experiment (OSE) [9-11]. The main idea behind this method can be illustrated as follows. Suppose we obtain the future state of a dynamical system by integrating the model equations from the initial time t_0 to a certain verification time t^f. For this experiment, the initial conditions were calculated via data assimilation utilizing all types of observations y^o. Let e^f be a quantitative measure of the forecast quality. Suppose that for the first experiment the calculated estimate of the forecast quality is e_1^f. In the second model run, data assimilation procedure utilizes all types of observations excluding y_s^o. The forecast quality of the second experiment is characterized by e_2^f. The difference between e_1^f and e_2^f faithfully quantifies the impact of observations y_s^o on the forecast skill. The observing system experiment, however, is computationally expensive. The use of the adjoint-based technique [12] allows assessing the impact of any or all available observations in a computationally efficient manner. This approach is very convenient since an adjoint model is one of the main parts of variational data assimilation schemes. Observation impact is calculated using, firstly, the sensitivity functions which are the adjoint sensitivity gradients of a certain cost function that characterizes the forecast errors [13], and, secondly, using the differences between observations and first-guess estimate of the dynamical system state projected into the observation space, which is known as innovations.

Quite often, dynamical systems applied in various applications (e.g. physics, engineering, economics, biology, meteorology and climate studies) are nonlinear and under certain conditions can exhibit a chaotic motion. The dynamics of these systems can be effectively emulated by low-dimensional models that describe the evolution of a reduced set of variables. For these simplified low-order models, computational costs are minor, however, obtained results are very beneficial from both the theoretical and practical viewpoint. The objective of this paper is as follows: using a coupled nonlinear dynamical system obtained by combining the fast and slow versions of the three-parameter Lorenz system, to consider the application of adjoint-based algorithm to calculate the forecast sensitivity with respect to observations and to estimate the impact of various subsets of observations on the forecast quality. The mathematical and computational instrument developed in this paper can be very helpful for exploring various aspects of numerical modeling and predicting the behaviour of complex dynamical systems arising, for example, in geophysical, environmental, biological, engineering and other branches of science.

2 Shadowing in Data Assimilation

One of the most rapidly developing components of the global theory of dynamical systems is the theory of pseudo-orbit shadowing in dynamical systems [14, 15]. Pseudo (or approximate) trajectories arise due to the presence of round-off errors and other truncation errors in computer simulation of dynamical systems. Consequently, in numerical

modeling we can compute a trajectory that comes very close to an exact solution and the resulting approximate solution will be a pseudotrajectory. The shadowing property (or POTP, pseudo-orbit tracing property) means that near an approximate trajectory there exists the exact trajectory of the system under consideration, such that it lies uniformly close to a pseudotrajectory. The shadowing property of dynamical systems is a fundamental feature of hyperbolic systems [16, 17].

Understanding the POTP is very important in data assimilation since data assimilation procedures shadow the true state of a system assuming that observations provide the truth. The formal definition of shadowing is as follows. Let (*X*, *d*) be a metric space, where $X \subseteq \mathbb{R}^n$ and *d* is a distance function. Let us consider a discrete-time dynamical system, generated by iterating a map $f : X \to X$, that is,

$$x_{k+1} = f\left(x_k\right), \ k \in \mathbb{Z}_+ . \tag{3}$$

Given the system state $x_0 \in X$ at time $t = 0$, we define the trajectory of x_0 under *f* to be the sequence of points $\left\{x_k \in X : k \in \mathbb{Z}_+\right\}$ such that $x_k = f^k\left(x_0\right)$, where f^k indicates the *k*-fold composition of f with itself, and $f^0\left(x\right) \equiv x$. Thus, given the map f and the initial condition x_0, equation (3) uniquely specifies the orbit of a dynamical system. If x_k is the state of dynamical system at time t_k, then the correct state at the next time t_{k+1} is given by $f\left(x_k\right)$. In practice, however, there is a discrepancy between x_{k+1} and $f\left(x_k\right)$ due to round-off errors and other truncation errors. Consequently, in computer simulations we can compute only a pseudo-orbit (or approximate trajectory) which can be shadowed by a real orbit. A set of points $\left\{x_k \in X : k \in \mathbb{Z}_+\right\}$ is a δ -pseudotrajectory ($\delta > 0$) of f if

$$d\left(x_{k+1}, f\left(x_k\right)\right) < \delta, \ k \in \mathbb{Z}_+ . \tag{4}$$

We say that f satisfies shadowing property if given $\varepsilon > 0$ there is $\delta > 0$ such that for any δ -pseudo-orbit $\left\{x_k\right\}_{k=o}^{\infty}$ there exists a corresponding trajectory $\left\{y_k\right\}_{k=o}^{\infty}$ such that

$$d\left(x_k, y_k\right) < \varepsilon, \ y_k = f^k\left(y\right), \ k \in \mathbb{Z}_+ . \tag{5}$$

The shadowing property plays a crucial role in the exploration of the stability of system dynamics. One of the most important theoretical results for hyperbolic dynamical systems is the shadowing lemma [18], which states that for each nonzero distance $\varepsilon > 0$, there exists $\delta > 0$ such that each δ -pseudotrajectory can be ε -shadowed. Informally, the shadowing lemma states that each pseudo-orbit stays uniformly close to a certain true trajectory.

The definition of pseudotrajectory and shadowing lemma for flows (continuous dynamical systems) are more complicated than for discrete dynamical systems [14, 15]. Let $f : X \to X$ be a vector field on *X* of class C^1. Let $\varphi : \mathbb{R} \times X \to X$ be a flow on *X* generated by *f*. A function $x : \mathbb{R} \to X$ is called a δ -pseudotrajectory if

$$d\left(\varphi\left(t, x\left(\tau\right)\right), x\left(\tau + t\right)\right) < \delta \tag{6}$$

for any $0 \le t \le 1$ and $\tau, \tau + t \in \mathbb{R}$. The "continuous" shadowing lemma ensures that for the vector field *f* generating the flow $\varphi(t, x)$, the shadowing property holds in a small neighborhood of a compact hyperbolic set for dynamical system $\varphi(t, x)$. However, the shadowing problem for continuous dynamical systems requires reparameterization of shadowing trajectories because for continuous dynamical systems close points of pseudo-orbit and true trajectory do not correspond to the same moments of time. A monotonically increasing homeomorphism $h : \mathbb{R} \to \mathbb{R}$ such that $h\left(0\right) = 0$ is called a reparameterization and denoted by *Rep*. For $\varepsilon > 0$, *Rep*(ε) is defined as follows [15]:

$$Rep\left(\varepsilon\right) = \left\{ h \in Rep : \left| \frac{h\left(t_1\right) - h\left(t_2\right)}{t_1 - t_2} - 1 \right| \le \varepsilon \right\}.$$

for any $t_1, t_2 \in \mathbb{R}$.

However, most physical systems are non-hyperbolic. Despite the fact that much of shadowing theory has been developed for hyperbolic systems, there is, however, evidence that non-hyperbolic dynamical systems also have shadowing property showing features of hyperbolic systems. In theory, this property should be verified for each particular dynamical system, but this is more easily said than done.

3 Variational Data Assimilation

Four dimensional variational data assimilation problem aims to define the initial condition estimate x^a, which is known as the analysis, such that the prior "first guess" information, which is usually the model state given by the previous forecast – the background state x^b, best fits the observations y^o within the finite time interval $\left[t_0, t_K\right]$, known as a data assimilation window. The background trajectory of the system $\left\{x_k^b\right\}_{k=0}^{K}$ is generated by the set of discrete-time model equations

$$x_{k+1} = \mathcal{M}_{0,k+1}(x_0),\ k = 0,\ldots,K-1, \qquad (7)$$

where the nonlinear operator $\mathcal{M}_{0,k+1}$ propagates the initial conditions to time t_k.

Suppose that the prior information is characterized by the probability density function $f^b(x)$. Then a background state x^b can be considered as the expected value of the probability distribution $f^b(x)$, i.e. $x^b = \mathrm{E}^b(x)$, where E is the statistical expectation operator, which is understood in accordance with the rules of multivariate statistical analysis as $\mathrm{E}(x) = \left[\mathrm{E}(x_1),\ldots,\mathrm{E}(x_n)\right]^{\mathrm{T}}$. Thus, we can write

$$x^b = x^t + \varepsilon^b.$$

Here ε^b is the random background errors, which are commonly considered to be unbiased and normally distributed:

$$\varepsilon^{\mathrm{b}} \in \mathcal{N}(0, \mathrm{B}),$$

where $\mathrm{B} = \mathrm{E}\left(\varepsilon^b \left(\varepsilon^b\right)^{\mathrm{T}}\right) \in \mathbb{R}^{n\times n}$ is the background error covariance matrix.

Let y^o be the vector of K observations measured at a discrete time $[t_0, t_K]$, k=0,…,K. Suppose that observations are linked to the system state by the nonlinear observation operator $\mathcal{H}$ and are influenced by random errors $\varepsilon_k^{\mathrm{o}}$, which are assumed to be white Gaussian noise:

$$y_k^{\mathrm{o}} = \mathcal{H}_k(x_k) + \varepsilon_k^{\mathrm{o}},\ \ \varepsilon_\kappa^o \in \mathcal{N}(0, \mathrm{R}_k),\ k = 0,\ldots,K,$$

where $\mathrm{R}_k = \mathrm{E}\left(\varepsilon_k^0 \left(\varepsilon_k^0\right)^{\mathrm{T}}\right) \in \mathbb{R}^{K\times K}$ is the observation error covariance matrix corresponding to a moment in time t_k. Observation errors at different times are assumed to be uncorrelated, i.e. $\mathrm{E}\left(\varepsilon_k^0 \left(\varepsilon_l^0\right)^{\mathrm{T}}\right) = 0$ if $k \neq l$.

The 4D-var data assimilation is considered as an optimal control problem in which initial conditions x_0 play the role of control variables. Given the observations y_k^o at time t_k and the corresponding observation error covariance matrices R_k, as well as the background initial state x_0^b and the error covariance matrix B_0, the 4D-var data assimilation seeks to minimize, with respect to x_0, a certain cost function $\mathcal{J}(x)$ expressing the misfit between observations and corresponding model state using the model equations as constraints:

$$x_0^a = \arg\min \mathcal{J}(x_0) \qquad (8)$$

subject to x satisfying the set of model equations (7).

Under the assumption that both the background errors and the observation errors are normally distributed, the minimization cost function can be written as

$$\mathcal{J}(x_0) = \underbrace{\frac{1}{2}\left\|x_0 - x_0^b\right\|_{\mathrm{B}_0}^2}_{\mathcal{J}^b} + \underbrace{\frac{1}{2}\sum_{k=0}^{K}\left\|\mathcal{H}_k(x_k) - y_k^0\right\|_{\mathrm{R}_k}^2}_{\mathcal{J}^o}. \qquad (9)$$

Thus, the optimization problem (8) is nonlinear with strong constraints, and a certain iterative minimization algorithm (for example, the gradient-based technique) is needed to obtain the solution, which is the minimum of the cost function. This requires the estimation of the cost function gradient with respect to the initial conditions:

$$\nabla_{x_0}\mathcal{J}(x_0) = \nabla_{x_0}\mathcal{J}^b(x_0) + \nabla_{x_0}\mathcal{J}^o(x_0).$$

Here

$$\nabla_{x_0}\mathcal{J}(x_0) = \mathrm{B}_0^{-1}\left(x_0 - x_0^b\right),$$

$$\nabla_{x_0}\mathcal{J}^o(x_0) = \sum_{k=0}^{K}\mathrm{M}_{0,k}^{\mathrm{T}}\mathrm{H}_k^{\mathrm{T}}\mathrm{R}_k^{-1}\left(\mathcal{H}_k(x_k) - y_k^o\right).$$

where $\mathrm{M}_{0,k}^{\mathrm{T}}$ is the adjoint of the linearized model operator $\mathrm{M}_{0,k} = \mathcal{M}'_{0,k}(x_k)$, known as a tangent linear (TL) model, and $\mathrm{H}_k^{\mathrm{T}}$ is the adjoint of the linearized observation operator $\mathrm{H}_k = \mathcal{H}_k'(x_k)$.

Let us emphasize that the performance of 4D-var data assimilation schemes depends on their key information components, such as the available observations y_k^o, estimates of the observation and background error covariances that are quantified by the matrices R_k and B_0, as well as the background state x^b. This means that y_k^o, R_k, B_0 and x^b strongly impact the accuracy of calculated initial conditions, thus influencing the forecast quality. In this context the assessment of model forecast sensitivity with respect to parameters of the data assimilation scheme represents one of the critical components of building and evaluating mathematical models.

4 Testing Data Assimilation Scheme

Testing the TL model and its adjoint is required to ensure the convergence of the minimization algorithm in data assimilation procedures. If $\zeta\delta x$ is a small perturbation of the model state, then

$$\mathcal{M}(x + \zeta\delta x) - \mathcal{M}(x) \approx \mathrm{M}(x)\zeta\delta x.$$

To verify the applicability of TL model on the finite time interval [t_0, t_K], the ratio [4]

$$e_R = \frac{\mathcal{M}(x+\zeta\delta x)-\mathcal{M}(x)}{\mathrm{M}(x)\zeta\delta x}$$

should be calculated. The TL model is valid if $e_R \to 1$ when $\zeta \to 0$. The results of numerical experiments showed that the TL model passed this test with e_R tending towards unity (Table 1).

Table 1. Results of verification of tangent linear model for $c = 0.8$ and $\delta x = 10^{-2} \times x_0$.

ζ	e_R
1	0.9182066027544249
10^{-1}	0.9997279965782743
10^{-2}	0.9999925468155463
10^{-2}	0.9999991929611531
10^{-4}	0.9999999534965012
10^{-5}	0.9999999911217883
10^{-6}	0.9999999914427087
10^{-7}	0.9999997435447022

Table 2. Results of verification of 4D-Var cost function gradient for $c = 0.8$ and $\delta x = 10^{-2} \times x_0$.

ζ	$\Psi(\zeta)$
10^{-4}	0.8727731981461396
10^{-5}	0.9975483775420343
10^{-6}	0.9998765512756632
10^{-7}	0.9999884562844123
10^{-8}	0.9999979865432855
10^{-9}	0.9999998912431426
10^{-10}	0.9999999244103234

The correctness of TL adjoint can be tested by verification of the inner product identity [4, 19]

$$\langle \mathrm{M}\delta x, \mathrm{M}\delta x\rangle = \langle \delta x, \mathrm{M}^{\mathrm{T}}\mathrm{M}\delta x\rangle .$$

It was found that this equality is essentially correct: the difference was observed only in the 7th digit, which is consistent with a round-off error. The second test to verify the adjoint model is the so-called gradient test [19], which aims to compare a finite difference representation of the gradient of 4D-var cost function (9) with the gradient obtained via adjoint model $\nabla J(x_0)$. A linear Taylor approximation of the cost function can be written as

$$J(x_0+\zeta\delta x) \approx J(x_0) + \zeta(\delta x)^{\mathrm{T}} \nabla J(x_0).$$

Let us introduce the following function [19]

$$\Psi(\zeta) = \frac{J(x_0+\zeta\delta x)-J(x_0)}{\zeta(\delta x)^{\mathrm{T}} \nabla J(x_0)}.$$

If the gradient is estimated correctly then the function $\Psi(\zeta) \to 1$ as $\zeta \to 0$. The perturbation vector δx is taken to be [19]

$$\delta x = \frac{\nabla J(x_0)}{\|\nabla J(x_0)\|},$$

where $\|\cdot\|$ is the L_2 norm. Table 2 manifests the success of the gradient test.

To examine the TL and adjoint models, we considered the dynamics of system on its attractor, which is obtained by numerical integration of the equation (24) (see below) started at $t_D = -20$ with the initial conditions

$$x(t_D) = (0.01,\ 0.01,\ 0.01,\ 0.02,\ 0.02,\ 0.02)^{\mathrm{T}}$$

and finished at $t_0 = 0$ to guarantee that the calculated model state vector $x_0 = x(0)$ is on the model attractor.

5 Observation Impact Estimation

The observation impact means variations in a certain metric of forecast quality caused by some changes in observational data. Assessing the impact of various observations on the forecast quality is a two-step procedure [12]: firstly, the sensitivity functions are computed, which are the adjoint sensitivity gradients of a certain cost function that characterizes the forecast errors, and, secondly, the desired observation impact is calculated by multiplying the obtained sensitivity functions on the innovations $y^o - \mathrm{H}x^b$. Let J^f be a metric of the forecast quality. The sensitivity functions required to assessing the observation impact, are the components of the gradient ∇J^f calculated with respect to the vector of observations y^o. The cost function is defined as [12]:

$$J^f = \frac{1}{2} e^f . \qquad (10)$$

Here e^f is a scalar forecast error norm

$$e^f = \langle (x^f - x^a),\ \mathrm{C}(x^f - x^a)\rangle , \qquad (11)$$

where x^a is the analysis at verification time t^f, C is a matrix of weighting coefficients, which gives e^f units of energy per unit mass, and $\langle \cdot,\cdot \rangle$ is a scalar product. The variable e^f can be interpreted as a total energy norm of the forecast error. From (10) and (11) we can obtain

$$\frac{\partial J^f}{\partial x^f} = \mathrm{C}(x^f - x^a).$$

To estimate the sensitivity functions $\partial J^f/\partial y^o$ we first need to calculate sensitivities with respect to the initial conditions of the predicted trajectory of dynamical system:

$$\frac{\partial J^f}{\partial x_0}=\mathrm{M}^{\mathrm{T}}\frac{\partial J^f}{\partial x^f}. \qquad (12)$$

Essentially, using sensitivity functions (12) we can study the influence of errors in the initial conditions on the forecast error cost function. The desired sensitivity functions are calculated as follows [12]:

$$\frac{\partial J^f}{\partial y^o}=\left[\mathrm{HB_0H^{\mathrm{T}}+R}\right]^{-1}\mathrm{HB_0}\frac{\partial J^f}{\partial x_0}. \qquad (13)$$

These functions provide the ability to estimate how the cost function J^f would be modified due to errors in various observations. Note that the gradient $\partial J^f/\partial x_0$ is evaluated on a gridded model space, while the gradient $\partial J^f/\partial y^o$ is estimated on an observation space.

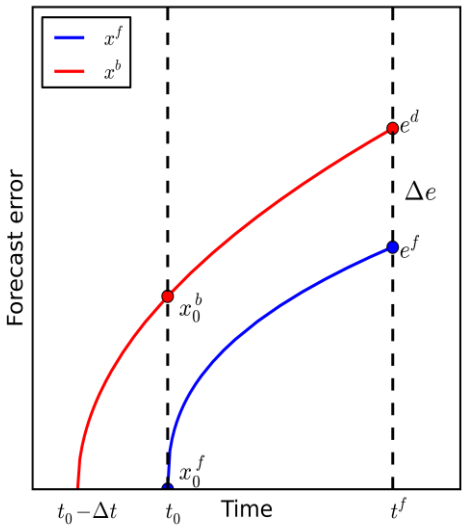

Fig. 1. Schematic representation of the discussed technique for observation sensitivity and impact assessment [12]

Using sensitivity functions $\partial J^f/\partial y^o$, we can calculate observation impact on the model forecast using the following approach [12]. Suppose we run the model twice to obtain two forecasts of length *f* and *d*. The verification time is the same to both of the forecasts. At the verification time t^f, a verifying analysis x^a is known. Forecast x^f starts from the initial conditions x_0, and forecast x^d starts Δt earlier than x^f, where Δt is a certain time interval. Usually this interval corresponds to the length of data assimilation window. In the numerical weather prediction, for example, the length of Δt is 6 or 12 hours. The forecast x^d can be considered as the background state x^b for the initial conditions x_0, which are used for x^f. Thus, hereinafter in this paper, we will write x^b instead of x^d. The difference $\Delta e=e^f-e^d$ characterizes the forecast error reduction due solely to observations. If the assimilated observations improve the forecast skill at the verification time t^f, then the forecast error is reduced, and the value Δe will be negative. However, if the assimilated observations diminish the forecast quality, the value Δe will be positive. Since

$$e^f=\left\langle\left(x^f-x^a\right),\mathrm{C}\left(x^f-x^a\right)\right\rangle, \qquad (14)$$

$$e^d=\left\langle\left(x^b-x^a\right),\mathrm{C}\left(x^b-x^a\right)\right\rangle, \qquad (15)$$

then

$$J^f=\frac{1}{2}e^f,\quad J^b=\frac{1}{2}e^b \qquad (16)$$

and

$$\frac{\partial J^f}{\partial x^f}=\mathrm{C}\left(x^f-x^a\right),\quad \frac{\partial J^b}{\partial x^b}=\mathrm{C}\left(x^b-x^a\right). \qquad (17)$$

Taking into account (17), from (14) and (15), we can obtain the expression for Δe:

$$\Delta e=\left\langle\left(x^f-x^b\right),\left(\frac{\partial J^f}{\partial x^f}+\frac{\partial J^b}{\partial x^b}\right)\right\rangle. \qquad (18)$$

Since the adjoint model used in 4D-var data assimilation schemes is linear, we can obtain the following approximate equation for calculating the Δe estimate [12]:

$$\delta e=\left\langle\delta x_0,\left(\frac{\partial J^f}{\partial x_0^f}+\frac{\partial J^b}{\partial x_0^b}\right)\right\rangle, \qquad (19)$$

where $\delta x_0=x_0^f-x_0^b$. An estimate of the forecast errors δe is evaluated in the grid point space of the model. However, in practice, the evaluation of δe is carried out in the observation space, because the vector of model state x is usually several orders of magnitude larger than that of the observations y^o. Since [2, 3]

$$\delta x_0\equiv x_0-x_0^b=\mathrm{K}\left(y^o-\mathrm{H}x_0^b\right),$$

where $\mathrm{K=B_0H^{\mathrm{T}}}\left[\mathrm{HB_0H^{\mathrm{T}}+R}\right]^{-1}$ is the Kalman gain matrix, the equation (19) can be rewritten as

$$\delta e=\left\langle\mathrm{K}\left(y^o-\mathrm{H}x_0^b\right),\left(\frac{\partial J^f}{\partial x_0^f}+\frac{\partial J^b}{\partial x_0^b}\right)\right\rangle. \qquad (20)$$

The operator K is substantially linear, thus we can rewrite the equation (20) by introducing the adjoint operator K^{T} satisfying $\left\langle\mathrm{K}y^o,x\right\rangle=\left\langle y^o,\mathrm{K}^{\mathrm{T}}x\right\rangle$:

$$\delta e = \left\langle \left(y^o - \mathrm{H}x_0^b \right),\ \mathrm{K}^{\mathrm{T}} \left(\frac{\partial J^f}{\partial x_0^f} + \frac{\partial J^b}{\partial x_0^b} \right) \right\rangle. \qquad (21)$$

The operator K^{T} is the adjoint operator of 4D-var data assimilation schemes that determines sensitivity functions in the observation space:

$$\frac{\partial J_b^f}{\partial y^o} = \mathrm{K}^{\mathrm{T}} \left(\frac{\partial J^f}{\partial x_0^f} + \frac{\partial J^b}{\partial x_0^b} \right). \qquad (22)$$

The substitution of equation (22) into equation (21) gives the expression for calculation of δe which uses only observation space variables:

$$\delta e = \left\langle \left(y^o - \mathrm{H}x_0^b \right),\ \frac{\partial J_b^f}{\partial y^o} \right\rangle. \qquad (23)$$

This equation provides an estimate of δe made by any or all observations. Indeed, if the observation errors between various subsets of observations y_1^o, $y_2^o, \ldots, y_L^o$ are independent from each other, then the objective function J_b^f can be represented as

$$J_b^f = \sum_{l=1}^{L} \left(J_b^f \right)_l,$$

where the cost function $\left(J_b^f \right)_l$ characterizes the observation impact on the forecast error reduction owing to the *l*-th subset of the observations.

Figure 1 shows the schematic representation of the discussed approach for evaluating the forecast sensitivity with respect to observations and assessing the impact of various observations on the forecast skill.

6 Coupled Dynamical System and Its Shadowing Property

Consider the nonlinear dynamical system obtained by coupling two versions of the L63 system [20] with distinct time scales differing by a factor ε [21, 22], which is represented by the following two sets of differential equations that describe:

a) The fast subsystem

$$\begin{aligned} \dot{x} &= \sigma(y - x) - cX, \\ \dot{y} &= rx - y - xz + cY, \\ \dot{z} &= xy - bz + cZ, \end{aligned} \qquad (24a)$$

b) The slow subsystem

$$\begin{aligned} \dot{X} &= \varepsilon\sigma(Y - X) - cx, \\ \dot{Y} &= \varepsilon(rX - Y - aXZ) + cy, \\ \dot{Z} &= \varepsilon(aXY - bZ) - cz. \end{aligned} \qquad (24b)$$

Thus, the state vector of the coupled model used in this study is $x = (x, y, z, X, Y, Z)^{\mathrm{T}}$, where lower case letters represent the fast subsystem and capital letters – the slow subsystem. The parameter vector is $\alpha = (\sigma, r, b, c, \varepsilon)^{\mathrm{T}}$. Note that σ, r and b are the parameters of L63 model, c is a coupling strength parameters. The parameter values are taken as

$\sigma = 10,\ r = 28,\ b = 8/3,\ \varepsilon = 0.1,\ c \in [0.1;\ 1.2]$.

Chosen values of σ, r and b correspond to the chaotic behaviour of the L63 model. The parameter $\varepsilon = 0.1$ indicates that the slow system is 10 times slower than the fast system. The larger parameter c, the stronger the coupling between two systems.

Since essential dynamical, correlation and spectral properties of system (24) as well as numerical integration scheme were presented in [23], in this paper we consider the influence of coupling strength parameter on the system dynamics. This parameter controls the interactions between fast and slow subsystems affecting the qualitative changes in the coupled system dynamics.

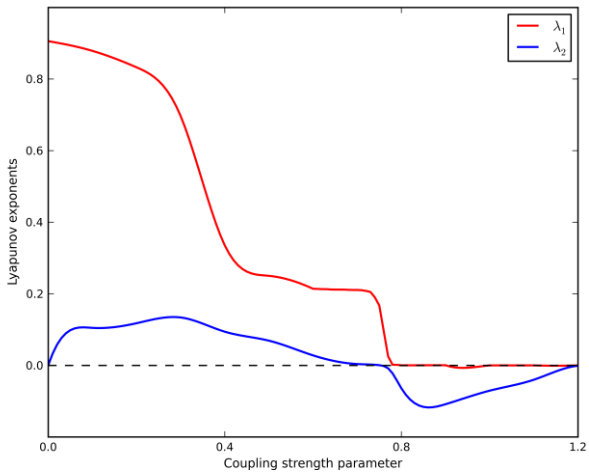

Fig. 2. Two largest conditional Lyapunov exponents as functions of coupling strength parameter

Qualitative changes in the dynamical properties of a system can be detected by determining and analyzing the system's spectrum of Lyapunov exponents. In the analysis of coupled dynamical systems we are dealing with conditional Lyapunov exponents that are normally used to characterize the synchronization with coupled systems. The largest Lyapunov exponent characterizes the average rate of exponential divergence (or convergence) of nearby trajectories in the phase space. The system (24) has six distinct exponents. If the parameter c tends to zero, then the system (24) has two positive, two zero and two negative Lyapunov exponents. The influence of parameter c on the two largest conditional Lyapunov exponents is illustrated in

Figure 2. The numerical experiments demonstrated that those, initially positive, exponents decrease monotonically with an increase in the parameter c. At about $c \approx 0.8$ they approach the x-axis and at about $c \approx 0.95$ - negative values. Thus, for $c > 0.95$ the dynamics of both fast and slow subsystems become phase-synchronous [24]. For $c > 1.0$, a limit circle dynamical regime is observed since all six exponents become negative.

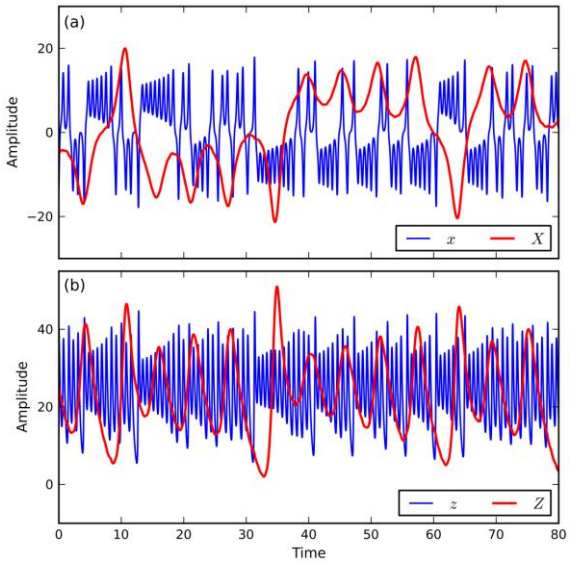

Fig. 3. Time evolution of fast and slow dynamic variables for c=0.15.

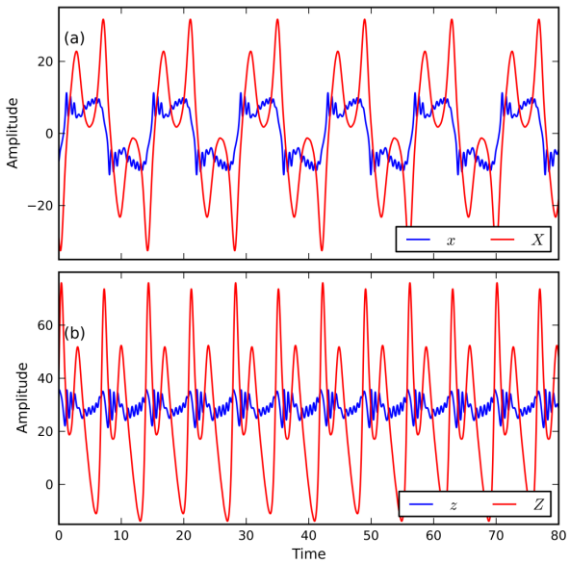

Fig. 4. Time evolution of fast and slow dynamic variables for c=0.8.

The calculated evolutions of the fast (x^t and z^t) and slow (X^t and Z^t) variables for weak and strong coupling are shown in Figure 3 and Figure 4 respectively. It is known that the L63 model produces chaotic oscillations of a switching type: the structure of its attractor contains two regions divided by the stabile manifold of a saddle point in the origin. For relatively small coupling strength parameter (c<0.5), the attractor for both fast and slow sub-systems maintains a chaotic structure, which is inherent in the original L63 attractor. As the parameter c increases, the attractor for both fast and slow sub-systems undergoes structural changes breaking the patterns of the original L63 attractor. Fast and slow subsystems affect each other through coupling terms, and at some value of the coupling strength parameter (c>0.5) a chaotic behaviour is destroyed and dynamic variables begin to exhibit some sophisticated motions which are not obviously periodic. Moreover, qualitative examination shows that the evolution through time of both subsystems becomes, to a large degree, synchronous (however, phase synchronization requires specific analysis which is not within the scope of this paper).

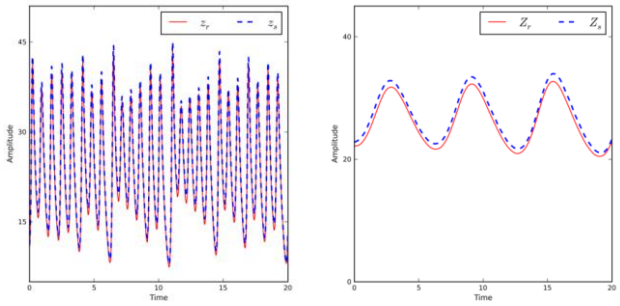

Fig. 5. Original orbit (in red) and pseudo-orbit (in blue) for fast z and slow Z variables for c=0.01.

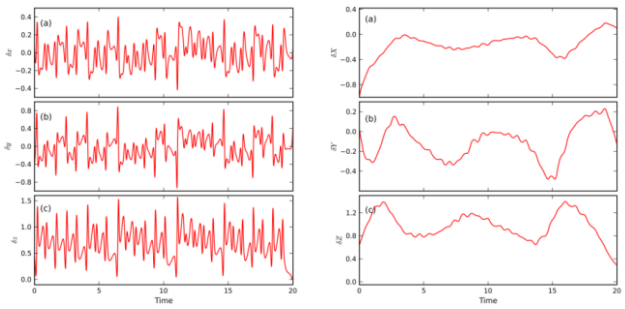

Fig. 6. Differences between variables that correspond to the original trajectory and pseudo-orbit for c=0.01.

Since data assimilation procedure shadows the "true" state of a system given by the observational data, it is important to know whether a system has the shadowing property. Using the approach that was developed for computing sensitivity coefficients in chaotic dynamical systems [25], we computed "pseudo-orbits" of the system (24) assuming that the numerical solution has no errors,

and "pseudo-orbits" were generated by variations in the system's parameters. We consider two sets of numerical experiments: weak coupling (c=0.01) and strong coupling (c=0.8) between fast and slow systems. Fast z and slow Z variables that correspond to the original and pseudo-orbits are shown in Figure 5 when the coupling strength parameter c=0.01. The differences between state variables corresponding to the original orbits and pseudotrajectories of the fast and slow systems are plotted in Figure 6. These figures show that the calculated pseudo-orbits are close to corresponding true trajectories over a specified time interval, demonstrating the shadowability. The strong coupling does not introduce significant qualitative and quantitative changes in the behavior of pseudo-orbits with respect to the true trajectories. The original and pseudo fast z and slow Z variables for c=0.8 are shown in Figure 7, and the differences between fast and slow state variables are presented in Figure 8.

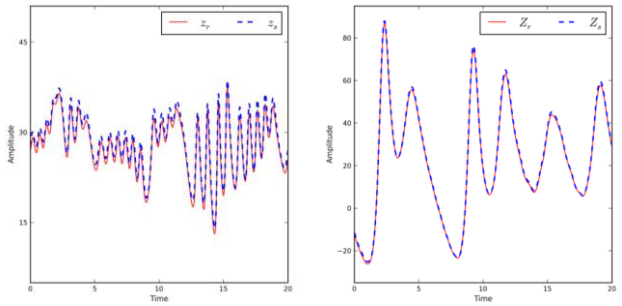

Fig. 7. Original orbit (in red) and pseudo-orbit (in blue) for fast z and slow z variables for c=0.8.

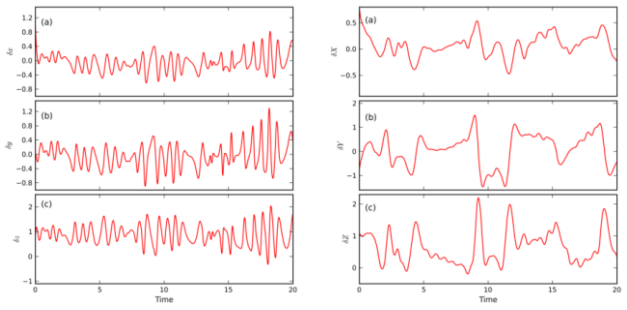

Fig. 8. Differences between variables that correspond to the original trajectory and pseudo-orbit for c=0.8.

7 Observation Impact Estimates

To estimate the impact of observations on the forecasts quality the following information is required:

(a) The "true" model state x^t, which is created by the numerical integration of equations (24) with the initial conditions x_0 taken on the system attractor;

(b) The background (first guess) forecast x^b, which is obtained by numerical integration of the model equations (24) with the initial conditions x_0^b;

(c) Observations, which are generated by adding the Gaussian random noise with zero mean and specified standard deviation σ_o to the true state.

The background state x_0^b at the initial time t_0=0 is defined as $x_0^b = x_0 + \delta x_0^b$, where δx_0^b is a random perturbation with standard deviation $\sigma_b = 0.2$ applied to all elements of the state vector. Observations are defined for every $2\delta t$ within the assimilation window, which has a total length of $30\delta t$. Here $\delta t = 0.005$ is the time step used for numerical solution of the model equations (24). The observed values of the fast model are defined as randomly perturbed values of the true state, with a standard deviation $\sigma_o^{(1)} = 0.05$ (the "accurate" observations experiment) and $\sigma_o^{(2)} = 0.1$ (the "inaccurate" observations experiment). Similarly, for the slow model, $\sigma_o^{(1)} = 0.1$ and $\sigma_o^{(2)} = 0.2$ were used. Since observation grid and model grid are the same, observation operator $\mathcal{H}$ is simply an identity mapping. To take into consideration the background covariances, for simplicity the assumption $\mathrm{B}_0 = \sigma_b^2 \mathrm{I}$, where σ_b^2 is the variance of background errors and I is the identity matrix, can be used. Under the assumption that the observation quality is the same for all fast variables and also the same for all slow variables, the observation covariance matrices can be defines as $\mathrm{R}_k = \mathrm{R} = \sigma_o^2 \mathrm{I}$.

Minimization of the cost function (9) was carried out using the conjugate gradient method [26], resulting in the analysis x_0^a at the initial time. The forecast trajectory is then calculated by numerically integrating the model equations (24) given initial conditions x_0^a.

The discrete-time dynamical system (7) obtained by finite-difference approximation of the continuous-time system (24) generates the discrete vector time series $\{x_k : k = 0,\ldots,K\}$, where $K = t^f / \Delta t$ and Δt is the integration time step. If $\{y_k^o : k = 0,\ldots,K\}$ is the corresponding verifying

(observed or analyzed) vector, then the forecast error is simply $E_k = x_k - y_k^0$, $k = 1,\ldots,K$. The two most commonly used error measures are mean absolute error (MAE) and root mean square error (RMSE) which are calculated as follows:

$$MAE = \sum_{k=1}^{K} |E_k|,$$

$$RMSE = \sqrt{\sum_{k=1}^{K} |E_k|}.$$

In our case, however, forecast errors are defined as the Euclidean distance between the forecast state and the "true" state (for each independent variable) because the latter has been previously calculated. Statistics were generated for over 100 different forecasts, and are shown in Table 3. In some applications, for example in the numerical weather prediction and climate modeling, a total energy norm is used as a measure of the forecast errors. In our study, the relative error in a total energy norm has been applied to measure the forecast accuracy [27]:

$$e_r = \left[\frac{\left(x^t - x^f\right)^{\mathrm{T}} \left(x^t - x^f\right)}{\left(x^t\right)^{\mathrm{T}} x^t} \right]^{1/2}. \qquad (25)$$

Table 3. Results of forecast verification for $c = 0.8$.

Variable	*MAD*	*RMSE*
x	0.0089	0.0110
y	0.0128	0.0157
z	0.0166	0.0196
X	0.0156	0.0199
Y	0.0336	0.0506
Z	0.0356	0.0512

Table 4 presents a summary of the forecast verifications for different time horizons using e_r as a measure of the forecast errors. Predicting the future state of dynamical system is an initial value problem, and the forecast quality strongly depends on the accuracy of initial conditions or in other words how accurately we are able to specify the initial state of the dynamical system. In general, the accuracy of initial conditions depends on the quality and density of observations, the forecasting model itself, and data assimilation system. In our numerical experiments the quality of observations is specified by the standard deviation σ_o. The results obtained are averaged over 10 sets of numerical experiments.

Table 4. Relative forecast errors (e_r^b - background forecast error; $e_r^{(1)}$ - "accurate" observation experiment; $e_r^{(2)}$ - "inaccurate" observation experiment) and forecast error reductions evaluated at a different verification time for "accurate" observations (Δe_1) and "inaccurate" observations (Δe_2).

	Verification time									
	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
Δe_1	-1.0126	-3.2149	-5.0829	-4.2464	-1.5882	-0.7142	-2.0250	-2.9439	-0.9128	-12.0190
Δe_2	-0.6975	-2.2817	-3.5146	-2.5495	-0.8494	-0.1767	1.0435	0.7010	1.1955	57.9270
$e_r^b \times 10^4$	7.77	13.40	12.91	10.74	8.13	4.90	12.43	32.39	11.59	117.76
$e_r^{(1)} \times 10^4$	0.73	1.18	1.16	1.25	1.11	0.94	3.92	9.71	4.66	70.90
$e_r^{(2)} \times 10^4$	2.94	4.73	4.44	5.04	4.37	3.92	16.82	37.86	20.66	342.81

Over the forecast time horizon t^f from 0.5 to 3 time units, both "accurate" and "inaccurate" observations demonstrate the positive impact on the forecast quality: the error in energy norm differences Δe_1 and Δe_2 are negative. However, the impact of "accurate" observations" is larger than the impact of "inaccurate" observations. For the forecasting time horizon $t^f > 3$, "inaccurate" observations have no positive impact on the forecast quality: Δe_2 is positive and the relative forecast error e_r is bigger than the background forecast error. The observation impact δe on the forecast skill is evaluated using equations (23). To calculate δe, the gradient $\partial J_b^f / \partial y^o$ is required. The adjoint model is used to obtain the gradient of the cost function J_b^f. The adjoint model is integrated backwards in time from t^f to t_0. Since the adjoint model is derived from a linearized forward propagation model, the observation impact estimate

δe is valid during a limited time horizon $t_{\lim}^{f}$, which is less than predictability limit of the original nonlinear model. The time horizon $t_{\lim}^{f}$ was found to be ~2.1 time units. Calculated observation impact estimates δe are shown in Table 5. The "true" values Δe are presented in Table 4.

Table 5 Observation impact estimates.

	Verification time			
	0.5	1.0	1.5	2.0
δe_1	-0.9821	-2.2948	-4.2183	-3.3952
δe_2	-0.6157	-1.8854	-2.7217	-1.2312

8 Conclusion

In this paper, we considered the application of adjoint-based technique developed in [12] to estimate the observational impact on the forecast quality using the coupled nonlinear chaotic dynamical system obtained by combining fast and slow versions of the Lorenz [20] system. This approach allows us to assess the impact of observing information coming from various sources of observations. Here, we take the term *forecast* to mean *a prediction of the future state* of any physical or socio-economic system produced by corresponding mathematical model.

We should not consider a forecasting numerical experiment to be completed until the forecast quality has been fully evaluated. To obtain a quantitative estimate of the forecast quality, the forecast is usually compared, or *verified*, against a corresponding observation of what actually occurred, or some good estimate of the true outcome. In this research, an energy norm of the forecast error was used as a measure of the forecast quality. Using this measure, we can estimate the forecast sensitivity with respect to observations, as well as to evaluate the impact of various types of observations on the forecast quality. Generally, the procedure of estimating sensitivity functions is quasi-linear. Therefore, the estimate of the observation impact δe is only valid during a limited time horizon that is smaller than the predictability limit of the original nonlinear model.

Since data assimilation procedure shadows the "true" state of a system given by the observational data, we computed "pseudo-orbits" of the system (24) assuming that the numerical solution has no errors, and "pseudo-orbits" were generated by variations in the system's parameters. It is important to underline that shadowing property of dynamical systems is a fundamental attribute of hyperbolic systems that was first discovered in [16, 17]. Since most physical systems are non-hyperbolic, the shadowing property should be verified for each particular dynamical system. However this problem is not trivial. We computed pseudotrajectories for the system under consideration using the approach developed for sensitivity analysis of chaotic systems [25].

References:

[1] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems*, Cambridge University Press, New York, 1995.

[2] R. Daley, *Atmospheric data analysis*, Cambridge University Press, Cambridge, 1991.

[3] E. Kalnay, *Atmospheric modelling: Data Assimilation and predictability*, Cambridge University Press, Cambridge, 2003.

[4] I.M. Navon, Data assimilation for numerical weather prediction: a review. In Park S.K. and Xu L. (Eds.), *Data assimilation for atmospheric, oceanic, and hydrologic applications*, Springer-Verlag, New York, 2009.

[5] V.V. Penenko and N.N. Obraztsov, Variational method for assimilation of meteorological fields, *Meteorology and Hydrology*, No. 11, 1976, pp. 3-16.

[6] F.-X. Le Dimet and O. Talagrand, Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects, *Tellus*, Vol.38A, No.2, 1986, pp. 97–110.

[7] P.Courtier, J.-N. Thépaut, A. Hollingsworth, A strategy for operational implementation of 4D-Var, using an incremental approach, *Quarterly Journal of the Royal Meteorological Society*, Vol.120, No.519, 1994, pp.1367 – 1387.

[8] J. Lions, *Optimal control of systems governed by partial differential equations*, Springer-Verlag, Berlin, 1971.

[9] J. Pailleux, Impact of various observing systems on numerical weather prediction, *Proceedings of CGC/WMO workshop, Geneva*, 7-9 April 1997, WMO Technical Report No.18, WMO/TD No. 868.

[10] C. Cardinali, Observing system experiments on the European conventional observing system, *Proceedings of the second CGC/WMO workshop on the impact of observing systems on NWP*, Toulouse, France, 6-8 March 2000, WMO Technical Report No.19.

[11] F. Bouttier and G. Kelly, Observing-system experiments in the ECMWF 4D-Var data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, Vol.127, 2001, pp.1469-1488.
[12] R.H. Langland and N.L. Baker, Estimation of observational impact using the NRL atmospheric variational data assimilation adjoint system, *Tellus*, Vol. 56A, 2004, pp.189-201.
[13] D.G. Cacuci, *Sensitivity and uncertainty analysis. Volume I: Theory*, CRC, Roca Baton, 2000.
[14] S.Yu. Pilyugin, *Shadowing in Dynamical Systems*, Lecture Notes in Mathematics, vol. 1706, Springer-Verlag, Berlin, 1999.
[15] K.J. Palmer, *Shadowing in Dynamical Systems. Theory and Applications*, Kluwer, Dordrecht, 2000.
[16] D.V. Anosov, On class of invariant sets of smooth dynamical systems, *Proceedings of the 5th International Conference on Nonlinear Oscillations*, Kiev, vol. 2, pp. 39-45, 1975.
[17] R. Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Mathematics, vol. 470, Springer, Berlin, 1975.
[18] D.V. Anosov, Geodesic flows on closed Riemann manifolds with negative curvature, *Proceedings of the Steklov Mathematical Institute*, vol. 90, 1967.
[19] I.M. Navon, X. Zou, J. Derrer and J. Sela, Variational data assimilation with an adiabatic version of the NMC spectral model, *Monthly Weather Review*, vol. 120, pp. 1433-1446, 1992.
[20] E.N. Lorenz, Deterministic nonperiodic flow, *Journal of the Atmospheric Sciences*, Vol. 20, 1963, pp. 130-141.
[21] G. Boffetta, A. Crisanti, A. Provenzale and A. Vulpiani, Slow and fast dynamics in coupled systems: A time series analysis view, *Physica D*, Vol. 116, 1998, pp. 301-312.
[22] M. Peña and E. Kalnay, Separating fast and slow models in coupled chaotic systems, *Nonlinear Processes in Geophysics*, Vol. 11, 2004, pp. 319-327.
[23] S. Soldatenko and D. Chichkine, Basic properties of slow-fast nonlinear dynamical system in the atmosphere-ocean aggregate modelling, *WSEAS Transactions on Systems*, Vol.13, 2014, pp.757-766.
[24] M. Rosenblum, A. Pikovsky and J. Kurth, Phase synchronization of chaotic oscillators, *Physical Review Letters*, Vol. 76, 1996, pp. 1804-1807.
[25] Q. Wang, Forward and adjoint sensitivity computation for chaotic dynamical systems, *Journal of Computational Physics*, vol. 235, no. 15, pp. 1-13, 2013.
[26] W.W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, *Pacific Journal of Optimization*, Vol. 2, 2006, pp. 35-58.
[27] R. Robert and C. Rosier, Long range predictability of atmospheric flows, *Nonlinear Processes in Geophysics*, Vol. 8, 2001, pp. 55-67.