Applied Mathematical Modelling 33 (2009) 2165–2174

Contents lists available at ScienceDirect

Applied Mathematical Modelling

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

A sequential data assimilation method based on the properties of a diffusion-type process

Clemente A.S. Tanajura [a,*], Konstantin Belyaev [a,b]

[a] Departamento de Física da Terra e do Meio Ambiente, Instituto de Física, Universidade Federal da Bahia (UFBA), Campus de Ondina, Travessa Barão de Jeremoabo, s/n, 40170-290 Salvador, BA, Brazil
[b] Shirshov Institute of Oceanology (SIO), Russian Academy of Science (RAS), Hakhimovsky 36, Moscow 217513, Russia

ARTICLE INFO

Article history:
Received 27 December 2006
Received in revised form 6 May 2008
Accepted 16 May 2008
Available online 24 May 2008

Keywords:
Sequence of Markov chains
Diffusion stochastic process
Data assimilation methods
Optimal filtering

ABSTRACT

Data assimilation method, as commonly used in numerical ocean and atmospheric circulation models, produces an estimation of state variables in terms of stochastic processes. This estimation is based on limit properties of a diffusion-type process which follows from the convergence of a sequence of Markov chains with jumps. The conditions for this convergence are investigated. The optimisation problem and the optimal filtering problem associated with the search of the best possible approximation of the true state variable are posed and solved. The results of a simple numerical experiment are discussed. It is shown that the proposed data assimilation method works properly and can be used in practical applications, particularly in meteorology and oceanography.

1. Introduction

In numerical modelling of geophysical systems, such as the ocean, the atmosphere or climate in general, data assimilation (DA) is a common and a popular tool. It produces the initial conditions for weather and climate models and complements monitoring by correcting model variables towards observations. Therefore, DA methods have a substantial impact in weather and climate model predictability and, consequently, in several human activities that rely on weather and climate forecasts, such as agriculture, water resources and others.

Such a correction of the model state by observational data is generally based on a scheme of the following type: a system of partial differential equations, usually represented in finite-difference or finite-element form, is considered on the time interval (t_0, T) and in a 3-d domain. In general, t_0 may be associated with 0, while T can be considered as finite or infinite. The interval (t_0, T) can be broken down into subintervals $(t_0, t_1), (t_1, t_2), (t_k, t_{k+1}) \ldots$ On any subinterval (t_k, t_{k+1}), the model starts at time t_k with the initial vector θ_k, also called as background state-vector, and it is integrated forward until t_{k+1}, when it produces (predicts) the state-vector θ^m_{k+1}. Hereafter, the superscript m indicates that the system state was obtained only by model integration without any other source of information. Over the time interval (t_k, t_{k+1}) and in some sub domain of the model, a series of observations of the state variable represented by the vectors $\xi^k_1, \ldots, \xi^k_l$ are made available. Each observational vector ξ^k_i may be composed by different state variables, e.g., temperature, salinity, etc. Usually, vectors ξ^k_i, $i = 1, \ldots, l$ are a subset of the state-vector θ_k, since only part of the model state may be observed. Then, the model output θ^m_{k+1} at time $k+1$ is corrected by observational information and the model according to the following scheme of correction:

* Corresponding author. Tel.: +55 71 3283 6646.
E-mail addresses: cast@ufba.br (C.A.S. Tanajura), kostia@ufba.br (K. Belyaev).

0307-904X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.apm.2008.05.021

$$\theta_{k+1} = \theta^m_{k+1} + \sum_{i=1}^{l} \alpha_i(\xi^k_i - H_i\theta^m_k). \qquad (*)$$

Here, $H_i\theta^m_k$ denotes the projection of model variables in the model domain onto the observational locations in each subinterval (t_k, t_{k+1}). The model and the measured variables are taken at the specific time in which the observational data are available during the considered time subinterval. The weight functions α_i also depend on time and space and may be either known a priori or determined by some appropriate algorithm. The corrected state-vector θ_{k+1}, the so-called objective analysis, is then taken as a initial condition for the next step of the model integration.

Under the name "data assimilation methods" (DA methods) or sequential DA methods various versions of this scheme are used in geophysics. In particular, observations and initial conditions may be considered as random, and the Kalman filter approach determines the optimal weight coefficient α_i utilising the statistical properties of observational data and the stochastic form of the model dynamics [1]. Alternatively observations may not be considered as random. In this case, the optimal weights may be determined according to a variational DA technique, by minimising the distance in a given metric between observations and model trajectory with respect to initial conditions [2]. The integration intervals can be previously defined, for instance, 24 h, and the DA is realised in the end of this integration. The assimilation intervals can also be considered as random, e.g., in modern coupled ocean–atmosphere models the correction can be applied when the model temperature difference between ocean and atmosphere exceeds a certain threshold.

Despite these and other differences, all sequential DA methods follow essentially the scheme mentioned above. Therefore, it turns out interesting to review these techniques from a different point of view. At the instant of assimilation, the time series of θ_k undergoes a jump of its trajectory. What will happen if the interval between two consecutive assimilations approaches to zero along with the values of the jump? How does the limiting behaviour of the trajectories depend on the number of state variables and their distributions? Under which conditions does the limiting distribution of θ_k exist as T goes to infinity? Also, what will the limit be? If it exists, the distribution is called stationary. These and similar questions attract interest not only from the pure theoretical point of view, but also for practical reasons. For instance, knowing the limiting behaviour of the time series for $\Delta t_k = t_{k+1} - t_k \to 0$, it becomes easy to calculate various parameters needed for the weather forecast, while knowing stationary distribution enhances the reliability of climate prediction. In addition, with this limit the optimisation problem for the weight coefficients, which generally are the extremum of some given function, may be simplified.

One of the goals of the present paper is to find out, under appropriate conditions, the limiting behaviour of trajectories of the objective analysis θ_k as a function of time when the interval between two consecutive assimilations approaches to zero along with the vanishing of the values of the jump. The limiting process has continuous trajectories with probability 1 and the distribution of these trajectories satisfies the Fokker–Planck equation. Based on the properties of the limiting trajectories, another goal of the present paper is to pose and to solve the optimisation problem for the best weight coefficients. Finally, only for illustration of the feasibility and usefulness of this method, a numerical experiment is performed. Also, it is shown that the known and popular optimal interpolation DA scheme can be considered as a particular case of the studied method under some additional conditions.

The proposed approach is a new contribution in the DA area and builds a bridge between the theory of stochastic process and DA techniques. This paper is a continuation of previous works by the authors [3,4]. Section 2 presents the formulation of the problem and the main notation. Section 3 is dedicated to the optimisation problem and to the reduction of the known estimation method to the considered scheme. Section 4 deals with the numerical experiment. The last section contains discussions and conclusions.

2. Main definitions and notations

Let the system of equations

$$\frac{\partial\theta(t)}{\partial t} = M(\theta, t) \qquad (1)$$

be considered on the time interval (t_0, T). Throughout the paper, time t is supposed to vary continuously. Without loss of generality, t_0 hereafter will be associated with 0, while T may be finite or infinite. In (1), $\theta(t)$ represents the random state-vector of dimension r defined on a given probability space, and $M(x, t)$ denotes a non-random, in general non-linear, operator acting in R^r. This operator does not explicitly involve temporal derivatives. The symbol $\prime$ denotes the transpose of a vector and/or matrix; the symbols $|\,|$ and $\|\|$ represent the vector and the matrix norms, respectively. A sequence of time series is considered, and for each series, the interval $(0, T)$ is broken down by the points $(0 = t_{0,n}, t_{1,n}, t_{2,n}, \ldots, t_{kT,n} = T)$. The first index in this series denotes time in chronological order, and the second index refers to the series number. It is supposed that in each series on the interval $\Delta t_{k,n} = t_{k+1,n} - t_{k,n}$ a number of random vectors $(\xi^{k,n}_1, \ldots, \xi^{k,n}_i, \ldots)$, $i = 0, 1, \ldots, \nu_{k,n}$, each vector with dimension q, $q \leqslant r$, are observed. Hereafter, $\nu_{k,n}$ is an integer random multi-index with the given distribution $p^n_i = P(\nu_{k,n} = i), i \geqslant 0$, independent of vectors $\xi^{k,n}_i$. So, this condition means that an arbitrary random set of observations can be involved in each interval. Their distribution may also vary.

Knowing: (i) the solution of system (1), i.e., the model forecast $\theta^m_{k,n}(t)$ for a given initial vector $\theta_{k,n}$ on the entire subinterval $\Delta t_{k,n}$; (ii) the observed variables $\xi^{k,n}_i$; and (iii) the value of the random index $\nu_{k,n}$, the innovation vectors for each series n in the subinterval can be introduced by