ECOLOGICAL MODELLING 217 (2008) 279–291

available at www.sciencedirect.com

ScienceDirect

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

Orthogonal decomposition methods for inclusion of climatic data into environmental studies

*Vladimir Penenko**, *Elena Tsvetova*

Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk 630090, Russia

ARTICLE INFO

Article history:
Published on line 9 July 2008

Keywords:
Environmental modeling
Scenario approach
Climatic data
Orthogonal decomposition
Principal component and factor analysis
Activity centers in the climatic system

ABSTRACT

Scenario approach is widely used in numerical modeling for the assessment of current and prospective states of environmental and ecological systems. The results of modeling essentially depend on the representation of the underlying hydrodynamics. We suggest a methodology for a quantitative description of the behavior of a dynamic system for a long time interval in a compressed, generalized form. Following it, the necessary information, intended to be used for the construction of scenarios, is extracted from a database containing measured and/or calculated data on hydrodynamic state functions.

The proposed methodology is a development of ideas of principal component analysis and factor analysis. Calculations are made with the help of a method of orthogonal decomposition of functional spaces formed by multi-variate, multi-dimensional state functions from the database. A two-level data structuring with allowance for given goal criteria is firstly produced. This provides an efficient realization of the methodology practically without restrictions upon the amount of data and component content of the database. The targeted structuring is just the element which differs the methodology proposed from the traditional approaches to data decomposition.

The NCEP/NCAR reanalysis database for 56 years (1950–2005) is used to demonstrate the possibilities of the methodology. The method of orthogonal decomposition results in the subspaces which correspond to the processes on different scales: from global climatic processes to weather noises. These subspaces serve as informative bases for analysis of the climatic system behavior. Moreover, the subspaces are key elements for the construction of deterministic–stochastic scenarios to obtain an atmospheric background for problems of environment protection and design, ecological risk/vulnerability assessment and control, etc.

1. Introduction

The idea of this study originates from the analysis of environmental forecasting and projecting problems, in which the assessment of possible nature changes for a long period of time has to be made.

A way to solve such problems is the development of scenario approach. In this case it is natural to formulate environmental protection problems considering not only the direct impacts of the specific objects, but also the indirect effects arising additionally due to climate change. As a matter of fact, the goal of our work is formulated in the following way. A

* *Corresponding author*. Tel.: +7 383 3306152; fax: +7 383 3308783.
E-mail address: penenko@sscc.ru (V. Penenko).
0304-3800/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2008.06.004

methodology should be created for incorporation of climatic data and some knowledge about assumed climate changes into environmental studies. To provide this, the climatic data should be presented in a compressed and generalized form. A deterministic "main" part extracted from the data under a given criterion of information completeness is implied here. To separate the scales of physical processes presented in the data, we propose a version of orthogonal decomposition method suited to the treatment of multi-dimensional, multi-variate fields of data. The ideas of principal component analysis (PCA) and factor analysis (FA) are essentially used here.

The fundamental and applied aspects of PCA and FA are of interest for many researchers from the different areas of knowledge. The origins of the methods are related to the end of the 19th and the beginning of the 20th centuries. The different ways of PCA and FA are actively used in meteorology and oceanography from the middle of the last century. The systematic description of the main statements and some applications can be found, for instance, in the monographs of Harman (1976), Mescherskaya et al. (1970), and Preisendorfer (1988). The voluminous bibliography containing more than a thousand of papers with surveys of the history, theoretical background and practical applications are presented there, too.

The limiting parameter for these methods is the dimension of the data analyzed. In practice, these algorithms are very sensitive to the increase of the dimensions of eigenvector problems (EVPs). Some existing approaches reducing the dimensions are discussed in detail by Mescherskaya et al. (1970) and Preisendorfer (1988). Among different algorithms, the eigenvector-partition methods should be mentioned. In such algorithms, the data are divided by parts and the total EVP is replaced by several EVPs of low dimensions. Then the solution is obtained by constructing the basis from the separate parts.

The problem of the informative bases construction holds a priority for many years. Lorenz (1965) used singular vectors of the linearized operator of a simple model to study the flow-dependent predictability. An application of biorthogonal decomposition method to the meteorological fields with the use of EVP of the linearized forward and adjoint operators of atmospheric hydrodynamics was considered by Marchuk (1968). Recently the other approaches for optimal representation of the perturbations have been intensively developed for the ensemble weather prediction. The singular vectors of the tangent linear operators associated with forecasting models were used by Molteni and Palmer (1992), Mureau et al. (1992), Ehrendorfer and Tribbia (1997), Gelaro et al. (1998), and Kim et al. (2004). Toth and Kalnay (1993) applied the bred vectors for generation of perturbations.

Our previous experience in these directions was acquired in the construction and use of orthogonal and biorthogonal bases for implementation of variational principles in the problems of atmospheric dynamics (Penenko, 1974, 1975, 1981). To construct such bases, a computational methodology was built. In this methodology, the eigenvectors and singular vectors of the linearized operators of the main and adjoint problems were calculated. Lanczos's algorithms as well as the other iterative procedures were applied there. The subspaces obtained by means of PCA and FA were also introduced. The development of a concept and methods of adjoint sensitivity analysis and variational data assimilation with adjoint problems were the most important fundamental results of that period.

In this paper we have been developing a methodology of orthogonal decomposition for analysis of the large databases which describe the behavior of the multi-dimensional dynamic systems. The kernel element of the methodology is the solution of EVP. It should be noted that our methodology differs from the above mentioned eigenvector-partition methods (Preisendorfer, 1988; Mescherskaya et al., 1970). We organize the data in such a way to avoid splitting the EVP and to solve the problem of searching for the main components without loss of information quality. A distinctive element of our approach is a two-level structuring of the analyzed database. This allows us to solve the decomposition problem practically without limitations on amount of data.

Quantitative information of atmospheric and ocean circulation for long time periods can be found in the high capacity databases such as reanalysis. For example, information on the basic state variables of the climatic system and some other characteristics are collected in the NCEP/NCAR reanalysis database (Kalnay et al., 1996). In this study, we have considered this database as the description of the concrete realization of the climatic system evolution. And a data subset for 56 years has been used here as the climatic background. In the presented example on decomposition of the reanalysis database, a two-level structuring with respect to time scales has been made. This gives the possibility of producing the time-dependent dynamical bases that is convenient for prognostic goals. In frames of our approach, the other variants of structuring are also possible. To realize the methodology, an efficient computational technology has been developed to be used for climate studies and for organizing the scenarios for environment protection and design.

2. Mathematical models, functionals, and variational principles

To solve the above-mentioned problems, we need to formulate some goal functionals for short description of the generalized characteristics of the state variables of complicated dynamic systems. The form of such functionals is conveniently chosen from the statements of mathematical models of the processes under discussion. Without going into details, let us write a model in which dynamics of the atmosphere as well as transport and transformation of impurities are presented. The structure of such a model can be given in the operator form:

$$L(\boldsymbol{\varphi}) \equiv B\frac{\partial \boldsymbol{\varphi}}{\partial t} + G(\boldsymbol{\varphi}, \mathbf{Y}) - \mathbf{f} - \mathbf{r} = 0, \qquad \boldsymbol{\varphi}^0 = \boldsymbol{\varphi}_0^0 + \boldsymbol{\xi}, \quad \mathbf{Y} = \mathbf{Y}_0 + \boldsymbol{\zeta}, \tag{1}$$

where $\boldsymbol{\varphi}(\mathbf{x},t)$ is the state vector that belongs to the real vector space $Q(D_t)$, B is a diagonal matrix, $G(\boldsymbol{\varphi},\mathbf{Y})$ is a nonlinear matrix differential operator, $\mathbf{f}$ is a source function, $D_t = D \times [0, \bar{t}]$, D is a domain of spatial coordinates $\mathbf{x}$, $t \in [0, \bar{t}]$ is a time interval. The vector $\mathbf{Y}$ describes the model parameters belonging to a range of admissible values of $R(D_t)$, $\boldsymbol{\varphi}^0$ is an initial state, $\boldsymbol{\varphi}_0^0$, $\mathbf{Y}_0$ are