Annual Reviews in Control 30 (2006) 169–182

Annual Reviews in Control

www.elsevier.com/locate/arcontrol

The data-based mechanistic approach to the modelling, forecasting and control of environmental systems

Peter C. Young [a,b,*]

[a] *Centre for Research on Environmental Systems and Statistics, Lancaster University, Lancaster, UK*
[b] *Centre for Resource and Environmental Studies, Australian National University, Canberra, Australia*

Received 3 March 2006; accepted 11 May 2006

Abstract

The paper presents a unified approach to the modelling, forecasting and control of natural and man-made environmental systems. The modelling approach exploits the author's Data-Based Mechanistic (DBM) modelling philosophy, combined with powerful methods of recursive statistical estimation. These provide the basis for two major stages of model building: first, the critical evaluation of the over-parametrized simulation models that are currently the most common vehicle used in environmental planning and management studies; and second, the adaptive, data-based estimation of parsimonious, 'top-down' models that can be used for adaptive forecasting and data assimilation, as well as operational control and management system design. The associated control system design methodology is based on the Non-Minimal State Space (NMSS) approach to the design of Proportional-Integral-Plus (PIP) control systems, based on the DBM models obtained at the previous modelling stage. The paper includes a case study concerned with the modelling and control of globally averaged levels of CO_2 in the atmosphere.

Keywords: Simulation modelling; Dominant mode analysis; Model order reduction; Data based mechanistic modelling; Management and control; Global carbon cycle; Global warming

1. Introduction

Throughout the 20th century, the hypothetico-deductive approach to scientific research, as extolled so lucidly by the philosopher Karl Popper (Popper, 1959), reigned supreme. The inductive approach, that had provided the cornerstone for scientific research in earlier centuries, became less attractive to the physicists and chemists who dominated science during the 20th century and whose well planned experimentation provided some of the stimulus for Popper's views. Moreover, the rise of the computer, with its ability to construct and solve large mathematical simulation models, provided a magnificent engine for the implementation of the hypothetico-deductive approach. Today, such computer-based simulation modelling has become straightforward, almost simple, with the availability of iconographic software, such as Matlab/Simulink TM, where a complex model can be assembled quickly from a built-in and comprehensive library of simulation objects.

All this would appear to be good news for environmental scientists involved in the management and planning of environmental systems. But is it? While acknowledging the virtues of simulation modelling, particularly when it is carried out in stochastic terms, this paper will discuss the limitations of this approach when it is used in the context of environmental systems, where planned experimentation is difficult, if not impossible, and where uncertainty about the nature of the processes involved sits uncomfortably with the deterministic models that appear to dominate simulation modelling practice.

This paper will outline a Data-Based Mechanistic (DBM) approach to modelling, forecasting and control that often starts with the construction and evaluation of a simulation model which reflects the scientists' perception of the physical, chemical and biological mechanisms that characterise the environmental system. However, these DBM studies are not simple exercises in simulation modelling: rather they constitute a critical evaluation of the model in both stochastic and response terms; an evaluation that marks only the beginning, not the end, of the modelling process.

Exploiting some of the tools of DBM modelling that are later applied to real data, this critical evaluation considers the

* Corresponding author. Tel.: +44 1524 59 3966; fax: +44 1524 5939 85.
E-mail address: p.young@lancaster.ac.uk.

1367-5788/$ – see front matter
doi:10.1016/j.arcontrol.2006.05.002

simulation model as a natural extension of the thought processes and scientific speculation that resides in the mind of the model builder. And, by providing insight into the strengths and limitations of the simulation model, it provides a prelude to the exercises in DBM modelling from real data that becomes possible when data are available on the response of the environmental system to natural or anthropogenically-induced perturbations.

Of course, in this environmental context, the real data required for DBM modelling are most often the result of monitoring studies, rather than planned experimentation. As a result, such data may not be available or, as is often the case, they may provide an insufficient basis for DBM modelling. Even in this data-deficient situation, however, we will see that the DBM modelling methodology can provide valuable insight into the strengths and limitations of the simulation model; insight that can radically effect the way in which the model is used as a tool in planning and management.

Whatever model emerges from the DBM modelling process, it should be a model that is well suited to the objectives of the study team. These may range from 'what-if' simulation, where the complexity of the simulation model is a clear advantage, to exercises in forecasting and operational control, where the over-parameterization that normally characterizes the large simulation model is a definite disadvantage. This paper will argue, therefore, that the construction of a single model that suits all purposes is normally impossible and, in any case, undesirable. Rather the objectives of the study should be clearly defined and a well integrated suite of models should be constructed, each designed to satisfy the requirements of these objectives.

It is suggested that, wherever possible, the parametrically efficient (or 'parsimonious') DBM model should provide a description of the core mechanisms that dominate the observed behaviour of the environmental system under study. And it should also provide the basis for the *final* construction of a *stochastic* simulation model that reflects this core behaviour but may involve other, more speculative elements that are required for 'what-if' simulation and planning exercises. The advantage of this DBM 'moderation' of the simulation model is that the relative confidence in the historically validated DBM core (when sufficient data are available) can be balanced with the reduced confidence in the more poorly validated speculative elements. Thus the results of any analysis can be better evaluated with these relative uncertainties in mind. For instance, if the model contains nonlinearities that have not been well-validated during DBM modelling over the historical period, but are thought to be of potential importance in the future, then the large level of uncertainty in this regard must be reflected clearly in any predictive application of the model.

The DBM approach to modelling and forecasting has been applied to a number of environmental systems. These include the modelling and control of water quality in rivers (see e.g. Young & Beck, 1974; Beck & Young, 1975; Young, Beck, & Singh, 1976; Wallis, Young, & Beven, 1989); the modelling of rainfall-flow processes, within the context of flood forecasting and warning (e.g. see Young, 2002 and the prior references therein); and the modelling and control of mass and energy transfer in agricultural buildings (Price, Young, Berckmans, Janssens, & Taylor, 1999). The present paper describes an environmental case study in a different environmental area which has involved most aspects of this DBM approach: namely, the modelling and control of globally averaged levels of atmospheric CO_2 and the possible implications of this research on global warming.

2. The fundamentals of data-based mechanistic modelling

For many years, the author and his co-workers have attempted to draw attention to the limitations of large deterministic models of the environment (e.g. Young, 1978, 1983, 1998b, 1999a; Beck, 1983; Young & Lees, 1993; Young, Parkinson, & Lees, 1996; Shackley, Young, Parkinson, & Wynne, 1998). In order to alleviate these limitations, the DBM approach to modelling is built on the assumption that, wherever possible, the dynamic modelling of environmental systems should not be based solely on such simulation models. Rather it should involve a rational combination of two basic model types: speculative simulation models which represent the current, state-of-the-art, scientific understanding of the environmental system; and DBM models, obtained initially from the analysis of observational time-series, but only considered credible if they can be interpreted in physically meaningful terms. The term 'data-based mechanistic modelling' was first used in (Young & Lees, 1993) but the basic concepts of this approach to modelling dynamic systems have developed over many years. For example, they were first applied seriously within a hydrological context in the early 1970s, with application to the modelling of water quality in rivers (Young & Beck, 1974).

The objective statistical derivation of these much simpler DBM models contrasts with the rather subjective formulation of the complex simulation models. However, the two, apparently quite different, types of model are brought together in a rather novel phase of the analysis where the DBM methodology is used to reduce the order of the complex simulation model, so producing a simplified model that exposes its 'dominant modes' of dynamic behaviour.

The five major phases in the DBM modelling strategy are as follows:

(1) The important first step in any modelling exercise is to define the objectives and to consider the types of model that are most appropriate to meeting these objectives. Since the concept of DBM modelling requires adequate data if it is to be completely successful, this stage also includes considerations of scale and the likely data availability at this scale, particularly as they relate to the defined modelling objectives.

(2) In the initial phases of modelling, it may well be that observational data will be scarce, so that any major modelling effort will have to be centred on simulation modelling, normally based on largely deterministic concepts, such as dynamic mass and energy conservation. In the DBM modelling approach, which is basically Bayesian