Environmental Modelling & Software 104 (2018) 81–93

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Extended State Dependent Parameter modelling with a Data-Based Mechanistic approach to nonlinear model structure identification

David A. Mindham, Wlodzimierz Tych*, Nick A. Chappell

Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, United Kingdom

ARTICLE INFO

Article history:
Received 22 May 2017
Received in revised form
7 February 2018
Accepted 14 February 2018

Keywords:
Nonlinear dynamic systems
Model identification
Data based mechanistic modelling
State-dependent parameters
Hydrological modelling
Information Criterion

ABSTRACT

A unified approach to Multiple and single State Dependent Parameter modelling, termed Extended State Dependent Parameters (ESDP) modelling, of nonlinear dynamic systems described by time-varying dynamic models applied to ARX or transfer-function model structures. Crucially, the approach proposes an effective model structure identification method using a novel Information Criterion (IC) taking into account model complexity in terms of the number of states involved. In ESDP, model structure involves not only the model orders, but also selection of the states driving the parameters, which effectively prevents the use of most current IC. This leads to a powerful methodology for investigating nonlinear systems building on the Data-Based Mechanistic (DBM) philosophy of Young and expanding the applications of the existing DBM methods. The methodologies presented are tested and demonstrated on both simulated data and on high frequency hydrological observations, showing how structure identification leads to discovery of dynamic relationships between system variables.

1. Introduction

The Data-Based Mechanistic methodology (DBM, Young, 1999a) is built on the premise that the model structure and parameters are to be determined through statistical analysis of observed data ('data-based') which then, along with model metrics leads to a physical interpretation of the model ('mechanistic').

The presented approach completes the nonlinear DBM modelling process by adding an objective identification stage to the nonlinear model selection. Multiple and single State Dependent Parameter (MSDP and SDP) modelling follows the DBM methodology by not parameterising the individual nonlinearities, however the selection of the model structure, including that of the nonlinear drivers, is the key element missing from the current method.

While MSDP employs a very different numerical engine to SDP, conceptually and in terms of the outcome, it is a multi-variable extension of the original SDP and thus, a generalisation of SDP that is not confined to one state dependency. However, both in SDP and then MSDP, the states' values were assumed equidistant, having the same distance in the state-space between each sample, which is a simplifying assumption. The solution introduced in this paper removes this assumption and makes the method fully general.

The new model structure identification procedure allows for the first time identification of nonlinear structural relationships in an objective manner using a robust and tested model form. This is demonstrated in the paper using high frequency hydrological observations, where the output variable is thought to be affected by more than one nonlinear process.

The terminology, explanation and clarification for the above are laid out below in a logical and methodical manner designed to lead the reader first through existing SDP and MSDP methodologies, then through the process of updating and extending these methodologies into one methodology described in this paper (ESDP – Extended SDP) with useful output tools. Finally, through the process of producing a generalised Model Structure Identification (MSI) procedure to identify the structure from a given data set for the application of ESDP. The MSI procedure is generalised in that it considers - no state dependency for each parameter (linear model) and one or more state dependencies for one or more parameters (nonlinear model).

1.1. Objectives and structure of this paper

This paper presents three key updates and additions to the SDP and MSDP methodologies leading to their unification and generalisation (items 1,…,3), and one major development (item 4) for applying the DBM approach to model structure identification in

* Corresponding author.
E-mail address: w.tych@lancaster.ac.uk (W. Tych).

https://doi.org/10.1016/j.envsoft.2018.02.015
1364-8152/

this new setting:

1. Transition to a true state-space for parameter estimation by moving from equidistant states to arbitrary-distant states, based on the state values. The terms and context of 'state-space' and 'states' are clarified and discussed in section 1.4 and onwards (Section 2.0).
2. Formation of multivariable parameter maps from M-dimensional state dependent parameters for the purposes of basic model validation and more importantly, for forecasting, scenario investigation and on-line simulation of live events (Section 2.1).
3. Use of model validation techniques to not only quantify the ability of the presented algorithms in parameter estimation, but also to validate any models identified from the model structure identification development step below (Section 2.2).
4. Development of a DBM approach for model structure identification (MSI) from a group of data sets for a given model type so that the data informs us which measured variables are more influential to the observed model output – importantly this methodology also considers a linear model, allowing for a 'pure' DBM approach (Section 3.0).

The whole approach is then applied to a hydrological example using a dynamic model of streamflow generation, thus forming objective 5 (Section 4.0).

1.2. Applications

The methodologies presented are general and can be applied to any system as long as time-series data for all the required variables (including inputs, outputs and additional states) are available. In terms of specific environmental applications, we have evaluated the approach for data in the two applications below, and present the former in this paper:

- Flood scenarios – how the flood response of a stream may be strongly affected by more than multiple nonlinear process, not solely the nonlinear effects of varying catchment wetness (Chappell et al., 2017).
- Water quality dynamics – how the dynamics of one output water quality variable (e.g., Dissolved Organic Carbon concentration) may be affected by more than one nonlinear process, related to separate effects of e.g., rainfall, soil temperature and solar radiation (Jones et al., 2014).

For clarity and in order to introduce the notation, this paper also briefly covers the progression from Transfer Function (TF) to SDP TF (for a more detailed account see Young, 2000) and MSDP TF, with the novel generalisation elements introduced. Significantly, the development of the structural identification methodology for this wide class of nonlinear models is then covered.

1.3. Background to SDP

There is extensive work on modelling input-output dynamic time-series data using Transfer Functions (TF or equivalent Auto-Regressive with eXogenous inputs, or ARX models) where linear or approximated linear relationships are used (Ockenden and Chappell, 2011; Tych et al., 2014; Ampadu et al., 2015; Chappell et al., 2017) as well as extensions into Time Varying Parameter (TVP) TF (Gou, 1990) and further extensions into State Dependent Parameter (SDP) TF (Young et al., 2001) with latter approaches using nonlinear functional relationships between states of the system and the ARX or TF parameters.

SDP modelling assumes that the system is truly nonlinear in that the TF parameters are time varying; importantly, the rate of change of the parameters is at a rate related to the rate of change within the state variables. This is unlike the more commonly seen time varying parameter TF models, where the parameters change smoothly. Here, the parameters are functions of the input or other states of the system under study. SDP, as originally published by Young (2000) bears the assumption that each parameter is a function of one variable only, and has been successfully applied to many nonlinear systems (e.g. Young et al., 2007a; Taylor et al., 2009; McIntyre et al., 2011). However, many systems, particularly in the natural environment, are complex dynamic systems with many variables that have interlinking relationships, e.g. water quality (Jones et al., 2014), atmospheric chemistry (Seinfeld and Pandis, 2016), and climate change (Ashkenazy et al., 2003; Young and Garnier, 2006). The parameters of models describing these environmental systems, or even just their specific aspects, could be functions of more than one variable and hence the need to generalise SDP modelling into the Multiple State Dependent Parameter (MSDP) modelling.

1.4. ARX - Transfer function (TF) – Time Invariant and Time varying parameter (TVP) issues

We begin with a simple linear discrete time dynamic model with time varying parameters (ARX or equivalent TF structure (Young, 1999b)) that relates a single input variable (u_t) to an output variable (y_t) and can be written as a difference equation (1). Due to the time-varying character of parameters the standard backward shift TF models are not applicable.

$$y_t = -a_{1,t}y_{t-1} - \ldots - a_{n,t}y_{t-n} + b_{0,t}u_{t-\delta} + b_{1,t}u_{t-\delta-1} + \ldots \\ + b_{m,t}u_{t-\delta-m} + e_t \tag{1}$$

where δ is a pure time delay (measured at this stage in sampling intervals), e_t is a zero mean, serially uncorrelated input with variance σ^2 and Gaussian amplitude distribution. The latter (Normality) assumption is usual, but not required for the Kalman Filter to function (Kalman, 1960 – the distribution needs to be finite and symmetric).

Expressing (1) as a vector equation we obtain the TVP observation equation:

$$y_t = \boldsymbol{z}_t^T \boldsymbol{p}_t + e_t \tag{2}$$

where,

$$\boldsymbol{z}_t^T = [-y_{t-1} \quad -y_{t-2} \cdots \ -y_{t-n} \ u_{t-\delta} \cdots u_{t-\delta-m}]$$

$$\boldsymbol{p}_t = [a_{1,t} \ a_{2,t} \cdots a_{n,t} \ b_{0,t} \cdots b_{m,t}]^T$$

When $\boldsymbol{p}_t$ changes slowly/smoothly there are methods to estimate these changes taking advantage of the smoothness assumption (see for example Dynamic Transfer Function or DARX models – Young, 2011).

However, many environmental systems can be described as complex and nonlinear, where the rates of change of the parameters vary at a rate commensurate to those of other, exogenous variables. This means the changes in $\boldsymbol{p}_t$ are too rapid to apply the smoothness assumption (slow parametric changes) and so other estimation methods are required. If the parameters are varying at a rate similar to that of the rate of another system variable, then that