Control Engineering Practice 52 (2016) 70–92

Contents lists available at ScienceDirect

Control Engineering Practice

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

# System identification of the upper part of Murray River

Hasan Arshad Nasir [a,*], Erik Weyer [b]

[a] National ICT Australia, Victoria Research Lab, Department of Electrical and Electronic Engineering, The University of Melbourne, 3010 Victoria, Australia
[b] Department of Electrical and Electronic Engineering, The University of Melbourne, 3010 Victoria, Australia

ARTICLE INFO

*Article history:*
Received 27 August 2015
Received in revised form
30 March 2016
Accepted 6 April 2016
Available online 28 April 2016

*Keywords:*
Rivers
Modelling
System identification
Parameter estimation
Grey box identification
Prediction error method
Continuous time system identification
Maximum likelihood
Data-based mechanistic modelling
Subspace identification

ABSTRACT

Data based modelling is an important tool for obtaining models of rivers. In this paper we consider several identification methods, namely Prediction Error Method, Maximum Likelihood, continuous time system identification, Refined Instrumental Variable method (used within the context of Data-Based Mechanistic modelling) and Subspace Identification Method, and apply them to real data from the Murray River in Australia. Both Multiple-Input Single-Output models where the output is the water level in a lake and Multiple-Input, Multiple-Output models where in addition a flow is also modelled, are considered. The models are compared in terms of their accuracy on validation data and on how easily the methods can incorporate prior knowledge.

## 1. Introduction

Water is a precious resource, and in the last few decades large efforts have gone into improving the management of water resources. Due to the rapid development in sensors and information and communication technology, operational data is now widely available. System identification and automatic control can now be used in the operation of large scale open water systems such as irrigation channels and rivers. The models obtained from system identification are simple ordinary differential or difference equations and are suitable for control design, prediction and fault detection purposes. In the last 10–15 years many works have appeared in the area of system identification and control of water systems. E.g. Cantoni et al. (2007), Litrico, Malaterre, Baume, Vion, and R-Bruno (2007), Litrico, Fromion, Baume, Arranja, and Rijo (2005), Malaterre and Baume (1998), Nasir and Muhammad (2011), Negenborn, Van Overloop, Keviczky, and De Schutter (2009), Ooi and Weyer (2008), Schuurmans, Hof, Dijkstra, Bosgra, and Brouwer (1999), Van Overloop, Schuurmans, Brouwer, and Burt (2005), Weyer (2001), and Weyer (2008) have demonstrated that system identification and control can improve the quality of service and water distribution efficiency for irrigation channels. Similarly, several works on modelling and control of rivers are also available, see e.g. Breckpot, Agudelo, Meert, Willems, and De Moor (2013), Foo, Ooi, and Weyer (2012), Foo, Ooi, and Weyer (2014), Glanzmann, von Siebenthal, Geyer, Papafotiou, and Morari (2005), Litrico (2002), Litrico and Pomet (2003), Maxwell and Warnick (2006), Papageorgiou and Messmer (1989), Pianosi, Castelletti, and Lovera (2012), Romanowicz, Young, and Beven (2006), Sahin and Morari (2010), Schoups and Vrugt (2010), Setz, Heinrich, Rostalski, Papafotiou, and Morari (2008), and Sohlberg and Sernfalt (2002).

In this paper we use several system identification techniques to find models of the upper part of Murray River in Australia. The methods we consider are Prediction Error Method (PEM), Maximum Likelihood (ML), continuous time system identification, Refined Instrumental Variable (RIV) method (used within the context of Data-Based Mechanistic (DBM) modelling) and Subspace Identification Method (SIM). Some of the methods have previously been applied to rivers, e.g. PEM was used in Foo et al. (2012, 2014) and Maxwell and Warnick (2006), SIM was used in Pianosi et al. (2012), and DBM was applied in Sohlberg and Sernfalt (2002), Young (2003, 2013), and Young, Castelletti, and Pianosi (2007).

Important factors in system identification of rivers are the accuracy of the obtained model and its suitability for the intended purpose, the ability to incorporate prior information such as water level-flow relationships (Bos, 1989), the ability to handle multiple

---

* Corresponding author.
*E-mail addresses:* hnasir@student.unimelb.edu.au (H.A. Nasir), ewey@unimelb.edu.au (E. Weyer).

http://dx.doi.org/10.1016/j.conengprac.2016.04.006
0967-0661/

inputs and outputs, and the availability of easy to use software.

The above listed methods have their own advantages and drawbacks. It is relatively easy to incorporate prior information in PEM and ML based approaches (Ljung, 1999; Söderström & Stoica, 1988). The Data-Based Mechanistic (DBM) approach has been successfully applied to many catchments and rivers (e.g. Young, 2003, 2013; Young et al., 2007), especially if rainfall-runoff effects need to be taken into account. Subspace identification (Overschee & Moor, 1996; Verhaegen & Verdult, 2007) is well suited for Multi-Input, Multi-Output (MIMO) systems, and rivers are often MIMO with inflows from a number of tributaries, and we are often interested in modelling many flows and water levels along the river. Continuous time system identification (e.g. Garnier, Mensler, & Richard, 2003) is also of interest since continuous time models are often preferred due to easy interpretations and usage. However, each method has drawbacks too, e.g. for optimisation based methods like PEM and ML, it can get computationally difficult to estimate a model with many parameters, and for subspace methods, it becomes challenging to incorporate available prior information if the number of outputs increases.

In this paper we compare the above identification techniques, and in particular we consider (i) simulation performance on validation data, (ii) ability to incorporate available prior information, and (iii) ease in identifying models. The intended use of the models is control under normal operating conditions (i.e. not under flood conditions) and we also put emphasis on the usefulness of the obtained models for control.

This paper is organised as follows. In Section 2 we describe the upper part of Murray River, the operational objectives and the available prior information. We also narrow down the phenomena to be modelled either due to lack of data or due to minor relevance for control. Furthermore we present the dataset used for identification and estimate key time delays in the system. In Section 3 we briefly discuss each identification method and apply them to Multi-Input Single-Output (MISO) models of the upper part of Murray River. Section 4 is dedicated to MIMO models. In Section 5 we compare the identified models before giving some concluding remarks in Section 6.

## 2. Upper part of Murray River

In this section we describe the upper part of Murray River, and we discuss the operational objectives and the available prior information. Furthermore, we present the data used for identification.

### 2.1. River description

Murray River is the longest river in Australia. Fig. 1 shows a sketch of the river stretch from Hume Reservoir to Lake Mulwala which has a river distance of 180 km and a straight line distance of 65 km. The release from Hume is measured at Heywoods. The maximum discharge capacity from Hume is approximately 600,000 ML/day ($\approx 6944\ \mathrm{m^3/s}$) at full supply level. Kiewa River joins Murray River just downstream of Heywoods as shown in Fig. 1. Inflow from Kiewa River is measured at Bandiana.[1] There are several measuring stations on Murray River upstream of Lake Mulwala, such as Doctor's Point, Albury, Howlong and Corowa. Ovens River joins Murray River just upstream of Lake Mulwala. Inflow from Ovens River is measured at Peechelba. Yarrawonga Weir which is situated at the downstream end of Lake Mulwala controls the release of water from the lake. Two irrigation canals also originate from the lake; Yarrawonga Main Channel and the Mulwala Canal with discharge capacities of 3170 ML/day ($\approx 37\ \mathrm{m^3/s}$) and 10,000 ML/day ($\approx 116\ \mathrm{m^3/s}$) respectively.

### 2.2. Operational objectives

There are two main operational objectives for the river stretch in Fig. 1.

1. The water level in Lake Mulwala should be kept between 124.65 and 124.9 mAHD (meter Australian Height Datum, which is relative to sea level). This is required in order to facilitate gravity fed diversion of water into Mulwala Canal and Yarrawonga Main Channel, safe boating and recreational activities in the lake.
2. The rate of fall in the water level at Doctors Point should be less than 15 cm/day to avoid river bank slumping. Doctors Point is just downstream of Heywoods (see Fig. 1).

During normal operations the river stretch is controlled from Hume Reservoir only, and the release from Yarrawonga Weir is used to meet downstream operational objectives. However, during flood operations Yarrawonga Weir is also used to achieve the objectives listed above.

In this paper we consider Multiple-Input, Single-Output (MISO) models and Multiple-Input, Multiple-Output (MIMO) models of the water level in Lake Mulwala and the flow at Doctors Points which are the two most important variables from an operational perspective. The purpose of the model is to use it for control design, and it is not intended for flood prediction. Control design is outside the scope of this paper, but some preliminary simulation results can be found in Nasir and Weyer (2014).

### 2.3. Available prior information

The following prior information is available for the upper part of Murray River,

1. The direction of the flows in Murray River as indicated in Fig. 1.
2. Flow releases from a storage only affect downstream water levels, downstream flows and the immediate upstream water level in the storage.
3. Lakes and reservoirs can often be modelled by using a volume balance. E.g. for Lake Mulwala we can use the following approximate volume balance as a starting point,

$$\begin{aligned}\frac{dV_{LM}}{dt} &= Q_H(t-\tau'_H) + Q_B(t-\tau'_B) + Q_P(t-\tau'_P) - Q_{DYW}(t)\\ &\quad - Q_{YMC}(t) - Q_{MC}(t),\end{aligned} \tag{1}$$

which simply says that the net change in volume of the water in the lake ($V_{LM}$) is equal to the sum of the inflows ($Q$) from Heywoods (H), Bandiana (B) and Peechelba (P) minus the sum of the outflows at (Downstream) Yarrawonga Weir (DYW), Yarrawonga Main Channel (YMC) and Mulwala Canal (MC). $\tau'_H$, $\tau'_B$ and $\tau'_P$ are the time delays from Heywoods, Bandiana and Peechelba to the lake. We consider Eq. (1) as a starting point for selecting a model class in a system identification setting. However, as mentioned in the previous sub-section, water level in the lake is the important variable in the river stretch, and because of that we want the water level as the output rather than the volume. Therefore, assuming direct proportionality between the volume and the water level in the lake, we get

---

[1] Strictly speaking, flows are not measured but they are calculated from water level measurements using rating curves.