Journal of Cleaner Production 154 (2017) 353–365

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Medium-term storage volume prediction for optimum reservoir management: A hybrid data-driven approach

Edoardo Bertone [a, *], Kelvin O' Halloran [b], Rodney A. Stewart [a], Guilherme F. de Oliveira [c]

[a] Griffith School of Engineering and Cities Research Institute, Griffith University, Gold Coast Campus, QLD, 4222, Australia
[b] Seqwater, Ipswich, QLD, 4305, Australia
[c] Griffith School of Engineering, Griffith University, Gold Coast Campus, QLD, 4222, Australia

ARTICLE INFO

Article history:
Received 10 December 2016
Received in revised form
27 March 2017
Accepted 1 April 2017
Available online 5 April 2017

Keywords:
Probabilistic forecasting
Drinking water treatment
Nonlinear regression
Water level prediction
Data-driven modelling

ABSTRACT

A hybrid regressive and probabilistic model was developed that is able to forecast, six weeks ahead, the storage volume of Little Nerang dam. This is a small elevated Australian drinking water reservoir, gravity-fed to a nearby water treatment plant while a lower second main water supply source (Hinze dam) requires considerable pumping. The model applies a Monte Carlo approach combined with nonlinear threshold autoregressive models using the seasonal streamflow forecasts from the Bureau of Meteorology as input and it was validated over different historical conditions. Treatment operators can use the model for quantifying depletion rates and spill likelihood for the forthcoming six weeks, based on the seasonal climatic conditions and different intake scenarios. Greater utilization of the Little Nerang reservoir source means a reduced supply requirement from the Hinze dam source that needs considerable energy costs for pumping, leading to a lower cost water supply solution for the region.

1. Introduction

Accurately predicting water level variations in lakes and reservoirs in response to hydroclimatological changes is crucial for efficient water demand – related decision-making (Güldal and Tongal, 2010) and eventually developing wise and sustainable water usage policies (Buyukyildiz et al., 2014). Understanding and forecasting water level fluctuations (WLFs) is also critical for a variety of water resource management operations such as flood control, local water supply management, shoreline maintenance, ecosystem sustainability, recreation, and economic development (Altunkaynak, 2014). In this research study, we have focused on the importance of predicting the water level of a drinking water reservoir to optimize the operations of the water treatment plant (WTP) receiving raw water from it. The study objective was to provide WTP operators with a support tool to inform decisions on the volume of reservoir raw water that could be withdrawn, considering both spill minimization and depletion risk scenarios. When WTP operators have two raw water source options available to withdraw from, an increased use of the source not requiring pumping has significant implications on the energy costs of the water supply of the city. Based on the electricity billing frequency and other contextual factors, a medium-term (i.e. 1–2 months) forecasting horizon was targeted for such prediction model. Although several studies exist that forecast WLF as detailed in the "Background" section, usually they aim at predicting either (1) in the very short term through very accurate, but computationally demanding models, or (2) in the longer term with more vague, statistical models mainly based on historical seasonal behaviors. For the context of forecast model development (i.e. targeted forecast horizon, data limitations and uncertainty) pertaining to this research study, an effective compromise between accuracy and computational time was required. As a result, we proposed a probabilistic model, founded on a Monte Carlo approach coupled with nonlinear regression analysis, whose outputs are based not only on historical lake behavior and planned outflows, but also on future long-term inflow predictions provided by the Australian Bureau of Meteorology (BoM). Such an integrative approach, detailed later in the paper, allows not only for exploiting existing data and other third-party prediction models to enhance accuracy in a computationally effective manner, but also for improved handling of uncertainty through probabilistic outputs and simulation of several different scenarios.

The paper is structured as follows: in Section 2 we provide background information regarding the importance of predicting

* Corresponding author.
E-mail address: e.bertone@griffith.edu.au (E. Bertone).

http://dx.doi.org/10.1016/j.jclepro.2017.04.003

WLFs as well as previous WLFs prediction models; we then identify challenges and limitations of existing models, and propose a modelling approach that best suits our case study. In Section 3, the research methodology is described in detail, starting with a description of the research domain, data collection and analysis procedure, and finally model development and validation. In Section 4, we describe the results obtained, and in Section 5 we conclude this paper by summarizing the findings and the relevance of this study.

2. Background

2.1. The importance of water level fluctuations

The role that WLFs play, either within and between years, in influencing water quality is not quite well understood, in spite of a rise in water level regulation (White et al., 2008). However, White et al. (2008) found that WLFs between years showed relevant concordance with water quality parameters, as has been proposed by other authors. It was demonstrated by Webster et al. (1996, 2000) that Wisconsin lakes of different landscape positions reacted distinctively to increases in magnesium and calcium concentrations after a drought that lasted 2 years. It has been noticed that water quality in aquatic ecosystems is decreasing, during low-water periods: for instance Arfi (2003) showed that a Mali reservoir shifted from an oligotrophic to a eutrophic state during low water levels. Similarly, Kangur et al. (2003) found that water level greatly affects the nutrient concentrations of an Estonian lake, and in turn the stability of its ecosystem. Nõges et al. (2003) confirmed this finding, determining that the amount of phytoplankton in another Estonian lake was proportional to the water level. Also, WLFs can directly affect the hydrology and ecology within the lake and the surrounding watersheds (Leira and Cantonati, 2008), with high levels affecting the lakeside plant and animal communities and possibly resulting in shoreline erosion (Meadows et al., 1997).

Moreover, depending on specific physical aspects of a body of water, the entire trophic structure of the ecosystem may be harmed by human-induced WLF, spoiling not only water quality but also endangering fish resource exploitation. Nevertheless, as Coops and Hosper (2002) and Roelke et al. (2003) have observed for, respectively, restoration of shallow lakes and phytoplankton control, WLF represents a means to improve water quality and may be utilized as a management tool for freshwater ecosystems. In the case of a dam, which is used as source of water for a water treatment plant, good-quality raw water would reduce the amount of chemicals dosage needed, eventually enhancing the water treatment plant efficiency.

2.2. Forecasting storage volumes

Forecasting future storage volumes can be challenging, as changes in lake water level are considerably complex outcomes of many hydrological factors such as rain falling on the lake surface or lake watershed, evaporation from the lake surface, direct and indirect runoffs from neighbour basins, humidity, air and water temperature, wind speed and groundwater change (Buyukyildiz et al., 2014), as well as hysteresis effects between storage volume and runoff (Wu et al., 2011). As a consequence, to build a reliable water level prediction model, it is crucial to determine the correlations between the change in lake water level and those hydro-meteorological variables (Kisi et al., 2012). However, a number of models have been developed, which could forecast lake water levels one day (e.g. Kisi et al., 2012; Afiq et al., 2013), month (e.g. Altunkaynak, 2007; Buyukyildiz et al., 2014), or year (e.g. Privalsky, 1992) in the future with different modelling approaches.

If all the required hydrological and hydro-meteorological data are available, the change in lake water level can be generally determined by water balance methods, which in essence estimate the difference between inflow to, and outflow from, the body of water. For instance, Crapper et al. (1997) made use of an existing soil moisture prediction model to estimate surface runoff flowing into an Australian lake, and along with rainfall and evaporation data, they could estimate monthly lake water level variations. However, evaporation was the only outflow considered and thus it was assumed that the model overestimates the real lake level. Potentially, this problem could have been solved by comparing observed and predicted levels and correlating the values (e.g. regression analysis).

Classical mathematical and statistical approaches have been also widely applied for lake water level forecasting (Goodarzi et al., 2014). A lot of research in relation to this topic has been performed in the Great Lakes region. Although some more recent studies have used deterministic models (e.g. Lofgren et al. (2002) deploying a hydrological modelling approach using the input of a global circulation model to assess climate change impacts), most of these studies applied conventional parametric time series modelling techniques. Examples are given by DeCooke and Meregian (1967), with multiple linear regression up to 6 months ahead; Cohn and Robinson (1976), based on historical cycles identified through spectral analysis; or Irvine and Eberthardt (1992) up to six months ahead using seasonal ARIMA. In addition, a seasonal autoregressive model (SAR) was developed by Privalsky (1992) for predicting the Lake Erie water level up to 12 months ahead based on over 60 years of historical mean monthly water level data. Despite not being able to take into account nonlinear behaviors, SAR is believed to be fast and easy to use in this context (Khan and Paulin Coulibaly, 2006). However, it could be argued that they offer a "naïve" forecast, largely based on historical seasonal behaviors, which might work well for large lakes, prone to proportionally much lower rates of level fluctuations over a period of 1–2 months, compared to a small reservoir such as the one of this study.

More recently, artificial intelligence techniques have been applied to develop increasingly accurate water level forecasting tools. Khan and Paulin Coulibaly (2006) investigated the potential of support vector machine application for long-term prediction of the water level in Lake Eire. Vaziri (1997) used a hybrid artificial neural network (ANN) and ARIMA model to predict water levels of the Caspian Sea; he observed that on average the ANN model underestimates the levels by 3 cm while the ARIMA model overestimates by the same amount. To develop the ANN he took 12 nodes in the input layer, representing the Caspian Sea water levels of the previous 12 months. Altunkaynak (2007) stated that, in the attempt to predict Lake Van water level one month ahead, neural networks could successfully model the complex relationship between rainfall and water levels; he also concluded that both ANN and ARMAX models can be effectively applied for short term predictions of the relevant time series data. Kakahaji et al. (2013), after developing different linear (i.e. ARX and Box-Jenkins approaches) and artificial intelligence (Multi-Layer Perceptron ANN and Local Linear Neuro-Fuzzy) models for water level prediction based on the water budget approach, concluded that intelligent methods are superior to traditional models. ANN alternatively have been also used to predict meteorological variables which could be input to a water level prediction model, such as solar radiation (Qazi et al., 2015). Neural Networks were also used by Güldal and Tongal (2010) as one of the models to predict the water level in Lake Egirdir, in Turkey. In Buyukyildiz et al. (2014), five different artificial intelligence methods were used to estimate monthly water level changes in the Turkish Lake Beysehir, with support vector regression being the best performing model. In other studies (Altunkaynak, 2014) ANN's have been combined with other