Computers & Geosciences 36 (2010) 620–627

Contents lists available at ScienceDirect

Computers & Geosciences

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

Sea water level forecasting using genetic programming and comparing the performance with Artificial Neural Networks

Mohammad Ali Ghorbani [a,*], Rahman Khatibi [b], Ali Aytek [c], Oleg Makarynskyy [d], Jalal Shiri [a]

[a] Water Engineering Department, Faculty of Agriculture, University of Tabriz, Tabriz, Iran
[b] Halcrow Group Ltd, Burderop Park, Swindon, UK
[c] Department of Civil Engineering, Hydraulic Division, Gaziantep University, 27310 Gaziantep, Turkey
[d] Asia–Pacific Applied Science Associates, P.O. Box 7650, Perth 6850, Australia

ARTICLE INFO

Article history:
Received 8 March 2009
Received in revised form
31 August 2009
Accepted 2 September 2009

Keywords:
Sea-level variations
Forecasting
Artificial Neural Networks
Genetic Programming
Comparative studies

ABSTRACT

Water level forecasting at various time intervals using records of past time series is of importance in water resources engineering and management. In the last 20 years, emerging approaches over the conventional harmonic analysis techniques are based on using Genetic Programming (GP) and Artificial Neural Networks (ANNs). In the present study, the GP is used to forecast sea level variations, three time steps ahead, for a set of time intervals comprising 12 h, 24 h, 5 day and 10 day time intervals using observed sea levels. The measurements from a single tide gauge at Hillarys Boat Harbor, Western Australia, were used to train and validate the employed GP for the period from December 1991 to December 2002. Statistical parameters, namely, the root mean square error, correlation coefficient and scatter index, are used to measure their performances. These were compared with a corresponding set of published results using an Artificial Neural Network model. The results show that both these artificial intelligence methodologies perform satisfactorily and may be considered as alternatives to the harmonic analysis.

1. Introduction

Water level forecasting at various time intervals using records of past time series is important in water resources for their planning, maintenance and operations. For instance, sea level affects groundwater tables in low-lying coastal areas, as well as hydrological regimes of some coastal rivers (Thain et al., 2004). Hence, a reliable forecast of sea-level variations is required in coastal engineering and hydrologic studies.

Variations in sea level are a complex outcome of many environmental forcings, such as lunar and solar gravitational attraction, ocean waves and currents, atmospheric pressure and wind forcing, as well the shape of the continental shelf. Conventional approaches are based on harmonic analysis by regarding tides as the superposition of different harmonics and deriving their frequencies from astronomy but other techniques are also used such as describing tides as propagation of long waves or using regression techniques. The applications of conventional approaches are limited due to their inherent restrictive assumptions on the following grounds. (i) Such methodologies used for sea-level predictions are initially data-intensive and several years of tidal observations have to be collected and processed for reliable sea-level estimates (e.g. Poutanen, 2000). (ii) Transfer of knowledge to a new location of interest with scarce tidal observations can be problematic, as the respective site may share only similar amounts of gravitational actions by the Moon and the Sun but differ in other components, which can be responsible for as much as 30% mismatch between measurements and harmonic estimates, Makarynska and Makarynskyy, 2008.

Since the 1990s, time-series methods employing the Genetic Programming (GP), Artificial Neural Network (ANN) and fuzzy logic methods have become viable, leading to the publication of a considerable amount of literature. The focus of this paper is on the application of a GP model to forecast sea level time series, which are data-driven modeling approaches, first proposed by Koza (1992), as a generalization of Genetic Algorithms (GAs; Goldberg, 1989).

The GP methods are wide-ranging, similar to the GA, where their details may be obtained from various textbooks (e.g. Koza, 1992). Generally they are robust applications of optimization algorithms and represent one way of mimicking nature. The techniques have the capability for deriving a set of mathematical expressions to describe the relationship between independent and dependent variables using operators such as mutation, recombination (or crossover) and evolution. To be elaborated later, these are operated in a population evolving in generations

* Corresponding author.
E-mail address: cusp2004@yahoo.com (M. Ali Ghorbani).

0098-3004/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2009.09.014

through a definition of fitness and selection criteria, where the subsequent techniques are data-driven. GP techniques are particularly applicable to cases where: (i) the interrelationships among the relevant variables are poorly understood or suspected to be wrong; (ii) finding the size and shape of the ultimate solution is itself a major part of the problem; (iii) conventional mathematical analyses are constrained by restrictive assumptions but approximate solutions are acceptable; (iv) small improvements in performance are routinely measured, easily measurable and highly prized and (v) the amount of data is large, e.g. satellite observation data, requiring examination, classification and integration (Banzhaf et al., 1998).

As applications of the GP have been widening, the focus of this paper is to apply it to forecast sea level three time steps ahead using time series averaged over 12 h, 24 h, 5 day and 10 day time intervals. The paper compares these results with those published by Makarynskyy et al. (2004) using an ANN. A sufficient amount of information is therefore reproduced here on the ANN but reference may be made to the published paper by Makarynskyy et al. (2004) for a detailed account.

2. Literature review

As tidal harmonic analysis conventionally used for sea-level predictions may suffer from 30% residual errors for catering only to the solar and lunar gravitational forcings and neglecting any hydrometeorological effects, it is not possible to compare directly the model results produced in this study with those produced by harmonic analysis in any meaningful way (Makarynska and Makarynskyy, 2008). Instantaneous measurements and averaged values of sea level are available temporally and/or spatially under the influence of variable tides, sea water temperature, etc. (e.g. Makarynskyy et al., 2004). They make up time series and some of the techniques used for their analysis are reviewed below.

Zaldivar et al. (1998) used chaos theory techniques for the detection of high water levels in Venice, Italy. Based on their study, non-linear approaches proved capable of simulating dynamic normal trend of water level. Livinia et al. (2003) applied stochastic models to estimate fluctuations of river discharges. Rahmstorf (2007) used a semi-empirical approach to study sea level fluctuations based on earth temperature changes.

Khu et al. (2001) applied the GP to real-time runoff forecasting for the Orgeval catchment in France and compared the findings with observed and calculated values using other methods such as the Kalman filter. Their results indicated an acceptable accuracy for the GP. Also, Babovic and Keijzer (2002), Muttil and Liong (2001), Liong et al. (2002) and Aytek and Alp (2008) applied the GP for rainfall-runoff modeling. Giustolisi (2004) determined Chezy resistance coefficient using the GP. Borelli et al. (2006) introduced an approach based on the GP for extracting trends in noisy data series. Kalra and Deo (2007) applied the GP for filling missing data in wave records along the west coast of India. Sheta and Mahmoud (2001) forecasted the Nile river flow in the Northern Sudan using the GP. Aytek and Kisi (2008) applied the GP for modeling suspended sediment in streams, concluding that the GP would improve over the conventional rating curves and multi-linear regression techniques and this approach would provide a useful tool in solving specific problems in water resources engineering. Ustoorikar and Deo (2008) used the GP for filling up gaps between data of wave heights. Gaur and Deo (2008) applied the GP for real-time wave forecasting.

ANNs can approximate any non-linear mathematical time series, so the prediction of water level would be achieved with an acceptable accuracy by using ANNs (Hornik, 1993). They have been used extensively for predicting water level fluctuations. Coulibaly et al. (2001) used an ANN model for predicting groundwater table fluctuations. Makarynskyy et al. (2004) used ANNs for forecasting sea level variations in Hillarys Harbor, Australia. Alvisi et al. (2006) predicted water levels using ANNs and fuzzy logic and found that the precision of ANNs over fuzzy logic would be higher whenever more reliable input data are used. Modeling the relationship between water surface and discharge has also been studied using ANNs by Bhattacharya and Solomatine (2005). Chang and Lin (2006) study multi-point tidal water-level prediction for sites with tidal characteristics similar to a reference site. They express tide generating functions in terms of a number of parameters based on essential physical concepts of tidal propagation and tide-generating forces. Using the ANN, they derive the parameters for various sites in terms of those of a reference site with the trained ANN model. Comparing their results with those from a global ocean tidal model, they conclude that their model is applicable if there is a similarity in tide types between the reference site and the application sites, but the applicability reduces as the bathymetric variations become complex.

3. Specification of the study

3.1. Data used

The present study is based on the recorded data used by Makarynskyy et al. (2004) at Hillarys Boat Harbor, Western Australia, at latitude 31.82° South and longitude 115.73° East. In this study, the hourly sea-water level measurements from December 1991 to December 2002 were used for training and validating a GP model of the time series of the site. Fig. 1 shows the geographical position of the studied site.

The measurements show a visible seasonal variability with annual minima during summers and maxima during winters, which are the results of solar and lunar gravitational effects. The recorded values range from 140 mm (December 1993) to 1680 mm (July 1995) with respect to an unspecified local datum. However, in a normal year the range of fluctuations does not exceed 1200 mm. Table 1 gives yearly statistical parameters of the data.

The initial time-series data of water levels were obtained at an hourly interval but different disciplines require the processing of

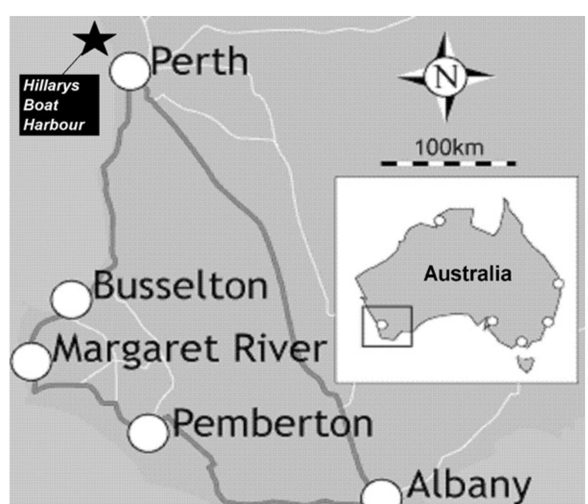

Fig. 1. Site location of single tide gauge at Hillarys Boat Harbor, Western Australia (Makarynskyy et al., 2004).