Engineering Science and Technology, an International Journal xxx (2016) xxx–xxx

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Engineering Science and Technology, an International Journal

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

JESTECH

Full Length Article

Augmented chaos-multiple linear regression approach for prediction of wave parameters

M.A. Ghorbani [a,*], H. Asadi [a], O. Makarynskyy [b], D. Makarynska [c], Zaher Mundher Yaseen [d]

[a] Water Engineering Department, University of Tabriz, Tabriz, Iran
[b] Australian Institute of Marine Science, Arafura Timor Research Facility, 23 Ellengowan Drive, Brinkin 0810, Northern Territory, Australia
[c] AECOM, 8/540 Wickham Street, Fortitude Valley 4006, Queensland, Australia
[d] Civil and Structural Engineering Department, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor Darul Ehsan, Malaysia

ARTICLE INFO

Article history:
Received 4 August 2016
Revised 30 November 2016
Accepted 1 December 2016
Available online xxxx

Keywords:
Chaos theory
Coastal
Prediction
Tasmania
Wave height
Wave period

ABSTRACT

Prediction of wave parameters is one of the significant component for several coastal applications; for instance, coastal erosion, inshore and offshore structures, wave energy and others. The current research investigates the potential of the Chaos theory integrated with multiple linear regression (Chaos-MLR) in prediction of wave heights and wave periods. The wave data were collected at four moorings in the coastal environment of Tasmania. In the first stage, reconstructing the phase space and determine the input data for Chaos-MLR model, the delay time and embedding dimension are computed using average mutual information and false nearest neighbors' analyses. The presence of chaotic dynamics in the used data is identified by the correlation dimension methods. In the second stage, the Chaos-MLR and pure MLR models are constructed for prediction model. Absolute error and best-fit-goodness diagnostic indicators are utilized to inspect the proficient of the proposed model in comparison with the pure MLR model.

The inter-comparisons demonstrated that the Chaos-MLR and pure MLR models yield almost the same accuracy in predicting the significant wave heights and the zero-up-crossing wave periods. Whereas, the augmented Chaos-MLR model is performed better results in term of the prediction accuracy vis-a-vis the previous prediction applications of the same case study.

1. Introduction

Parameters prediction of wind waves are important for coastal engineering applications, navigation, recreational activities, for developing alternative energy techniques, as well as for better understanding near-shore and coastal ecosystems [20].

The ability of producing metocean parameter predictions based on time series and with minimum site information requirements lead to the artificial intelligence, data-driven techniques becoming increasingly popular within the applied science community. These data-driven are based on learning machine methods and are characterized by flexibility, accurate and nor required a prior assumption. In the light of the literature, prediction of wave parameters were conducted using several data-driven for instance artificial neural network (ANN), genetic programming (GP), M5 model Tree (e.g., [2,4–7,19–21,24,27,33,35,36] among many others). However, the accuracy of those conceptual models influenced by certain critical variables like model structure, correlated input parameters, the flexibility of method mechanism to mimic the complicated relationship, and many others criteria.

Speaking within the scope of the current research, as a prior step in accomplishing an accurate model, selecting appropriate inputs is one of the most important step in the development of a data driven model. There have been a number of methods developed to complete this step, such as sensitivity analysis, the Gamma test, partial mutual information, hybrid independent component analysis, input variable selection filter, principal component analysis and cross-correlation analysis (e.g. [17]), but the selection of input structures is still insufficiently supported from the theoretical point of view [29].

Chaos theory is relatively modern tool in analysis chaotic time series data set that characterized by highly non-linearity and non-stationary patterns [12]. In this research, the chaos theory used as a sufficient tool for studying wind waves. According to the chaos theory concept, a dynamic system can be represented by a set of effective variables and can be modeled by a phase space

* Corresponding author.
E-mail address: Ghorbani@tabrizu.ac.ir (M.A. Ghorbani).
Peer review under responsibility of Karabuk University.

http://dx.doi.org/10.1016/j.jestch.2016.12.001
2215-0986/© 2016 Karabuk University. Publishing services by Elsevier B.V.

diagram such that each point on the diagram represents the system's behavior at a specific time. A method for reconstructing phase space from an observed time series has been presented by Takens [30]. Determination of delay time and embedding dimension is the basis of phase space reconstruction. Systems with different embedding dimensions can be reconstructed in different phase spaces, therefore, will correspond to different prediction results [25].

Selecting appropriate inputs is one of the most significant steps in the development of a data driven model. There have been a number of methods developed to complete this step, such as sensitivity analysis, the Gamma test, partial mutual information, principal component analysis and cross-correlation analysis (e.g. [17]), but the selection of input structures is still insufficiently supported from the theoretical point of view (e.g. [29]).

The chaos theory may be a useful tool for studying river flow. According to the chaos theory concept, a dynamic system can be represented by a set of effective variables and can be modeled by a phase space diagram such that each point on the diagram represents the system's behavior at a specific time [15]. Determination of delay time and embedding dimension is the basis of phase space reconstruction (PSR). Systems with different embedding dimensions can be reconstructed in different phase spaces; therefore, will correspond to different prediction results [25].

So far, combing the phase space reconstruction with data driven methods has been successfully used in time series predictions such as prediction of network traffic [18], sea level [29] and evaporation simulations [3], modeling of exchange rates [12], and prediction of stock market prices [14]. The above studies demonstrated that combined with chaos theory models perform better than the pure models. This assumption is tested in the present study, the objective of which is to combine the phase space reconstruction with multiple linear regression model for predicting wave heights and wave periods in the nearshore environments of Tasmania. The remainder of this article is organized as follows. In the second section, the methodology is presented including phase space reconstruction, the proposed model, and the case study. Section 3 indicates the application and analysis with detailed discussion. Conclusions are displayed in Section 4.

2. Proposed PSR-MLR model

In this paper, four different methods are used to characterize chaotic approach: (1) phase space reconstruction; (2) False Nearest Neighbour (FNN) algorithm; (3) Mutual Information Function (MIF) and (4) Correlation Dimension (CD). Finally, Chaos-MLR is used to predict the wave parameters. The above mentioned methods are explained in detail in the following subsections.

2.1. Chaos theory

A chaotic system reveals a relatively complex behavior through the dynamic of a non-linear system. The orbits of the system attract a complex higher-dimensional subset called a strange attractor. The importance of studying chaotic behavior lies in the fact that chaotic behavior is much more widespread, and may even be the norm in the real world [12]. In this paper, four different methods are used to characterize chaotic approach: (1) phase space reconstruction which requires determination of the embedding parameters for preparing the input data; (2) False Nearest Neighbour (FNN) algorithm for finding the embedding dimension; (3) Mutual Information Function (MIF) to determine the optimum time delay and (4) Correlation Dimension (CD) to determine the presence of chaotic dynamics.

2.2. Phase space reconstruction (PSR)

The concept of chaos theory is a powerful tool for characterizing dynamic systems (e.g. De Domenico [4]). Each dynamics system may be stochastic, deterministic or chaotic which can be identified using the phase space concept (e.g. [36]).

Phase-space reconstruction theory is the basis for chaotic time series prediction. As for a chaotic system, the phase space can be used to reconstruct a univariate time series, because all the variable information in this dynamic system may be contained in the univariate time series (e.g., [16]). Every point in a phase space shows a state of the system and every trajectory represents the time evolution of the system corresponding to different initial conditions. Points or a set of points in a phase space compose a unique pattern which attract trajectories onto itself. These kinds of patterns are called attractors. A phase space reconstruction is necessary in order to quantitatively estimate the attractor's complexity of the system and then determine whether the observed dynamic behaviors are complex or not. From a one dimensional time series, a phase space can be constructed using the Takens time-delay embedding theorem [30]. The theorem paved the way for the analysis of chaotic time series. The theorem establishes that, given a scalar time series $S_t = (X_1, X_2, X_3, \ldots, X_n)$ from a chaotic system, it is possible to reconstruct a phase space in terms of the phase space vectors X_t expressed as [30]:

$$X_t = (X_t, X_{t-\tau}, X_{t-2\tau}, \ldots, X_{t-(m-1)\tau}) \quad t = 1, 2, \ldots, M; M = n - (m-1)\tau \tag{1}$$

here m is the embedding dimensions or dimension of phase-space reconstruction, τ is a time delay and M is the number of phase points of reconstructed phase-space. Phase space diagram can give information about the dynamics of a system through its trajectory [28].

When unfolded into the phase space, the underlying structures in the chaotic time series can be revealed. The input signals of the network can be selected as the components of the current phase space vectors, i.e. $X_t, X_{t-\tau}, X_{t-2\tau}, \ldots, X_{t-(m-1)\tau}$, while keeping the future value X_{t+T} as the desired response.

Given a time series S_t, an m-dimension phase space (PhS) can be extended as follows [30]:

$$PhS = \begin{bmatrix} x_1 & x_{1-\tau} & x_{1-2\tau} & x_{1-3\tau} & \cdots & x_{1-(m-1)\tau} \\ x_2 & x_{2-\tau} & x_{2-2\tau} & x_{2-3\tau} & \cdots & x_{2-(m-1)\tau} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{n_m} & x_{n_m-\tau} & x_{n_m-2\tau} & x_{n_m-3\tau} & \cdots & x_{n_m-(m-1)\tau} \end{bmatrix} \tag{2}$$

An important step in reconstructing a suitable phase space is selecting an optimal embedding dimension m and delay time τ.

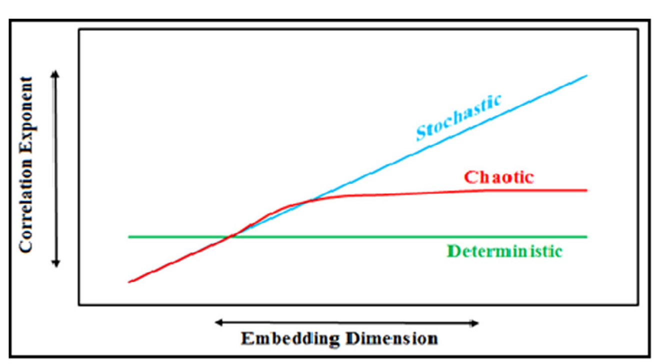

Fig. 1. Plot differentiating stochastic, chaotic and deterministic systems.

Please cite this article in press as: M.A. Ghorbani et al., Augmented chaos-multiple linear regression approach for prediction of wave parameters, Eng. Sci. Tech., Int. J. (2016), http://dx.doi.org/10.1016/j.jestch.2016.12.001