Ocean Engineering 132 (2017) 11–24

Contents lists available at ScienceDirect

Ocean Engineering

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

# A new non-parametric correction model and its applications to hindcasting wave data

Lvqing Wang[a,b], Bingchen Liang[a,b,*], Huajun Li[a,b]

[a] Shandong Province Key Laboratory of Ocean Engineering, Ocean University of China, 238 Songling Road, Qingdao 266100, China
[b] College of Engineering, Ocean University of China, 238 Songling Road, Qingdao 266100, China

ARTICLE INFO

*Keywords:*
Non-parametric correction model
Numerical wave model
Machine learning
Wave
Error correction
WAVEWATCH III

ABSTRACT

In those oceans where measured wave data are not available, numerical wave models are usually adopted to hindcast wave parameters in order to define design waves for marine structures. To utilize these hindcating data, it is very important to perform error corrections of model results for accurate estimation of the appropriate wave parameters. In this paper, a new non-parametric correction model is established to improve wave model accuracy through modifying a previous approach released by Caires and Sterl in 2005. The new correction model introduces a kernel algorithm to learn error information from both value magnitude and series trend through training datasets, and utilizes the information to correct potential errors in model outputs. It is shown that the two-dimensional learning method is more effective than the previous one-dimensional which only learns error information from the value magnitude. Furthermore, an error constraint parameter is initially adopted in the new correction model to decrease the possibility of overcorrection. The new correction model performs better than its predecessor, especially when modeling wave period and altimeter synchronized wave height. Though this paper evaluates the model correcting performance with WAVEWATCH III outputs, the modified model can be adopted to correct other kinds of time-series data.

## 1. Introduction

Wave force is a key environment parameter for designing of the costal and offshore structures. Generally, designing waves are defined as return period values which are calculated following extreme value theories based on hindcasting wave data or long historical measurements (Anthony et al., 2015; Linbin et al., 2015). Measurement data, such as buoy data, are always top priority to act as sampling data for calculation of return values. However, measurement data is not always readily available in the locations of interest.

As a substitute for measurement data, numerical wave models are widely adopted to hindcast wave time-series in kinds of studies. Currently, the third generation wave models highlighted with nonlinear wave interaction dominate wave hindcasting or forecasting operations (Peter and Janssen, 2008) such as WAM(TheWAMDI Group, 1988), WAVEWATCH III (Tolman, 2014) and **S**imulating **WA**ve **N**earshore ( Booij et al., 1999, hereafter as SWAN), all of which simulate wave propagation based on the energy or wave action equation. In these models the wave physics are parameterized through source terms and the parameters have been tuned elaborately according to diverse measured data in the past decades resulting in a robust performance with good agreement with measured data (Tolman, 2014). Generally, WW3 and WAM are designed especially for oceanic or global applications (Amrutha et al., 2016; Bernier et al., 2016; Charles et al., 2012; Chawla et al., 2013; Dykes and Rogers, 2011; Hanafin et al., 2012; Li et al., 2016). On the contrary, SWAN is more efficient and accurate in coastal areas since some nearshore wave evolution processes have been incorporated into the model with parameterized terms, such as depth induced wave breaking and triad wave interactions (Amrutha et al., 2016; Booij et al., 1999; SWAN Team, 2013). Considering its reliable performance and wide applications in the oceanic basin scale, WW3 V4.18 is adopted to hindcast wave data in the South Sea of China (SSC) in the present work.

It is unavoidable that there are biases between the numerical model outputs and the real waves, which are mainly caused by input wind field (Liu et al., 2002; Peter and Janssen, 2008) although numerical approach for the energy equation and parameterization of the wave physics also partly contribute to the errors (Hanson et al., 2009; Tolman and Chalikov, 1996). Researchers have revealed that there are some definite tendencies in the WW3 model errors (Hanson et al., 2009) and different ocean areas are featured with different error tendency (Chawla et al., 2013). The most efficient approach to reduce

* Corresponding author at: College of Engineering, Ocean University of China, 238 Songling Road, Qingdao 266100, China.
*E-mail address:* bingchenliang@aliyun.com (B. Liang).

http://dx.doi.org/10.1016/j.oceaneng.2017.01.010
Received 25 August 2016; Received in revised form 5 January 2017; Accepted 7 January 2017

WW3 model error is to improve quality of wind input data since whose accuracy dominates the credibility of model output, for example making a forecasting wind filed with kinds of source data (Sampson et al., 2013) or with regional higher resolution wind field (Alves et al., 2014).

Over the past few decades, researchers have released different parameterization proposals for wave physics with expectation of improving model performance (Ardhuin et al., 2010; Babanin et al., 2010; Chawla et al., 2013; Hanafin et al., 2012; Peter and Janssen, 2008; Seemanth et al., 2016; Tolman, 1992; Tolman and Chalikov, 1996). Additionally, some studies introduce statistical techniques to evaluate the simulation of wave spectra (Peter and Janssen, 2008) and further to improve the accuracy of simulation. However, it seems that "there may be an underlying limitation to further improvement of models based on the concept of a wave energy spectrum"(Liu et al., 2002).

Though numerous efforts have been made to minimize errors through improvement of model performance, errors always exist in numerical model outputs. Consequently, some researchers introduce error correction on the model results. A simple method is to assume that there are linear relationships between the errors and the model parameters under which all model output data can be corrected with linear or constant coefficients (Mackay et al., 2010; Xiaoli Guo et al., 2015). However, it has been revealed that modeling waves seldom share a universal error ratio (Chawla et al., 2013; Hanson et al., 2009). Differentiating from the linear correcting methods, some researchers attempted to correct errors based on nonlinear methods (Caires and Sterl, 2003, 2005) which had been utilized to correct EAR-40 reanalysis data. Hadadpour et al. (2013) introduced an artificial neural network approach, a widely used machine learning (ML) method, to correct hindcasting results of SWAN model which is proven to be efficient. According to these researches, it is a feasible method to improve hindcasting performance of wave model outputs through correcting model results.

It has been proven strong consistency of estimators of the conditional distribution function and conditional expectation of a future observation of a discrete time stochastic process given a fixed number of past observations (Caires and Ferreira, 2005). Based on the strong consistency, a non-parametric method is proposed to correct ERA-40 wave height (Caires and Sterl, 2005). Because of its superiority without predefined parameters, non-parametric methods are widely used to estimate best fitted values (Zhang and Singh, 2006). The non-parametric method proposed by (Caires and Sterl, 2005) can correct wave height data selectively based on external knowledge learned from a training dataset avoiding arbitrarily identical correction on all wave values. Because the non-parametric correction method is initially released by **CA**ires and **ST**erl in 2005, this method will be abbreviated as CAST model in the following sections.

Although was not explicitly classified as ML model in its debut, the CAST model does essentially perform ML functions according to definition of ML (Lantz, 2015), for example it can learn error information through training datasets and use the learned information to correct errors. In fact, rather than just proposing an error correction method, the initial CAST (Caires and Sterl, 2005) has developed a framework to correct model errors with ML technique.

Though it performs good corrections on ERA-40 wave data, practical operations prove that the previous CAST model fails to conduct effective corrections to the WW3 outputs notably that wave periods are frequently overly corrected referring to the Section 4 of this paper.

The present work proposes an Improved CAST model (ICAST) based on the framework of the previous CAST model aiming to improve accuracy of the results of numerical wave models. WW3 is selected as the example model to be investigated. Generally, the ICAST model shares similar principle with the CAST, that is, firstly to learn error information from a short duration of training datasets and secondly to perform error correction on longer duration of time series. However, the ICAST model adopts a new kernel algorithm. According to the studies in this paper, the ICAST model can perform better correction for WW3 results than the CAST model, especially wave periods can be corrected effectively. The present work is structured as follows: Section 2 gives brief introductions to the datasets used in this paper. Section 3 introduces the formulas of the previous CAST and the present ICAST models. Detailed comparison of the two models are presented in the Section 4. Discussions and conclusions are given in Section 5 and Section 6 respectively.

## 2. Wave data

There are two types of data used in this paper. The first type of data consists of measured data including buoy data and altimeter data, whereas, hindcastng wave data by WW3 model constitute the second type of wave data.

### 2.1. Hindcasting wave data

The hindcasting wave model in this paper is established based on the third generation wave model WW3 (V4.18) (Tolman, 2014). The modeling area covers latitude from 5°N~45°N and longitude from 100°E~ 145°E with grid resolution of 10′×10′and numeric obstacles have been embedded to simulate effect of numerous islands. The NCDC blended sea winds data (Zhang, 2006) (1989–2008, http://nomads.ncdc.noaa.gov/data/seawinds/) and the GFSANL wind data (2009–2014, http://nomads.ncdc.noaa.gov/data/gfsanl) are adopted as input wind forcing the hindcastng model. Classical T & C package (Tolman and Chalikov, 1996) is adopted to simulate wave physics with quaternate time steps set up as '1800s,900s,180 s,40 s′. More than 20 years of wave time series are obtained.

This present paper takes the South China Sea (SCS) as the ocean of interest. Two types of model data are extracted from the hindcasting WW3 model. One type of data is outputted from the WW3 model at the locations of three operational buoys B1~B3, referring to the Fig. 1, which can be regarded as buoy synchronized data. All of the three buoys are moored in open ocean where water depth is approximately 30–50 m. The buoys are approximately 50–85 km apart away from the coastline.

Another set of wave outputs of the WW3 is extracted following synchronizing track of altimeters whose orbiting information can be accessed by the GlobWAVE project (http://globwave.ifremer.fr/). Because of the numeric discretization for solution of partial equations, exactly synchronization between altimeter and model is

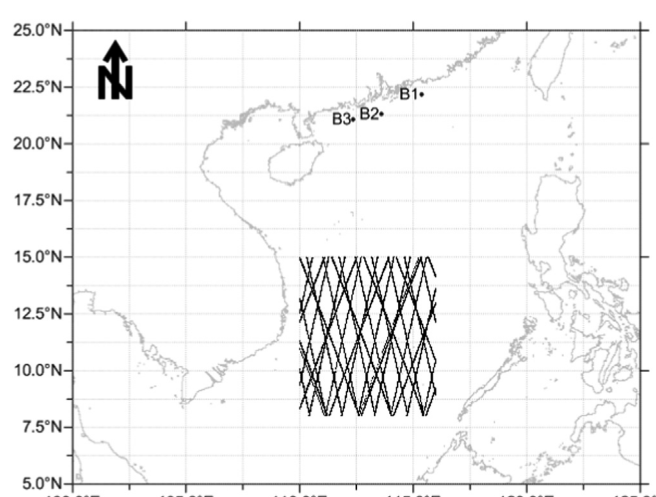


**Fig. 1.** Locations of buoys and covering range of the altimeter wave data (ALWD).