Ocean Modelling 94 (2015) 128–140

Contents lists available at ScienceDirect

Ocean Modelling

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

Significant wave height record extension by neural networks and reanalysis wind data

D.J. Peres*, C. Iuppa, L. Cavallaro, A. Cancelliere, E. Foti

Department of Civil Engineering and Architecture, University of Catania, Via Santa Sofia 64, 95123 Catania, Italy

ARTICLE INFO

Article history:
Received 18 May 2015
Revised 1 August 2015
Accepted 14 August 2015
Available online 21 August 2015

Keywords:
Stochastic models
Soft computing
Italian Sea Monitoring Network
NOAA CFSR
ERA-Interim

ABSTRACT

Accuracy of wave climate assessment is related to the length of available observed records of sea state variables of interest (significant wave height, mean direction, mean period, etc.). Data availability may be increased by record extension methods.

In the paper, we investigate the use of artificial neural networks (ANNs) fed with reanalysis wind data to extend an observed time series of significant wave heights. In particular, six-hourly 10 m a.s.l. $u-$ and $v-$wind speed data of the NCEP/NCAR Reanalysis I (NRA1) project are used to perform an extension of observed significant wave height series back to 1948 at the same time resolution. Wind for input is considered at several NRA1 grid-points and at several time lags as well, and the influence of the distance of input points and of the number of lags is analyzed to derive best-performing models, conceptually taking into account wind fetch and duration.

Applications are conducted for buoys of the Italian Sea Monitoring Network of different climatic features, for which more than 15 years of observations are available. Results of the ANNs are compared to those of state-of-the-art wave reanalyses NOAA WAVEWATCH III/CFSR and ERA-Interim, and indicate that model performs slightly better than the former, which in turn outperforms the latter. The computational times for model training on a common workstation are typically of few hours, so the proposed method is potentially appealing to engineering practice.

1. Introduction

Availability of wave climate data represents a fundamental requisite in coastal and offshore engineering design. Recordings from moored measuring buoys belonging to sea monitoring networks managed by national and international climate monitoring centers represent a primary source of information. Commonly, such wavemeters provide records of sea state variables covering a period not longer than 40–50 years, and generally of about 15–30 years. For instance, the eldest buoy of the U.S. National Oceanic and Atmospheric Administration's National Data Buoy Center (http://www.ndbc.noaa.gov/) started collecting data in the 1978, while in Italy the national meteorological service sea monitoring network (known as RON, Rete Ondametrica Nazionale; http://www.idromare.it/) started to gather data on July of the year 1989. In many cases, typically in the estimation of design waves of high return period, the length of the observed series should be longer in order to have a reliable assessment of the wave climate at the location of interest (U.S.A.C.E., 2006).

A way to increase the amount of information available from sea monitoring networks is to use past wind data as input to models of wave prediction (hindcasting). To this end, several models can be applied, ranging from empirical to the most sophisticated numerical third generation models (Bretschneider, 1951; Kinsman, 1965; Wilson, 1965; Hasselmann et al., 1973; Donelan et al., 1985; The WAMDI Group, 1988; Bouws et al., 1998; Booij et al., 1999; Ris et al., 1999; Goda, 2003). Wind field data produced by reanalysis projects, such as the ERA 40 (Uppala and et al., 2005) (http://apps.ecmwf.int/datasets/) by the European Centre Medium-range Weather Forecast (ECMWF), the JRA-55 (Ebita and et al., 2011) (http://jra.kishou.go.jp/JRA-55/index_en.html) of the Japanese Meteorological Society, as well as the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) Reanalysis I (Kalnay and et al., 1996) (http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html), may be suitable for such a purpose. Website http://www.reanalyses.org/ provides tools for comparisons of the various data sets (Smith et al., 2014). Among the listed projects, the NCEP/NCAR Reanalysis I (Kalnay and et al., 1996) (hereafter indicated as NRA1) is the constantly-updated data set that covers the largest period, which extends from

* Corresponding author. Tel.: +390957382729; fax: +390957382748.
E-mail address: djperes@dica.unict.it (D.J. Peres).

http://dx.doi.org/10.1016/j.ocemod.2015.08.002
1463-5003/

present back to 1948. NRA1 data have been analyzed and used to perform hindcasts in various studies through physically-based models (Sterl et al., 1998; Swail and Cox, 2000; Cox and Swail, 2001; Cieślikiewicz and Paplińska-Swerpel, 2008; Musić and Nicković, 2008; Pilar et al., 2008; Ratsimandresy et al., 2008; Rusu et al., 2008). Nonetheless, use of numerical models may be limited by their complexity, high computational requirements and the need of accurate bathymetric data (Browne et al., 2007).

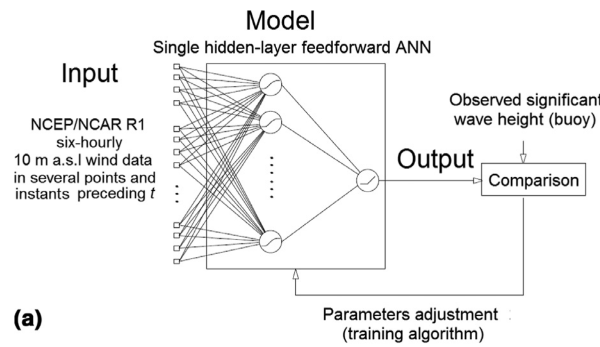
An alternative approach may be based on the use of data-driven models, such as Artificial Neural Networks (hereafter indicated as ANNs), which have the ability to map complex non-linear relationships between input and output variables once a data set of sufficient length of realizations of the process is available (Haykin, 1999). ANNs have found several applications in ocean and coastal engineering (cf. Jain and Deo, 2006; Deo, 2010; Krasnopolsky, 2013), wave reflection estimation (Garrido and Medina, 2012; Zanuttigh et al., 2013) and prediction of sediment suspension in the surf zone (Yoon et al., 2013), among others. Many applications aim at wave forecasting (Deka and Prahlada, 2012; Deo et al., 2001; Deo and Sridhar Naidu, 1998; Etemad-Shahidi and Mahjoobi, 2009; Jain et al., 2011; Londhe and Panchang, 2006; Makarynskyy et al., 2005; Mandal and Prabaharan, 2006; Paplinska-Swerpel et al., 2008; Reikard et al., 2011; Tsai et al., 2002; Zamani et al., 2008). Reikard et al. (2011) suggested that statistical models, such as ANNs, may be preferred to physically-based models, when only very short horizons (less than 6 h) of forecasting are needed. Reikard et al. (2015a, 2015b); have extended the applications to wave energy forecasting. Agrawal and Deo (2004) used ANNs to model the relationship between significant wave height and other sea state variables. Günaydin (2008) applied ANNs and regression methods for the estimation of monthly mean significant wave heights. ANNs have been applied for filling missing data gaps in the measured time series as well. Works on this issue have been developed by Arena and Puca (2004); Hidalgo et al. (1995); Makarynskyy and Makarynska (2007). In all these works ANNs have been used to fill missing values at one buoy, based only on wave observations of nearby buoys. Wave hindcasting based on wind data through ANNs has been investigated to a less extent (Rao and Mandal, 2005; Mahjoobi et al., 2008; Malekmohamadi et al., 2008), and furthermore, in all studies, no real hindcast was carried out, for lack of wind data for extension (no additional data than that used in model development was available).

In this paper, we propose an approach based on ANNs driven by NRA1 six-hourly 10 m a.s.l. sea wind data to extend an existing record of significant wave heights. In particular, we use fully-connected feedforward ANNs to relate significant wave height at a measuring buoy to the NRA1 $u-$ and $v-$winds at 10 m a.s.l. in several points and lagged in time to take into account wind fetch and duration. Early-stopping and Bayesian regularization are applied to ensure good generalization capabilities of the networks. A specific focus is given to the assessment of model performance during storms and at the storm peak, which still represent the real challenge in wave modeling (Cavaleri, 2009). The method is applied to three buoys of the Italian sea monitoring network RON, exploring a range of wave generation areas, in order to evaluate the proposed methodology. The proposed approach is further evaluated, by comparing the ANN data with state-of-the-art global wave hindcast data, and in particular the ECMWF ERA Interim (Dee and et al., 2011) and the NOAA WAVEWATCH III/CFSR (Chawla et al., 2013; Saha et al., 2010).

2. Record extension by artificial neural networks

2.1. Neural network modeling

Artificial neural networks are a type of data-driven models, or black-box models, whose structure and calibration procedure have some analogies with brain neural networks and their learning process



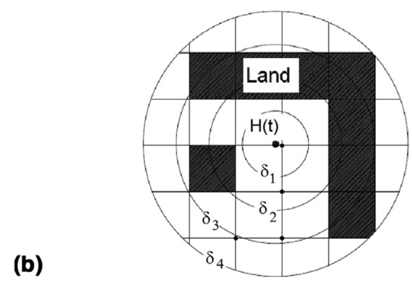


Fig. 1. Sketch indicating: (a) the artificial neural network model and training process (adapted from Demuth et al., 2008) and (b) the wind input point selection and classification method.

(Wenzel and Schröter, 2010). From a mathematical standpoint, they can be considered as multiple nonlinear regression methods able to capture hidden complex nonlinear relationships between input and output variables (Krasnopolsky et al., 2002). ANNs are based on elements named neurons, which transform through a nonlinear transfer function a linear combination of the inputs/preceding layer. The coefficients of the linear combination are named synaptic weights, and together with the network biases they represent the free parameters of the model. Several books focused on ANN modeling do exist, e.g., Freeman and Skapura (1991), Haykin (1999) and Demuth et al. (2014), as well as software toolboxes, devised, e.g., for R (Bergmeir and Benítez, 2012) and MATLAB (Demuth et al., 2008) platforms.

Fig. 1 shows the ANN modeling framework adopted in this work. The significant wave height $H(t)$ at the buoy location P_0 and time t is modeled by an ANN as a function of the six-hourly $u-$ and $v-$wind at 10 m a.s.l. of the NRA1 project, in several grid points P_i within the wave generation area and time instants preceding current time t. This is conceptually done to take into account the fetch and the duration of wind. A systematic analysis of the influence of the extension of the area considered for the wind input as well for its duration is carried out. In particular, the wind input is characterized by two parameters: the maximum distance class R and the length of the temporal window L (indicated as the number of six-hourly temporal values of wind preceding current time t) – we denote the corresponding model as $F_{R,L}$. Six-hourly NRA1 $u-$ and $v-$wind 10 m a.s.l. data are available at the spatial resolution of 2.5° × 2.5°, which roughly corresponds to 220 km in the $u-$direction (W-E) and to 280 km in the $v-$direction (S-N); based on such a resolution, the wind points P_i within the wave generation area are grouped by a distance interval of 200 km (see Fig. 1b). A given point P_i is said to belong to distance class $\delta_r, r = 1, \ldots, R$, if $200(r-1)$ km $< \mathrm{dist}(P_i, P_0) \le 200r$ km. For all points within class δ_r, the wind is considered at time instants $t-(r-1)\Delta t, t-r\Delta t, \ldots, t-(r+L-1)\Delta t$, where the temporal resolution Δt is 6 h. As it can be seen, the first instant is $(r-1)\Delta t$ h before t (for instance for the class $\delta_r = 2$ wind is considered starting from $t-6$ h). This is done to take into account the fact that waves need some time to propagate from the point in which they are generated to the point of interest (buoy location). In particular,