Ocean Engineering 104 (2015) 530–539

Contents lists available at ScienceDirect

Ocean Engineering

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

Short-term forecasting of the wave energy flux: Analogues, random forests, and physics-based models

Gabriel Ibarra-Berastegi [a,*], Jon Saénz [b,c], Ganix Esnaola [d], Agustin Ezcurra [e], Alain Ulazia [f]

[a] University of the Basque Country (UPV/EHU), NE & Fluid Mechanics Department, Faculty of Engineering, Alda, Urkijo, s/n, 48013 Bilbao, Spain
[b] University of the Basque Country (UPV/EHU), Applied Physics II Department, Leioa, Spain
[c] Plentziako Itsas Estazioa (PIE), University of the Basque Country (UPV/EHU), Plentzia, Spain
[d] University of the Basque Country (UPV/EHU), NE & Fluid Mechanics Department Donostia, Spain
[e] University of the Basque Country (UPV/EHU), Applied Physics II Department, Vitoria-Gasteiz, Spain
[f] University of the Basque Country (UPV/EHU), NE & Fluid Mechanics Department, Eibar, Spain

A R T I C L E I N F O

Article history:
Received 12 January 2015
Accepted 26 May 2015
Available online 16 June 2015

Keywords:
Wave energy flux
Forecasting
Random forest
Machine learning
Fluid mechanics
Applied physics

A B S T R A C T

This paper analyzes the performance of three types of statistical models and a well-known physics-based model for forecasting the wave energy flux. The forecasts are run over horizons of 1–24 h at five buoys located in the Bay of Biscay. The data resolution is hourly. The 1999–2005 timeframe is used to train the models. The forecasts are run and evaluated over the six-year period from 2006 to 2012. The statistical forecasting models use three techniques: analogues, random forests (a machine learning algorithm) and a combination of the two. The physics model is the Wave Model (WAM). The forecasts are compared at a 95% confidence level with the simplest prediction – Persistence – and also with the nearest grid point of the WAM forecasts. Over horizons between 3 and 16–19 h at locations near the coast (where wave farms may be installed), the random forests models outperform the others, including WAM and Persistence. These models exploit the inherent predictability associated with the strong autocorrelation present in ocean energy values. The additional prognostic capabilities that random forests models provide over Persistence, are due to their ability to sucessfully incorporate the information that both, atmospheric and sea-state variables provide.

1. Introduction

The last decade has seen increasing interest in ocean wave energy as a renewable power source. In 2008, the first generation of commercial ocean energy devices came into operation. One key issue in grid integration is forecasting, particularly over short horizons, on the order of a few hours. Wave energy is usually expressed in terms of the flux, a function of the significant wave height and the period, and is denominated in kW per metre of crest length (kW/m).

As with other types of renewable energy, wave energy can be difficult to forecast, mainly because of its intermittency (Esteban et al., 2010; Esteban and Leary, 2012). In particular, because the wave height (Hs) and period (Tm) combine multiplicatively in the flux, the forecast errors for the flux can be expected to be higher than for Hs and Tm individually (Pinson et al., 2012). An in-depth review of the current state of the art regarding the short-term prediction of the wave energy flux can be found in the Nitsure et al. (2012), and references therein.

There have been two main approaches to wave forecasting. Large-scale physics-based models are operated by the European Centre for Medium-Range Weather Forecasts (ECMWF), and the National Oceanographic and Atmospheric Administration in the United States. These models are the WAM and WAVEWATCH III respectively (The Wamdi group, 1988; Jansen, 2007; Bidlot et al., 2007; Richardson et al., 2009 and references therein). A smaller physics model, SWAN (Simulating Waves Near Shore), which is adapted for shallow water, was developed in Booij et al. (1999). These models forecast the wave height and period, as well as other parameters such as the wave direction. The energy can easily be calculated from these outputs.

The second branch of the literature has used time series models. These include regressions, neural networks and genetic programming (Deo and Naidu, 1999; Deo et al., 2001; Agrawal and Deo, 2002; Kamranzad et al., 2011; Surabhi and Deo, 2008; Nitsure et al., 2012; Rao et al., 2013; Hadadpour et al., 2014). Other techniques like Support Vector Machines and M5 model trees, have also been used (Etemad-Shahidi and Mahjoobi, 2009; Javad and Ehsan, 2009). It is worth mentioning that a pre-processing step of dimensionality reduction is often taken. Instead of the original variables, scores from the leading

* Corresponding author.

http://dx.doi.org/10.1016/j.oceaneng.2015.05.038
0029-8018/

Empirical Orthogonal Functions (EOFs) are used to feed the models (Rao et al., 2013).

Studies have found that statistical methods forecast more accurately at short horizons, while physics-based models predict more accurately at longer horizons, of about 6 h and beyond. Combining both techniques may forecast better than either one individually (Reikard et al., 2011; Reikard and Rogers, 2011; Pinson et al., 2012). Further, the errors from statistical and physics models evolve differently as the prediction horizon increases. In the case of statistical models, the error increases rapidly over the first few hours. In the physics models, the error increases gradually over a period of days.

The objectives of this research were as follows:

1. Use statistical algorithms and only the data available at time t to obtain hourly forecasts of flux levels for time $t+k$ ($k=1,\ldots,24$) at five locations (buoys) in the Bay of Biscay. The following statistical techniques have been used (i) analogues, (ii) random forests and (iii) a combination of both. The models are fed with both buoy data and ECMWF ocean and atmosphere information, after a previous stage of dimensionality reduction using extended EOF.
2. Evaluate the models' performance according to the standard indicators as customarily used in the literature. The results yielded for each buoy by the different statistical techniques will be compared among them and also with two additional references: (i) the forecasts yielded by the physics-based WAM model at the nearest gridpoint from the buoy and (ii) the persistence of flux values. As a result, a ranking of performances will be obtained for each buoy and forecasting horizon.

In the following section, the data used and the methodology followed will be explained in detail. Results and conclusions will be presented in the final sections.

2. Data and methodology

2.1. Data

This study has been carried out on the Bay of Biscay (Fig. 1). The average values of the wave energy flux obtained from the WAM analysis in the 1999–2012 period range from 15 to 50 kW/m with values increasing from east to west (Fig. 2). These values are similar to those reported in the literature (Iglesias et al., 2009; Iglesias and Carballo, 2010a, 2010b, 2010c; Castro et al., 2014; Gonçalves et al., 2014).

The data are derived from two sources corresponding to the 1999–2012 period:

1) Buoy records
   1.1. Hourly directional wave data from three directional, near-shore buoys (Tables 1 and 2, Fig. 3) operated by the Spanish Port Authority http://www.puertos.es/oceanografia_y_meteorologia/redes_de_medida/index.html
   1.2. Hourly non-directional or scalar wave data from two open-sea buoys (Tables 1 and 2, Fig. 3) operated by the UK MetOffice http://www.metoffice.gov.uk/weather/marine/observations/.
2) Retrospective simulations of the ECMWF atmospheric and wave models as follows:
   2.1 ECMWF (www.ecmwf.int) ERA-Interim atmospheric reanalysis (Dee et al., 2011) data in analysis mode. The selected variables have been mean sea level pressure (MSL), zonal (U10) and meridional (V10) components of the surface

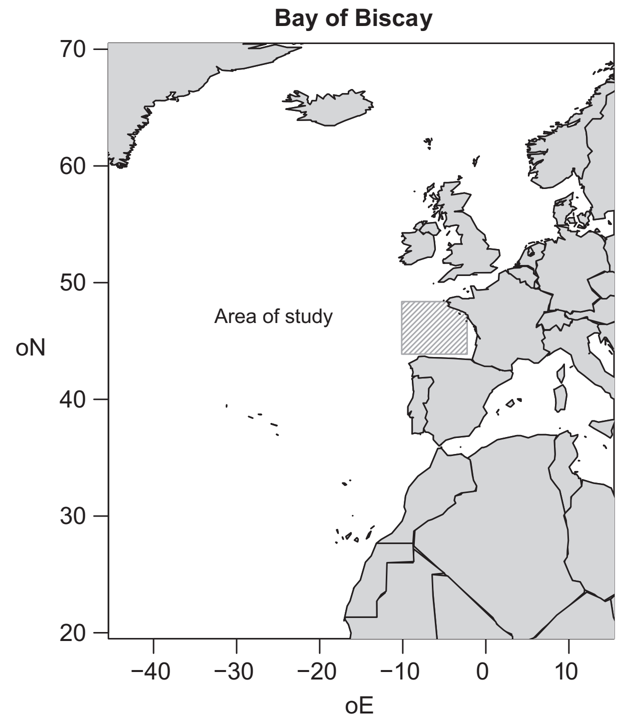

Fig. 1. Area of study.

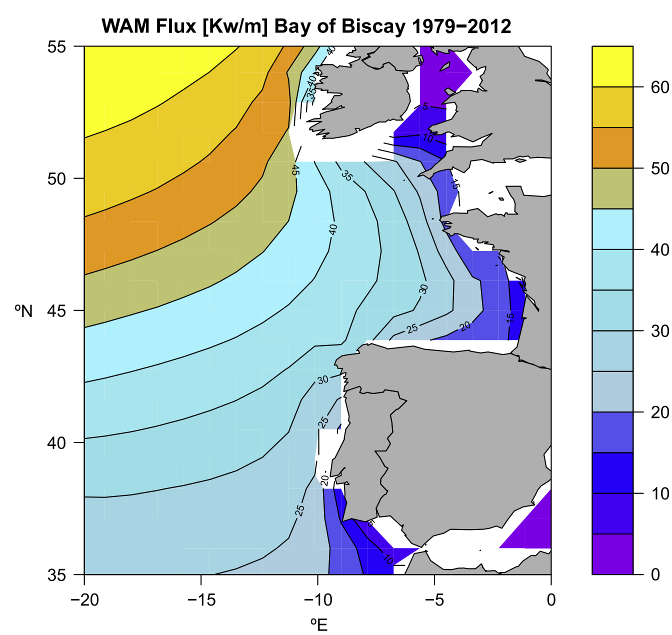

Fig. 2. Module of the wave energy flux [kW/m] in the Bay of Biscay.

Table 1
Number of cases used in this study for training and testing and average wave energy flux values (kW/m).

Buoy #	Name	1999–2005 #Cases for training	2006–2012 #Cases for testing	Flux (kW/m)
1	Villarino Sisargas	7093	8894	24.1
2	Estaca de Bares	7030	6957	22
3	Cabo Peñas	7029	8802	15.5
4	Gascogne	9601	8555	30.7
5	Brittany	8904	8790	40.50