Peer-Reviewed Technical Communication

Wave Energy Forecasting at Three Coastal Buoys in the Bay of Biscay

Gabriel Ibarra-Berastegi, Jon Sáenz, Ganix Esnaola, Agustin Ezcurra, Alain Ulazia, Naiara Rojo, and Gorka Gallastegui

***Abstract*—In 2008, the first commercial wave farm came online in Portugal. As with other types of renewable energy, the electricity obtained from waves has the drawback of intermittency. Knowing a few hours ahead how much energy waves will hold can contribute to a better management of the electricity grid. In this work, three types of statistical models have been used to create up to 24-h forecasts of the zonal and meridional components of wave energy flux (WEF) levels at three directional buoys located off the coast in the Bay of Biscay. Each model's performance has been compared at a 95% confidence level with the simplest prediction (persistence of levels), along with the forecasts provided by the physics-based WAve Modeling (WAM) wave model at the nearest grid point. The results indicate that for forecasting horizons between 3 and roughly 16 h ahead, the statistical models built on random forests (RFs) outperform the rest, including WAM and persistence.**

***Index Terms*—Applied physics, Bay of Biscay, fluid mechanics, forecasting, random forests (RFs), wave energy flux (WEF).**

I. Introduction

THE LAST decade has seen increasing interest in ocean waves as a renewable source of energy. On the one hand, there is exponential growth in the number of prototypes being tested at the numerous facilities recently built for that purpose in many countries. However, in 15–20 years' time only a reduced number of them will successfully combine the necessary efficiency and economic feasibility required to harvest energy from waves at industrial level.

Whatever the final outcome of this current competition among different designs, all of them will have to interact with ocean waves, and so the problem of intermittency in electricity production will also have to be tackled. An accurate understanding of current and forthcoming wave energy levels may therefore help to address this problem by developing real-time grid management strategies [1].

Wave energy is usually expressed in terms of the wave energy flux (WEF), a function of the significant wave height (Hws) and the mean wave period (Tz), being stated in kilowatts per meter of crest length (kW/m)

$$\text{WEF[W/m]} = \rho g^2 (64\Pi)^{-1}\, Hws^2 Tz \sim 0.491\, Hws^2 Tz \quad \text{[kW/m]}. \tag{1}$$

WEF is a vectorial magnitude, so a complete prediction of this variable involves forecasting not only its magnitude, but also its zonal (WEFu) and meridional (WEFv) components. Since WEF is never measured directly, it is usually derived by combining the measurements taken of Hws, Tz, and mean wave direction. WEF is affected by the bathymetry of the area, and if the depth is more than half the wavelength, it is considered that the deep-water hypothesis has been met, and bed shoaling, partial reflection, and diffraction effects can be ignored [2]. For a value of $g = 9.81$ m/s^2, assuming a seawater density of $\rho \sim 1025$ kg/m^3 and according to the deep-water hypothesis, Hws and Tz combine according to (1) to yield the magnitude of WEF [2]. Physics-based models solve the energy balance equation for different frequencies within the directional wave spectrum. Despite not being specifically designed to yield WEF forecasts, WAve Modeling (WAM), WAVEWATCH III, or similar physics-based models such as SWAN [3]–[6] can be used to obtain ocean wave energy predictions. These models are fed with (near)-real-time data from a wide variety of oceanic and atmospheric surveillance sources worldwide. These physics-based models assimilate observations, and then solve the equations of the highly nonlinear mechanisms involved, as generally described by the laws of physics and fluid mechanics-dynamics. The output is a set of geographical points regularly projected onto a grid that covers the area of interest.

Another approach is to learn from the past to forecast the future. Under this approach, the problem of forecasting is treated as a "black box" in which a statistically based transfer function is fitted onto historical records relating current and

Manuscript received May 28, 2015; revised July 10, 2015 and December 12, 2015; accepted February 08, 2016. Date of publication April 20, 2016; date of current version October 11, 2016. This work was supported by the University of the Basque Country and the Spanish Ministry of the Economy and Competitiveness (MINECO) under the following Contracts: Basque Government ELKARTEK 2015 call, the Esnaola UPV/EHU 2013 Postdoctoral Contract, UPV/EHU (GIU14/03), UPV/EHU (UFI11/55), and the MINECO Contract (CGL2013-45198-C2-1-R).

Associate Editor: K. von Ellenrieder.

G. Ibarra-Berastegi is with the NE & Fluid Mechanics Department, University of the Basque Country (UPV/EHU), Bizkaia 48013, Spain (e-mail: gabriel.ibarra@ehu.es).

J. Sáenz is with the Applied Physics II Department, University of the Basque Country (UPV/EHU), Leioa 48940, Spain (e-mail: jon.saenz@ehu.es).

G. Esnaola is with the NE & Fluid Mechanics Department, University of the Basque Country (UPV/EHU), San Sebastian-Donostia 20010, Spain (e-mail: ganix.esnaola@ehu.eus).

A. Ezcurra is with the Applied Physics II Department, University of the Basque Country (UPV/EHU), Vitoria-Gasteiz 01006, Spain (e-mail: agustin.ezcurra@ehu.eus).

A. Ulazia is with the NE & Fluid Mechanics Department, University of the Basque Country (UPV/EHU), Eibar 20600, Spain (e-mail: alain.ulazia@ehu.eus).

N. Rojo and G. Gallastegui are with the Environmental and Chemical Engineering Department, University of the Basque Country (UPV/EHU), Vitoria-Gasteiz 01006, Spain (e-mail: naiara.rojo@ehu.eus; gorkajavier.gallastegui@ehu.eus).

Digital Object Identifier 10.1109/JOE.2016.2529400

future values of WEF at a given location [7]–[10]. The methods most widely used so far have been time series and different types of neural networks [10], with the target variable being the WEF magnitude.

In this work, three different mathematical methods have been used accordingly: analogs, a machine learning technique [random forests (RFs)], and a combination of the two. RF is a development of the classical regression trees [11] in which a large number of trees (typically > 1000) are replicated after each one of them has been perturbed by presenting a sample of cases and inputs randomly selected from among the total number available. This group of trees constitutes a forest. The outputs obtained by each one of these perturbed trees are averaged to yield the RF's overall output. A very interesting aspect of this technique is that it is always free from overfitting [11]–[13]. An in-depth description of the mathematical aspects of RF can be found in the literature [12]–[15].

Previous studies on the matter have focused on the WEF magnitude [16], but its vectorial nature means that a complete prediction of wave energy involves forecasting its zonal and meridional components. Therefore, this study aims to fill this gap, and forecasts have been obtained separately for the zonal (WEFu) and meridional components (WEFv). WEFu and WEFv have been derived from (1) by incorporating the information held in the mean wave direction (MDIR)—another variable measured in vectorial buoys—and by projecting the vector's magnitude onto a parallel (zonal) and a meridian (meridional). WEFu zonal projections have been considered positive/negative for incoming flows from west/east, while positive/negative WEFv meridional values correspond to flows from south/north. The area of study is the Bay of Biscay, where the three buoys are located (Figs. 1 and 2).

II. Data and Methods

A. Data

This study has been carried out at three directional buoys located near the northwest coast of the Spanish region of Galicia on the Bay of Biscay (Figs. 1 and 2). Accordingly, 14 years of hourly data corresponding to the 1999–2012 period have been used. On average, the buoys were missing values in about 20% of the cases. The original database was divided into two sets: hourly cases for the 1999–2005 period were used for training the statistical models, and the 2006–2012 period was used to test the models and draw conclusions on their performance. This study combines several databases from different sources:

1) hourly data from three directional buoys located off the Spanish coast (Fig. 2);
2) European Centre for Medium-Range Weather Forecast (ECMWF) ERA-Interim meteorological model (every 6 h);
3) ECMWF WAM model in analysis mode (every 6 h);
4) ECMWF WAM model in prognostic mode (every 12 h).

An initial preprocessing stage was needed for each buoy to establish a single coherent structure following the same timeline for the data from all the above databases. The total number of hourly cases present for the different buoys ranges between approximately 7000 and 9000. The prevailing direction of the WEF at the three buoys is northwest (Fig. 3), with average values ranging from 15 to 24 kw/m for the 1999–2012 period.

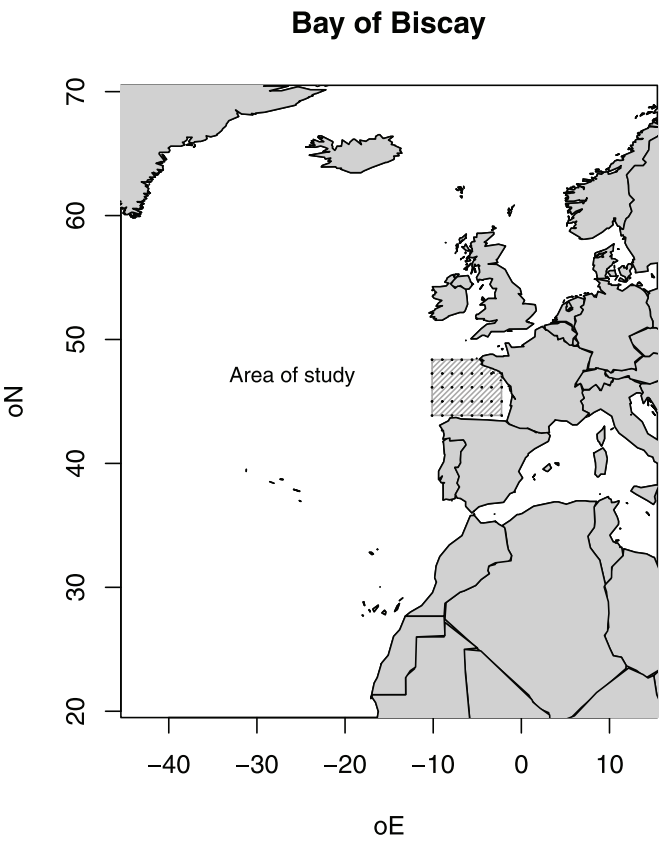

Fig. 1. Area of study. Bay of Biscay.

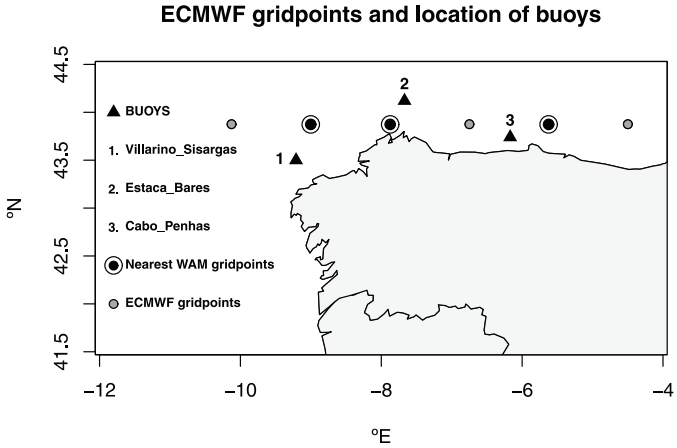

Fig. 2. Directional buoys used for this study and nearest ECMWF grid points.

B. Methodology

1) Extended EOF: An extended empirical orthogonal function (ExtEOF) is a development of the classical EOF, also known in other scientific fields as principal component analysis. ExtEOFs are usually used with variables with a strong time autocorrelation. ExtEOFs are used in the field of wave

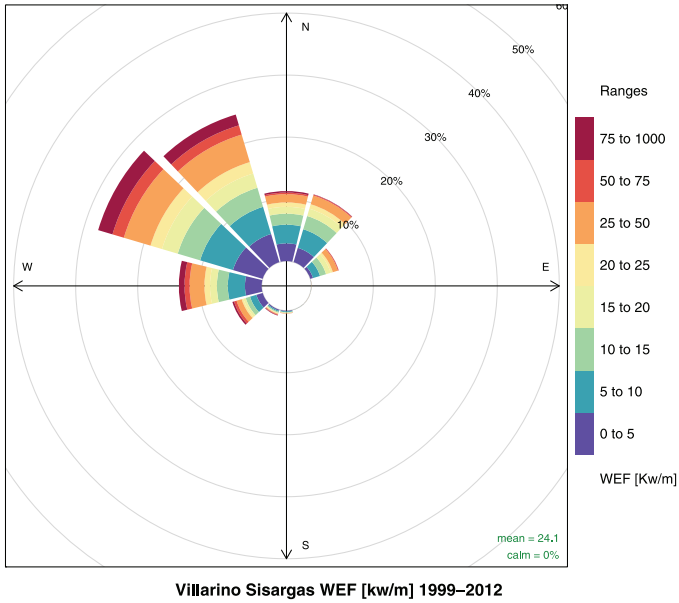

Fig. 3. WEF rose at buoy #1, Villarino Sisargas.

energy forecasting [16], [17], as well as in many other geophysical studies [18]–[25]. To calculate ExtEOFs, the realizations of the same variable several steps back in time are in themselves considered as variables. When applied to a set of grid points throughout an area of interest, ExtEOFs are expected to capture and represent the most relevant features of the variable being analyzed for that time and space window. Using scores from ExtEOFs significantly reduces the number of variables, while retaining most of the variability. Since ExtEOFs are orthonormal, using them as inputs avoids the problem of multicollinearity, to which several techniques, such as linear regression, are very sensitive.

In this work, ExtEOFs have been calculated for the ECMWF atmospheric and oceanic variables corresponding to the 38 grid points in the area of study (Fig. 1) and with three 6-h steps back in time (18 h). The atmospheric and oceanic variables involved were as follows: sea level pressure (MSL), zonal wind speed at 10 m above the sea surface (U10), meridional wind speed (V10), the magnitude of the flux, derived from (1), and its zonal and meridional components. The final number of ExtEOFs retained was 21, and they were selected under the condition of retaining at least 90% of the original variance. This allowed a dramatic reduction in the number of variables used, while still holding most of the information on the atmosphere/sea state in the Bay of Biscay.

2) Building the Models: A statistical modeling approach has been used at the three buoys to build a transfer function to predict the WEFu and WEFv values k hours ahead ($k = 1, \ldots, 24$) with the following two types of inputs:

a) the zonal (WEFu) and meridional (WEFv) components of the WEF observed at the three buoys at time t;
b) the ExtEOFs that hold the highest fractions of variance (ExtEOFs, see below) describing the atmosphere/sea state from t to $t - 18$ hours back in the whole Bay of Biscay (38 grid points, Fig. 1).

The general structure of the models built can be summarized in

$$\begin{aligned}&\text{WEFu\_buoy}[t+k]\\&\quad = F_1(\text{extEOF}_1[t,\ldots,t-18],\ldots,\\&\qquad \text{extEOF}_{21}[t,\ldots,t-18],\text{WEFu\_buoy}[t]) \qquad (2)\\&\text{WEFv\_buoy}[t+k]\\&\quad = F_2(\text{extEOF}_1[t,\ldots,t-18],\ldots,\\&\qquad \text{extEOF}_{21}[t,\ldots,t-18],\text{WEFv\_buoy}[t]) \qquad (3)\end{aligned}$$

where F_1 and F_2 represent the general transfer functions that were fitted, and $k = (1, \ldots, 24)$ is the number of hours ahead: Accordingly, three types of statistical models have been built to forecast zonal and meridional WEF levels at the three buoys analyzed: 1) the analog technique; 2) the analog technique followed by an RF regression stage; and finally, 3) RFs. The reason for using these techniques is that the mechanisms involved in the evolution of WEF a few hours ahead are known to be highly nonlinear, and these algorithms, particularly RFs, are capable of successfully capturing nonlinear relationships between a set of inputs and outputs.

The analog technique is perhaps the simplest one. It is based on the selection of similar historical states of the sea/atmosphere system (analog) from a historical database. After standardization, similar cases to those with the smallest Euclidean distances to the current case are identified. The rationale here is that the evolution observed in those past analogs can now be expected to be similar to the current case. This model will be denoted here as "analogs." A further development is to add a second stage of regression based on RF, in which a transfer function connecting WEF values observed from t to $t - 18$ hours and forthcoming WEF values at time $t + k$, $k = (1, 2, \ldots, 24)$ is fitted only on past analogs. This transfer function will now be fed with current observations to yield predictions up to 24 h ahead. This model will be referred to here as "analogs + RF."

The third technique considered was the straightforward use of RFs with all the historical records available, and not only the most similar ones. Here, this approach will be called "RF." The inputs used to feed the models were the ExtEOFs calculated for the whole area (with a maximum lag of 18 h back), corresponding to the sea (ECMWF WAM) and atmospheric variables (ECMWF ERA-Interim), plus observed values from buoys at time t. The three approaches were applied to forecast the zonal and meridional values of the wave energy flux (WEFu and WEFv) at the three buoys up to 24 h ahead. As the number of test cases at each buoy was 8894 (#1), 6957 (#2), and 8802 (#3), respectively, the three approaches involved building 1 183 344 "analogs"-type models plus 1 183 344 "analogs + RF"-type models, and 144 "RF"-type models. This means that in total 2 366 832 models have been built and tested in this study.

3) Evaluation and Intercomparison of Models: For model performance intercomparison purposes, some measure of error is customarily adopted [7]–[10], with the most widely used indicator being the mean absolute log difference times 100.

However, in this study, the zonal and meridional components of the WEF can have negative values, so this indicator was not an option due to its logarithmic nature.

Therefore, the criterion adopted for the model intercomparison of WEF forecasts was the mean absolute error. This is defined as the average value of [Abs(forecast-observed)]. To correctly identify the convergence of forecasting horizons, confidence intervals of this indicator have been calculated to assess differences in errors at a 95% confidence level. The performance of the different models has been compared with a) the simplest model (persistence of levels); and b) the forecasts yielded by the physics-based WAM model at the nearest grid points (Fig. 2). The reason is that any modeling effort should yield better results than both of these. The identification of preferential forecasting horizons for one model or another will be the objective of this intercomparison.

4) Calculations: All the calculations have been carried out within the framework of R [26]. The main packages used were: "randomForest" [27], "FactoMineR" for EOF calculation [28], and "sp," "maps," and "rgdal" for mapping data [29], [30], [31]. In the "analogs" model, the predictions are calculated as the average of the past analogs observed. After some tests at the initial stages of this study, a figure of 50 was identified as the optimum number of similar records used to calculate that average, and the same value was also applied for the "analogs + RF" approach. The basic implementation of RF is available in R [27], but its application both after the "analogs" stage and in standalone mode involved the extensive development of specific scripts. Under these approaches, hourly forecasts were calculated for both the zonal and meridional components of WEF at the three buoys, as well as for forecasting horizons ranging from 1 to 24 h. These predictions were compared with the forecasts available from ECMWF WAM, as well as persistence. Finally, 95% confidence boundaries were calculated for all the statistical performance indicators. It is important to stress that over 2×10^6 models were built, tested, and evaluated.

III. Results and Discussion

A. Results

The main findings of the extensive analysis provided by the intercomparison of this large number of models can be summarized as follows.

1) Models tend to behave similarly at the three locations, so the results will be presented in an aggregated manner (average errors from all buoys), as shown in Fig. 4 for zonal (WEFu) and Fig. 5 for meridional (WEFv) forecasts.
2) Quite the opposite to their behavior when applied to other geophysical variables [19], analogs perform poorly. "Analogs + RF" performs better, and among the statistical models "RF" outperforms all the others.
3) The WAM model's error is higher for the zonal component than for the prediction of the meridional WEF.
4) RF errors, however, tend to be more balanced in all directions.
5) Until 2–3 h ahead, persistence is the best model.
6) In all cases, the WAM model's error remains constant at a 95% confidence level for 12- and 24-h forecasts, while the different statistical models increase their error with the forecasting horizon. This difference in error behavior between statistical and physics-based models has also been reported in the literature [7]–[9] for different environments and locations. Assuming a constant error, and with the aim of clearly identifying the convergence point with the statistical models in Figs. 4 and 5, 95% confidence boundaries of the WAM error have been interpolated/extrapolated.
7) For zonal prediction (Fig. 4), the following preferential windows emerge for each type of model:
   a) from 1- to 2–3-h forecasts, persistence is the best option;
   b) from 3 to 19–20 h, "RF" outperforms all the others, and between 8 and 20 h, "RF" and "analogs + RF" record a similar performance;

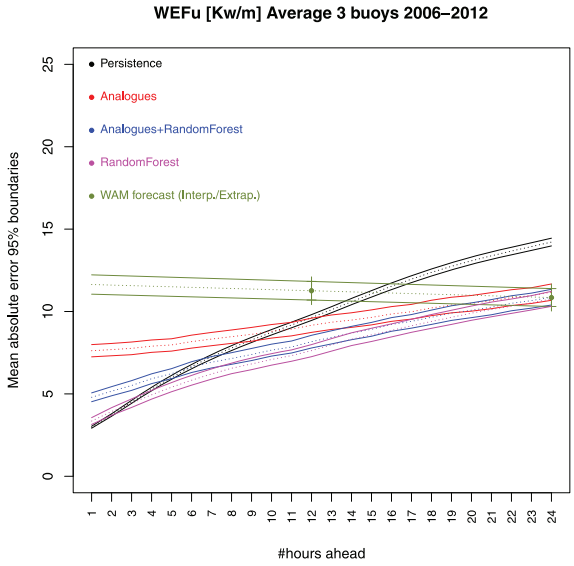

Fig. 4. Mean absolute error for WEFu forecasts up to 24 h ahead. Average for the three buoys.

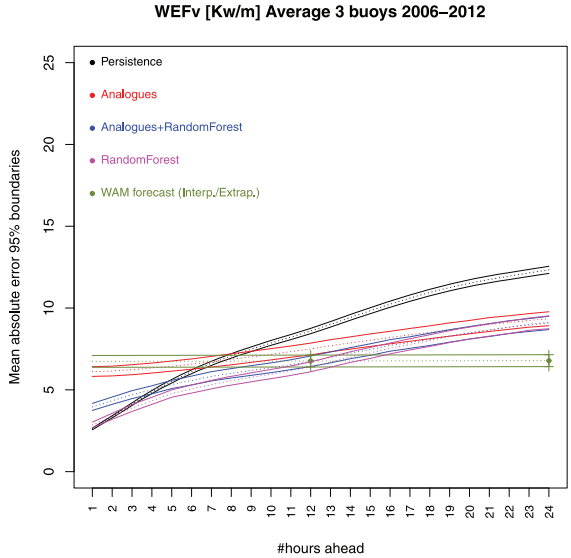

Fig. 5. Mean absolute error for WEFv forecasts up to 24 h ahead. Average for the three buoys.

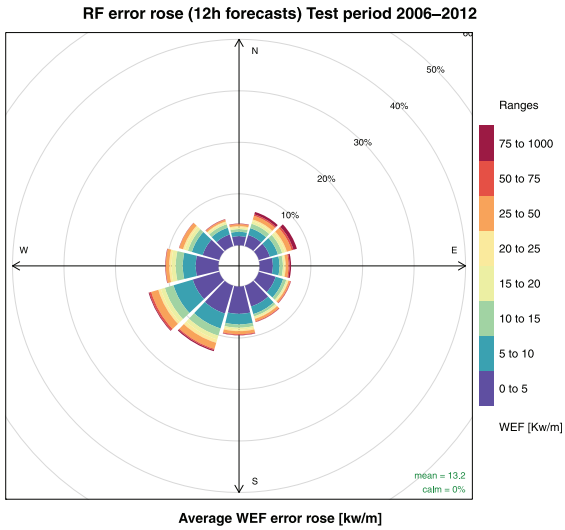

Fig. 6. Error rose for the "RF" model's 12-h forecasts at buoy #1 (Villarino-Sisargas).

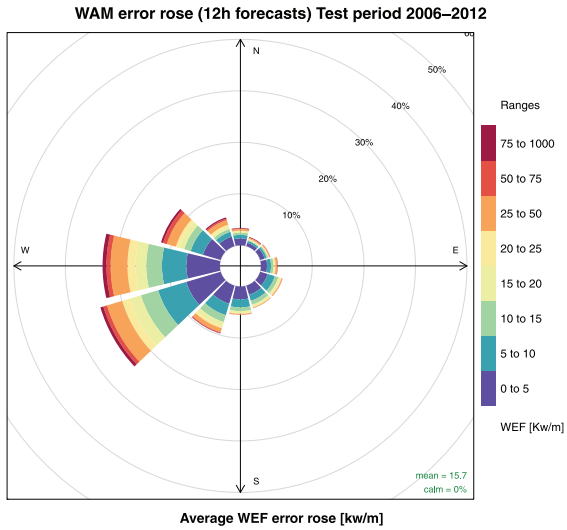

Fig. 7. Error rose for the WAM model's 12-h forecasts at buoy #1 (Villarino-Sisargas).

c) above 20 h, WAM, "analogs," "analogs + RF," and "RF" record a similar performance.

8) For meridional prediction (Fig. 5), the preferential windows are as follows:
   a) again, from 1- to 2–3-h forecasts, persistence is the best option;
   b) from 3 to 11 h, "RF" outperforms all the others;
   c) above 12 h, WAM outperforms all the others.

When WEFu and WEFv are analyzed jointly, the error roses indicate that RF errors (magnitude + direction) tend to be more homogeneously distributed than in the case of WAM, whose error rose stretches westwards. To illustrate this common behavior at the three buoys, error roses corresponding to buoy #1 (Villarino-Sisargas) are shown for "RF" (Fig. 6) and WAM (Fig. 7).

B. Discussion

The literature suggests that a complex bathymetry may render physics-based models less accurate at short horizons, and more specifically in shallow coastal sites, while statistical models are reported to work better at short horizons [7], [8], [9]. However, at least one of the buoys in this study (buoy #1, −1800 m) is not located in shallow waters. In these deep-water cases, the literature indicates that the results seem to depend on local conditions. Along these lines, it is worth mentioning the case of a buoy located in deep waters (Yucatan Basin, 19.874′ N, 85.06′ W), where the statistical models are reported to outperform WAM forecasts up to a convergence point of 9–12 h [9]. An analysis of the Yucatan Basin suggests that the combination of a complex bathymetry with an undersea mountain range and the presence of shorelines at mid-distances in almost all directions can generate effects that are difficult to simulate accurately, and to a certain extent deteriorates the WAM's prognostic capabilities.

With the grid size it uses here, the WAM model makes the forecasts considering the following depths for the [1.125° × 1.125°] pixels corresponding to three nearest grid points: −990 m (buoy #1), −223 m (buoy #2), and −186 m (buoy #3). However, the effective depths at the buoys are as follows: −386 m (buoy #1), −1800 m (buoy #2), and −450 m (buoy #3). Fig. 8 shows that the three buoys are located at places where bathymetry records a steep gradient, while the WAM model considers a flat sea bottom for each pixel of the grid with constant depths.

This means that with the resolution it uses, the WAM is probably unable to successfully simulate the effects associated to the complex bathymetry below the buoys. Additionally, coastal effects may not be captured accurately since significant proportions of the cells are land and not sea. Fig. 8 also suggests that the bathymetry of the area can originate a channeling effect along the east/west direction. This can also explain why the WAM error is almost twice as high for the zonal component than for the meridional one, while the error in the statistical models is roughly the same for both components and is not affected by all these local factors.

IV. Conclusion and Future Outlook

A. Conclusions

In recent years, wave energy farms have been attracting increasing interest. A shortcoming associated to this source of energy is that intermittency might cause grid management problems. For this reason, accurate knowledge of how much energy ocean waves will hold a few hours ahead is a major issue.

This work has presented a statistical approach based on different models. Among them, "RF" performs best. "RF" is a machine learning algorithm that from a historical database can capture, under a black box approach, the major patterns regarding the evolution of WEF in the timescale of hours. "RF" also outperforms readily available persistence and WAM models for different forecasting horizons.

It is important to stress that due to its vectorial nature, an accurate prediction of the WEF vector involves forecasting both components: zonal (WEFu) and meridional (WEFv). As the "RF" model's preferential forecasting horizons are different for the two components (WEFu [3–20 h], WEFv [3–11 h]), a reasonable combined preferential forecasting horizon for WEF

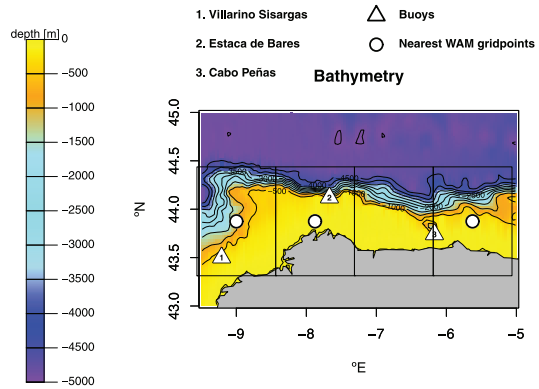

Fig. 8. WAM cells and bathymetry of the area studied.

would be between 3 and 16 h. "RF" also allows identifying the most influential variables involved in the forecasts. For predictions up to 12 h ahead, the most relevant variable is the observed value of WEFu, or WEFv at time t, followed by two groups of variables: the ExtEOFs of WEFu and WEFv calculated for the area, and the ExtEOFs corresponding to MSL, U10, and V10. Beyond 12 h ahead, the same inputs are again the most influential, but the observed value of WEFu or WEFv loses relevance, and its impact is smaller than the two groups of oceanic and atmospheric ExtEOFs.

This indicates that RF models take advantage of the inherent predictability associated to the strong autocorrelation present in the local WEF values observed. The additional prognostic capabilities that "RF" models provide, when compared with persistence, are due to the information that both atmospheric and sea state variables incorporate into the "RF" model. This information is held by the most significant ExtEOFs retained, and describes the major features corresponding to atmospheric and oceanic variability in the Bay of Biscay in a time window ranging from time t to time $t - 18$ hours. The models obtained for this work are useful locally, but the methodology can be readily extrapolated to other environments.

The buoys in this study are located near the coast, between 17 and 36 km offshore. These distances represent the maximum boundaries from the coast that could be expected for any wave farm. A common aspect of ocean wave farms is that they are not located in open-sea areas, but at distances from the coast that are similar to or shorter than those shown in this study. The results shown here can hence be useful for other tentative near-shore locations. It is in these areas where WAM error seems to deteriorate heavily, and statistical models work better. However, this work indicates that neither of the WEF components is affected by all these factors in the same way, which is another good reason to take a vectorial approach for forecasting purposes.

Although a few studies have focused on forecasting WEF magnitudes using statistical techniques, to the best of our knowledge a vectorial approach like the one shown here has never been applied. Likewise, the use of a machine learning technique like "RF" has never been reported for statistical WEFu and/or WEFv forecasting purposes. Nevertheless, the relatively small number of cases reported in the literature suggests that more studies are needed, whereby a comparison with other cases worldwide may contribute to a better understanding of certain effects, such as the roles of bathymetry and distance to the coast.

B. Future Outlook

Further research is currently being carried out along these lines by the authors of this study, albeit for other environments, with the aim being to expand the database of similar cases, where zonal and meridional WEF operational forecasts might be of interest. Furthermore, gaining a better understanding of the factors that affect WAM's performance (complex bathymetry, grid resolution, and coastal effects, among others) is an issue of interest for a clearer delimitation of the forecasting horizons and general operational conditions in which one or other should be used. This will allow a better understanding of the applicability of these techniques for operational WEF forecasting purposes.

Considering the high number of models tested and cases used in this study, the conclusions may be considered sound enough for the three buoys analyzed here. Unfortunately, a database as large as the one used here is not always available, which currently represents another limitation for similar studies.

Acknowledgment

ECMWF ERA-Interim data used in this study have been obtained from the ECMWF-MARS Data Server thanks to agreements with ECMWF and AEMET. The authors would like to thank the Spanish Port Authorities (Puertos del Estado) for kindly providing data for this study.

References

[1] M. Esteban and D. Leary, "Current developments and future prospects of offshore wind and ocean energy," *Appl. Energy*, vol. 90, pp. 128–136, 2012.
[2] B. Multon, *Marine Renewable Energy Handbook*, New York, NY, USA: Wiley, 2012, p. 112
[3] The WAMDI Group "The WAM model—A third generation ocean wave prediction model," *J. Phys. Oceanogr.*, vol. 18, pp. 1775–1810, 1988.
[4] P. A. E. Jansen, "Progress in ocean wave forecast," *J. Comput. Phys.*, vol. 227, pp. 3572–3594, 2007.
[5] J. R. Bidlot, D. J. Holmes, P. A. Wittmann, R. Lalbeharry, and H. S. Chen, "Inter comparison of the performance of operational ocean wave forecast systems with buoy data," *Weather Forecast*, vol. 17, pp. 287–310, 2002.
[6] N. Booij, R. C. Ris, and L. H. Holthuijsen, "A third-generation model for coastal regions. Part 1: Model description and validation," *J. Geophys. Res.*, vol. 104, pp. 7649–7666, 1999.
[7] G. Reikard, "Integrating wave energy into the power grid: Simulation and forecast," *Ocean Eng.*, vol. 73, pp. 168–178, 2013.
[8] G. Reikard, P. Pinson, and J. R. Bidlot, "Forecast ocean wave energy: The ECMWF wave model and time series methods," *Ocean Eng.*, vol. 38, pp. 1089–1099, 2011.
[9] G. Reikard and W. E. Rogers, "Forecast ocean waves: Comparing a physics-based model with statistical models," *Coast. Eng.*, vol. 58, pp. 409–416, 2011.
[10] S. Hadadpour, A. Etemad-Shahidi, and B. Kamranzad, "Wave energy forecasting using artificial neural networks in the Caspian Sea," *Proc. Inst. Civil Eng., Maritime Eng.*, 2014, vol. 167, no. 1, pp. 42–52.
[11] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
[12] U. Grömping, "Variable importance assessment in regression: Linear regression versus random forest," *Amer. Stat.*, vol. 63, pp. 308–319, 2009.
[13] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning, Data Mining, Inference, and Prediction*, New York, NY, USA: Springer-Verlag, 2001, p. 48.

[14] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News.*, vol. 2, no. 3, pp. 18–22, 2002.
[15] D. S. Siroky, "Navigating random forests and related advances in algorithmic model," *Stat. Surv.*, vol. 3, pp. 147–163, 2009.
[16] G. Ibarra-Berastegi, J. Saenz, G. Esnaola, A. Ezcurra, and A. Ulazia, "Short-term forecasting of the wave energy flux: Analogues, random forests, and physics-based models," *Ocean Eng.*, vol. 104, pp. 530–539, 2015.
[17] A. D. Rao, M. Sinha, and S. Basu, "Bay of Bengal wave forecast based on genetic algorithm: A comparison of univariate and multivariate approaches," *Appl. Math. Model.*, vol. 37, pp. 4232–4244, 2013.
[18] A. Hannachi, I. T. Jolliffe, and D. B. Stephenson, "Empirical orthogonal functions and related techniques in atmospheric science: A review," *Int. J. Climatol.*, vol. 27, pp. 1119–1152, 2007.
[19] G. Ibarra-Berastegi *et al.*, "Downscaling of surface moisture flux and precipitation in the Ebro Valley (Spain) using analogues and analogues followed by random forests and multiple linear regression," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 6, pp. 1895–1907, 2011.
[20] J.-M. Chen and P. A. Harr, "Interpretation of extended empirical orthogonal function (EEOF) analysis," *Monthly Weather Rev.*, vol. 121, pp. 2631–2636, 1993.
[21] K. H. Seo and Y. Xue, "MJO-related oceanic kelvin waves and the ENSO cycle: A study with the NCEP global ocean data assimilation system," *Geophys. Res. Lett.*, vol. 32, 2005, L07712.
[22] F. T. Tangang, B. Y. Tang, A. H. Monahan, and W. W. Hsieh, "Forecast ENSO events: A neural network extended EOF approach," *J. Climatol.*, vol. 11, no. 1, pp. 29–41, 1998.
[23] B. C. Weare and J. C. Nasstrom, "Examples of extended empirical orthogonal function analyses," *Monthly Weather Rev.*, vol. 110, pp. 481–485, 1982.
[24] J. Peters, B. De Baets, N. E. C. Verhoest, R. Samson, S. Degroeve, P. De Becker, and W. Huybrechts, "Random forests as a tool for ecohydrological distribution modeling," *Ecol. Model.*, pp. 304–318, 2007.
[25] K. Fraedrich, J. L. McBride, W. M. Frank, and R. Wang, "Extended EOF analysis of tropical disturbances: TOGA COARE," *J. Atmos. Sci.*, vol. 54, pp. 2363–2372, 1997.
[26] R Development Core Team "R: A language and environment for statistical computing," R Found. Stat. Comput., Vienna, Austria, 2012, http://www.R-project.org/ ISBN 3-900051-07-0.
[27] F. Husson, J. Josse, S. Le, and J. Mazet, "FactoMineR: Multivariate exploratory data analysis and data mining with R," R package version 1.25, 2013, http://CRAN.R-project.org/package=FactoMineR
[28] R. S. Bivand, E. Pebesma, and V. Gomez-Rubio, *Applied Spatial Data Analysis with R*, 2nd ed., New York, NY, USA: Springer-Verlag, 2013, p. 39 [Online]. Available: http://www.asdar-book.org/
[29] R. Bivand, T. Keitt, and B. Rowlingson, "rgdal: Bindings for the Geospatial Data Abstraction Library," R package version 0.8-11, 2013, [Online]. Available: http://CRAN.R-project.org/package=rgdal
[30] E. J. Pebesma and R. S. Bivand, "Classes and methods for spatial data in R," *R News*, vol. 5, no. 2, 2005, [Online]. Available: http://cran.r-project.org/doc/Rnews/
[31] G. Ibarra-Berastegi, J. Saénz, G. Esnaola, A. Ezcurra, and A. Ulazia, "Short-term forecasting of zonal and meridional wave energy flux in the Bay of Biscay using random forests," in *Proc. IEEE OCEANS Conf.*, 2015, DOI: 10.1109/OCEANS-Genova.2015.7271404
[32] Y. Fukutomi and T. Yasunari, "Structure and characteristics of submonthly-scale waves along the Indian Ocean ITCZ," *Clim. Dyn.*, vol. 40, no. 7–8, pp. 1819–1839, 2013.

Gabriel Ibarra-Berastegi received the Ph.D. degree in engineering.

Currently, he is an Associate Professor at the Faculty of Engineering, University of the Basque Country (UPV/EHU), Bilbao, Spain. He has a long experience in geophysical and environmental fluids. His current research focus is ocean wave energy. He coordinates the activities of "EOLO" research group.

Jon Sáenz received the B.Sc. degree in physics of the atmosphere from Complutense University, Madrid, Spain, in 1986 and the Ph.D. degree in physics from the University of the Basque Country (UPV/EHU), Bilbao, Spain.

Currently, he teaches physics at the University of the Basque Country. His main research interests are in the field of atmospheric physics, either using mesoscale atmospheric models or analyzing climate data, including several processes involved in the atmosphere–ocean coupling. Climate data derived from observations (surface, satellite), global or regional models, and reanalyses are his main source of research material.

Ganix Esnaola received the Ph.D. degree in physics.

He is an Assistant Professor at the University of the Basque Country (UPV/EHU), San Sebastian-Donostia, Spain. He specializes in fluid mechanics and geophysical fluid dynamics. His research activities include air–sea interaction problems, ocean modeling, reconstruction of missing satellite data, and data-assimilation techniques.

Agustin Ezcurra received the Ph.D. degree in physics.

He works as a teacher and researcher at the University of the Basque Country, (UPV/EHU), Bilboa, Spain. He has a long experience in geophysical fluids and environmental issues.

Dr. Alain Ulazia graduated in astrophysics from the University of La Laguna, San Cristóbal de La Laguna, Spain, in 1997 and received the Ph.D. degree in logic and philosophy of science from the University of the Basque Country (UPV/EHU), Bilbao, Spain.

He currently works as a teacher and researcher at the University of the Basque Country.

Naiara Rojo received the Ph.D. degree in engineering.

She is an Assistant Professor at the University of the Basque Country (UPV/EHU), Vitoria-Gasteiz, Spain. She specializes in the treatment of polluted gaseous streams by biofiltration.

Gorka Gallastegui received the Ph.D. degree in engineering.

He is an Assistant Professor at the University of the Basque Country (UPV/EHU), Vitoria-Gasteiz, Spain. He specializes in environmental engineering. His main research activity is related to air-pollution control and biotreatment of contaminated waste gas streams.