Journal of
Marine Science and Engineering

MDPI

Article

Estimation of Coastal Currents Using a Soft Computing Method: A Case Study in Galway Bay, Ireland

Lei Ren [1,2], Jianming Miao [1,*], Yulong Li [1,*], Xiangxin Luo [1,*], Junxue Li [3,*] and Michael Hartnett [4]

1 School of Marine Engineering and Technology, Sun Yat-sen University, Guangzhou 510275, China; renleicomeon@163.com
2 Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China
3 College of Science and Technology, Hebei Agricultural University, Cangzhou 061100, China
4 College of Engineering and Informatics, National University of Ireland Galway, H91 TK33 Galway, Ireland; michael.hartnett@nuigalway.ie
* Correspondence: miaojm@mail.sysu.edu.cn (J.M.); liylong7@mail.sysu.edu.cn (Y.L.); luoxx6@mail.sysu.edu.cn (X.L.); lijunxue1125@163.com (J.L.)

Received: 1 April 2019; Accepted: 14 May 2019; Published: 20 May 2019

check for updates

Abstract: In order to obtain forward states of coastal currents, numerical models are a commonly used approach. However, the accurate definition of initial conditions, boundary conditions and other model parameters are challenging. In this paper, a novel application of a soft computing approach, random forests (RF), was adopted to estimate surface currents for three analysis points in Galway Bay, Ireland. Outputs from a numerical model and observations from a high frequency radar system were used as inputs to develop soft computing models. The input variable structure of soft computing models was examined in detail through sensitivity experiments. High correlation of surface currents between predictions from RF models and radar data indicated that the RF algorithm is a most promising means of generating satisfactory surface currents over a long prediction period. Furthermore, training dataset lengths were examined to investigate influences on prediction accuracy. The largest improvement for zonal and meridional surface velocity components over a 59-h forecasting period was 14% and 37% of root mean square error (RMSE) values separately. Results indicate that the combination of RF models with a numerical model can significantly improve forecasting accuracy for surface currents, especially for the meridional surface velocity component.

Keywords: coastal surface currents; soft computing; radar; sensitivity experiments; numerical model

1. Introduction

Interaction between atmospheric forces such as wind, river inflows and tide drive the movement of coastal water bodies. The horizontal phenomenon related to this movement is surface current. Good understanding of coastal surface currents is of great importance for many coastal economic operations, in particular marine renewable energy exploration/exploitation. In general, numerical models and observation platforms are powerful and conventional tools to study patterns of coastal surface currents, providing useful information [1]. However, model approximations and simplifications in defining initial and boundary conditions, model grid structure and other factors lead to a reduction in model performance and accuracy. Model prediction accuracy over the long term is hard to guarantee. Oceanic observation tools such as satellites and radars based on electro communication technologies are useful means to record surface flow information over a large coastal domain [2]. However, these observation tools cannot provide forecasts.

In order to accurately forecast states of parameters of interest by making the best use of available observations, soft computing approaches such as artificial neural networks (ANNs), support vector machines (SVMs) and random forests (RF) have been widely used as an alternative in a number of fields [3–5]. Soft computing models are computationally more efficient since they construct the relationship between input variables and output variables directly. In soft computing models, predictions are generated from a trained soft computing model, which is based on relationships between inputs variables and output variables. According to some case studies such as the prediction of renewable energy sources [6], forecasting urban water demand [7] and ecological prediction [8], RF is a powerful and efficient soft computing approach used to establish internal relationships among datasets and ultimately provide forecasting states with high accuracy. Since the RF algorithm is capable of dealing with large numbers of predictor variables even in the presence of complex interaction, it has been successfully applied in genetics, clinical medicine and bioinformatics within the past decades [9]. Currently, it has become increasingly a more popular approach in oceanographic engineering. For example, Ibarra-Berastegi et al. [6] applied the RF algorithm to produce short-term forecasting of the wave energy flux in comparison to using a physical model—wave model (WAM) and analogues. Their comparative results indicate that the RF model outperforms other statistical techniques when compared with the WAM, and a window of forecasting horizons emerges in which the use of RF outperforms any other solutions. Lahouar and Slama [10] applied the RF algorithm to predict electrical load demand of the day ahead, finding that RF coupled with expert selection was able to capture complex load behavior and solve some special cases that are specific to culture, high temperature, religious events and moving holidays owing to appropriate input variable structure. Catani et al. [11] used a RF algorithm to estimate landslide susceptibility and found that the dimension of parameter space, the mapping scale and the training process strongly influenced the classification accuracy and the prediction process. They also showed that the choice of the training set was of key importance for obtaining accurate results. Mahjoobi and Mosabbeb [12] used present time and previous time data as input variables to predict significant wave heights using a support vector machine. Moreover, Balas et al. [13] used historical data to predict missing wave parameters using a soft computing approach. Soft computing models are commonly established based only on observational data. In this research, a three-dimensional coastal hydrodynamic model i.e., environment fluid dynamics code (EFDC) of Galway Bay has been set up by Ren et al. [14]. This is the first time that the RF algorithm has been used to predict surface currents and the first time RF has been used in combination with high frequency coastal radar data. The predictive model described herein was developed in order to further improve forecasting capability of surface currents in this area and to provide continuous and operational predictions in the future. Outputs of surface currents $u(t)$ and $u(t-1)$ from the EFDC model were chosen as main input variables to establish RF models in this research. Reasons for including previous observations lies in that the variation of surface current patterns is continuous in time at each physical location, development of surface currents at the present time step is related to previous states. Galway Bay is located on the west coast of Ireland, tidal water elevation and winds are known to be the main driving forces that generate surface currents in the bay. Tidal water elevation, wind speeds and wind directions were examined as input variables to examine the input variable structure of the RF model developed.

In this research, the soft computing approach RF was adopted to predict surface currents at the 1/4-scale marine energy test site Galway Bay area using surface current data measured by a high frequency coastal ocean dynamic application radar (CODAR) system and outputs from a numerical model. Surface current speeds at the present time step are denoted by $u(t)$, and surface currents i hour(s) before the present time are denoted as $u(t-i)$.

The main objective of this research was to use model data and observations from radars to run a soft computing model for coastal surface current prediction. The layout of this paper is as follows: Section 2 describes the methodologies used and outlines the study domain; this section summarizes the numerical model, the high frequency radar (HFR) system, the random forests algorithm and criteria assessment skills. Results are presented in Section 3, followed by a discussion in Section 4. Research conclusions are presented in Section 5.

2. Methodologies

2.1. Study Domain

Galway Bay is located on the west coast of Ireland with an entrance opening onto the northeast Atlantic Ocean. The bay is semi-enclosed as it is partially shielded from the harsh Atlantic conditions by three small islands. Hydrodynamics within the bay are mainly affected by oceanic flows to the bay from the adjacent continental shelf and wind-driven currents. Figure 1 shows the extents of the model domain and the measurement deployment locations. The water depth of the inner bay covered by the radar system ranged from 10–40 m. Meteorological conditions in Galway Bay are mainly influenced by Atlantic weather systems [15]. Effects of wind forcing on surface flow patterns has been studies in details by Ren et al. [16].

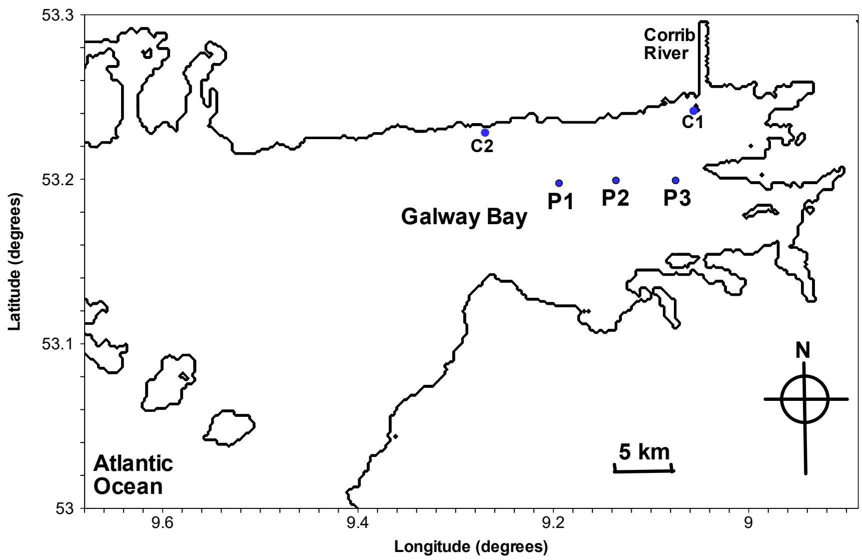

Figure 1. Galway Bay domain; (C1 and C2 indicate radar stations; P1–P3 indicate three analysis points).

2.2. Numerical Model

EFDC is a finite difference based model which can simulate three-dimensional flows and transport processes in surface water systems, rivers, lakes, estuaries, wetlands and coastal areas [17]. The model structure includes four major modules: (1) A hydrodynamic module; (2) a water quality module; (3) a sediment transport module; and (4) a toxics module. The EFDC model solves the three dimensional, vertically hydrostatic, free surface, turbulent-averaged equations of motion for a variable density fluid and can simulate both barotropic and baroclinic circulation. The model uses a stretched (sigma) vertical coordinate system and a Cartesian or curvilinear orthogonal horizontal coordinate system [18–20]. The EFDC model has been successfully applied in a number of areas [21–23].

A three-dimensional computational model of Galway Bay had been developed previously. A bathymetric model of the Galway Bay was developed using data from the Integrated Mapping for the Sustainable Development of Ireland's Marine Resource (INFOMAR) program. The deepest area is approximately 70 m near the south west corner of the open boundaries. The mean water depth in the domain is 30.7 m. Tidal elevation for boundary conditions obtained from the Oregon Tide Prediction Software (OTPS), wind data obtained from the Europe Centre for Mid-Range Weather Forecasts (ECMWF) and River Corrib flow data provided by the Irish Open Public Work (OPW) were used to drive the models. Detailed EFDC modeling setup for Galway Bay such as wind forcing and vertical layer structure is described in Ren et al. [14], O'Donncha et al. [24] and Ren et al. [25].

2.3. Radar Data

There are two commonly used types of HFR observation systems: Beam-forming (e.g., CODAR and ocean state measuring and analyzing radar (OSMAR) system) and direction-finding (e.g., WERA HFR system) [26–28]. The former HFR system uses a distributed array of elements to electronically scan ocean surface with a relatively narrow beam; while the latter HFR system uses a small-aperture antenna to from a quite broad beam for obtaining sea echoes [29]. A land-based CODAR HFR observation system consisting of two radars has been deployed in the Galway Bay area since 2011, as shown in Figure 1. Each radar station is capable of monitoring radial surface currents toward or away from the station. The raw ocean surface information is obtained from radar signals, which are scattered in 360°, measurement information returns to the radar receiver when the radar signal scatters off a wave that is exactly half of the transmitted signal wavelength [30–32]. The information in the radio-wave backscatter exploited from the ocean surface is used to infer movement of the near surface layer, including parameters of surface currents, waves and winds [33–36]. Surface flow fields in the area covered by both radars were combined based on the radial current maps for at least two stations [37,38]. Measurements obtained from the HFR systems are available in near real time. Temporal and spatial resolution of surface flow fields are sixty minutes and 300 m, respectively. The operating frequency is 25 MHz for both radars. The measured surface currents at the three analysis points (P1, P2 and P3) by the HFR system were taken as target fields to train and assess the RF models. Validation of radar surface currents have been undertaken by Ren et al. [14], O'Donncha et al. [24] and Ren et al. [25], and reasonably good agreement between radar data and ADCP observations provided confidence in using radar data for other applications such as establishing RF models.

2.4. Random Forests

The RF algorithm proposed by Breiman [39] is an improved version of the decision tree learning approach, which integrates the prediction of multiple uncorrelated decision trees [40]. The RF algorithm is based on bagging that builds a large collection of de-correlated trees, and then averages them [41,42]. The RF algorithm has been implemented in a number of fields due to its robustness and satisfactory predictions with high accuracy.

The RF methodology does not only produce one decision tree, but produces a variety of trees using subsets of a training dataset. The RF algorithm can be expressed in the following steps [43].

(1) Select n_{RF} multiple bootstrap samples from the dataset;
(2) Develop an unpruned regression tree for each of the bootstrap samples by randomly sampling m_{try} of the predictors and select the best split among those variables;
(3) Forecast new data by aggregating the predictions of the n_{RF} trees; for a regression analysis, the averaged value is taken as output.

For the classification case, a random forest obtains a class vote from every tree, and then classifies them using a majority vote; for the regression case, predictions from every tree are averaged as final outputs [41]. The RF approach is to build n_{RF} multiple decision trees and merge them together to generate a prediction with higher accuracy. In each decision tree, n_{RF} features were selected and used.

In this research, the RF algorithm was applied to establish regression models, input variables including historical outputs of surface currents from both the EFDC model and radar observations were taken as the target field was viewed as a full dataset.

The n_{RF} trees were separately developed. Each tree includes m samples, which were randomly selected. In the selection process for all trees, the number of predictors m_{try} was the same. The recommend values of m_{try} for the classification and regression were given as [11]:

Classification:

$$\mathrm{m}_{try} = \sqrt{\mathrm{p}} \tag{1}$$

Regression:

$$m_{try} = \frac{p}{3} \tag{2}$$

where, p is the number of predictors.

Each tree was developed by training the selected samples. R packages named "*randomForest*" from the Comprehensive Archive R Network (CRAN) was used in this research [8,43]. An estimate of the error rate can be obtained in RF approach based on the training data: Firstly, out of bag (OOB) data not used in the training were used for prediction using the tree grown with the bootstrap sample; secondly, the OOB predictions were aggregated and the error rate computed which was called the OOB estimate of error rate [43,44].

Advantages of the RF algorithm are: (a) RF is capable of dealing with high-dimensional datasets, twice randomly sampling reduces the dimension of the datasets; (b) bootstrapping sampling process results in around one third data are out of bag (OOB) samples, OOB samples are used to compute the unbiased error rate and variable importance; (c) the prediction only depends on one user-selected parameter, the number of predictors are chosen randomly at each node [8].

In this research, outputs of surface currents from EFDC model, tidal water elevation, wind speed and wind direction were considered as input variables to establish RF models. River discharge was not considered here because it had weak impacts on the three analysis points located far from the estuary. Selection of input variable structure was presented in detail in Section 3.1.

2.5. Criteria Skills

In order to quantify performances of the RF models developed in a quantitative way, multiple statistics including correlation (R), bias, root mean square error (RMSE) and scatter index (SI) were computed using the following Equations (3)–(6) [45]. The correlation coefficient is an indicator of the linear relationship between two datasets; bias indicates the trend of a measurement process to systematically over- or underestimate the magnitude of a predicted parameter; RMSE is an error index presenting an overall error distribution; SI gives the percentage of expected error for the parameter.

$$bias = \overline{y} - \overline{x} \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - x_i)}{n}} \tag{4}$$

$$\mathrm{SI} = \frac{RMSE}{\overline{x}} \tag{5}$$

$$\mathrm{R} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}} \tag{6}$$

where,

x_i is an observed value;
y_i is a predicted value;
n is the number of observations;
$\overline{x}$ and $\overline{y}$ is the mean of x and y, respectively.

Statistics were computed for each RF model at every analysis point respectively.

3. Results

3.1. RF Models

A full dataset consisting of 305 hourly values starting from Julian Day 220 01:00, 2013 was used in this research. Based on division principles proposed by Aydogan et al. [46], the full dataset was subdivided as follows: (i) A training dataset accounting for 60% of the total; (ii) a test dataset accounting for 20% of the total and (iii) a prediction dataset accounting for 20% of the total. Surface currents from the EFDC model were outputted at the same time steps as the HFR observations. Tidal elevations and wind stresses were used as boundary conditions for the numerical model simulations; in order to further explore the relationship between surface flows and the dominant forcing factors tidal elevations, wind speeds and wind directions were adopted as input variables in establishing the RF models as well as surface current outputs from the numerical model. RF models for predicting surface current components at each of the three analysis points were established separately. The functional equation for RF models is expressed as:

$$\mathrm{URF}(t) = f(UEF(t), UEF(t-1), TWL(t), WS(t), WD(t)) \tag{7}$$

where,

$\mathrm{URF}(t)$ is the output of surface current components from RF models at time step t (cm/s);
$f(\bullet)$ indicates the random forests function;
$UEF(t)$ is the EFDC model surface current output at time step t (cm/s);
$UEF(t-1)$ is the EFDC model surface current output at time step $(t-1)$ (cm/s);
$TWL(t)$ is the tidal water level from OTPS at time step t (m);
$WS(t)$ is the ECMWF wind speed at time step t (m/s);
$WD(t)$ is the ECMWF wind direction at time step t (°).

RF models were established for predicting surface currents based on training the relationship between two datasets. A schematic of the development process of proposed models is shown in Figure 2. HFR data were only used as target variables during the procedure of training RF models; HFR data were used to evaluate the RF modeling performance when the test dataset was adopted; HFR data were not applied in the RF models during the forecasting period.

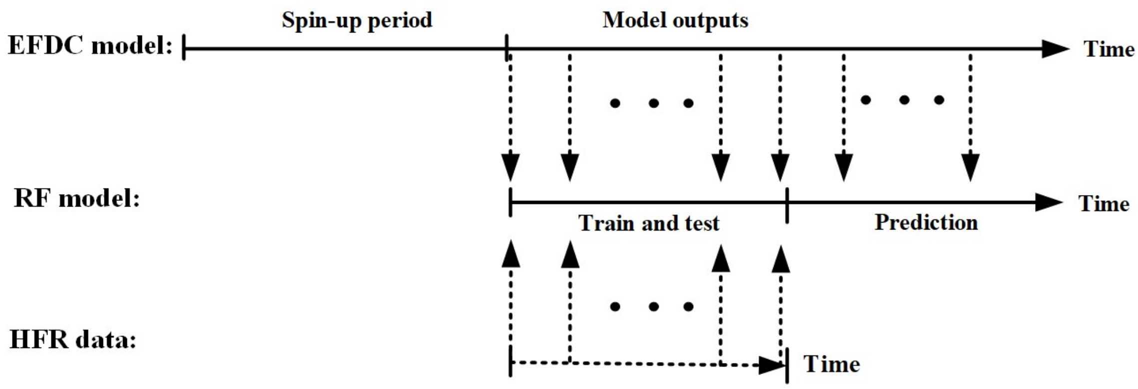

Figure 2. Schematic of the proposed prediction method.

To define an appropriate input variable structure, sensitivity experiments were performed to determine the appropriate number of input variables. Tidal water elevation, wind speed and wind direction were linearly interpolated to the radar data observation time step. Mean and range of these variables of the training dataset, test dataset and forecasting dataset are presented in Tables 1–3.

Table 1. Mean and range of training dataset (60% of total dataset).

Point	Variable	Mean	Minimum	Maximum
P1	U(HFR) (cm/s)	16.29	0.81	35.63
	UEF (cm/s)	14.43	1.26	36.89
	TWL (m)	0	−2.12	2.08
	WS (m/s)	5.35	1.81	7.98
	WD (degrees)	186.39	102.40	255.01
P2	U(HFR) (cm/s)	17.19	1.20	35.80
	UEF (cm/s)	15.55	0.78	32.64
	TWL (m)	0.00	−2.12	2.08
	WS (m/s)	5.36	1.81	7.98
	WD (degrees)	185.77	97.35	255.01
P3	U(HFR) (cm/s)	16.71	2.41	34.54
	UEF (cm/s)	17.87	2.54	35.53
	TWL (m)	0.00	−2.12	2.08
	WS (m/s)	5.36	1.81	7.98
	WD (degrees)	185.77	97.35	255.01

Note that U(HFR) indicates total surface velocity from high frequency radar; TWL indicates tidal water elevation from the Oregon Tide Prediction Software (OTPS) prediction model; WS and WD indicates wind speed and wind direction from the European Centre for Medium-Range Weather Forecasts (ECMWF) forecasting model respectively.

Table 2. Mean and range of testing dataset (20% of total dataset).

Point	Variable	Mean	Minimum	Maximum
P1	U(HFR) (cm/s)	14.23	2.03	32.34
	UEF (cm/s)	10.87	1.15	25.61
	TWL (m)	−0.03	−1.51	1.44
	WS (m/s)	6.21	3.38	8.41
	WD (degrees)	198.72	171.64	247.65
P2	U(HFR) (cm/s)	14.69	2.76	27.18
	UEF (cm/s)	11.88	0.98	25.66
	TWL (m)	−0.02	−1.51	1.44
	WS (m/s)	6.13	3.38	8.41
	WD (degrees)	198.15	166.09	247.65
P3	U(HFR) (cm/s)	13.91	4.96	24.59
	UEF (cm/s)	12.14	2.01	21.89
	TWL (m)	−0.02	−1.51	1.44
	WS (m/s)	6.13	3.38	8.41
	WD (degrees)	198.15	166.09	247.65

Tables 1–3 show mean, minimum and maximum values of the three datasets at each analysis point (P1, P2, and P3). The mean values, minimum and maximum values varied in these datasets at each analysis point. Variations among these datasets is of great significance when examining and assessing the effectiveness of the proposed RF models. The datasets are important in determining whether or not the proposed RF models are robust enough to generate accurate predictions.

Sensitivity experiments of the input variable structure were performed for surface velocity components separately at each analysis point, as presented in Tables 4 and 5. Table 4 presents statistics of each RF model on testing input variable structure for the zonal surface velocity component; statistics for the meridional surface velocity component are presented in Table 5. The test dataset, accounting for 20% of the total dataset, was used to evaluate the performance of each RF model based on statistics between RF model results and radar data.

Table 3. Mean and range of prediction dataset (20% of total dataset).

Point	Variable	Mean	Minimum	Maximum
P1	U(HFR) (cm/s)	17.44	2.94	34.91
	UEF (cm/s)	16.51	1.92	32.79
	TWL (m)	−0.04	−2.53	2.36
	WS (m/s)	6.09	4.21	8.84
	WD (degrees)	211.88	167.31	250.24
P2	U(HFR) (cm/s)	17.65	3.20	31.83
	UEF (cm/s)	17.74	1.55	30.75
	TWL (m)	0.07	−2.43	2.36
	WS (m/s)	6.02	4.21	8.84
	WD (degrees)	210.12	167.31	250.24
P3	U(HFR) (cm/s)	16.49	1.68	28.13
	UEF (cm/s)	17.74	3.14	32.42
	TWL (m)	0.07	−2.43	2.36
	WS (m/s)	6.02	4.21	8.84
	WD (degrees)	210.12	167.31	250.24

Table 4. Statistics of zonal surface velocity component (testing dataset).

Model	Input Variables	R1	R2	Bias	RMSE1 (cm/s)	RMSE2 (cm/s)	SI
UP1RF1	UEF(t), UEF(t − 1)	0.9259	0.9429	−3.3166	5.8804	7.1152	0.7323
UP1RF2	UEF(t), UEF(t − 1),TWL(t)	0.9204	0.9429	−3.5441	6.1199	7.1152	0.7621
UP1RF3	UEF(t), UEF(t − 1),WS(t)	0.9457	0.9429	−3.0766	5.2315	7.1152	0.6515
UP1RF4	UEF(t), UEF(t − 1),WD(t)	0.9337	0.9429	−2.3496	5.1282	7.1152	0.6396
UP1RF5	UEF(t), UEF(t − 1),TWL(t),WS(t)	0.9420	0.9429	−3.4323	5.5389	7.1152	0.6897
UP1RF6	UEF(t), UEF(t − 1),TWL(t),WD(t)	0.9271	0.9429	−2.8268	5.5479	7.1152	0.6909
UP1RF7	**UEF(t), UEF(t − 1),WS(t),WD(t)**	**0.9450**	**0.9429**	**−2.2880**	**4.7701**	**7.1152**	**0.5940**
UP1RF8	UEF(t), UEF(t − 1),TWL(t),WS(t),WD(t)	0.9428	0.9429	−2.7379	5.0978	7.1152	0.6348
UP2RF1	UEF(t), UEF(t − 1)	0.8733	0.8446	0.9801	6.2389	7.1729	0.8376
UP2RF2	UEF(t), UEF(t − 1),TWL(t)	0.8729	0.8446	1.6002	6.1501	7.1729	0.8256
UP2RF3	**UEF(t), UEF(t − 1), WS(t)**	**0.8862**	**0.8446**	**1.0590**	**5.9567**	**7.1729**	**0.7997**
UP2RF4	UEF(t), UEF(t − 1),WD(t)	0.8831	0.8446	1.4787	6.0854	7.1729	0.8170
UP2RF5	UEF(t), UEF(t − 1),TWL(t),WS(t)	0.8823	0.8446	1.1141	6.0047	7.1729	0.8061
UP2RF6	UEF(t), UEF(t − 1),TWL(t),WD(t)	0.8744	0.8446	1.4671	6.2523	7.1729	0.8394
UP2RF7	UEF(t), UEF(t − 1),WS(t),WD(t)	0.8851	0.8446	1.4156	6.0472	7.1729	0.8118
UP2RF8	UEF(t), UEF(t − 1),TWL(t),WS(t),WD(t)	0.8821	0.8446	1.1639	6.1089	7.1729	0.8201
UP3RF1	UEF(t), UEF(t − 1)	0.8661	0.8764	−4.3970	6.9525	6.2198	1.0892
UP3RF2	UEF(t), UEF(t − 1),TWL(t)	0.8682	0.8764	−4.8322	7.2423	6.2198	1.1346
UP3RF3	UEF(t), UEF(t − 1),WS(t)	0.9029	0.8764	−4.2053	6.2616	6.2198	0.9810
UP3RF4	UEF(t), UEF(t − 1),WD(t)	0.8696	0.8764	−3.7031	6.4570	6.2198	1.0116
UP3RF5	UEF(t), UEF(t − 1),TWL(t),WS(t)	0.9004	0.8764	−4.6492	6.6429	6.2198	1.0407
UP3RF6	UEF(t), UEF(t − 1),TWL(t),WD(t)	0.8700	0.8764	−4.1360	6.7427	6.2198	1.0564
UP3RF7	**UEF(t), UEF(t − 1),WS(t), WD(t)**	**0.8988**	**0.8764**	**−3.4843**	**5.8521**	**6.2198**	**0.9168**
UP3RF8	UEF(t), UEF(t − 1),TWL(t),WS(t),WD(t)	0.8995	0.8764	−3.9344	6.1614	6.2198	0.9653

Note that R1 indicates the correlation coefficient between radar data and RF model results; R2 indicates the correlation coefficient between radar data and numerical model outputs only; RMSE1 indicates root mean square error between radar data and RF model results; RMSE2 indicates root mean square error between radar data and numerical model outputs; SI indicates the scatter index between radar data and RF model results; bias is the difference between radar data and RF model results. The same parameters are used as follows.

Table 4 shows that model UP1RF7, which used EFDC model outputs, wind speed and wind direction as input variables, produced the minimum root mean square error (RMSE) value (RMSE1) for the zonal surface velocity component at point P1; the same input variable structure for point P3 generated the minimum RMSE (RMSE1); while the best input variables for point P2 were EFDC model outputs and wind speed. This indicates that inclusion of wind speed was a highly significant parameter and improves RF model performance for the zonal surface velocity component. However, the effect of wind directions varied from location to location. Improvement of RMSE1 values for P1, P2 and P3 were 33%, 17% and 6% in comparison with RMSE2 respectively. Additionally, correlation (R1) between RF model results and radar data was enhanced as well at these locations comparing with correlation (R2) between EFDC model results and radar data. The maximum improvement in correlation was at point P2, by 5%. Table 4 shows that using (i) EFDC model outputs at two time-steps; (ii) wind speeds and (iii) wind directions to establish RF models can further improve estimation of zonal surface velocity components.

Table 5. Statistics of meridional surface velocity component (testing dataset).

Model	Input Variables	R1	R2	Bias	RMSE1 (cm/s)	RMSE2 (cm/s)	SI
VP1RF1	VEF(t), VEF(t − 1)	0.1722	0.3670	−0.2246	5.7549	5.5092	−23.8347
VP1RF2	VEF(t), VEF(t − 1),TWL(t)	0.3217	0.3670	0.1495	5.2536	5.5092	−21.7586
VP1RF3	VEF(t), VEF(t − 1),WS(t)	0.2996	0.3670	0.8936	5.5466	5.5092	−22.9719
VP1RF4	VEF(t), VEF(t − 1),WD(t)	0.6852	0.3670	0.7174	4.1340	5.5092	−17.1215
VP1RF5	VEF(t), VEF(t − 1),TWL(t),WS(t)	0.4922	0.3670	1.0752	4.9408	5.5092	−20.4630
VP1RF6	**VEF(t), VEF(t − 1),TWL(t),WD(t)**	**0.7941**	**0.3670**	**0.9206**	**3.6745**	**5.5092**	**−12.2185**
VP1RF7	VEF(t), VEF(t − 1),WS(t),WD(t)	0.5934	0.3670	1.9665	4.9099	5.5092	−20.3348
VP1RF8	VEF(t), VEF(t − 1),TWL(t),WS(t),WD(t)	0.7270	0.3670	1.8684	4.2173	5.5092	−17.4663
VP2RF1	VEF(t), VEF(t − 1)	0.1779	0.5097	−0.0881	7.4797	6.9956	17.7816
VP2RF2	VEF(t), VEF(t − 1),TWL(t)	0.1350	0.5097	1.5291	7.7613	6.9956	18.4510
VP2RF3	VEF(t), VEF(t − 1),WS(t)	0.4335	0.5097	0.9096	7.0880	6.9956	16.8503
VP2RF4	VEF(t), VEF(t − 1),WD(t)	0.7307	0.5097	0.2646	5.1449	6.9956	12.2310
VP2RF5	VEF(t), VEF(t − 1),TWL(t),WS(t)	0.4789	0.5097	1.8376	6.8769	6.9956	16.3485
VP2RF6	**VEF(t), VEF(t − 1),TWL(t),WD(t)**	**0.8180**	**0.5097**	**1.3854**	**4.7285**	**6.9956**	**11.2410**
VP2RF7	VEF(t), VEF(t − 1),WS(t),WD(t)	0.6424	0.5097	0.8763	5.8199	6.9956	13.8357
VP2RF8	VEF(t), VEF(t − 1),TWL(t),WS(t),WD(t)	0.7440	0.5097	1.5349	5.2029	6.9956	12.3688
VP3RF1	VEF(t), VEF(t − 1)	0.0127	0.4749	2.4416	8.9829	9.9814	6.4104
VP3RF2	VEF(t), VEF(t − 1),TWL(t)	0.0479	0.4749	1.6222	8.9451	9.9814	6.3835
VP3RF3	VEF(t), VEF(t − 1),WS(t)	0.4449	0.4749	−0.4086	7.4830	9.9814	5.3401
VP3RF4	VEF(t), VEF(t − 1),WD(t)	0.8133	0.4749	−1.3446	5.0184	9.9814	3.5812
VP3RF5	VEF(t), VEF(t − 1),TWL(t),WS(t)	0.4157	0.4749	−0.1224	7.4912	9.9814	5.3459
VP3RF6	**VEF(t), VEF(t − 1),TWL(t),WD(t)**	**0.8670**	**0.4749**	**−0.9184**	**4.4955**	**9.9814**	**3.2081**
VP3RF7	VEF(t), VEF(t − 1),WS(t),WD(t)	0.7648	0.4749	−0.5450	5.2767	9.9814	3.7656
VP3RF8	VEF(t), VEF(t − 1),TWL(t),WS(t),WD(t)	0.8191	0.4749	−0.4309	4.7590	9.9814	3.3962

Sensitivity experiments for the meridional surface velocity component, presented in Table 5, showed that RF models using EFDC model outputs, tidal water elevation and wind direction yielded the best estimations at each of the three locations. Improvements in RMSE values for the meridional surface velocity component between RF model estimations and radar data were 33%, 32% and 55% respectively at points P1, P2 and P3 in comparison with RMSE2. Correlations between RF model results and radar data (R1) were also significantly improved by 116%, 60% and 83% respectively for points P1, P2 and P3 in comparison with correlation between EFDC model results and radar data (R2).

Improvements were generally more significant in the meridional surface velocity component than the zonal surface velocity component at all points based on statistics as presented in Tables 4 and 5. The meridional component is difficult to predict as its signal generally is quite low and is more induced by transient wind stresses than deterministic tidal forcing. Results show that the proposed RF models can improve prediction accuracy for surface velocity components, and it is especially significant for difficult to predict meridional surface velocity components.

3.2. Comparison of Predictions

RF models at the three analysis points, using the best input variable structure, were applied to produce predictions using the forecasting dataset, and compared against "future" radar data. Time series of surface velocity components at the three analysis points are presented in Figures 3–8 respectively.

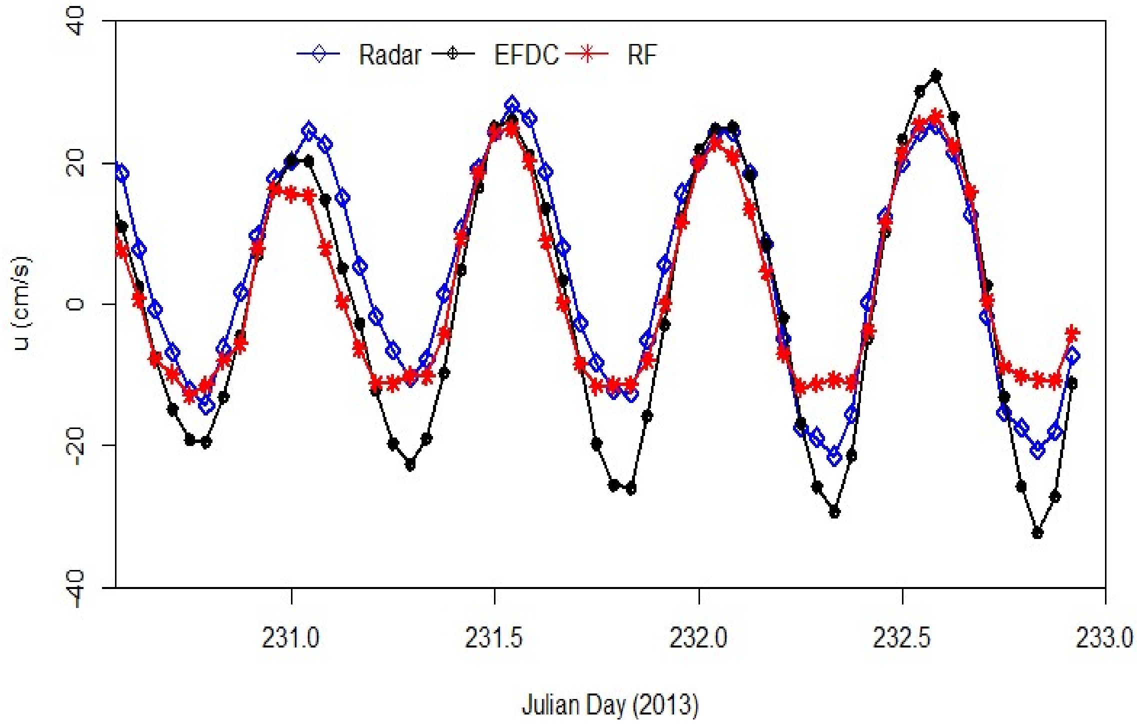

Figure 3. Predictions of zonal surface velocity component at point P1.

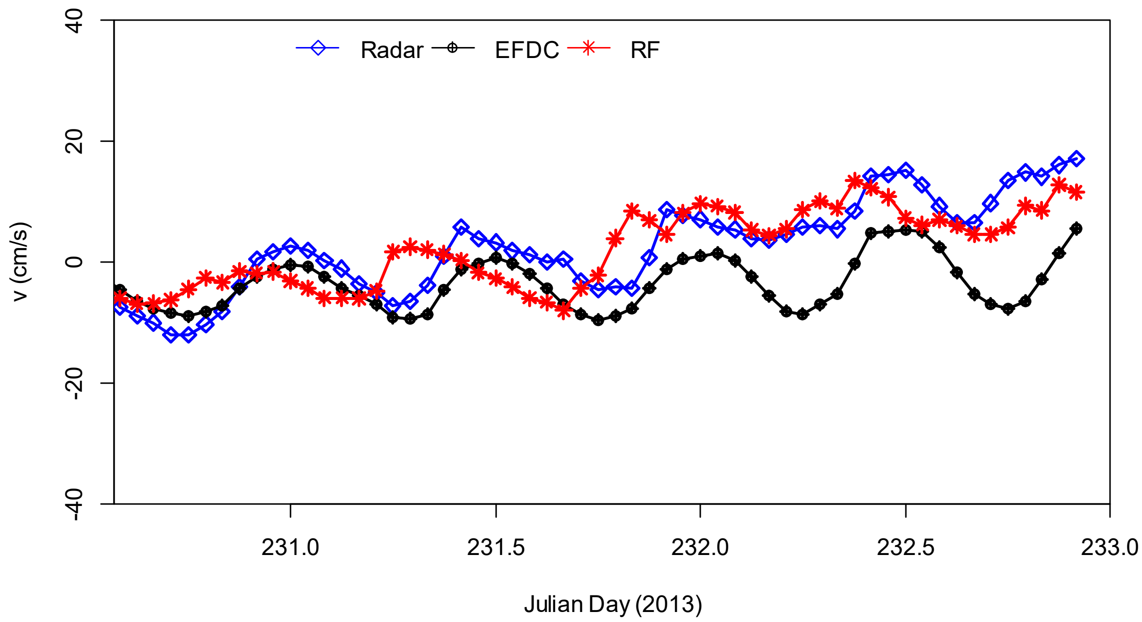

Figure 4. Predictions of meridional surface velocity component at point P1.

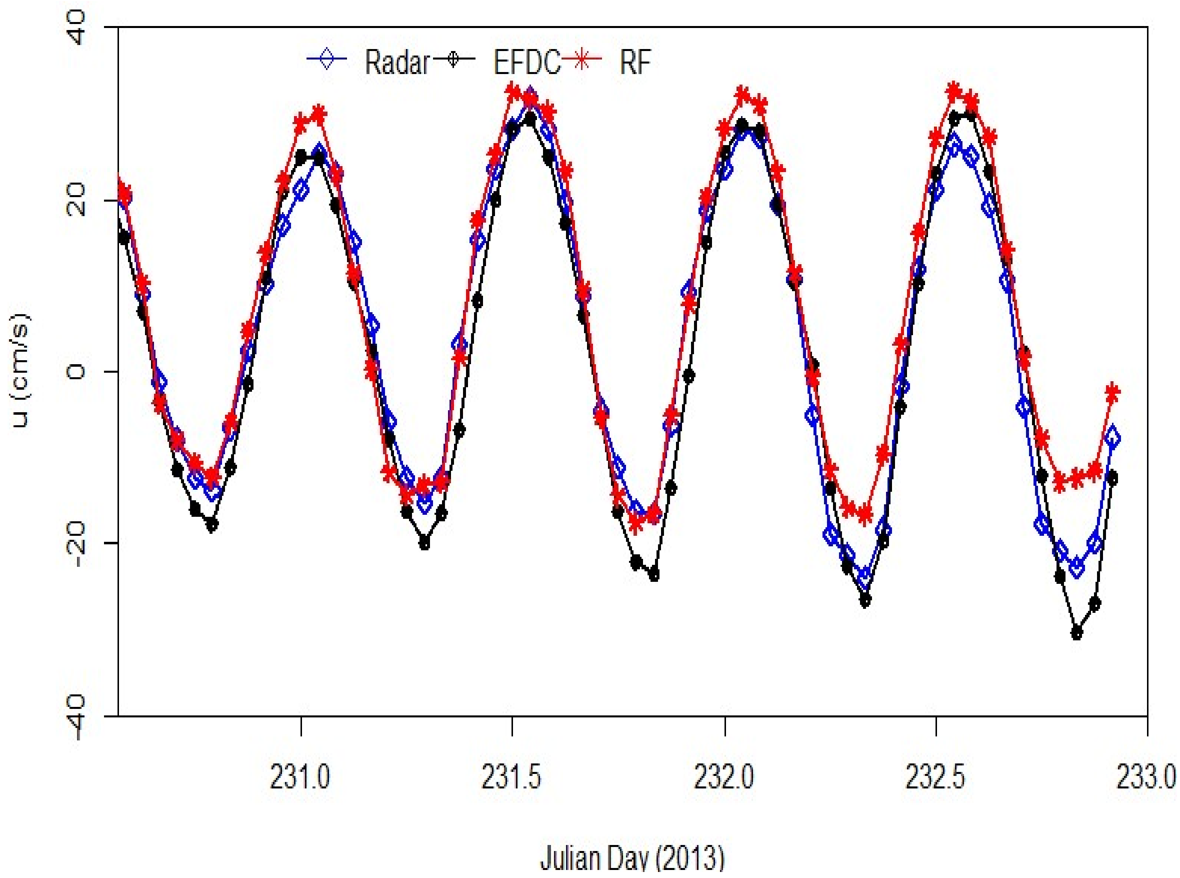

Figure 5. Predictions of zonal surface velocity component at point P2.

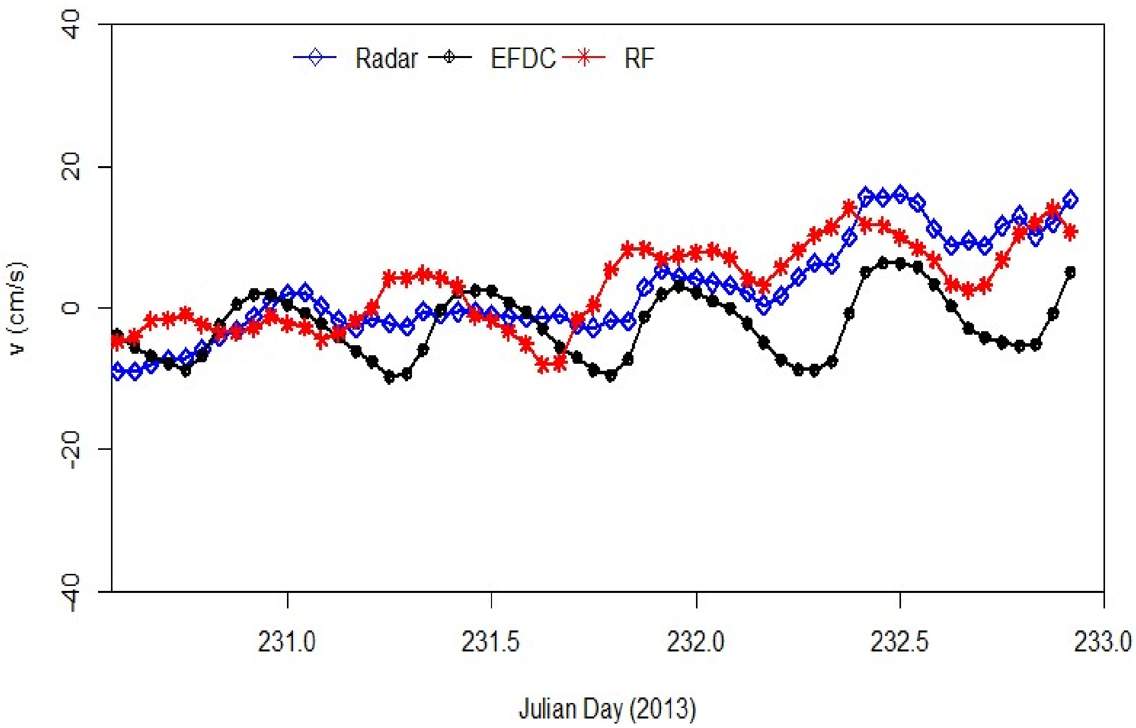

Figure 6. Predictions of meridional surface velocity component at point P2.

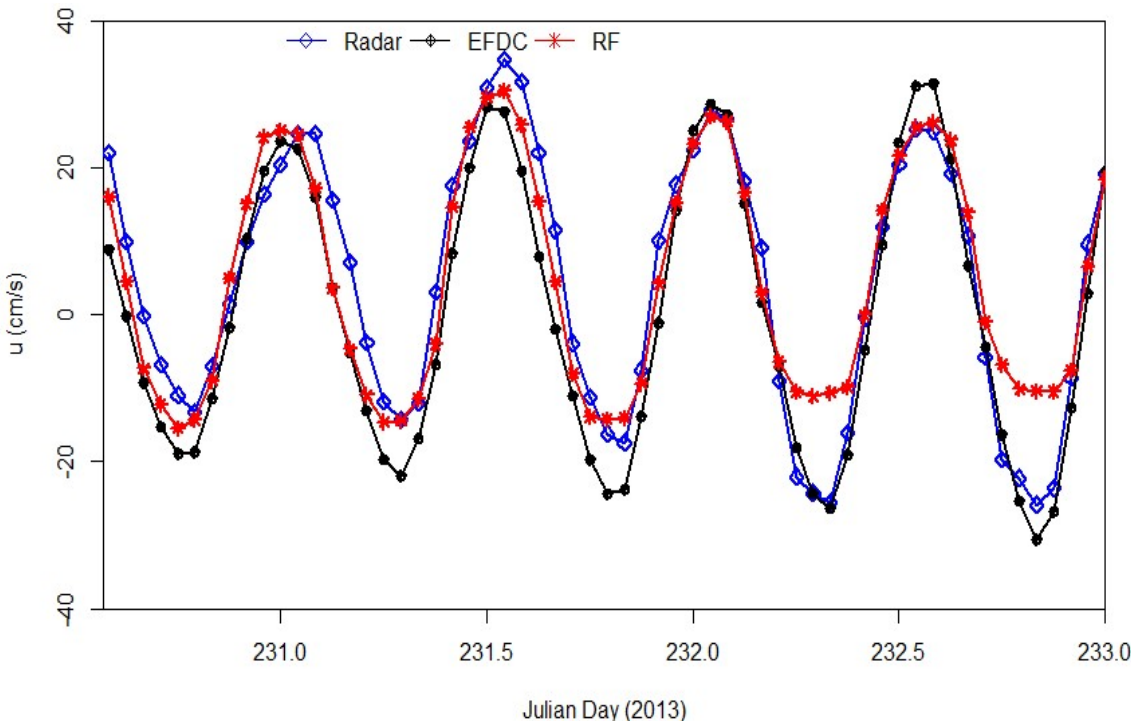

Figure 7. Predictions of zonal surface velocity component at point P3.

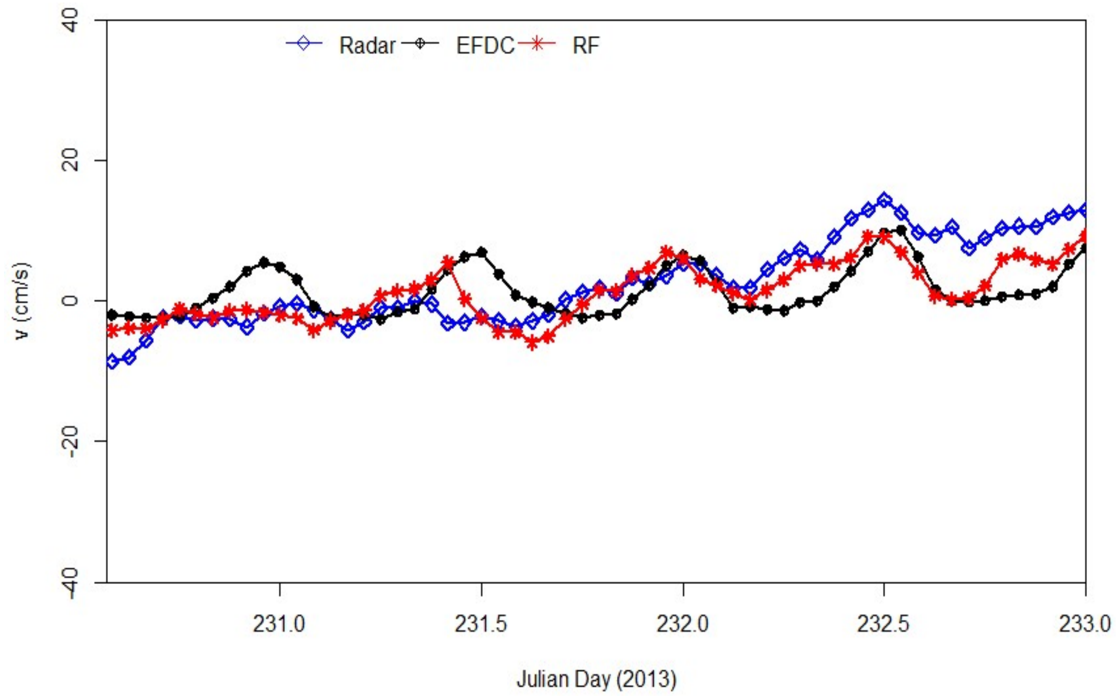

Figure 8. Predictions of meridional surface velocity component at point P3.

Figures 3–8 show that established RF models can produce predictions which have a better agreement with radar data in comparison with EFDC model results at the three analysis points. Statistics between HFR data and RF model predictions, and EFDC results were computed and are presented in Table 6.

Table 6. Statistics of forecasting dataset (59 h).

Location	Component	R1	R2	Bias	RMSE1 (cm/s)	RMSE2 (cm/s)	SI
P1	u	0.9326	0.9589	0.2477	6.3775	6.7774	1.4291
P1	v	0.8250	0.4551	−1.3209	3.7671	5.5251	1.3993
P2	u	0.9772	0.9795	2.7250	4.5843	4.2732	0.9378
P2	v	0.7650	0.4937	0.7659	4.4600	7.5313	2.0941
P3	u	0.9335	0.9710	−2.0553	5.9289	6.9026	1.1886
P3	v	0.7395	0.6713	−0.4820	5.1840	8.1931	2.1957

Table 6 shows that the RF models generally generated better prediction of surface velocity components than EFDC models based on RMSE values over 59 h forecasting period; the improvements are statistically highly significant for meridional surface velocity components. Improvement of the RMSE value for the meridional surface velocity components between the RF model predictions and radar data was greater than 32% at the three analysis points; the maximum RMSE improvement of the meridional surface velocity component was at point P2 at 41%, while improvement of RMSE value for the zonal surface velocity component was greater than 6% at locations P1 and P3. Although correlation between RF model predictions and radar data was quite similar to the correlation between EFDC results and radar data, the RMSE value between RF model predictions and radar data (RMSE1) increased by 7% in comparison with RMSE2. It is also clearly seen from Figures 5, 6 and 8 that RF models were much closer to HFR data than EFDC output at surface current maxima and minima; this is highly significant in terms of accurate prediction of tidal energy and transport effects for operational purposes at marine renewable energy sites. In general, established RF models can improve prediction of surface current components; improvement was more significant in the meridional surface velocity component than the zonal surface velocity component.

4. Discussion

The above analysis shows that the newly developed RF models were capable of producing improved surface velocity components, especially for the meridional surface velocity component. In the above RF models, the length of training dataset at the three analysis locations accounted for 60% of the total dataset i.e., 184 h. The RF models generally produced highly satisfactory predictions of surface velocity components over the 59-h forecasting period. The authors were also interested in assessing the effect of the length of RF training datasets on model predictions. Sensitivity analyses on the length of training dataset were performed at the three analysis locations respectively. Seven versions of training dataset lengths, from one day (24 h) to one week (168 h), were performed. Statistics over a 59-h prediction period between predicted results from RF models and HFR data were computed and are presented in Table 7.

Table 7 shows that RF models generated forecasts closer to HFR data than the EFDC model, in terms of RMSE values, except for models with a 24-h training length. In addition, correlation coefficients between RF model results and HFR data were comparable to those between EFDC model results and HFR data. High correlation existed in all RF models for the zonal surface velocity component, since the EFDC model produced satisfactory results for the zonal surface velocity component. The authors focused on improving forecasting accuracy for the meridional surface velocity component in this research. For the meridional surface velocity component, statistical values of the correlation coefficient between RF forecasts and HFR data generally had an increasing trend as the length of the training dataset increased. This indicated that larger training datasets produced better forecasts than using a small training dataset. This was because the larger variation range of each input variable used for training can be more comprehensively captured by RF models, and then the developed RF model is more robust to deal with the forecasting dataset which may have a large variation range. Additionally, all correlation coefficients R1 between RF models and radar data were greater than the correlation coefficient R2 between EFDC model results and radar data. This illustrated that the proposed approach that combines a numerical model with the RF model is effective in enhancing forecasts for the meridional surface velocity components. Moreover, RMSE values between RF model results and HFR data generally decreased as training datasets over longer periods were used. This proved that RF models generated forecasts closer to HFR data when a larger dataset was adopted during training phase. Improvement of the meridional surface velocity component by the RF model in comparison with EFDC model results was 36% based on RMSE values when the training dataset length was 168 h. Moreover, computational cost of using the RF algorithm based on R software was less than one hour for these models, which is much faster than a numerical model with data assimilation.

Table 7. Averaged statistics for training length tests (59-h prediction, averaged at three points).

Length (hours)	Variable	R1	R2	Bias	RMSE1 (cm/s)	RMSE2 (cm/s)	SI
24	u	0.95	0.97	1.60	6.40	5.98	1.35
	v	0.73	0.54	−1.36	5.20	7.08	2.20
48	u	0.95	0.97	0.86	5.63	5.98	1.18
	v	0.74	0.54	−0.74	4.84	7.08	2.05
72	u	0.95	0.97	−0.80	5.50	5.98	1.16
	v	0.76	0.54	−0.77	4.63	7.08	1.96
96	u	0.94	0.97	0.14	5.68	5.98	1.19
	v	0.76	0.54	−0.80	4.66	7.08	1.98
120	u	0.95	0.97	0.41	5.23	5.98	1.10
	v	0.76	0.54	−0.91	4.62	7.08	1.96
144	u	0.95	0.97	0.03	5.57	5.98	1.17
	v	0.69	0.54	−1.74	5.23	7.08	2.20
168	u	0.94	0.97	0.58	5.83	5.98	1.23
	v	0.77	0.54	−0.41	4.54	7.08	1.92

In short, application of RF method based on a numerical model and HFR data improve both forecasts of surface velocity components. But forecasting performance of meridional surface velocity component was more sensitive to training dataset length than the zonal surface velocity component.

5. Conclusions

Marine renewable energy sites require the most accurate data and forecasts possible for resources assessment, project planning and operational purposes. In this research, outputs of surface current components from a coastal hydrodynamic EFDC model, along with tidal elevations, wind speeds and directions were taken as input variables to establish RF models for components of surface water velocity. Input variable structures of each RF model was examined at three analysis points in the model domain within Galway Bay. Influences of training dataset lengths on model performances were also examined. Improvements in RF models were assessed based on several skills criteria: Correlation, RMSE, bias and SI in comparison with HFR data.

The main conclusions from this research are:

1. The best input variable structure in establishing RF models for the zonal surface velocity component was obtained using outputs u(t) and u(t − 1) from the EFDC model and wind speeds as input variables at the three analysis points, wind direction was needed at some locations. Correlation coefficients of zonal surface velocity components between RF model results and HFR data were improved and greater than 0.89 during testing.
2. The best RF models for the meridional surface velocity component used outputs v(t) and v(t − 1) from the EFDC model, tidal elevations and wind directions as input variables at the three analysis points. Correlation coefficients of the meridional surface velocity component between RF model

results and HFR data were significantly improved and greater than 0.79 for the three analysis points during testing. This is a significant result as the meridional component of velocity at Galway Bay is a low signal and strongly influenced by transient wind conditions.

3. The RF models produced comparable, or improved, forecasts over 59 h for both surface velocity components. Improvement was more significant for the meridional surface velocity component than for the zonal surface velocity component in terms of RMSE values at the three analysis points. The maximum improvement in RMSE values between RF model results and radar data for the zonal surface velocity component was 14% at point P3 and 37% for the meridional surface velocity component at point P2.
4. The RF models were robust and efficient enough to generate high-accuracy predictions using a very short-term training dataset, even less than 48 h. Correlation of both surface velocity components between HFR data and RF model prediction over a 59-h period was greater than 0.73 at the three analysis points when only a 24-h dataset was used to train RF models. This is of great importance for various realistic applications such as operations of a marine renewable energy site as improvement in the meridional surface velocity component, by RF model in comparison with EFDC model results, was 36% based on RMSE values when the training dataset length was 168 h.
5. Along with the improved forecast statistics, maximum and minimum velocities are also significantly improved using RF models; this is clearly seen by inspecting the zonal velocities of Figures 5, 6 and 8. At P1, Figure 5, the minimum values of HRF and RF were almost the same, whereas the minimum value from the EFDC model output was about 50% lower that the measured HFR data. This is a recurring feature of improved predictions using RF. Minimum and maximum values are very significant in terms of resource assessment and site operations, so this improved predictability is very important to developers of marine renewable energy extraction sites.

In short, the established RF models were capable of improving prediction accuracy for coastal surface currents. The RF model might become a robust, efficient and promising approach to provide timely and useful surface current information especially for the meridional surface velocity component of the Galway Bay area. This research is a test and case study to develop a soft computing model using an RF method to improve forecasting accuracy of surface currents. Future research will focus on combining the RF algorithm into the EFDC model to generate predictions over a large domain.

Author Contributions: L.R. and J.M. wrote the original draft; Y.L. visualized data; X.L. undertook analysis; J.L. prepared methodology and M.H. reviewed and edited the manuscript.

Funding: This research was funded by Sun Yat-sen University Basic Research Funding [76170-18841203].

Acknowledgments: We would like to thank ECMWF for providing wind data and University of Oregon for providing the OTPS tide prediction software. The authors gratefully acknowledge financial support of Sun Yat-sen University Basic Research Funding [76170-18841203].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Peng, M.; Schmalz, R.A., Jr.; Zhang, A.; Aikman, F., III. Towards the Development of the National Ocean Service San Francisco Bay Operational Forecast System. *J. Mar. Sci. Eng.* **2014**, *2*, 247–286. [CrossRef]
2. Ren, L.; Hartnett, M. Comparative Study on Assimilating Remote Sensing High Frequency Radar Surface Currents at an Atlantic Marine Renewable Energy Test Site. *Remote Sens.* **2017**, *9*, 1331. [CrossRef]
3. He, Z.; Wen, X.; Liu, H.; Du, J. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *J. Hydrol.* **2014**, *509*, 379–386. [CrossRef]
4. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]

5. Malek, S.; Gunalan, R.; Kedija, S.Y.; Lau, C.F.; Mosleh, M.A.A.; Milow, P.; Lee, S.A.; Saw, A. Random forest and Self Organizing Maps application for analysis of pediatric fracture healing time of the lower limb. *Neurocomputing* **2018**, *272*, 55–62. [CrossRef]
6. Ibarra-Berastegi, G.; Saénz, J.; Esnaola, G.; Ezcurra, A.; Ulazia, A. Short-term forecasting of the wave energy flux: Analogues, random forests, and physics-based models. *Ocean Eng.* **2015**, *104*, 530–539. [CrossRef]
7. Herrera, M.; Torgo, L.; Izquierdo, J.; Pérez-García, R. Predictive models for forecasting hourly urban water demand. *J. Hydrol.* **2010**, *387*, 141–150. [CrossRef]
8. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* **2006**, *9*, 181–199. [CrossRef]
9. Strobl, C.; Malley, J.; Tutz, G. An introduction to Recursive Partitioning: Rationale, Application and Characteristics of Classification and Regression Trees, Bagging and Random Forests. *Psychol. Methods* **2009**, *14*, 323–348. [CrossRef]
10. Lahouar, A.; Slama, J.B.H. Day-ahead load forecast using random forest and expert input selection. *Energy Convers. Manag.* **2015**, *103*, 1040–1051. [CrossRef]
11. Catani, F.; Lagomarsino, D.; Segoni, S.; Tofani, V. Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2815–2831. [CrossRef]
12. Mahjoobi, J.; Mosabbeb, E.A. Prediction of significant wave height using regressive support vector machines. *Ocean Eng.* **2009**, *36*, 339–347. [CrossRef]
13. Balas, C.E.; Koc, L.; Balas, L. Predictions of Missing Wave Data by Recurrent Neuronets. *J. Waterw. Portcoast. Ocean Eng.* **2004**, *130*, 256–265. [CrossRef]
14. Ren, L.; Nash, S.; Hartnett, M. Observation and modeling of tide- and wind-induced surface currents in Galway Bay. *Water Sci. Eng.* **2015**, *8*, 345–352. [CrossRef]
15. Wen, L. Three-Dimensional Hydrodynamic Modelling in Galway Bay. Ph.D. Thesis, University College Galway, Galway, Ireland, 1995.
16. Ren, L.; Nagle, D.; Hartnett, M.; Nash, S. The Effect of Wind Forcing on Modeling Coastal Circulation at a Marine Renewable Test Site. *Energies* **2017**, *10*, 2114. [CrossRef]
17. Hamrick, J.M. *A Three-Dimensional Environmental Fluid Dynamics Computer Code: Therotical and Computatonal Aspects*; Virginia Institute of Marine Science: Gloucester Point, VA, USA, 1992.
18. Hamrick, J.M. *User's Manual for the Environmental Fluid Dynamics Computer Code*; Department of Physical Sciences, School of Marine Science, Virginia Institute of Marine Science, College of William and Mary: Williamsburg, VA, USA, 1996.
19. Hamrick, J.M. *EFDC Technical Memorandum*; Tetra Tech: Fairfax, VA, USA, 2006.
20. Hamrick, J.M. *The Environmental Fluid Dynamics Code Theory and Computation Volume 1: Hydrodynamics and Mass Transport*; Tetra Tech, Inc.: Fairfax, VA, USA, 2007; p. 60.
21. Wang, C.; Sun, Y.; Zhang, X. Numerical Simulation of 3D Tidal Currents Based on the EFDC Model in Jiaozhou Bay. *Period. Ocean Univ. China* **2008**, *38*, 833–840.
22. Zhang, Q.; Tan, F.; Han, T.; Wang, X.; Hou, Z.; Yang, H. Simulation of sorting sedimentation in the channel of huanghua harbor by using 3D multi-sized sediment transport model of EFDC. In Proceedings of the 32nd Conference on Coastal Engineering, Shanghai, China, 30 June–5 July 2010; pp. 1–11.
23. Hayter, E.J. *Literature Review of EFDC Applications Demonstrating Capability for Use in the Jacksonville Harbor Feasibility Study*; US Army Engineer Research and Development Center: Vicksburg, MS, USA, 2014; pp. 1–36.
24. O'Donncha, F.; Hartnett, M.; Nash, S.; Ren, L.; Ragnoli, E. Characterizing observed circulation patterns within a bay using HF radar and numerical model simulations. *J. Mar. Syst.* **2015**, *142*, 96–110. [CrossRef]
25. Ren, L.; Nash, S.; Hartnett, M. Forecasting of Surface Currents via Correcting Wind Stress with Assimilation of High-Frequency Radar Data in a Three-Dimensional Model. *Adv. Meteorol.* **2016**, *2016*, 8950378. [CrossRef]
26. Gurgel, K.W.; Antonischski, G.; Essen, H.H.; Schlick, T. Wellen Radar (WERA): A new ground-wave HF radar for ocean remote sensing. *Coast. Eng.* **1999**, *37*, 219–234. [CrossRef]
27. Shay, L.K.; Martinez-Pedraja, J.; Cook, T.M.; Haus, B.K.; Weisberg, R.H. High-Frequency Radar Mapping of Surface Currents Using WERA. *J. Atmos. Ocean. Technol.* **2007**, *24*, 484–503. [CrossRef]
28. Zeng, Y.; Zhou, H.; Lai, Y.; Wen, B. Wind Direction Mapping with a Modified Wind Spreading Function by Broad-Beam High-Frequency Radar. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 679–683. [CrossRef]
29. Lai, Y.; Zhou, H.; Zeng, Y.; Wen, B. Quantifying and Reducing the DOA Estimation Error Resulting from Antenna Pattern Deviation for Direction-Finding HF Radar. *Remote Sens.* **2017**, *9*, 1285. [CrossRef]

30. Wang, J.; Dizaji, R.; Ponsford, A.M. Analysis of clutter distribution in bistatic high frequency surface wave radar. In Proceedings of the Conference on Electrical & Computer Engineering, Niagara Falls, ON, Canada, 2–5 May 2004; pp. 1301–1304.
31. Haus, B.K.; Wang, J.D.; Rivera, J.; Martinez-Pedraja, J.; Smith, N. Remote Radar Measurement of Shelf Currents off Key Largo, Florida, U.S.A. *Estuar. Coast. Shelf Sci.* **2000**, *51*, 553–569. [CrossRef]
32. Paduan, J.D.; Washburn, L. High-Frequency Radar Observations of Ocean Surface Currents. *Annu. Rev. Mar. Sci.* **2013**, *5*, 115–136. [CrossRef]
33. Shulman, I.; Paduan, J.D. Assimilation of HF radar-derived radials and total currents in the Monterey Bay area. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **2009**, *56*, 149–160. [CrossRef]
34. Haines, S.; Seim, H.; Muglia, M. Implementing Quality Control of High-Frequency Radar Estimates and Application to Gulf Stream Surface Currents. *J. Atmos. Ocean. Technol.* **2017**, *34*, 1207–1224. [CrossRef]
35. Rubio, A.; Mader, J.; Corgnati, L.; Mantovani, C.; Griffa, A.; Novellino, A.; Quentin, C.; Wyatt, L.; Schulz-Stellenfleth, J.; Horstmann, J.; et al. HF Radar Activity in European Coastal Seas: Next Steps toward a Pan-European HF Radar Network. *Front. Mar. Sci.* **2017**, *4*, 8. [CrossRef]
36. Hisaki, Y.; Kashima, M.; Kojima, S. Surface current patterns observed by HF radar: Methodology and analysis of currents to the north of the Yaeyama Islands, East China Sea. *Ocean Dyn.* **2016**, *66*, 329–352. [CrossRef]
37. Sun, Y.; Chen, C.; Beardsley, R.C.; Ullman, D.; Butman, B.; Lin, H. Surface circulation in Block Island Sound and adjacent coatsal and shelf regions: A FVCOM-CODAR comparison. *Prog. Oceanogr.* **2016**, *143*, 26–45. [CrossRef]
38. Liu, Y.; Weisberg, R.H.; Merz, C.R.; Lichtenwalner, S.; Kirkpatrick, G.J. HF Radar Performance in a Low-Energy Environment: CODAR SeaSonde Experience on the West Florida Shelf. *J. Atmos. Ocean. Technol.* **2010**, *27*, 1689–1710. [CrossRef]
39. Breiman, L. *Random Forests-Random Features*; University of California: Berkeley, CA, USA, 1999; pp. 1–28.
40. Niu, D.; Pu, D.; Dai, S. Ultra-Short-TermWind-Power Forecasting Based on the Weighted Random Forest Optimized by the Niche Immune Lion Algorithm. *Energies* **2018**, *11*, 1098. [CrossRef]
41. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning:Data Mining, Inference, and Prediction*; Springer: Los Altos, CA, USA, 2009.
42. Liu, C.; Tang, F.; Bak, C.L. An Accurate Online Dynamic Security Assessment Scheme Based on Random Forest. *Energies* **2018**, *11*, 1914. [CrossRef]
43. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
44. Huang, N.; Lu, G.; Xu, D. A Permutation Importance-Based Feature Selection Method for Short-Term Electricity Load Forecasting Using Random Forest. *Energies* **2016**, *9*, 767. [CrossRef]
45. Mahjoobi, J.; Etemad-Shahidi, A. An alternative approach for the prediction of significant wave heights based on classification and regression trees. *Appl. Ocean Res.* **2008**, *30*, 172–177. [CrossRef]
46. Aydogan, B.; Ayat, B.; Öztürk, M.N.; Özkan Çevik, E.; Yüksel, Y. Current velocity forecasting in straits with artificial neural networks, a case study: Strait of Istanbul. *Ocean Eng.* **2010**, *37*, 443–453. [CrossRef]