Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 107 (2017) 345 – 350

Energy Procedia

3rd International Conference on Energy and Environment Research, ICEER 2016, 7-11 September 2016, Barcelona, Spain

Prediction of Surface Currents Using High Frequency CODAR Data and Decision Tree at a Marine Renewable Energy Test Site

Lei Ren*, Michael Hartnett

Marine Renewable Energy Ireland (MaREI) Research Centre, Galway, Ireland
Department of Civil Engineering, National University of Ireland, Galway, Ireland
Ryan Institute, Galway, Ireland

Abstract

In this study, Decision Tree (DT) was employed to predict surface currents in a ¼-scaled marine renewable energy test site—Galway Bay. In training and testing models, wind speed, wind direction and tidal water elevation from a forecasting model, and observations of surface velocity components during previous hours were taken as input variables; surface velocity components were taken as the output variable. Appropriate value of Complexity Parameter (CP) in decision tree models was determined based on experiments producing the minimum Root-Mean-Square-Error (RMSE) values compared with the radar data. Statistics including RMSE, bias, correlation (R) and Scatter Index (SI) were computed between predictions and radar data to assess predictions. Results indicated that the DT model can produce satisfactory predictions of surface currents. Good performance of DT model indicated that it can be regarded as a promising approach for future applications.

Peer-review under responsibility of the scientific committee of the 3rd International Conference on Energy and Environment Research.

Keywords: surface currents; decision tree; marine renewable energy; prediction; Galway Bay; scatter index; correlation; CODAR; radars

* Corresponding author. Tel.: +353 89 2363 957.
E-mail address: leirencomeon@gmail.com

Peer-review under responsibility of the scientific committee of the 3rd International Conference on Energy and Environment Research.
doi:10.1016/j.egypro.2016.12.171

1. Introduction

Accurate real-time information of surface currents is becoming more and more important for a sea of aspects. Good understanding of surface currents can help human to better deal with matters such as oil spill treatment, marine renewable energy extraction and climate forecasting and so on. In addition, green energies such as marine renewable energy is currently becoming a hot topic owing to pollution issues in air, water and soil using traditional energies.

In general, two approaches—numerical models and observation platforms based on remote sensors are widely used to investigate dynamics of water movement in coastal areas. However, numerical model which approximates dynamic processes in a mathematic way inevitably results in model errors. These errors mainly result from the discretization of model grids in space, simplification of initial and boundary conditions. Although numerical models can produce forecasting information, quality of predictions is not easily met. Observation platforms such as radars and satellites can monitor water parameters over a large spatial domain even global with short observation window, but they can not provide information of future status of water parameters of interest. An alternative approach has been proposed and widely applied in a variety of fields due to development of Artificial Intelligence (AI). Decision Tree (DT), Artificial Neural Networks (AANs) and Deep Learning (DP) and so on have been employed to predict states of interest in many fields, such as Partal, Cigizoglu and Kahya [1], Timofeev [2], Makarynskyy, Pires-Silva, Makarynska and Ventura-Soares [3], Solomatine and Xue [4]. In this work, authors focused on predicting surface currents for a 1/4-scaled marine renewable test site—Galway Bay. Observation of surface currents from the CODAR observation system, wind speed and wind direction from a forecasting model and tidal water elevation from an inversion model were applied to develop DT models. Three locations with high coverage by the CODAR system were used for analysis. Surface east-west and north-south velocity components were predicted separately using DT models.

Structure of this paper is organized as: Section 2 introduction the High Frequency radar observation system. Section 3 presents the Decision Tree algorithm. Results are presented in Section 4, followed by conclusions in Section 5.

2. High frequency radar data

The total surface current velocity vector is determined by summing surface currents radial velocity components from at least two radar locations. There are various applications of radar data, including search and rescue support, oil-spill mitigation in real time and larval population connectivity assessment when viewed over many years [5]. Two CODAR SeaSonde high frequency radar stations have been deployed in Galway Bay since summer in 2011 as shown in Fig. 1; they are located at Mutton Island Waste Water Treatment Plant (C1 in Fig. 1) and Spiddal Pier (C2 in Fi. 1). This radar system provides 300 metres horizontal resolution of surface currents data every hour.

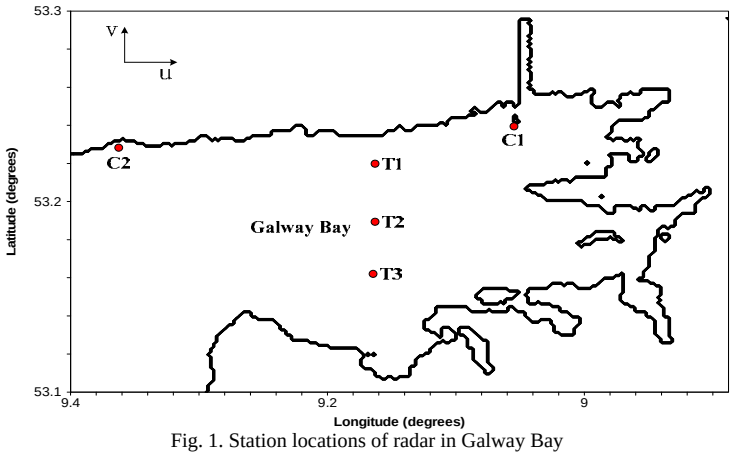


Fig. 1. Station locations of radar in Galway Bay

(C1 and C2 indicate radar stations; T1-T3 indicate three analysis locations)

3. Decision tree

A decision tree which describes the different classes or values that an output may take in terms of a set of input features or values [6]. Decision tree can be used to develop highly interpretable models that can effectively capture effects and nonlinearities [7]. Mahjoobi and Etemad-Shahidi [8] applied Classification And Regression Tree (CART) algorithm to build a Decision Tree model for predictions of significant wave heights. Based on comparisons of results from Artificial Neural Networks (ANNs) model, they found that the CART algorithm and ANNs algorithm had nearly similar error statistics. They concluded that CART algorithm can be successfully used to predict significant wave heights with an acceptable range of error. One significant advantage of decision tree is that rules are represented in contrast to ANNs. In this work, the decision tree CART algorithm which produces binary decision trees was used for predictions. Input features or values are dichotomously split from root node onto the bottom nodes (called "leaves") in a DT model. Each internal node for binary decision tree has exactly two outgoing edges from root node to bottom nodes: left child and right child [6]. A test function to be used to the incoming dataset was stored in each split node. Every node in a tree performs a test of some characteristics of the instance, and each branch descending from the node corresponds to one of the possible values for this characteristics [8]. The CART algorithm aims at yielding subsets of data which have the maximum homogeneity with respect to the target field. The least squared deviation (LSD) impurity measure is used for splitting rules and goodness of fit criteria for regression trees. The splitting process is recursively employed until one of the stopping rules is met.

The above split procedure performs in a recursively fashion until a stopping rule is reached. Construction of a CART tree is finished when all nodes can no longer be split and set as "terminal nodes". Once a CART tree is built, fitted values are sent to each terminal node based on inputs that land in that node. For regression case, the fitted value is a weighted average of responses values [7]. The most importantly practical attribute of CART algorithm is that its classification and regression structure in invariant with respect to monotone transformations of independent variables [2].

4. Results

In this work, authors focused on predicting surface currents at three spatial locations using available measurements from the CODAR system. The task can be divided into twofold steps: firstly, DT model was developed at each location for east-west and north-south surface velocity component separately. Secondly, the developed DT model was employed to generate predictions of surface current components.

Scatter plots between CODAR measurements and predictions from DT models are used for comparison. In addition, statistics: bias, RMSE, Scatter Index (SI) and correlation (R) were computed and used to evaluate the prediction performance.

In order to ensure the developed DT models be efficient and powerful, original datasets are categorized as three parts: training dataset accounting for 60%; testing data set accounting for 20% and forecasting dataset accounting for 20%. R package *"rpart"* was applied to develop DT models in this work.

4.1. Tests

First step is to develop training dataset to build up a DT model via tuning the CP parameter. Second step is to test model performance through comparing statistics i.e. RMSE here to determine the best value of CP. Last step is to employ the trained and tested DT model to prediction surface currents. DT models are separately developed for east-west (u) and north-south (v) velocity components at three different locations. In total, there are six DT models. Statistics—bias, RMSE, SI and R are calculated using the following equations.

$$bias = \bar{y} - \bar{x} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - x_i)^2}{n}} \tag{2}$$

$$SI = \frac{RMSE}{\bar{x}} \tag{3}$$

$$R = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y - \bar{y}_i)}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4}$$

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

Results of testing data are shown in Table 1. Table 1 shows that all developed DT models can generate good performance of surface velocity components with correlation greater than 0.84. Best results are obtained at location T3 for east-west velocity component. Correlation of east-west velocity component between the CODAR data and predictions from DT models is greater than values of north-south velocity components at the three analysis locations. This may due to the main flow trend in Galway Bay domain is east-west direction and magnitudes of east-west velocity components are generally greater than north-south velocity components.

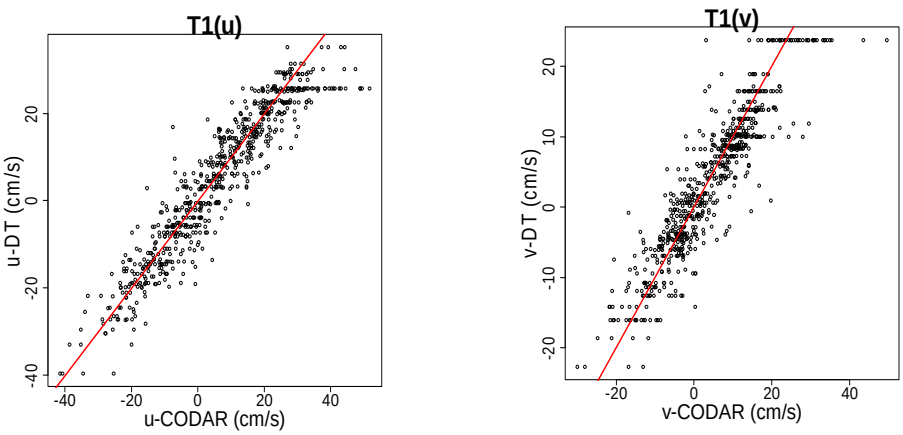
Table 1. Sensitivity tests of cp (cm/s)

cp	T1(u)	T2(u)	T3(u)	T1(v)	T2(v)	T2(v)
0.1	12.96	12.09	11.29	6.31	7.18	7.89
0.01	8.68	8.04	7.21	5.07	5.24	6.43
0.001	6.78	6.42	5.72	4.82	**5.12**	6.07
0.0001	6.20	5.68	**5.17**	**4.79**	5.16	**6.05**
0.00001	**6.18**	**5.67**	5.17	4.80	5.17	6.05
0.000001	6.18	5.67	5.17	4.80	5.17	6.05

Table 1 shows that the best CP value is 0.00001 for building up DT models to predict east-west components at T1 and T2. The best CP value is 0.0001 for developing DT models to predict east-west component at T3 and north-south components at T1 and T3. The best value of CP for north-south component DT model is 0.001.

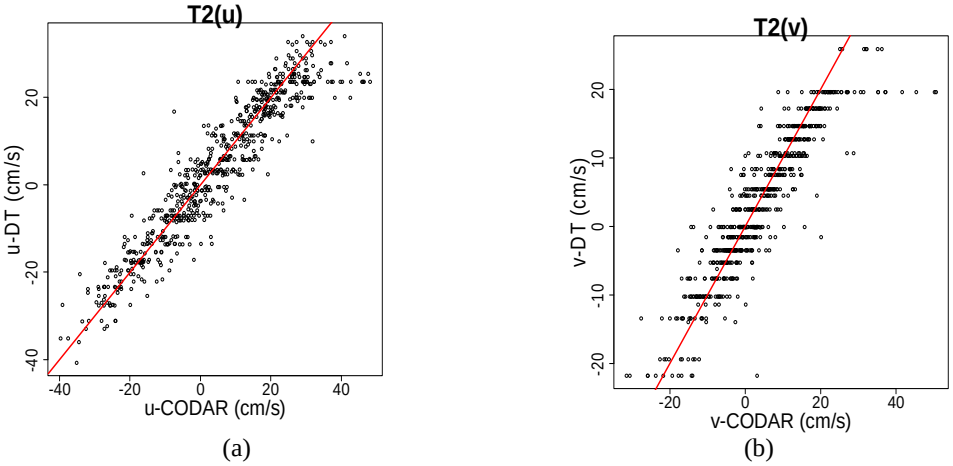
4.2. Prediction

Predictions of both surface velocity components using the best developed DT models are shown in Fig. 2- Fig. 4.



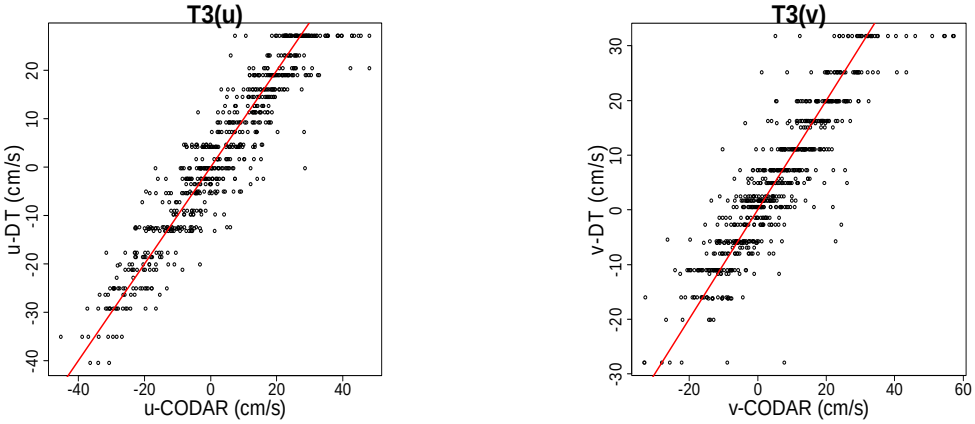
(a) (b)

Fig. 2. Predictions at T1



(a) (b)

Fig. 3. Predictions at T2



(a) (b)

Fig. 4. Predictions at T3

In order to assess performance of predictions from the DT models, statistics including bias, RMSE, Scatter Index (SI) and correlation (R) are computed and shown in Table 2.

Table 2. Statistics of predictions

Location	Variable	Bias (cm/s)	RMSE (cm/s)	SI	R
T1	u	-1.12	6.07	1.00	0.94
T1	v	-0.53	4.68	1.17	0.91
T2	u	-1.02	5.93	1.23	0.94
T2	v	-0.61	5.01	1.23	0.91
T3	u	-0.75	6.20	1.81	0.94
T3	v	-0.46	6.44	1.19	0.89

Table 2 shows that correlation between predictions from DT models and CODAR data is greater than 0.89 for both surface velocity components. This indicates that the developed DT models taking tidal water elevation, wind speed, wind direction and historical surface velocities over six hours can generate good forecasts. Correlation is very significant for east-west velocity component at the three analysis locations with 0.94. In general, correlation of east-west component is greater than north-south component at the three analysis locations.

5. Conclusions

A soft computing approach—Decision Tree was employed to predict surface currents in a ¼-scaled marine renewable test site—Galway Bay using the observations from the CODAR system for the first time. Wind speed, wind direction, tidal water elevation and historical observation of surface velocities were used as input variables and future states of surface currents were taken as output variables. Main conclusions are listed as follows.

(a). Performance of testing dataset and forecasting dataset shows that it is reasonable to take tidal water elevation, wind speed, wind direction and historical observation as input variables.

(b). The developed DT models can generate satisfactory results for both surface velocity components at three analysis location using forecasting dataset. Correlation between the CODAR data and outputs from DT models is greater than 0.89. Correlation of east-west velocity component is greater than north-south velocity component.

(c). Computational time of using DT model is much less than conventional hydrodynamic models. All the above models only need few minutes to produce predictions using R package. It is very useful to provide forecasting information for a specified location, such as tidal energy turbine test site.

(d). It is reasonable to use 60% data to develop DT models and 20% to test DT models based on strong correlation between the CODAR data and predictions from DT models.

(e). Decision Tree is a promising soft computing approach to predict surface velocities. They can provide accurate forecasting information, which is of great importance for exploring marine renewable energy and providing forecasting data for oil-spill treatment and search and rescue and so on.

Acknowledgements

This work is supported by the MaREI project. We would like to thank ECMWF for providing the wind data and OSU for providing the OTIS tide prediction software.

References

[1] T. Partal, H.K. Cigizoglu, E. Kahya, Stochastic Environmental Research and Risk Assessment, 29 (2015) 1317-1329.
[2] R. Timofeev, Center of Applied Statistics and Economics, Humboldt University, 2004, pp. 40.
[3] O. Makarynskyy, A.A. Pires-Silva, D. Makarynska, C. Ventura-Soares, Computers & Geosciences, 31 (2005) 415-424.
[4] D. Solomatine, Y. Xue, Journal of Hydrologic Engineering, 9 (2004) 491-501.
[5] J.D. Paduan, L. Washburn, Annual Review of Marine Science, 5 (2013) 115-136.
[6] A. Lahouar, J. Ben Hadj Slama, Energy Conversion and Management, 103 (2015) 1040-1051.
[7] J. Bleich, University of Pennsylvania, 2015, pp. 262.
[8] J. Mahjoobi, A. Etemad-Shahidi, Applied Ocean Research, 30 (2008) 172-177.