Correcting Predictions from Simulating Wave Nearshore Model via Gaussian Process Regression

1. Xiaoyu Zhang
College of Oceanography and Space Informatics
China University of Petroleum (East China)
Qingdao, China
zhangxiaoyuupc@163.com

2. Song Gao
North Sea Marine Forecast Center
of State Oceanic Administration
Qingdao, China
gaosong@bhfj.gov.cn

3. Tingwei Wang
Department of Electrical and
Electronic Engineering, University of
Surrey, Guildford, UK
tingwei.wang@surrey.ac.uk

4. Yongqing Li
College of Control Science and Engineering
China University of Petroleum (East China)
Qingdao, China
lyq_upc@163.com

5. Peng Ren
College of Oceanography and Space Informatics
China University of Petroleum (East China)
Qingdao, China
pengren@upc.edu.cn

Global Oceans 2020: Singapore - U.S. Gulf Coast | 978-1-7281-5446-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/IEEECONF38699.2020.9389333

***Abstract*—Accurately predicting wave height has a significant impact on offshore production and marine transportation. Numerical model predictions are the most commonly used wave height prediction method. But all numerical models are hard to describe the rules of ocean physics numerical model completely. Machine learning methods learn the inherent rules of data and are widely used in various predictions. In this paper, we try to use machine learning to correct errors in numerical model predictions. We count wave height numerical model predictions and observations for one year and find that the residuals of them are subject to a Gaussian distribution. Therefore, we propose a method for correcting wave height predictions from the Simulating Wave Nearshore (SWAN) model based on Gaussian process regression (GPR). To this end, we use residuals of wave height observations and the predictions of the SWAN model as input. We then train a GPR model for correcting the next time step wave height. Experimental results reveal that the proposed method predicts the wave height more accurately compared with the maritime numerical model.**

***Index Terms*—Gaussian process regression, wave height prediction, numerical model, residuals correcting.**

I. Introduction

Maritime disasters like tsunami and storm surges which are caused by typhoons, temperate cyclones, and undersea earthquakes threaten offshore activities [1]. Maritime disasters cause huge loss of life and property to coastal residents. Wave height is an important feature to monitor the development of tsunami and storm surge. Therefore, predicting ocean wave height accurately is essential for marine disaster warning in advance [2].

The wave height variation process is a highly complex non-linear stochastic system that is affected by many factors such as offshore wind field, bottom terrain, and human activities [3]. Wave height prediction is a hot and difficult issue in marine science. One prediction method is on the basis of numerical models which consider wave generation, propagation, and dissipation. SWAN model [4] is a widely used wave spectrum model based on energy conservation and is suitable for different regions because it considers the offshore physical processes such as bottom friction and submarine topography. However, the model is constrained by fixed equations of ocean dynamics and it is hard to make accurate predictions in a complex marine environment.

Machine learning is an effective way of learning the internal rules of data [5]. Therefore, many machine learning methods have been used for wave height prediction. Mandal *et al.* [6] proposed a recurrent neural network with backpropagation algorithm for wave forecasting. James *et al.* [7] developed a multi-layer perceptron (MLP) framework to estimate wave height and period. The prediction accuracy of MLP framework is comparable to the SWAN model. However, the training model of machine learning method is uninterpretable. When faced with different marine environments, the data adaptation of machine learning methods is not stronger than the numerical model methods.

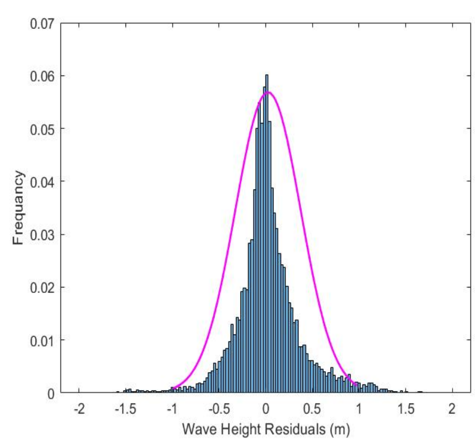


Fig. 1: Residuals distribution between wave height observations and SWAN model predictions.

978-1-7281-5446-6/20/$31.00 ©2020 IEEE

To address the aforementioned limitations, we develop a method based on machine learning to correct the predictions of the SWAN model. Fig. 1 is a histogram of the residuals between the SWAN model predictions and the wave height observations. The pink curve represents the normal distribution curve which fitted by the mean and variance of wave height residuals. From the histogram and the fitted normal distribution curve, we notice that the wave height residuals are subject to a Gaussian distribution. Therefore, GPR [8] is able to predict wave height residuals. In our work, we train a GPR model to predict the residuals between the wave height predictions and the wave height observations for correcting the SWAN model. Experimental results show that our method improves the accuracy of the wave height predictions of the SWAN model effectively.

II. Related Work

A. Simulating Wave Nearshore Model

The SWAN model is a discrete spectral numerical model of wind and waves in shallow offshore waters [9]. The model is based on the action balance equation as follows:

$$\frac{\partial P}{\partial t}+\frac{\partial c_a P}{\partial a}+\frac{\partial c_b P}{\partial b}+\frac{\partial c_\gamma P}{\partial \gamma}+\frac{\partial c_\theta P}{\partial \theta}=\frac{U}{\gamma}. \quad (1)$$

Here P denotes the action density and t denotes time, a, b denotes the components of the geospatial coordinates . γ and θ denotes the relative frequency and wave direction. c_a, c_b, c_γ and c_θ are the propagation velocity of the action density in the a-direction, b-direction, γ-space, and θ-space, respectively. The left-hand side of the equation (1) reveals the kinematic component of the wave. U is the source function expressed in spectral density. The source function contains many physical processes such as wave growth by wind, wave-wave interactions, the dissipation with regard to wave breaking and wave decay caused by bottom friction or white capping.

We solve the equation (1) by fully implicit finite difference and obtain the action density P. Then the wave height prediction S is calculated as follows:

$$S=4(\int_0^\infty \int_{-\pi}^{\pi} P\gamma \mathrm{d}\gamma \mathrm{d}\theta)^{\frac{1}{2}}. \quad (2)$$

The SWAN model is an extension of the third generation wave numerical model and accurately represents the various physical processes of the wave generation. The model is suitable for predicting the wave height and period in coastal, lakes or other offshore areas [10].

B. Gaussian Process Regression

For prediction problems, the training data is mostly nonlinear. GPR is an effective method for dealing with nonlinear problems [11]. For N data pairs $\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$, we assume that training data is subject to a Gaussian distribution. When x_{n+1} is given, n+1 data is also subject to a joint Gaussian distribution. Therefore, the GPR method gets the value of y_{n+1}. GPR has a strictly theoretical basis for statistical learning. It has good adaptability and strong generalization ability to deal with complex problems such as high dimensions, small samples, and nonlinearity. Compared with the neural network method, GPR has fewer parameters and is easier to implement, so it is widely used in various fields [12] [13].

GPR for time series prediction is mainly divided into two categories [14]. The first is to use time t as the independent variable to generate a GPR model. The second is to build input vectors from time series. The first method requires the construction of complex kernel functions, and we use a data-driven method for wave height residual prediction.

III. Correcting SWAN Model Predictions Via GPR

Fig. 2 shows the implementation process of our wave height correcting method. In this paper, we define O_t and S_t as wave height observations and predictions for each time t. O_t are obtained by an offshore buoy and S_t are obtained by the equation (2). The wave height residuals r_t is given as follows:

$$r_t=O_t-S_t. \quad (3)$$

For the residual learning phase, we use a sequential of N historical residuals $x_t=[r_t,r_{t+1},\cdots,r_{t+N-1}]$ to predict the residual r_{t+N}. Then, we construct a residual feature vector $X=[x_1,x_2,\cdots,x_t]^{\mathrm{T}}$, where T is the transpose operation. The training model takes the feature vector X as the input and generates a residual prediction sequential $R=[r_{N+1},r_{N+2},\cdots,r_{t+N}]^{\mathrm{T}}$ as the output. We assume that the training data is subject to Gaussian distribution which is formulated as follows:

$$R=f(X)\sim\mathcal{N}(0,K). \quad (4)$$

Here K is a covariance matrix formulated as follows:

$$K=\begin{bmatrix} k(x_1,x_1) & k(x_1,x_2) & \cdots & k(x_1,x_t)\\ k(x_2,x_1) & k(x_2,x_2) & \cdots & k(x_1,x_t)\\ \vdots & \vdots & \ddots & \vdots\\ k(x_t,x_1) & k(x_t,x_2) & \cdots & k(x_t,x_t)\end{bmatrix}, \quad (5)$$

where $k(x,x')$ denotes a covariance function. In our work, we choose Gaussian kernel function as a covariance function:

$$k(x,x')=\sigma_f^2\exp\left[\frac{-(x-x')^2}{2l^2}\right], \quad (6)$$

where σ_f^2 denotes maximum allowable covariance. If $x=x'$, then $k(x,x')$ approaches this maximum. If x is distant from x', we have instead $k(x,x')=0$. l is a parameter that controls the degree of separation.

In the prediction phase, according to (4), when giving a residual prediction sequence $x_\tau=[r_{\tau-N},r_{\tau-N+1},\cdots,r_{\tau-1}]$, we get the joint distribution of training data R and predicting result which is estimated r_τ as follows:

$$\begin{bmatrix} R\\ r_\tau\end{bmatrix}\sim\mathcal{N}\left(0,\begin{bmatrix} K & K_\tau^{\mathrm{T}}\\ K_\tau & K_{\tau\tau}\end{bmatrix}\right), \quad (7)$$

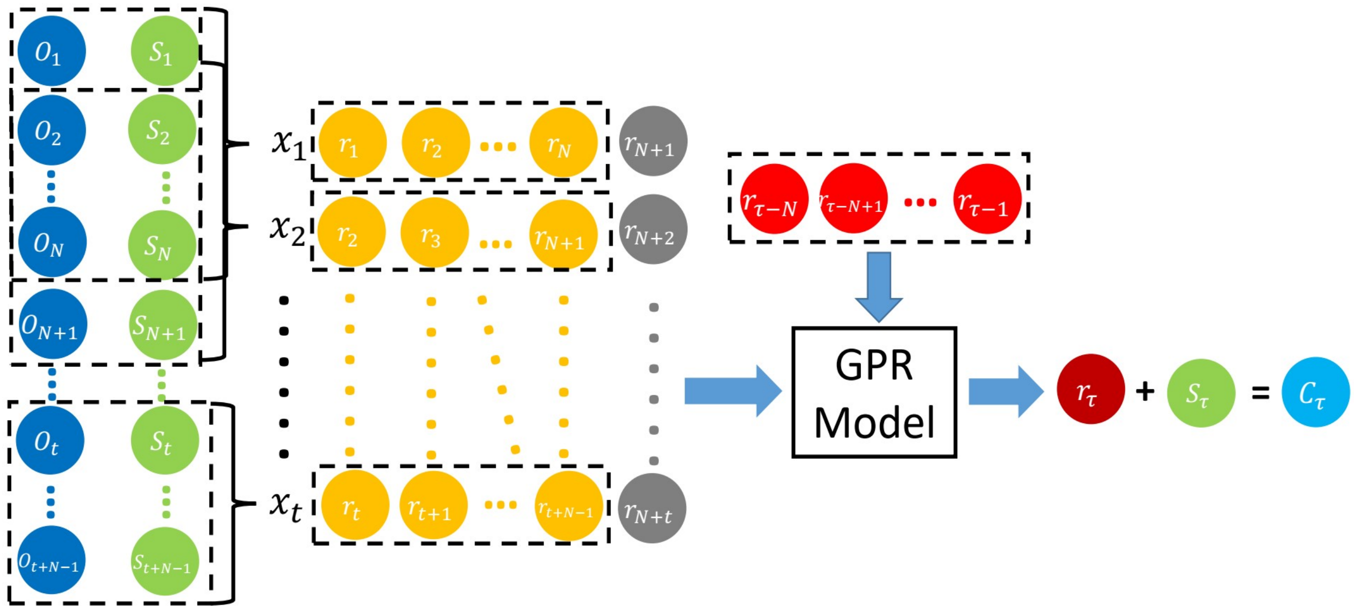


Fig. 2: The framework of correcting wave height residuals by GPR.

where $K_\tau = [k(x_\tau, x_1), k(x_\tau, x_2), \cdots, k(x_\tau, x_t)]$, $K_{\tau\tau} = k(x_\tau, x_\tau)$. According to the bayes formula, the prediction equation of GPR is written as follows:

$$p(r_\tau | R, X, x_\tau) = \mathcal{N}(K_\tau K^{-1} R, K_{\tau\tau} - K_\tau K^{-1} K_\tau^{\mathrm{T}}), \quad (8)$$

where $K_\tau K^{-1} R$ is the mean of the prediction equation, i.e. the residual prediction of GPR. The corrected SWAN model prediction C_τ is given as follows:

$$C_\tau = S_\tau + r_\tau. \quad (9)$$

We put a sequence of N wave height residuals into the trained GPR model and obtain a predicted wave height residual for the (N+1)th time step. But the GPR method only corrects one residual for once prediction. If correcting 3 hours or more time steps wave height residuals are needed, we add (N+1)th time step prediction in the sequence and get rid of the first wave height residual. Then we can obtain the predicted residual for the (N+2)th time step. The prediction process continues until the required time is reached. Finally, we add the residuals to wave height predictions from the SWAN model by (9).

IV. Experimental Results

To validate the effectiveness of our proposed model, we test our method on the predictions from the SWAN model [4] and the observations which are collected from a buoy in Bohai Sea, China. The SWAN model predictions and the observations are lasting for one year. The minimum time step for prediction is one hour. We compare the performances of our method with the extreme learning machines(ELM) method [15] and correct the predictions over a range of three hours, six hours, twelve hours and twenty-four hours.

TABLE I: Prediction accuracy improvement.

Skill Score		Time range for prediction (hour)			
$\rho(\%)$		3h	6h	12h	24h
Method	ELM residual learning	37.54	58.18	55.47	39.47
	GPR residual learning	46.20	75.54	64.58	55.11

We adopt a skill score to evaluate the relative improvement of our proposed method. The skill score ρ is defined as follows:

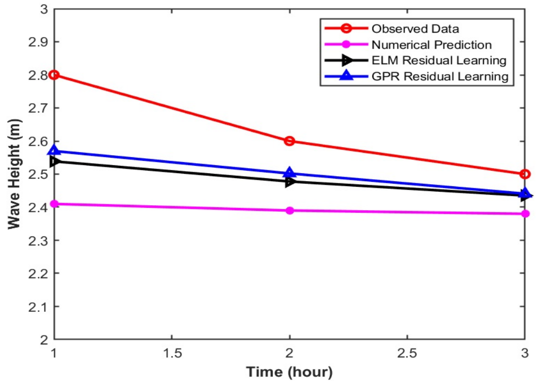
$$\rho = \frac{e_a - e_b}{e_a} \times 100\%, \quad (10)$$

where e_a is the mean absolute error between SWAN model predictions and wave height observations. e_b is the mean absolute error between correcting wave height predictions and wave height observations.

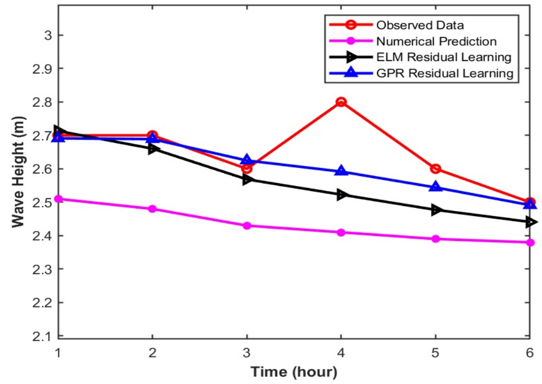
As is illustrated in Fig. 3, both two methods improve the prediction accuracy of the SWAN model and the GPR residual learning results are closer to the observations. In addition, we compute the prediction accuracy improvement of ELM method and GPR method. The results are given in Table.I. Table.I shows that the effect of GPR method is better than ELM method. Furthermore, the predictions of our method are more stable than ELM method because the weights and biases which are needed in ELM method are generated randomly.

V. Conclusion

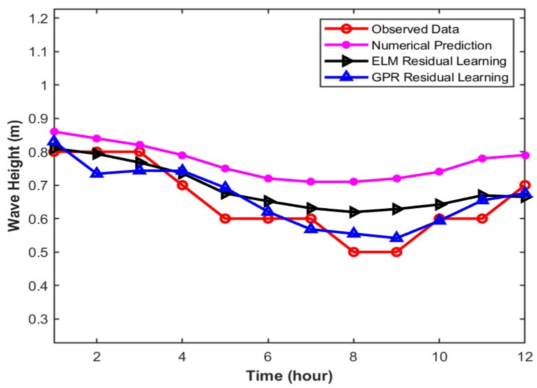
In this paper, we develop a wave height correction method that combines numerical model and machine learning. According to the discovery that the residuals between the SWAN model predictions and the observations are subject to a



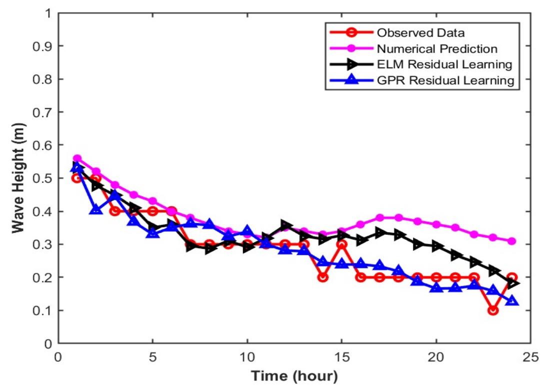
(a) Three hours



(b) Six hours



(c) 12 hours



(d) 24 hours

Fig. 3: Wave height predictions from the SWAN model and two residual correction methods.

Gaussian distribution, we use GPR as a training model to predict wave height residuals. In contrast to the existing correcting methods which are based on the statistical property of historical observations or just learn an end-to-end mapping between observations and errors. Our method makes full use of ocean dynamic equations and the laws of data. Compared with the ELM method, experimental results show that our method corrects the wave height predictions better. However, the method we proposed is difficult to capture the abrupt change of wave height accurately. In the future work, we will pay more attention to the abrupt changes for wave height prediction.

References

[1] S. C. Yim, "Modeling and simulation of tsunami and storm surge hydrodynamic loads on coastal bridge structures," in *21st US-Japan Bridge Engineering Workshop*, 2005, pp. 3–5.

[2] A. Zamani, D. Solomatine, A. Azimian, and A. Heemink, "Learning from data for wind–wave forecasting," *Ocean engineering*, vol. 35, no. 10, pp. 953–962, 2008.

[3] I. R. Young and A. Ribal, "Multiplatform evaluation of global trends in wind speed and wave height," *Science*, vol. 364, no. 6440, pp. 548–552, 2019.

[4] W. E. Rogers, J. M. Kaihatu, L. Hsu, R. E. Jensen, J. D. Dykes, and K. T. Holland, "Forecasting and hindcasting waves with the swan model in the southern california bight," *Coastal Engineering*, vol. 54, no. 1, pp. 1–15, 2007.

[5] E. Alpaydin, *Introduction to machine learning*. MIT press, 2014.

[6] S. Mandal and N. Prabaharan, "Ocean wave forecasting using recurrent neural networks," *Ocean engineering*, vol. 33, no. 10, pp. 1401–1410, 2006.

[7] S. C. James, Y. Zhang, and F. O'Donncha, "A machine learning framework to forecast wave conditions," *Coastal Engineering*, vol. 137, pp. 1–10, 2018.

[8] C. K. Williams and C. E. Rasmussen, "Gaussian processes for regression," in *Advances in neural information processing systems*, 1996, pp. 514–520.

[9] N. Booij, L. Holthuijsen, and R. Ris, "The" swan" wave model for shallow water," in *Coastal Engineering 1996*, 1997, pp. 668–676.

[10] W. E. Rogers, P. A. Hwang, and D. W. Wang, "Investigation of wave growth and decay in the swan model: three regional-scale applications," *Journal of Physical Oceanography*, vol. 33, no. 2, pp. 366–389, 2003.

[11] T. Mutanen, L. Sirro, and Y. Rauste, "Tree height estimates in boreal forest using gaussian process regression," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2016, pp. 1757–1760.

[12] D. Wang, Y. Wu, and Z. Xiao, "A gaussian process regression method for urban road travel time prediction," in *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, July 2017, pp. 890–894.

[13] K. Blix, G. Camps-Valls, and R. Jenssen, "Sensitivity analysis of gaussian processes for oceanic chlorophyll prediction," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 996–999.

[14] S. Brahim-Belhouari and J. M. Vesin, "Bayesian learning using gaussian process for time series prediction," in *Proceedings of the 11th IEEE Signal Processing Workshop on Statistical Signal Processing*, Aug 2001, pp. 433–436.

[15] T. Wang, S. Gao, J. Xu, Y. Li, P. Li, and P. Ren, "Correcting predictions from oceanic maritime numerical models via residual learning," in *2018 OCEANS - MTS/IEEE Kobe Techno-Oceans (OTO)*, May 2018, pp. 1–4.