

## Short-Term Forecasting for Harbor Waterway Currents Speeds

Cheng Gong<sup>1</sup>, Yan Lv<sup>1</sup>, Chunjiang Zhang<sup>2</sup>, Xiyuan Wang<sup>1</sup>, Wei Huangfu<sup>1\*</sup> and  
Zhongshan Zhang<sup>1</sup>

<sup>1</sup>*Beijing Engineering and Technology Research Center for Convergence Networks  
and Ubiquitous Services,*

*University of Science and Technology Beijing (USTB), Beijing 100083, China*

<sup>2</sup>*Qinhuangdao Beacons, Tianjin Maritime Safety Administration of the People's  
Republic of China*

*Email: cgong1986@gmail.com, lvyuan159357@126.com, 13303352526@163.com,  
xiyuanwang.ee@gmail.com, huangfuwei@ustb.edu.cn, zhangzs@ustb.edu.cn*

### **Abstract**

*The ocean currents speeds in the harbor waterway are directly related to the ability of the ship to in or out the harbor. Accurately predict the speeds can assist the ship to choose the right time for sailing. To solve this problem, we chose two models of linear and non-linear prediction. We had set sensors in Qinhuangdao for a long time, then using the collected data for training. Our test is using a lot of random data to train and predict with different steps and orders. The results show that both methods can use less original data to train the model, and finally achieve preferably prediction. According to the characteristics of Qinhuangdao harbor, Auto-Regressive (AR) model is more appropriate than Support Vector Regression (SVR) model.*

**Keywords:** *Currents speeds prediction; AR; SVR; Short-term forecasting*

### **1. Introduction**

Ocean accounts for 71% of the earth's surface, is an important component of the global life-support systems. Meanwhile, ocean is valuable space, which can help human to achieve sustainable economic and social development. Currently 19% of the global bulk cargo is shipping to China, 20% of container traffic from China. While among the new bulk cargo marine transportation, there are 60 to 70 percent is shipped to China. China's harbor cargo throughput and container throughput already ranked first in the world. However, our harbor can only complete 60% of the actual cargo throughput [1].

Harbor waterway is the channel in the harbor waters for safe navigation of ships. The speeds and directions of ocean currents in waterways will affect the traffic capacity of the ships. In order to ensure the safety and effectiveness of harbor traffic, it is necessary for monitoring ocean currents in waterways. The traditional method for determining the time of entering or leaving the harbor is by human judgment. Experienced sailors can make accurate judgments. Navigation-aid Information System can get real-time currents information via underwater sensors [2]. Monitoring waterway status can improve the safety of navigation. An accurate and reliable hydrological information prediction is one of the most effective solutions to schedule the transportation in harbor.

Notwithstanding, finding a statistically appropriate model for all hydrological information time series is not straightforward. Our focus here is on speeds of currents rather than the other parameters. In general, the water flow forecasting methods include Back Propagation (BP)

neural network, Auto-Regressive and Moving Average Model (ARMA), Support Vector Machine (SVM), *etc.*, [3-6] to solve practical problems in specific region, this paper compares two models Auto-Regressive (AR) and Support Vector Regression (SVR) for currents-speeds forecasting. The effectiveness of the models is verified with the real output of the harbor waterway in Qinhuangdao harbor.

The outline of this paper is as follows. Section 2 reviews related works. Section 3 describes the forecasting algorithm. Section 4 is devoted to the presentation of AR and SVR models experiments on original data. Section 5 provides a conclusion and outlook.

## 2. Related Works

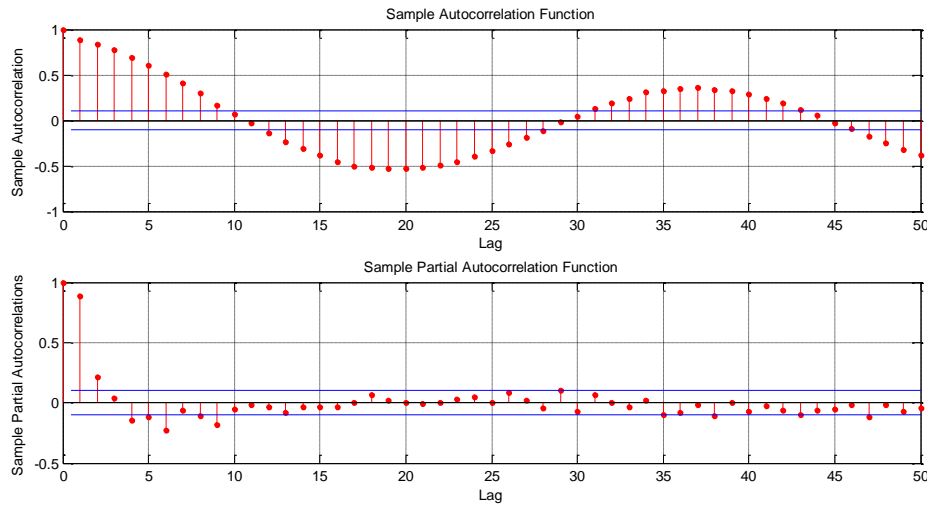


**Figure 1. The Data Collection System**

Our data collection system is mainly as shown in Figure 1. The farthest part is a sensor buoy, which consist of power supply, various sensors, beacons, and also satellite communication modules, working in harbor waterways for day and night navigation aids. The currents speeds, sampled every ten minutes at 15 different depths (1.5, 2.5, 3.5 ... 15.5 meters), was acquired by Nortek Aquadopp Profiler measuring instrument that can make supervision in multiple currents layers. The Beidou satellite system will transmit the data to the ground data center, which was collected by sensor buoy. Currents speeds were predicted in the data center, so users can query through that database.

## 3. Short-term Forecasting

In most situations, the autocorrelation and partial autocorrelation contain the sequential characteristics of the time series needed for forecasting. So, the two statistic characteristics are analyzed at first, and the result is shown in Figure 2. With the increase of the Lag, the autocorrelation is tailing and oscillating attenuation. Whereas, under the assumption that the value between the two blue lines (too small) can be recognized zero, the partial correlation can be considered censored when Lag equals 6. Therefore, AR (p) can be used in terms of the linear model to predict.



**Figure 2. The Autocorrelation and Partial Autocorrelation of Data**

### 3.1. Auto-Regressive

Auto-Regressive (AR) model is a traditional linear prediction method which incorporate into forecast only past and present values, Its basic idea is that the variation regular of past will continue into future. The standard structure of the r-order AR model can be written as [7]:

$$X_t = C + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t$$

Where C is the constant term,  $\varepsilon_t$  is a random error with the mean equaling 0, and standard deviation equaling  $\sigma$  which is supposed not change whatever the t is.  $\varphi_i$  is the linear prediction coefficient and can be estimated in real time.  $X_{t-1}$ ,  $X_{t-2}$ ,  $X_{t-i}$  are present and past observations. The output  $X_t$  is the linear combination of the past and present sampled values.

- 1) AR has the advantage of less information and predicting using variable sequence of them, but when using AR model, the following points need paying attention to:
- 2) AR is applicable to predict stationary series when the observations have a strong correlation with prophase of themselves, once affected greater by other variables, the prediction will be not ideal.
- 3) Autocorrelation must existed, if the coefficient of autocorrelation is less than 0.5, AR is not suitable, otherwise, the forecast result will be highly inaccurate
- 4) For AR(p) model, with the increase of k, the autocorrelation function is trailing, tending to zero, partial correlation is censored, namely, when  $k=p$ , there is a censored point.

### 3.2. Support Vector Regression

Support Vector Machines and Support Vector Regression (SVR) are based on statistical learning theory, or VC theory (VC - Vapnik, Chervonenkis), developed over the last several decades [8]. Besides solving classification problems SVM can also be used to deal with regression problems, it is called SVR. In regression, each instance corresponding to a label is a continuous real number rather than discrete categories.

Prediction functions for non-linear regression:

$$f(x) = (w \cdot \phi(x)) + b$$

$x(t)$  means a time series,  $t = \{0, 1, 2, \dots, N - 1\}$ ,  $\phi(x)$  is to map  $x(t)$  into a higher dimension space.

Regularized risk function:

$$R_{\text{reg}}(f) = R_{\text{emp}}(f) + \frac{\lambda}{2} \|w\|^2 = \sum_{i=1}^l C(f(x_i) - y_i) + \frac{\lambda}{2} \|w\|^2$$

$C$  denote cost,  $\lambda$  denote a regularization constant. In the higher space, we can possibly deal with the problem using linear methods. So the question changed into:

$$\min \frac{\lambda}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i$$

Subject to:

$$y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0; i = 1, 2, \dots, l; C > 0$$

Based on the Karush-Kuhn-Tucker conditions [9] that:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) (\phi(x_i) \cdot \phi(x)) + b$$

In the high dimensional space,  $\phi(x_i)$  can be replaced by  $K(x_i, x)$ , which called kernel functions.

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b$$

Some popular kernels are:

1) radial basis function (RBF) kernel:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

2) polynomial kernel:

$$K(x_i, x_j) = (1 + x_i \cdot x_j)^d$$

$\gamma$  and  $d$  are parameters.

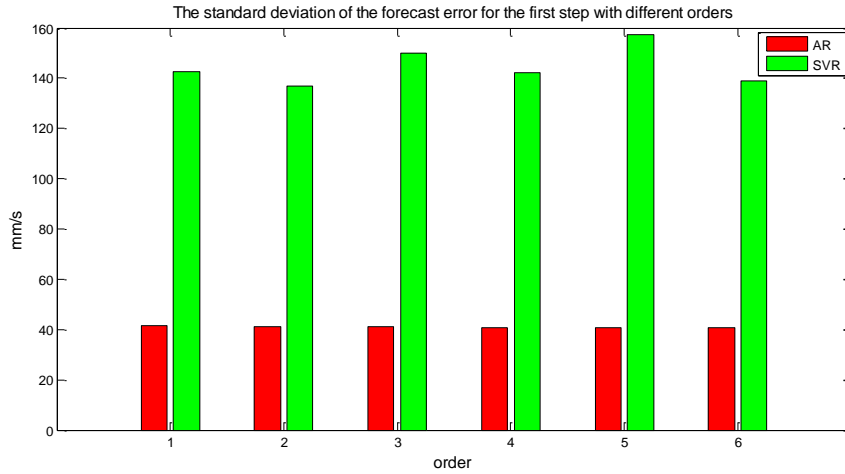
## 4. Experiments

In this paper, we use MATLAB Version R2012b as the experiment platform. A toolbox named Libsvm is used to implement training and testing on the sample data. In the following, we will use the same data for the two prediction methods, AR and SVR. Especially for the SVR, it's better to change the historical data between  $[-1, +1]$ .

In this study, the currents data we used was from 2009 of Qinhuangdao harbor underwater at 6.5 meters depth. Practically, the data at the sampled point may be lost sometimes, and we replace the null value with the previous one. In order to fully reflect the general results of experiment, we randomly selected 400 points in all of the data among 4 months and then the test was repeated 1000 times. The test is about different orders and different steps. Orders means use how many historical data to predict an unknown data. And the steps indicate to predict data after how much time interval. Give an example, 3-orders-4steps means use  $\text{data}_{t-3}$ ,  $\text{data}_{t-2}$ ,  $\text{data}_{t-1}$  to predict  $\text{data}_{t+3}$ . SVR model is selected RBF kernel function,  $\gamma = 2.8$ ,  $\varepsilon = 0.01$ ,  $C = 2.2$ . After the prediction, do inverse transform to the normalized data then calculate standard deviation.

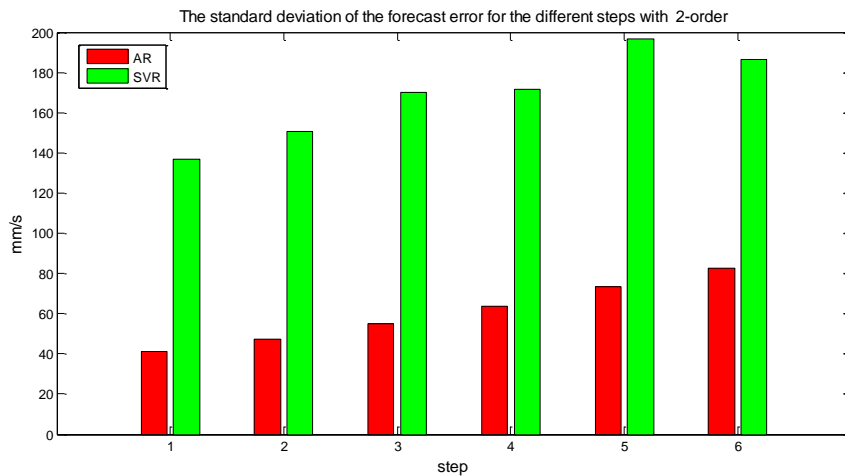
After comparing the 72000 sets experimental results, we selected the two most representative charts. Obviously, the closer the prediction step is, the higher the accuracy will

be got. By search experimental results also verify that, the prediction of 1-step have the smallest standard deviation.



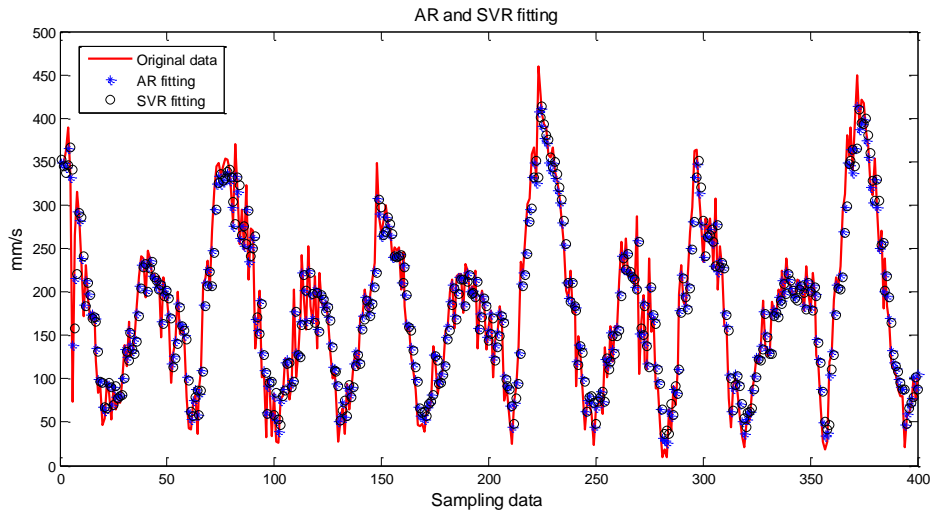
**Figure 3. 1-step Prediction Result**

Figure 3 shows the standard deviation of AR model and SVR model with different orders. For SVR model, even order prediction slightly less than the odd order. But AR model has almost the same results. Another significant result is that standard deviation of AR model is only one-third of the SVR model.



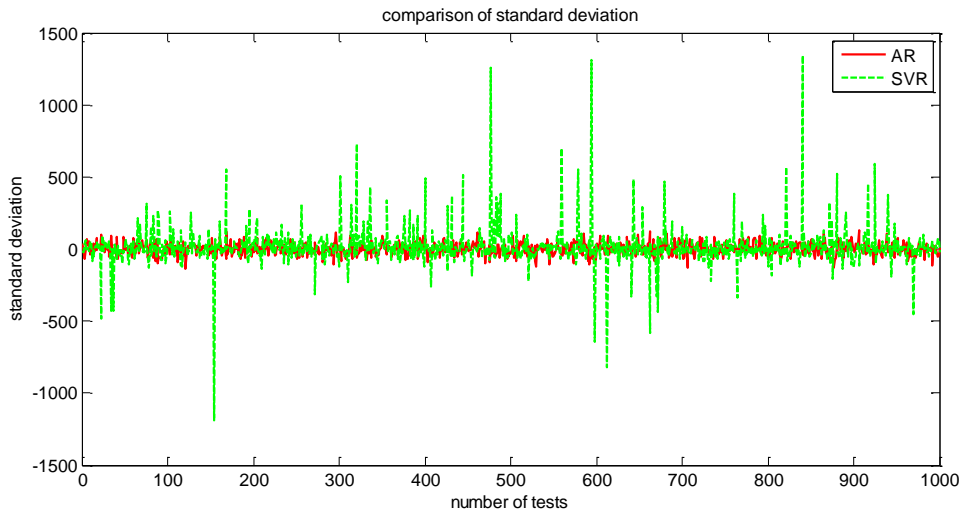
**Figure 4. 2-order Prediction Result**

Then select 2-order to contrast the steps, in that order the SVR model has the smallest error. Both models have the same trend that, with the growth of step, standard deviation is correspondingly larger. Figure 4 also reveals that 1-step prediction is the best way to use.



**Figure 5. Fitting Comparison of the Two Methods**

Figure 5 illustrates the fitting comparison of the two methods. In general, both methods are able to achieve better prediction. In contrast, the performance of AR model is better. Using 2-order-1step standard deviation as an example, this shows in Figure 6. Standard deviation of AR model is more stable. On the contrary, SVR model will occasionally produce large errors.



**Figure 6. 2-order-1step Standard Deviation**

## 5. Conclusion and Outlook

The main conclusions are shown as follows.

- 1) Both methods can use less original data to train the model, and finally achieve good prediction.
- 2) Performance of the AR model shows more stable and more accurate than SVR model.
- 3) The prediction for small amounts of short-term data, which like the speeds of currents, should consider using a linear model.

In the future, the research will be extended to the forecasting of directions and speeds of the currents in harbor waterway. Simultaneously, we are going to use existing data to make predictions for more complex data, such as visibility.

## Acknowledgments

This work was supported by the IOT (Internet of Things) Fund of Ministry of Industry and Information Technology, China (No.12121002B) “Information Acquirement and Publish System of Shipping Lane in Harbor”, the fund of Beijing Science and Technology Commission “Network Monitoring and Application Demonstration in Food Security”, the National Natural Science Foundation of China (No. 61172050) the Program for New Century Excellent Talents in University (NECT-12-0774) and the China Postdoctoral Science Foundation (2013M530528).

The corresponding author is Wei Huangfu.

## References

- [1]. X. Li and W. Ren, “Trend Characteristic and Apocalypse of the Foreign Port Logistics Development”, China Port, vol. 02, (2007), pp. 27-28+43.
- [2]. X. Zhang and S. Zhou, “Realization of Integrated Navigation-aid Information System”, Journal of Shanghai Maritime University, vol. 01, (2013), pp. 15-18
- [3]. Z. Yang and L. Kang, “Forecasting and Trend Analysis Method for Danjiangkou Reservoir Runoff”, Water Resource and Power, vol. 02, (2011), pp. 1-3.
- [4]. S. Dong, “Current Prediction Research Based on ARIMA-BP Neural Network”, China Science and Technology Information, vol. 02, (2014), pp. 86-88.
- [5]. J. Xiong and Z. Li, “Forecasting Sand Carrying Capacity of Water based on Support Vector Machine”, ShuiLi XueBao, vol. 10, (2005), pp. 1171-1175.
- [6]. V. N. Vapnik, “The Nature of Statistical Learning Theory”, Springer-Verlag, New York, (1999), pp. 156–160.
- [7]. Z. Huang and Z. S. Chalabi, “Use of Time-series Analysis to Model and Forecast Wind Speed”, Journal of Wind Engineering and Industrial Aerodynamic, vol. 56, (1995), pp. 311-322.
- [8]. N. I. Sapankevych and R. J. Sankar, “Time Series Prediction Using Support Vector Machines: A Survey”, Computational Intelligence Magazine, IEEE, vol. 4, no. 2, (2009), pp. 24-38.
- [9]. Y. Liu, J. Shi, Y. Yang, *et al.*, “Short-Term Wind-Power Prediction Based on Wavelet Transform–Support Vector Machine and Statistic-Characteristics Analysis”, Industry Applications, IEEE Transactions, vol. 48, no. 4, (2012), pp. 1136-1141.
- [10]. F. Fusco and J. V. Ringwood, “Short-term Wave Forecasting with AR Models in Real-time Optimal Control of Wave Energy Converters”, Industrial Electronics (ISIE), IEEE International Symposium on IEEE, (2010).
- [11]. E. Mangalova and E. Agafonov, “Time Series Forecasting Using Ensemble of AR Models with Time-varying Structure”, IEEE conference on evolving and adaptive intelligent systems (EAIS), (2012).
- [12]. M. Martin, L. V. Cremades and J. M. Santabarbara, “Analysis and Modelling of Time Series of Surface Wind Speed and Direction”, International journal of climatology, vol. 19, no. 2, (1999), pp. 197-209.
- [13]. P. J. Brockwell and R. A. Davis, “Introduction to Time Series Analysis and Forecasting”, People Post Press, Beijing, (2009).
- [14]. P. J. Brockwell and R. A. Davis, “Time Series: Theory and Methods”, Springer-Verlag, New York, (2001).
- [15]. J. Contreras, R. Espinola, F. J. Nogales and A. J. Conejo, “ARIMA models to predict next-day electricity prices”, Power Systems, IEEE Transactions on, vol. 18, no. 3, (2003), pp. 1014-1020.
- [16]. J. Contreras, R. Espinola, F. J. Nogales and A. J. Conejo, “ARIMA models to predict next-day electricity prices”, Power Systems, IEEE Transactions on, vol. 18, no. 3, (2003), pp. 1014-1020.

