Predict Maneuvering Indices Using AIS Data by Ridge Regression

Hao Rong[1,2], Jun-min Mou[1,3]
[1] School of Navigation, Wuhan University of Technology, Wuhan, China, email: ronghao1230@hotmail.com
[2] Hubei Key Laboratory of Inland Shipping Technology, Wuhan, China
[3] Key Laboratory of High Performance Ship Technology of Ministry of Education (Wuhan University of Technology) , Wuhan, China, email: moujm@whut.edu.cn.

***Abstract*—Since Automatic Identification Systems (AIS) was widely used in navigation industry, it has not only greatly benefited ship operations to keep encountered vessels visible to each other, but also fosters the relevant researches. Because AIS message contain vessel's identity, position, course, speed and so on, the application of AIS data make it possible to investigate accurate and actual behaviors of ships during ship maneuvering. The ridge regression is presented for identify maneuvering indices in Nomoto's model, and the result indicates that the method is robust and does not rely on initial estimation. For selecting appropriate AIS data for maneuvering indices predicting, a frequency domain identification method is presented. Integrated with the predicted indices, the ship maneuvering mathematical model can be used to simulate the vessel traffic flow and assess the risk of collision.**

Keywords—Ship Maneuvering; Nomoto model; AIS data;

1. INTRODUCTION

The Automatic Identification System (AIS) for ships was introduced as a real-time system augmenting ship-borne radar to aid ship collision avoidance. It is compatible to the Vessel Traffic Services, so that it is a powerful tool for vessel traffic observation. Due to ship traffic with position and speed measurements from uniquely identified ship maneuvering, the AIS data available from the system can then be used to predict maneuvering indices by System Identification (SI) method.

The area of ship maneuvering has been extensive application of a variety of SI methods. Some of the established SI methods in this area are indirect model reference adaptive systems (Van Amerongen 1982), extended Kalman filter (Abkowitz 1980), neural networks (Haddara & Wang 1999). These methods focused on ship control problems. To various levels of accuracy, these derivatives are usually obtained from model tests, empirical and semi-empirical formulae, formulae based on statistical regression analysis, and computational fluid dynamics (CFD) tools. The model test route requires carrying out a large number of sophisticated experiments in such laboratories as towing tanks with planar motion mechanism (PMM), or rotating arm system. Of the three test facilities, the latter two, namely, the PMM and the rotating arm system are very costly. The CFD tools are also costly and time consuming and require highly qualified personnel. Thus, SI is an interesting alternative to estimate these indices directly from either free-running model tests or full-scale maneuvering trial data. The objective of this paper is to present an SI method for estimation of the maneuvering indices based on AIS data.

Since AIS is wireless connection between ships or ships with shore stations, the reliability of AIS message is disputable. To investigate the quality of the data in AIS message, some surveys have been undertaken. By far the most errors occurred in the categories are: Destination and Draught. Errors in Destination include misspelling, empty data fields, incomprehensible abbreviations and references to the previous port. By analysis (Harati Mokhtari et al. 2007), errors in AIS message are ranging from data corruption, erroneous MMSI number, target swaps, faulty position reports and errors in rate-of-turn data. Especially, rate of turn and rudder angle are essential to predict maneuvering indices in this research. However, rudder angle is not included in AIS data and some cases rate of turn only indicates positive or negative turn. (Hasegawa, 2012) calculated rate of turn by the difference of two previous values of heading angle, and Quasi-Newton method is presented for optimize the K-T indices.

Inspired by Hasegawa's idea, the author apply ridge regression for identify maneuvering indices in Nomoto's model. And in order to select appreciate data for study a frequency domain method is presented.

PREDICT MANEUVERING INDICES BY RIDGE REGRESSION

The identification study will be based on the Nomoto transfer function. Nomoto's first order model is the simplest mathematical model to describe ship manoeuvres, and the model is a compromise between the demand for a simple mathematical model and a fair approximation of the actual ship maneuvering.

The discretization of Nomoto's first order model

The first order model of Nomoto's can be expressed as Eq. 1.

$$T*\ddot{\psi}+\dot{\psi}=K*\left(\delta-\delta_r\right) \tag{1}$$

Where:

$\ddot{\psi}$ ---- Yaw angular acceleration

$\dot{\psi}$ ---- Yaw angular velocity or rate of turn (deg/s2 or rad/s2)

ψ ---- Yaw angle (deg/s or rad/s)

δ ---- Actual rudder angle (deg or rad)

$\delta-\delta_r$ ---- Effective rudder angle (deg or rad)

K ---- Proportionality constant (1/sec)

T ---- Time constant (s)

And the corresponding linear discrete equation is expressed as Eq. 2.

$$y\left(k+2\right)=ay(k+1)+bu(k)+c \tag{2}$$

Where:

$y\left(k+2\right)=\psi\left(k+2\right)-\psi\left(k+1\right)$

$u\left(k\right)=\delta\left(k\right)$

$c=b*\delta_r$

$T=\dfrac{h}{1-a}$; $K=\dfrac{b}{(1-a)h}$; h is the sampling period.

Given time series $\{\delta\left(k\right),r\left(k\right)\}_{k=0,1,\dots,N}$ and $r\left(N+1\right)$, the 3 parameters can be calculated by the system identification method.

Ridge Regression

Ridge regression is the most commonly used method of regularization of ill-posed problems. It is related to the Levenberg-Marquardt algorithm for non-linear least squares problems (Tikhonov A. N, 1977). Ridge regression is an improved least square estimation substantially.

Given the observed rate of turn data $r_i\left(i=1,2,3,......,n\right)$, and suppose observation error is Δr_i . Therefore, rate of turn can be defined as:

$$r_i = f\left(\theta, t_i\right) + \Delta r_i \tag{3}$$

The left side of Eq. 3 is the observed data and $f\left(\theta, t_i\right)$ is the calculated data. Ridge Regression starts at an initial guess $\theta_0 = \{\alpha_1(0), \alpha_2(0), \alpha_3(0)\}$, and according to the Taylor expansion $f\left(\theta, t_i\right)$ can be expressed as Eq. 4,

$$r_i = f(\theta_0, t_i) + \left.\frac{\partial f(\theta, t_i)}{\partial \alpha_1}\right|_{\theta=\theta_0} * \Delta\alpha_1 + \left.\frac{\partial f(\theta, t_i)}{\partial \alpha_2}\right|_{\theta=\theta_0} * \Delta\alpha_2 + \left.\frac{\partial f(\theta, t_i)}{\partial \alpha_3}\right|_{\theta=\theta_0} * \Delta\alpha_3 + \Delta\alpha_i \tag{4}$$

The matrix equation of Eq. 3 is

$$\Delta R = J_r * \Delta\theta + E_r\,, \tag{5}$$

Where, $E_r = \begin{bmatrix} \Delta r_1 \\ \Delta r_2 \\ ..... \\ \Delta r_3 \end{bmatrix}$, $\Delta R = \begin{bmatrix} r_1 - f(\theta_0, t_1) \\ r_2 - f(\theta_0, t_2) \\ .... \\ r_n - f(\theta_0, t_n) \end{bmatrix}$, $\Delta\theta = \begin{bmatrix} \Delta\alpha_1 = \alpha_1 - \alpha_1(0) \\ \Delta\alpha_2 = \alpha_2 - \alpha_2(0) \\ \Delta\alpha_3 = \alpha_3 - \alpha_3(0) \end{bmatrix}$, $J_r = \begin{bmatrix} \frac{\partial f(\theta, t_1)}{\partial \alpha_1} & ... & \frac{\partial f(\theta, t_1)}{\partial \alpha_3} \\ ... & & ... \\ \frac{\partial f(\theta, t_n)}{\partial \alpha_1} & ... & \frac{\partial f(\theta, t_n)}{\partial \alpha_3} \end{bmatrix}$

Define quadratic objective function minimize the observation error:

$$\min\{J(\Delta\theta) = E_r^T E_r\}, \tag{6}$$

Then $$\Delta\theta = (J_r^T J_r + \xi I)^{-1}(J_r^T \Delta R)\,, \tag{7}$$

In the Eq. 7, ξ is damping factor, and if ξ =0 the algorithm above is the classical least squares. Given any sufficiently small positive number ε , if

$$\max\{|\Delta\alpha_1|, |\Delta\alpha_2|, |\Delta\alpha_3|\} \le \varepsilon \tag{8}$$

end of the algorithm. Otherwise assign each coefficient and iterate until conditions of satisfaction. The assignment of each coefficient:

$$\theta_{j+1} = \theta_j + \Delta\theta_j\,,\ j = 0,1,2,...... \tag{9}$$

Compared with Gauss-Newton method, Ridge Regression is a biased estimate, but accuracy has been greatly improved.

PREDICT MANEUVERING INDICES BY AIS DATA

The data broadcasted from an AIS transponder is divided into static, semi-static, and dynamic ones.

- Static data: Ship identification number (MMSI number), length and breadth.
- Semi-static data: Ship destination, hazard level of cargo and ship draft.
- Dynamic data: Time of broadcast, ship speed, rate of turn, course over ground and position.

The rate of AIS data transmission is listed in Table I.

TABLE III. RATE OF AIS TRANSMISSION

Maneuvering Situation	Sample Rate
At Anchor	3 min
Speed 0-14 knots	12 s
Speed 0-14 knots and changing course	4 s
Speed 14-23 knots	6 s
Speed 14-23 knots and changing course	2s
Speed > 23 knots	3 s
Speed > 23 knots and changing course	2 s

Meanwhile, static data is refreshed every 6 minutes or when information has changed.

The application of AIS data make it possible to investigate accurate and actual behavior of ships which can be used in ship maneuvering forecast.

A. Calculate rudder angle from AIS data

The AIS reports are not in the form of ship position reports in a sequence, but as a stream of packets from different ships as received. To apply the data from AIS in maneuvering analysis, the data for individual ships must first be reconstructed. This reconstruction is achieved by use of the unique MMSI number and the time of broadcast with which every data frame is marked and allows the sorting of data from AIS into time series for each different MMSI number.

There are various error sources in AIS data as reported in (Harati Mokhtari et al., 2007) and (Norris, 2007), ranging from data corruption, erroneous MMSI number, target swaps, faulty position reports and errors in rate-of-turn data. Especially, rate of turn and rudder angle are essential to predict maneuvering indices in this research. However, rudder angle is not included in AIS data and some cases rate of turn only indicates positive or negative turn, because it is calculated by the difference of two previous values of heading angle (Nakano. Hasegawa, 2012). In this research, the algorithm that predicts the K-T indices just applies to specific AIS data which include rate of turn.

There are 3 steps in predicting the K-T indices based on AIS data:

1) Find appropriate AIS data that can be used for maneuvering indices predicting. The specific methodology and technology will be discussed in the next chapter.
2) Calculate rudder angle. Because the Nomoto's equation reflects the relationship between rudder angle and rate of turn, in this step we use the rate of turn data included in AIS and appropriate initial value of K and T.

The initial K and T are on the basis of empirical formula (Chen, 2009).

$$K' = 0.428419 + 0.19936L/B \tag{10}$$

$$T' = 1.94599 - 0.04104L/B \tag{11}$$

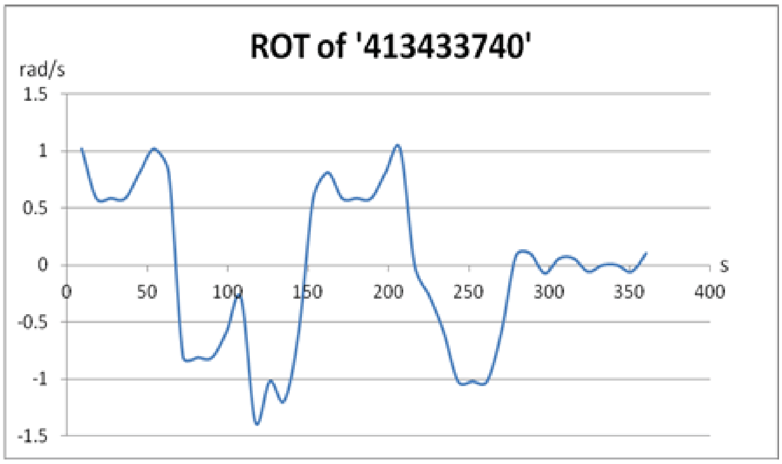

FIG. 1 RATE OF TURN OF '413433740'

MMSI	Ship type	Length over all(m)	Beam (m)	Draught(m)	Speed(knot)
413433740	Cargo	97	16	6	12.75

According to the step 2, we determine the initial K-T indices are 0.2152 and 12.9118 respectively. Rudder angle is updated each time by above Nomoto's equation. Fig. 2 shows the calculated rudder angle of '413433740'.

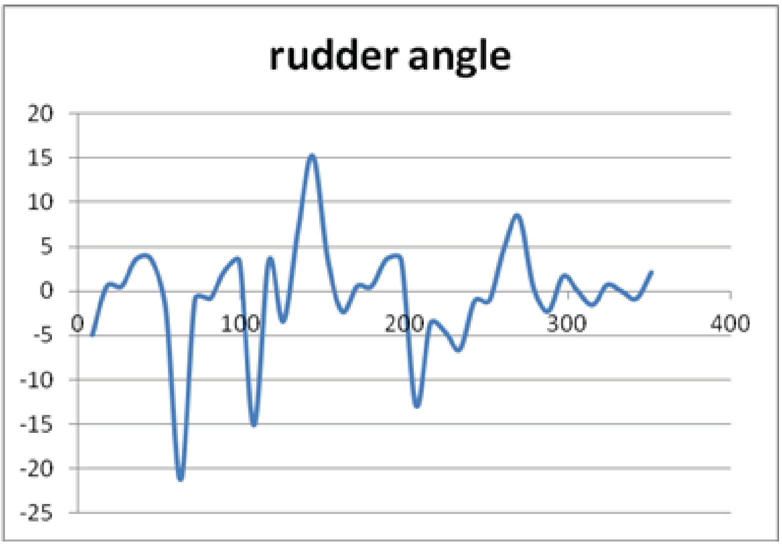

FIG. 2 RUDDER ANGLE OF '413433740'

3) In the second step, rudder angle is calculated in each time. Then apply ridge regression mentioned above to determine the K-T indices. In this step, K and T is refreshed each time until find the best K-T indices that can match the rate of turn and rudder angle data. The predicted maneuvering indices are illustrated in Table.Ⅲ.

TABLE IV. PREDICTED MANEUVERING INDICES OF '413433740'

MMSI	K	T	Speed(knot)
413433740	0.2518	15.1746	12.75

The above 3 steps is how to predict K-T indices by AIS data. Initial value of K and T must be determined appropriately, because in the third step, rudder angle is calculated according to the initial K and T. That means the ridge regression just can optimize the initial K and T. Therefore, it is important to determine the appropriate initial K-T data.

B. A case of study

In order to validate the proposed methodology, the calculated rudder angle and manoeuvrability is compared with the 'Mariner' ship's simulation data. Table. IV shows the characteristics of 'Mariner'.

TABLE IV. TABLE.3 CHARACTERISTICS OF SHIP

Length overall	160(m)
Beam	23.2(m)
Draught	7.467(m)
Block coefficient	0.588
Area of Rudder	14.407(m2)
Diameter of propeller	6.707(m)

TABLE V. TABLE. 4 MANOEUVRABILITY OF SHIP

Experiment condition	*K'*	*T'*
10(degree)	2.848	2.549

In this case, the initial value of K, T is 0.1792 and 16.7341 respectively. These maneuvering indices are calculated by using result of zigzag simulation test. The 'Mariner' have been numerically simulated by solving equation (1) using MATLAB. The solvers used were ode-45 as well as a fourth-order Runge-Kutta method-based solver.

For a typical zigzag maneuvering test, the x-y plot is shown in Fig. 1, $\psi(t)$ and $\delta(t)$ plots are shown in Fig. 2. The data used in this maneuver are service speed $\bar{u}_0 = 15.9$ kn, preselected rudder angle $(\delta_s) = \pm 20\deg$.

Fig. 3 shows the results for rudder angle. The dot line is simulation rudder angle and the solid line is calculated rudder angle. It is clear that the tendency of these two lines is similar while the calculated rudder angles are slight bigger than simulation rudder angle. It is because that the initial K determined will be smaller than actual one..

Table.VI shows the calculated manoeuvrability of 'Mariner', and Fig. 4 shows the comparison of heading angle. The line 2 is Mariner's simulation heading angle and the line 1 is the calculated one.

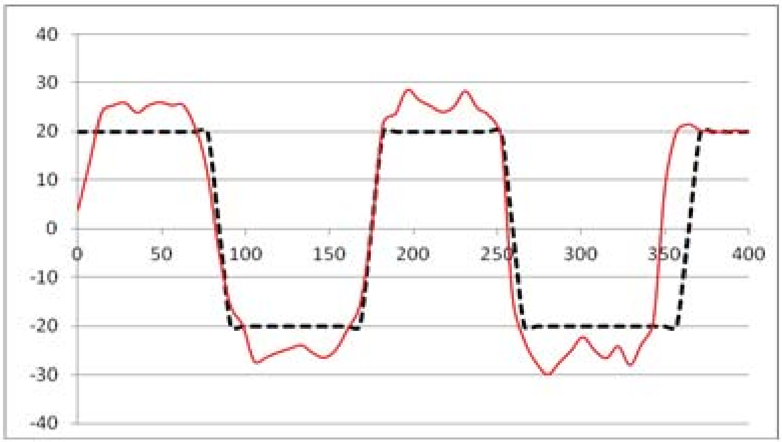

FIG. 3 COMPARISON OF RUDDER ANGLE

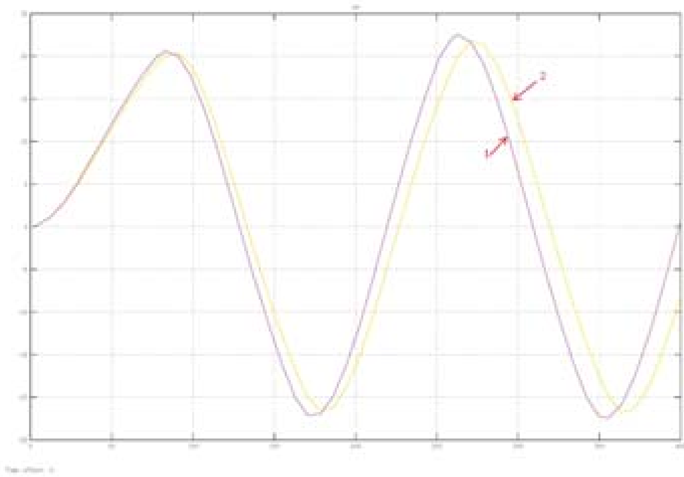

FIG. 4 COMPARISON OF HEADING ANGLE

TABLE VI. CALCULATED MANOEUVRABILITY OF 'MARINER'

Maximum rudder angle	T(s)	K(1/s)	T'	K'	Velocity(knot)
20(degree)	19.4539	0.2174	1.9332	1.9454	15.9

Compare this value with actual manoeuvrability on Table. V, this optimized result seems to be slightly smaller. This is because regarding T', it is affected by initial value of T and this may cause small differences (Hasegawa, 2011). Even though there are small differences between actual manoeuvrability with calculated indices, the simulation results seem good enough to predict manoeuvrability indices. Thus, it is also applied to actual AIS data, and an automatic system that can find appropriate data is proposed.

4. A FREQUENCY DOMAIN IDENTIFICATION METHOD FOR SELECTING APPROPRIATE ROT DATA

Reliability and lost data are the biggest problems of AIS. Harati-Mokhtari et al (2007) use three different datasets to investigate errors and inaccuracies in the different fields of AIS message. In one of the used datasets ('Data-mining study'), 8% from a total of 400,059 reports contained errors concerning MMSI number, IMO number, position, course over ground (COG), and speed over ground (SOG). Especially, rate of turn and rudder angle are essential data to predict manoenvring indices in this research, but the data also exist errors. Besides data errors, how to select appropriate data for system identification is also we concerned.

In statistics, identifiability is a property which a model must satisfy in order for inference to be possible. We say that the model is identifiable if it is theoretically possible to learn the true value of this model's underlying parameter after obtaining numbers of observations from it. In this research, we obtain the K-T indices from analyzing the AIS data, and as known the Nomoto's model is the transfer function relating the heading angle ψ to the rudder angle δ . In another word if we want to identify the K-T indices, we must find the AIS series when ships are steeling. A power spectrum estimation algorithm is presented for select appropriate AIS data.

A. Power Spectrum Estimation

Suppose x_n , from $n=0$ to $N-1$ is a time series (discrete time) with zero mean. Suppose that it is a sum of a finite number of periodic components:

$$x_n = \sum_k [a_k \cos(2\pi v_k n) + b_k \sin(2\pi v_k n)].$$

The variance of x_n is, for a zero-mean function as above, given by $\frac{1}{N}\sum_{N=0}^{N-1} x_n^2$. If these data were samples taken from an electrical signal, this would be its average power.

Apply Fast Fourier Transform (FFT), estimate Fourier coefficient:

$$p_k = \sum_{i=1}^{N} C_j \exp\left[\frac{2\pi k_j \sqrt{-1}}{N}\right]$$

Where, p_k represent the effect that k component weight.

And the power spectrum is $\overline{p}_k = a_k^2 + b_k^2$.

Where $a_k = \frac{1}{N}\sum_{t=1}^{N} x_i \cos\left[\frac{\pi ik}{N}\right]$, and $b_k = \frac{1}{N}\sum_{i=1}^{N} x_i \sin\left[\frac{\pi ik}{N}\right]$.

B. Estimation with AIS data

The rate of turn data which is very important for predicting maneuvering indices in AIS message always indicates 0 or -128(means not available). The first step is getting rid of the wrong rate of turn data. In this research we just analyze the situation that rate of turn is changing of the time.

As a case study, we choose the following two ships. Fig. 5 shows the time series plot of the ROT data of these two ships (MMSI are 413433740 and 413352570 respectively). The data of 413433740 is observed at Yangtze Estuary from 5:17:55PM on 4/18/2010 to 5:24:35 PM. And data of 413352570 is observed at Yangtze Estuary from 11:36:56 AM on 4/18/2010 to 11:43:36 AM. It is conveniently done by calculating the power spectrum $E(f)$ (in one dimension, the ensemble average of the square of the Fourier amplitudes as a function of the frequency f). Fig. 6 shows the power spectrum $E(f)$ of the ROT time series of the two ships.

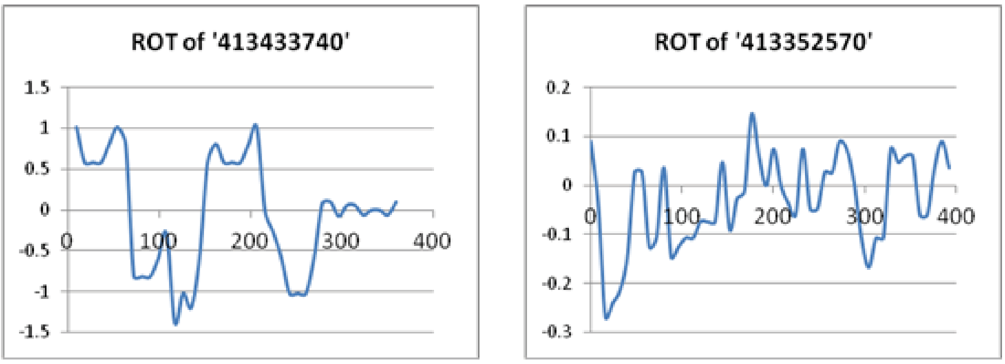

FIG. 5 ROT OF TWO SHIPS

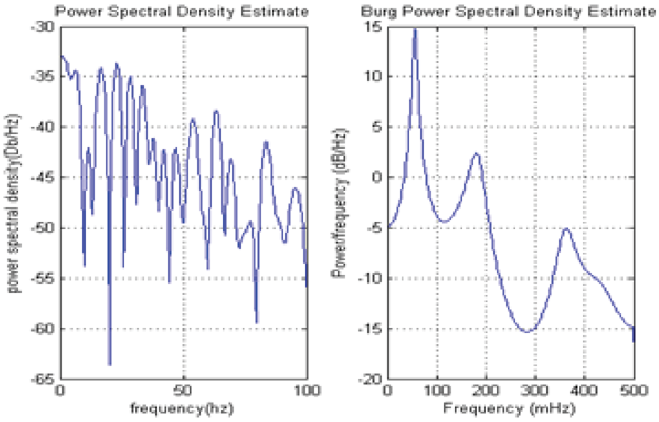

FIG. 6 POWER SPECTRAL OF ROT OF THE TWO SHIPS

The left one of Fig. 6 is power spectral of 413352570, and the spectrum has been averaged over logarithmically spaced frequency intervals. The Max spectrum appears at 1Hz and the power is extreme sporadic. On the contrary, the power spectrum of 413433740 is focus on 50 mHz (T=20S).

As illustrated in FIG. 6, power-laws are observed at multiple timescales in the ROT spectra. At low frequencies, ship steeling is in the highest flight, and the frequencies associated with ships' maneuverability. At higher frequencies, the power spectrum reflects the influence of external environment conditions.

The algorithm that selecting the appropriate AIS data for maneuvering indices predicting is explained as:

Given rate of turn data $\{r_1, r_2 ......r_{60}\}$, and r_i is unequal to -128.

Repeat

1. Calculating the power spectrum of r_i.

2. If 0.01 Hz< $\overline{p_k}$ <0.1Hz, return r_i.

r_i is unequal to -128 because in AIS data -128 means the data is unavailable. It is a pilot study, and by the algorithm illustrated above, the ROT data that can be used for K-T identify are selected.

CONCLUSIONS

This paper presented a detailed method for predicting maneuvering indices by using AIS data. In this research, ridge regression is applied for identify the K-T indices in Nomoto's model. This method is turned out to be reliable by the simulation of Mariner.

The presence of a scaling regime in the power spectrum of the ROT time series indicated the possibility of selecting appropriate data for SI automatically.

ACKNOLEDGMENT

The work presented in this paper is financially supported by NSFC and NOW (Sino-Dutch Cooperation, Grant No.51061130548), and Key Laboratory of High Performance Ship Technology of Ministry of Education (Grant No. 2013033103)

REFERENCES

Van Amerongen, J. 1982 Adaptive Steering of ships-A Model Reference Approach to Improved Maneuvering and Economical Course Keeping, Ph.D. thesis, Delft University of Technology, Netherlands.

Abkowitz, M. A. 1980 Measurement of hydrodynamic characteristics from ship maneuvering trials by system identification, Transactions of the Society of Naval Architects and Marine Engineers, 88, 283-318.

Chen, L. 2009. Study on simulation calculation for inland ship maneuverability K and T indexes, master thesis, Jiaotong University of Chongqing, China.

Haddara, M., AND WANG, Y. 1999 Parametric identification of manoevring models for ships, International Shipbuilding Progress, 46, 445, 5-27.

Harati Mokhtari, A., Wall. A., Brooks, P., Wang, J. 2007 Automatic Identification System (AIS): Data Reliability and Human Error Implications, Journal of Navigation 60: 373-389.

Tikhonov A.N., Goncharsky A.V., Stepanov V.V., Yagola A.G., 1995, Numarical Methods for the Solution of Ill-Posed Problems, Kluwer Academic Publishers.

Chatfield C., 1989, The Analysis of Time Series—An Introduction. Chapman and Hall, London. P. 1. ISBN 0-412+31820-2.

Luo WL., Zou ZJ., 2009, Ship Maneuvering Modeling based on Ridge Regression, SHIP&OCEAN ENGINEERING, Vol. 38, No. 6: 17-19.