Prediction Model for Ship Traffic Flow Considering Periodic Fluctuation Factors

Zhenguo Ding [a], Shukui Zhang [b]

Navigational Department Jiangsu Maritime Institute, Jiangsu Nanjing, China

[a]37163204@qq.com, [b]2091391569@qq.com

Keywords: ship traffic flow; periodic fluctuation; linear growth model; prediction.

Abstract: In order to improve prediction accuracy of ship traffic flow, an improved linear growth model is developed to predict ship traffic flow, by taking into all periodic fluctuation factors, such as seasonal changes, climate impact, and so on, then the Bayesian estimation and prediction are used to solve the new model, and ship traffic flow is predicted using the historical data of ship traffic flow. A case is carried out to compare the prediction effect of the models, and results show that, compared with the linear growth model, the prediction results with the improved model are more in line with the actual situation of ship traffic flow; besides, the mean absolute error of monthly ship flow decrease 3.56%, and the standard deviation decrease 3.79%.

1. Introduction

In order to further improve the level of water transportation services, it is necessary to accurately predict the growth trend and periodic variation of ship traffic flow in the channel. At present, ship traffic flow prediction methods mainly include gray prediction model [1], linear regression analysis [2], etc. These prediction methods require that the time series be relatively stable, with little or no consideration of the non-stationary of the time series, so the prediction error is large, the prediction accuracy is relatively low.

This paper introduces dynamic linear model prediction method into ship traffic field. The model has the advantages of not requiring a relatively stable time series and solving the Bayesian method, and can predict the development trend of traffic flow.

2. Prediction Model

2.1 Linear Growth Model

The linear growth model is a special Bayesian dynamic linear model that is widely used in traffic flow prediction. [5]. the general Bayesian dynamic linear model (DLM) is:

Observation equation:

$$\boldsymbol{y}_{t_i} = \boldsymbol{F}_{t_i}\boldsymbol{\theta}_{t_i} + v_{t_i}, v_{t_i} \sim N_j(0, \boldsymbol{V}_{t_i}) \tag{1}$$

State Equation:

$$\boldsymbol{\theta}_{t_i} = \boldsymbol{G}_{t_i}\boldsymbol{\theta}_{t_i} + w_{t_i}, w_{t_i} \sim N_k(0, \boldsymbol{W}_{t_i}) \tag{2}$$

Here, $\boldsymbol{y}_{t_i}$ represents the ship traffic flow observed at time t_i, which is an r dimensional vector, $\boldsymbol{\theta}_{t_i}$ represents traffic status at time t_i, which is an p dimensional vector. $\boldsymbol{F}_{t_i}$ is $p \times r$ dynamic regression matrix, $\boldsymbol{G}_{t_i}$ is $p \times p$ state transition matrix, $\boldsymbol{V}_{t_i}$ is $r \times r$ variance matrix of observation error, $\boldsymbol{W}_{t_i}$ is $p \times p$ variance matrix of state error, $\boldsymbol{F}_{t_i}, \boldsymbol{G}_{t_i}, \boldsymbol{V}_{t_i}, \boldsymbol{W}_{t_i}$ determines the DLM at any time .

v_{t_i}, w_{t_i} is normal zero mean error term, and satisfy the independence hypothesis. Assume that $\boldsymbol{D}_{t_0}$ is the initial information, $\boldsymbol{D}_{t_i}=\{\boldsymbol{D}_{t_{i-1}},\boldsymbol{y}_{t_i}\}$ represents all information at time t_i and before.

On the basis of the general dynamic linear model, when the dynamic slope η_{t_i} of the state μ_{t_i} is increased, the dynamic model becomes a linear growth model, at this point the state equation changes to:

$$\begin{cases} \boldsymbol{\theta}_{t_i}=\boldsymbol{G}_{t_{i-1}}\boldsymbol{\theta}_{t_{i-1}}+(s_i/s_{i-1})\eta_{t_{i-1}}+w_{t_{i,1}} \\ \eta_{t_i}=(s_i/s_{i-1})\eta_{t_{i-1}}+w_{t_{i,2}} \\ w_{t_i}=\begin{bmatrix} w_{t_{i,1}} \\ w_{t_{i,2}} \end{bmatrix}\sim N_2(0,\boldsymbol{W}_{t_i}) \end{cases} \tag{3}$$

Here, η_{t_i} is rate of change of state parameters from time t_{i-1} to time t_i , $s_i=t_i-t_{i-1}$ is sampling interval. Model parameters are:

$$\boldsymbol{F}_{t_i}=\begin{bmatrix}1 & 0\end{bmatrix},\quad \boldsymbol{\theta}_{t_i}=\begin{bmatrix}\mu_{t_i}\\ \eta_{t_i}\end{bmatrix},\quad \boldsymbol{G}_{t_i}=\begin{bmatrix}1 & s_i/s_{i-1}\\ 0 & s_i/s_{i-1}\end{bmatrix}\quad \boldsymbol{W}_{t_i}=\begin{bmatrix}W_{t_{i,1}} & 0\\ 0 & W_{t_{i,2}}\end{bmatrix}.$$

Since the sampling period is equal, so $s_i/s_{i-1}=1$, therefore, the linear growth model can be expressed as:

$$\begin{cases} \boldsymbol{y}_{t_i}=\mu_{t_i}+v_{t_i}, v_{t_i}\sim N(0,\boldsymbol{V}), \\ \mu_{t_i}=\mu_{t_{i-1}}+\eta_{t_{i-1}}+w_{t_{i,1}} \\ \eta_{t_i}=\eta_{t_{i-1}}+w_{t_{i,2}} \\ w_{t_i}=\begin{bmatrix} w_{t_{i,1}} \\ w_{t_{i,2}} \end{bmatrix}\sim N_2(0,\boldsymbol{W}_{t_i}) \end{cases} \tag{4}$$

2.2 Periodic Fluctuation Model

Assume that the fluctuation period is l (at this point, the fluctuation model has only l-1free states. l-1), then the parameters of the periodic fluctuation model are [6]:

$$\boldsymbol{F}_{t_i}=\begin{bmatrix}1 & 0 & \cdots & 0\end{bmatrix},\qquad \boldsymbol{\theta}_{t_i}=\begin{pmatrix}\beta_{t_1} & \beta_{t_2} & \cdots & \beta_{t_{l-1}}\end{pmatrix}^T,$$

$$\boldsymbol{G}_{t_i}=\begin{bmatrix}-1 & -1 & \cdots & -1 & -1\\ 1 & 0 & \cdots & 0 & 0\\ 0 & 1 & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & 1 & 0\end{bmatrix}\boldsymbol{W}_{t_i}=\begin{bmatrix}\varepsilon^2_{w_{t_{i,1}}} & 0 & \cdots & 0\\ 0 & \varepsilon^2_{w_{t_{i,2}}} & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & \varepsilon^2_{w_{t_{i,l-1}}}\end{bmatrix}$$

Here, $\beta_{t_i}=y_{t_i}-\frac{1}{l}\sum_{q=1}^{l}y_{t_q}$ $(i=1,2,\cdots,l-1)$, represents the deviation of the i th observation from the mean, means estimated value of state fluctuation, it satisfy $\sum\beta_{t_i}=0$, $\varepsilon_{w_{t_i}}$ represents of variance. $\boldsymbol{\theta}_{t_i}$ is $(l-1)$ dimensional state column vector, $\boldsymbol{F}_{t_i}$ is $1\times(l-1)$ dimensional dynamic regression matrix, $\boldsymbol{G}_{t_i}$ is $(l-1)\times(l-1)$ dimensional state transition matrix, $\boldsymbol{W}_{t_i}$ is $(l-1)\times(l-1)$ dimensional state error variance matrix.

3. Simulation Calculation and Evaluation Index of Prediction Model

3.1 Simulation Calculation

The calculation steps are as follows:

Analyze various factors affecting ship traffic flow and screen out factors with periodic fluctuation characteristics;

Considering the cyclical fluctuation factors, construct a ship traffic flow prediction model and solve it by Bayesian estimation and prediction method;

Enter historical observation data of ship traffic flow and establish $\boldsymbol{y}_{t_i}$;

Determine the values of respective parameters of $\boldsymbol{F}_{t_i}, \boldsymbol{G}_{t_i}$ based on the linear growth model and the improved model, the value of l depends on the time period of the observed data, for example, l=12 in monthly traffic flow prediction model and l=4 in quarterly ship traffic flow prediction model ;

Using MATLAB advanced calculation software to calculate the ship traffic flow and give the results.

3.2 Evaluation Index

Correlation coefficient, effect coefficient, consistency index, average absolute error and relative root mean square error can all reflect the fitting degree of predicted value and actual observation value. Assume that the predicted traffic flow of the ship is d_{fore}, actual observation value is d_{mea}, correlation coefficient of d_{fore} and d_{mea} is C, effect coefficient is E, consistency indicator is A,average absolute error is D_{MAPE}, relative mean square root error is $D_{R-RMPSE}$,then:

$$C = \frac{\sum_{i=1}^{n}(d_{\text{fore}i} - \overline{\mathrm{d}_{\text{fore}}})(d_{\text{mea}i} - \overline{\mathrm{d}_{\text{mea}}})}{\sqrt{\sum_{i=1}^{n}(d_{\text{fore}i} - \overline{\mathrm{d}_{\text{fore}}})^2 \sum_{i=1}^{n}(d_{\text{mea}i} - \overline{\mathrm{d}_{\text{mea}}})^2}} \tag{5}$$

$$E = 1.0 - \frac{\sum_{i=1}^{n}(d_{\text{fore}i} - d_{\text{mea}i})^2}{\sum_{i=1}^{n}(d_{\text{for}i} - \overline{\mathrm{d}_{\text{mea}}})^2} \tag{6}$$

$$A = 1.0 - \frac{\sum_{i=1}^{n}(d_{\text{fore}i} - d_{\text{mea}i})^2}{\sum_{i=1}^{n}(\left|\overline{\mathrm{d}_{\text{fore}}} - d_{\text{mea}i}\right| + \left|d_{\text{fore}i} - \overline{\mathrm{d}_{\text{fore}}}\right|)^2} \tag{7}$$

$$D_{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\frac{\left|d_{\text{fore}i} - d_{\text{mea}i}\right|}{d_{\text{mea}i}} \tag{8}$$

$$D_{R-RMPSE} = 100\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\frac{d_{\text{fore}i} - d_{\text{mea}i}}{d_{\text{mea}i}})^2} \tag{9}$$

Standard deviation is selected to judge the stability of prediction error in the paper.

$$\delta = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_{\mathrm{rt}i} - \overline{\mathrm{D}_{\mathrm{rt}}})^2} \tag{10}$$

Here, d_{rt} is relative error of predicted value at time t, $\overline{\mathrm{D}}_{\mathrm{rt}}$ is average value of d_{rt}.

4. Practical Example Verification and Data Analysis

The actual ship traffic flow data of a certain channel from 2010 to 2016 is used as observation data to predict monthly ($l = 12$)ship traffic flow of 2017, and compared prediction results with actual ship traffic flow data to verify the validity of the model.

As shown in Table 1. Improved model is significantly higher than that of the linear growth model, and variation of prediction error range is also small.

Table 1. Comparison of monthly ship traffic flow prediction results between model

Observations value	Linear growth model			Improved model		
	predictive value / ship times	absolute error / ship times	relative error /%	predictive value / ship times	absolute error / ship times	relative error /%
58882	59586	704	1.20	60142	1260	2.14
53195	59896	6701	12.60	54526	1331	2.50
55456	60217	4761	8.59	56913	1457	2.63
67368	60532	6836	10.15	66112	1256	1.86
66724	60872	5852	8.77	67774	1050	1.57
68963	61204	7759	11.25	70243	1280	1.86
55144	61562	6418	11.64	55515	371	0.67
58362	61938	3576	6.13	56327	2038	3.49
59464	62352	2888	4.86	60246	782	1.32
67424	63792	4632	6.87	68518	1094	1.62
67023	63213	3810	5.68	68832	1809	2.70
68021	63683	4338	6.38	69572	1551	2.28
Average value		5000.71	7.97	Average value	1474.42	2.41

The value of each evaluation index is calculated according to formulas (5) to (10), and results are shown in table 2. It can be seen from table 2 that the value of t C, E and A of improved model are larger than corresponding linear growth model, besides, the value of D_{MAPE}, $D_{R-RMPSE}$ and δ of improved model are both are smaller than corresponding linear growth model.

Table 2. Test of goodness foe fit with different models

Model	Evaluation index					
	C	E	A	D_{MAPE}	$D_{R-RMPSE}$	δ
Linear growth model	0.852	0.813	0.832	5.97%	7.46%	5.32%
Improved model	0.912	0.863	0.894	2.41%	2.68%	1.53%

5. Conclusion

Example verification shows that improved model can not only reflect the overall trend of ship traffic flow well, but also fully reflect the volatility characteristics of actual reflect the volatility characteristics of actual ship traffic flow data; Compared with linear growth model, improved model has higher prediction accuracy and better stability.

Acknowledgements

This work was supported by “Qing Lan Project” of Jiangsu.

References

[1] ZHANG Shukui XIAO Ying-jie. Grey neural network model for ship traffic flow prediction [J]. Journal of Shanghai Maritime University, 2015,36(1):46-49.

[2] LI Jun. The research of maritime accidents based on the SVM [J]. Wuhan: Wuhan University of Technology,2008.

[3] ONDER E, KUZU S. Forecasting air traffic volumes using smoothing techniques[J]. Journal of China Aeronautics and Space Technologies,2014,7(1):65-70.

[4] MENON P K. New approach for modeling, analysis, and control of air traffic flow[J]. Journal of Guidance, Control, and Dynamics, 2014,27(5):737-744.

[5] BOUGAS C. Forecasting air passenger traffic flows in Canada: an evaluation of time series models methods[D]. Quehec ,Canada: Constlantinos Bougas,2013.