Transportation Research Part D 65 (2018) 817–824

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Predicting ship fuel consumption based on LASSO regression

Shengzheng Wang, Baoxian Ji*, Jiansen Zhao, Wei Liu, Tie Xu

Merchant Marine College, Shanghai Maritime University, Shanghai 201306, PR China

A R T I C L E   I N F O

Keywords:
LASSO regression
Fuel consumption prediction
Voyage optimization
Energy conservation
Weather routing

A B S T R A C T

During the voyage, predicting fuel consumption of ships under different sea-states and weather conditions has been a challenging and far-reaching topic, because there are a great number of feature variables affecting the fuel consumption, including main-engine status, cargo weight, ship draft, sea-states and weather conditions, etc. Data driven statistical models have been employed to model the relationship between fuel consumption rate and these voyage parameters. However, some of the feature variables are highly correlated, e.g. wind speed and wave height, air pressure and wind force, cargo weight and draft etc., thus a typical multiple collinearity problem arises so that the fuel consumption cannot be accurately calculated by using the traditional multiple linear regression. In this study, the LASSO (Least Absolute Shrinkage and Selection Operator) regression algorithm is employed to implement the variable selection for these feature variables, additionally, it guides the trained predictor towards a generalizable solution, thereby improving the interpretability and accuracy of the model. On the basis of the LASSO, a novel ship fuel consumption prediction model is proposed. Experimentally, the superiority of the proposed method was confirmed by comparing it with some existing methods on predicting the fuel consumption. The proposed method is a promising development that improves the calculation of the fuel consumption.

1. Introduction

The control of fuel consumption has aroused extensive attention since the Ship Energy Efficiency Management Plan (SEEMP) was issued by International Maritime Organization (IMO) in MARPOL convention (IMO, 2009, 2012a,b,c,d,e, 2014). In order to carry out this plan, speed optimization has become a typical way to improve the energy efficiency because it would lead to engine power or fuel consumption decline at a rate of three time. Low-horse power engine or slow steaming was commonly used to reduce the fuel consumption, and hence reduce carbon emissions (Psaraftis and Kontovas, 2013). Over the past decade, some researchers have focused on the speed optimization of the entire voyage for reducing the fuel consumption (Fagerholt et al., 2010; Norstad et al., 2011; Wang and Meng, 2012; Fukasawa et al., 2016).

Actually, slow steaming is more often used by shipping companies. However, during the voyage, only depending on slow steaming is a limited approach to improve the energy efficiency due to the complicated requirements resulting from safety and shipping schedule. Therefore, many research efforts have been dedicated to optimizing the sailing speed by combining the ship's sailing schedule, the hydro-meteorological conditions, the sea-states (e.g. waves, swells, currents etc.) and the shipping operation states under the premise of ensuring ship safety so as to achieve the purpose of saving fuel consumption.

Recently, the electronic devices based on the high acquisition rate sensors are more and more widely used in the modern ships to develop the realtime and continuous data collection systems. Therefore, a large number of data available in practice are used to build

* Corresponding author.
E-mail address: jibaoxian123@163.com (B. Ji).

http://dx.doi.org/10.1016/j.trd.2017.09.014

Available online 28 October 2017

a data driven statistical model for modeling ship energy efficiency under the different loading conditions, weather conditions etc. (Petersen et al., 2011a). The well-known Holtrop-Mennen method is a useful regression approach based on model experiments and full-scale data to estimate the resistance and required power in the initial design stage (Holtrop and Mennen, 1982; Holtrop, 1984). To predict the ship performance, Journe et al. proposed an online data collecting system. On the basis of the physics and hydrodynamic formulas, a mathematical model was constructed and a decision-making system was implemented to find an optimal performance with regard to fuel consumption through changing the trim, heading and speed of the ship (Journe et al., 1987; Journe, 2003). Petersen et al. presented a novel statistical model to improve the fuel efficiency of the ship propulsion. They investigated and compared two statistical models: Artificial Neural Network (ANN) and Gaussian Processes (GP), and then further discussed the interpretations of the operational data in relation to the underlying physical system (Petersen and Winther, 2011; Petersen et al., 2011b). Lu et al. proposed a semi-empirical ship operational performance prediction model by taking into account the added resistance caused by wave and wind for a specific ship type (Lu and Turan, 2015). In addition, on the basis of the real shipping log data, Meng et al. developed a ship fuel efficiency regression analysis model for modeling the relationships between the fuel consumption of a specific container ship and the feature variables (Meng et al., 2016). Bocchetti et al. attempted to find a solution through a statistical approach based on multiple linear regression, which allows for fuel consumption prediction of a specific voyage given its operational conditions (e.g., ship speed, steaming distance, wind speed, wind direction, cumulative docking time, displacement, stabilizer fin operating time, and engine operation mode). This approach also allows for prediction interval estimation that can be used to control the actual fuel consumption and to identify potential anomalous voyages in terms of fuel consumption or to test the effectiveness of a specific efficiency improvement operation. The proposed approach overcame the speed-power curves, which are usually utilized in the naval architecture and predict fuel consumption only through the bi-dimensional relation between power or consumption and speed (Bocchetti et al., 2015).

Although existing approaches have produced impressive results, which demonstrate the effectiveness of the data driven statistical models, the existing methods still lack good interpretability and high accuracy resulting from the complexity of voyage process. Nevertheless, the common purpose of all researches is to minimize the total energy consumption on the premise of maintaining the safety and efficiency of the voyage. Therefore, the key technique is to accurately calculate the fuel consumption for different ships under various surrounding environments. Only based on this, it is possible to build the optimization model to calculate the appropriate sailing speed and the engine power.

For fuel consumption prediction, the ANN-based method is a common and widely used prediction model. Bal Besikci et al. used the noon report data to build an ANN-based fuel prediction model for modeling the relationship between the engine RPM (revolutions per minute) and the feature variables outside of the engine, which was used to analyze the bunker fuel efficiency of a tanker (BalBesikci et al., 2016). Wang et al. established a real-time optimization model based on the wavelet neural network (WNN) for optimizing energy efficiency. The optimal engine speed can be determined by the WNN-based model under different working conditions and navigation environments, which provides theoretical guidance for optimizing ship's sailing and improving energy efficiency (Wang et al., 2016a). Although the ANN is cheap computationally, there is no a rule to select the feature variables and avoid the overfitting phenomenon during the training. In addition, the Support Vector Regression (SVR) has been widely used for prediction, but the SVR has higher memory requirements and poorer CPU running efficiency. Another disadvantage of the SVR is not able to handle high-dimensional data very well, which may affect the prediction accuracy. Therefore, the SVR would not be a good choice for an onboard system if these problems were not solved in some manner.At the same time, some other related regression approaches can also be found in Lepore et al. (2017), Wang and Yang (2015) and Wang et al. (2016b).

Therefore, our main research task is to study a novel model from a new perspective to describe the fuel consumption of a specific ship as a function of the ships states and surrounding environments. After carefully analysis for various feature variables affecting the fuel consumption, such as weather conditions, sea-states, main engine status, cargo weight, and ship draft etc., it is clear that there are a large number of feature variables, and some of the feature variables are highly correlated, e.g. wind speed and wave height, air pressure and wind force, cargo weight and draft etc., thus a typical multiple collinearity problem arises so that the traditional multiple linear regression method cannot be used to correctly calculate the fuel consumption. In this paper, the LASSO regression algorithm is employed to implement the variable selection of these feature variables, additionally, it guides the trained predictor toward a generalizable solution, and based on that, a novel ship fuel consumption prediction model is proposed to improve the interpretability and accuracy.

With this in mind, our main contribution is to employ LASSO regression algorithm to construct a novel framework to predict the ship fuel consumption. This paper has three main components.

(1) We present a novel framework of predicting the fuel consumption from a new perspective.
(2) We present the LASSO-based fuel consumption prediction model.
(3) We compare our proposed method with existing popular approaches.

The rest of this paper is organized as follows. A brief overview of the proposed scheme for fuel consumption prediction is presented in Section 2. A LASSO-based fuel consumption model is proposed in Section 3, while in Section 4, we evaluate the effectiveness of the proposed approach by comparing with some typical algorithms. Section 5 provides the conclusion.

2. Approach overview

The system framework for fuel consumption based on the proposed LASSO regression method is shown in the Fig.1, which can

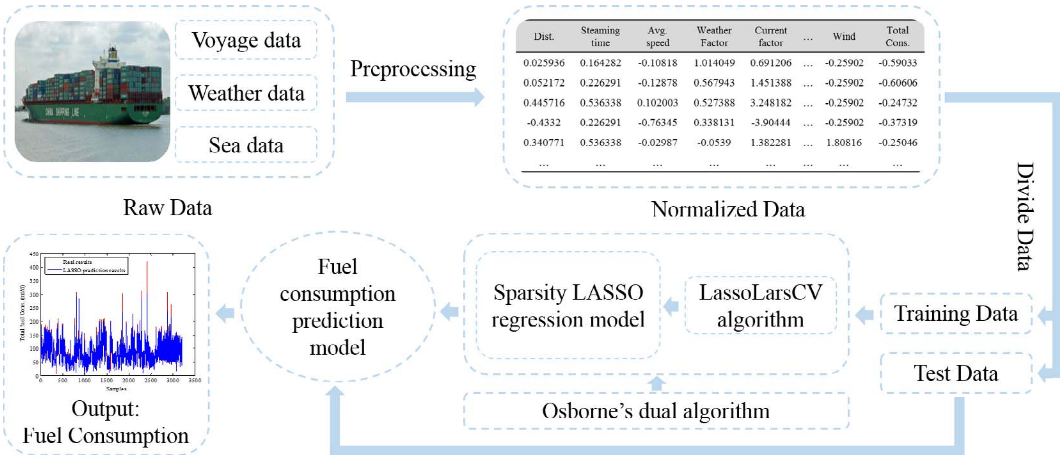

Fig. 1. The system framework for fuel consumption based on the proposed approach.

summarize our point of view. The core of the system is the LASSO algorithm described in detail in Section 3.1. After completing the building and training for the fuel consumption prediction model, we use the test set to verify the performance of the model, and the results are shown in Section 4.3.

3. Ship's fuel consumption prediction model based on LASSO regression

3.1. LASSO regression model

LASSO presented by Tibshirani (1996) is an innovative variable selection method for regression by minimizing the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant, and it is a well-known sparse regression method which regularizes the parameter β under sparse assumption. It was originally introduced in the context of least squares. The basic framework is summarized as follows. Considering a sample consisting of N cases, each of which consists of p covariates and a single outcome. Supposing y_i is the response variable and $x_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$ is the covariate vector for the i^{th} case, $\boldsymbol{\beta} = (\beta_1, \beta_1, ..., \beta_p)^T$, so the objective of LASSO is to solve the optimization problem:

$$\arg\min_{\beta_0, \boldsymbol{\beta} \in \mathscr{R}^p} \frac{1}{N} \sum_{i=1}^{N} (y_i - \beta_0 - x_i^T \boldsymbol{\beta})^2 \quad (1)$$

$$s.\ t. \quad \sum_{j=1}^{p} |\beta_j| \leqslant t$$

where $t \geqslant 0$ is a pre-specified free parameter that determines the amount of regularization. If t is large, all the coefficients are almost zero. For smaller values of t, the LASSO shrinks some of the estimated coefficients equal to zero. Supposed that X represents the $N \times p$ covariates matrix, N is the number of samples, p is the number of the covariates, y represents a response vector as excepted output. Formula (1) can be wrote more compactly as

$$\arg\min_{\beta_0, \boldsymbol{\beta} \in \mathscr{R}^p} \frac{1}{N} \|y - \beta_0 I - X\boldsymbol{\beta}^T\|_2^2 \quad (2)$$

$$s.\ t. \quad \|\boldsymbol{\beta}\|_1 \leqslant t$$

where $\|Z\|_p = (\sum_{i=1}^{N} |Z_i|^p)^{1/p}$ is the standard l^p norm. Since $\hat{\beta}_0 = \bar{y} - \bar{x}^T \boldsymbol{\beta}$, so that

$$y_i - \hat{\beta}_0 - x_i^T \boldsymbol{\beta} = y_i - (\bar{y} - \bar{x}^T \boldsymbol{\beta}) - x_i^T \boldsymbol{\beta} \quad (3)$$

$$= (y_i - \bar{y}) - (x_i - \bar{x})^T \boldsymbol{\beta},$$

we can rewrite (2) as

$$\arg\min_{\boldsymbol{\beta} \in \mathscr{R}^p} \left\{ \frac{1}{N} \|y - X\boldsymbol{\beta}^T\|_2^2 \right\} \quad (4)$$

$s.\ t. \quad \|\boldsymbol{\beta}\|_1 \leqslant t$

The LASSO estimator $\hat{\beta}$ (Belloni et al., 2011; Efron et al., 2004; Tibshirani, 1996; Zou, 2006) can be represented by the following Lagrangian form:

$$L(\boldsymbol{\beta},\lambda) = \min_{\boldsymbol{\beta}\in\mathscr{R}^p}\left\{\frac{1}{N}\|y-X\boldsymbol{\beta}^T\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1\right\} \tag{5}$$

where the tuning parameter $\lambda \geqslant 0$ is used to balance the empirical error and the sparsity of model parameter, and where the exact relationship between t and λ is data dependent.

3.2. Computation of the Lasso solutions

The k-fold cross-validation combined with the LARS algorithm (*i.e.* LassoLarsCV) (Tibshirani, 1996; Efron et al., 2004; Osborne et al., 2000) is generally used to estimate the regularization parameter λ of the Lasso regression model. In this study, we adopt *10*-fold cross-validation to estimate the regularization parameter λ or t. The constant parameter λ is estimated by minimizing the loss function according to the Eq. (5), where the regression coefficient vector β is estimated using the LARS algorithm, continuously reducing the residual error until it is small enough or less than or equal to a constant.

In k-fold cross-validation, the training set is split into k approximately equal sized sets. We then perform k training runs using in turn one of the sets for validation. The average test error over the k runs is regarded as the test error for the regression model trained on $N-N/k$ samples. Finally, we can get a range of λ values, and then select the optimized λ value that corresponds to the lowest estimated generalization error.

Osbornes dual algorithm is employed to compute the LASSO estimates based on the above parameter value in this study.

3.3. Fuel consumption prediction

The dataset of ship reports mainly consists of: length of overall (LOA), beam, global ship position, steaming distance, steaming time, cargo weight, average draft, trim angle, initial metacentric height (GM), displacement at departure, revolutions per minute (RPM), main engine load, true course, average speed over ground (SOG), loading condition, weather factor, current factor, performance speed, main engine consumption, auxiliary engine consumption, beaufort scale, wind waves direction, wind waves period, wind waves height, swell direction, swell period, swell height, CO_2 efficiency, wind direction, significant wave height, fuel efficiency etc., as shown in Table 1.

Table 1 shows that the standard of each data field are not consistent, for instance, some feature variables contain positive and negative value. If these data items are directly used for training model, it maybe cause the unstability of the solver.

Therefore, normalization is a prerequisite for the improvement of model prediction accuracy. Data needs to be normalized preliminarily. There are two methods commonly used for normalization: max-min normalization and zero mean standardization (Z-score standardization). The Z-score standardization method is introduced into this study to pre-process the original data.

Supposing there is a dataset $X = \{x_i\}, i = 1,2,\ldots,n, \bar{x}$ is the mean value of the whole observations. So firstly, the observations of the dataset X are centered as follows:

$$x_i' = x_i-\bar{x} \tag{6}$$

Z-score standardization can be expressed as

$$z_i = \frac{x_i'}{S_i} \tag{7}$$

where

$$S_i = \sqrt{\frac{\sum_{i=1}^{n}(x_i')^2}{n-1}} \tag{8}$$

For the standardized variable Z corresponding to Eq. (7), the mean is 0 and the standard deviation is 1.

As the basis of the above measured features, we use the LASSO method to build a regression model for predicting the fuel consumption. Using the multiple linear regression equation (Montgomery et al., 2012), the fuel consumption is modelled as

$$Y = \boldsymbol{\beta}X + b \tag{9}$$

Table 1
Partial original dataset of the container ships.

ID	Fuel Consumption (mt/d)	LOA (m)	Beam (m)	SOG (kts)	Trim Angle (deg)	Beaufort Scale	Swell Height (m)	…
x_1	73.7	300	48.3	17.65	1.2	5	1.3	…
x_2	88.8	334.1	42.8	15.8	0.6	4	1.43	…
x_3	77.6	261.1	32.3	18.75	0.1	4	1.75	…
x_4	94.7	349.1	45.6	16.99	−0.1	6	0.81	…
x_5	69.8	280	39.8	16.75	1.7	5	1.62	…
…	…	…	…	…	…	…	…	…

Table 2
Partial standardization dataset of the container ships.

ID	Fuel Cons.	LOA	Beam	SOG	Trim Angle	Beaufort Scale	Swell Height
x_1	−0.12294	−0.26059	1.00778	0.60757	0.94011	0.45725	−0.70577
x_2	0.28780	0.68098	0.13650	−0.11799	0.11698	−0.30763	−0.56006
x_3	−0.01686	−1.33470	−1.52684	1.03899	−0.56897	−0.30763	−0.20138
x_4	0.44829	1.09516	0.58006	0.34872	−0.84335	1.22214	−1.25500
x_5	−0.23447	−0.81283	−0.33874	0.25460	1.62605	0.45726	−0.34709
…	…	…	…	…	…	…	…

where, $X = (x_1,x_2,...,x_n)$ is the sample, x_i is the i^{th} sample corresponding feature vector. $Y = (y_1,y_2,...,y_n)$ is the output vector, y_i is the i^{th} sample corresponding response variable, specifically referring to the fuel consumption. b are intercepts. $\boldsymbol{\beta} = (\beta_1,\beta_2,...,\beta_p)$ is regression coefficient vector, which is computed by using the method in Section 3.2, β_j is j^{th} regression coefficient.

4. Experiments

4.1. Database and evaluation criterions

In order to evaluate the performance of the proposed prediction model, the fuel consumption prediction was performed on the dataset derived from the COSCON (COSCO SHIPPING Lines) fleet management system.

The original dataset includes 884 voyages for 97 container ships, which spans from October 2014 to March 2017. In our study, the original dataset is used to train and evaluate our model, which consists of 14,334 samples and 21 feature variables. The Z-score standardized dataset is shown in the Table 2.

For the evaluation criterias of the different models, the mean absolute error (MAE), the root mean square deviation (RMSD), and the cumulative score (CS) are used. The MAE is defined as the average value of the absolute errors between the predicted results and the real values for the test dataset.

The RMSD is defined as:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n} (y_i - y_i')^2}{n}} \tag{10}$$

where y_i is a predicted value, y_i' is a real value, n is the number of samples.

The CS is defined as:

$$CS(\delta) = \frac{M_{e<\delta}}{M} \times 100\% \tag{11}$$

where M is the number of the test samples and $M_{e<\delta}$ represents the number of the test samples whose absolute errors e are less than δ mt/d.

4.2. Experimental setup

To prove the effectiveness and reliability of the prediction model based on LASSO regression, as mentioned in the Section 4.1, the subset of the original dataset has been divided into two subsets: the training dateset and the test dateset. The 80% of the dataset was used for the training dataset, which has been generated by selecting randomly from the original dataset. The remainder was used as the test dataset. All feature variable values were pre-processed by using the Z-score standardization method, that is, the mean value $E(x_i) = 0$, the standard deviation $D(x_i) = 1$, and the response variable was centered before dividing the dataset.

In this study, the fuel consumption was regarded as the model response, while other feature variables were considered as the inputs of the model. The optimal model was obtained by calculating the best parameter λ corresponding to the minimum mean squared error. The regularization parameter λ of the LASSO regression model was estimated by performing *10*-fold cross-validation as mentioned in Section 3.2. The training datasets were divided into 10 approximately equal sized datasets. We then performed 10 training runs using in turn one of the sets for validation. Finally, the best parameters obtained by parameter learning are $\lambda = 0.02397591$ and $b = -131.1047$.

The feature variables corresponding to the nonzero regression coefficients are the results of the variable selection, as shown in the Table 3. It can be seen that the features were reduced to 20 based on the feature selection, including LOA, beam, cargo weight, average draft, trim angle, GM, RPM, main engine load, SOG, current factor, performance speed, beaufort scale, wind waves height, swell height, CO_2 efficiency, fuel efficiency, ship course, wind direction, significant wave height, swell direction. Furthermore, most of coefficients are very small, which illustrates the correlation between each feature variable and the fuel consumption. The correlation between feature variables also can be illustrated in Fig. 2. The $\boldsymbol{\beta}$ corresponding to the selected features, the parameters λ and b were used in the LASSO-based prediction model of the fuel consumption.

Table 3
Regression coefficients of the fuel consumption prediction model based on LASSO model.

Features	β_j	Features	β_j
LOA	0.07769331	Beam	0.2324890
Cargo weight	−0.00007323914	Average draft	0.3756406
Trim angle	0.02061014	GM	−0.2837404
Displacement	0	RPM	0.3556936
ME Load	0.1241662	SOG	4.710010
Current Factor	−1.423200	Performance speed	0.3121529
Beaufort scale	0.4816946	Wind waves height	−0.2165611
Swell height	0.2870997	CO_2 efficiency	−0.006012160
Fuel efficiency	0.2975090	Ship course	0.004961683
Wind direction	0.003454263	Significant wave height	0.2010854
Swell direction	−0.001673432		

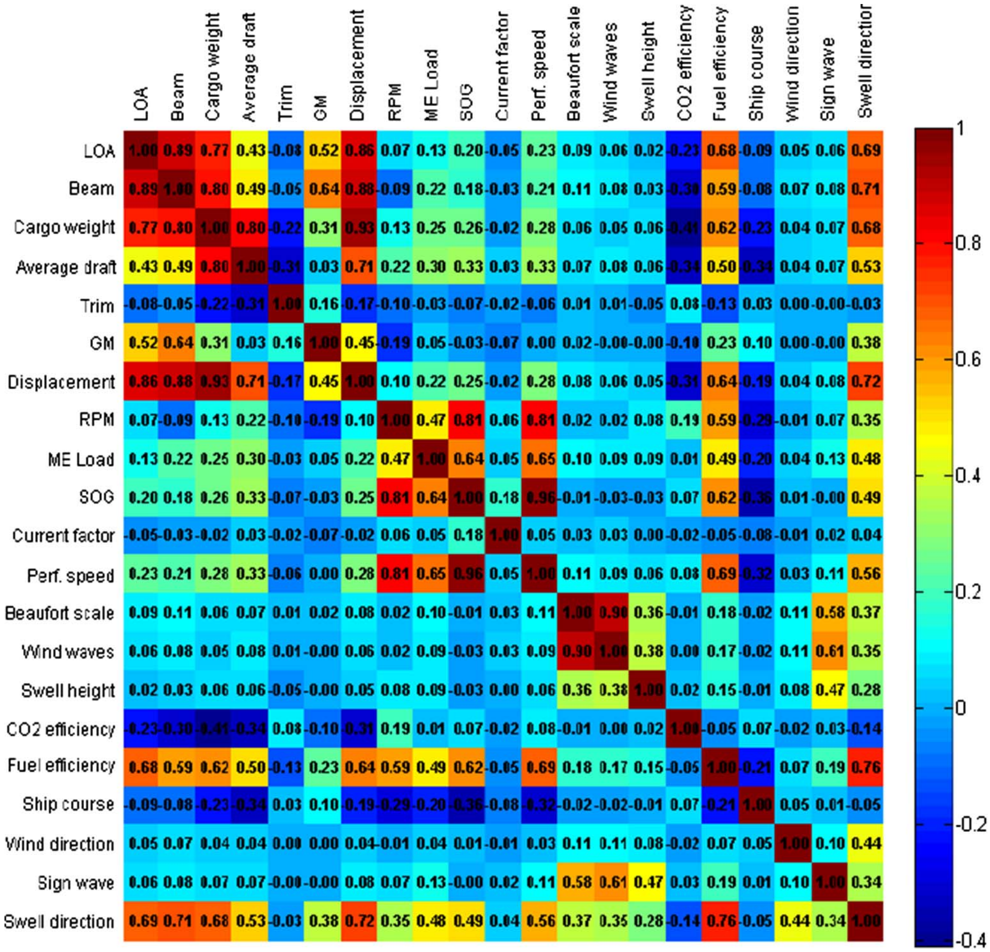

Fig. 2. The correlation analysis of the feature variables.

4.3. Experimental results

To demonstrate the performance of the proposed LASSO-based fuel consumption prediction model, we compared it with the typical regression methods such as ANN, SVR and GP according to the evaluation criterias presented in the Section 4.1 respectively. The fitting performance for these four algorithms are shown in Fig. 3. The predicted and the real fuel consumption values are marked by blue and red lines respectively. Apparently, the proposed LASSO-based model outperformed other three models. Fig. 3 shows that the prediction results of the LASSO-based model are able to fit accurately the real values in the most of situations.

In addition, the comparison results of the CS of the proposed method with the state-of-the-art ANN, SVR, and GP are shown in Fig. 4. The proposed LASSO-based model outperformed the other state-of-the-art methods, which further demonstrates the

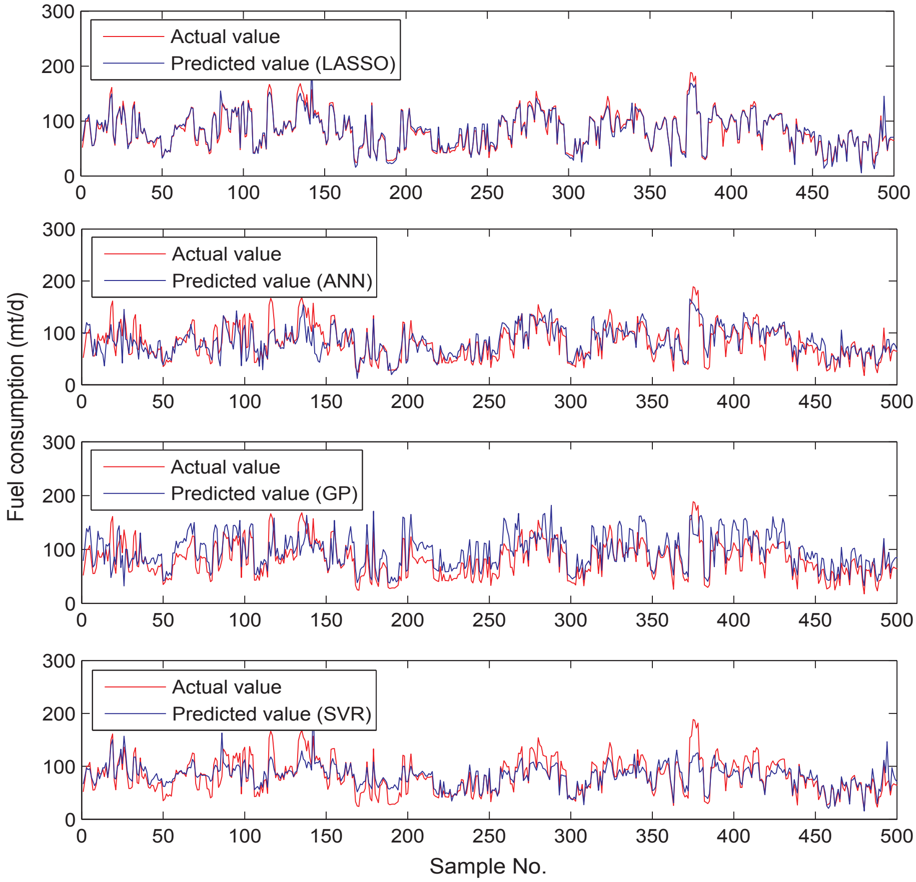

Fig. 3. The predicted results of four models based on the same test set.

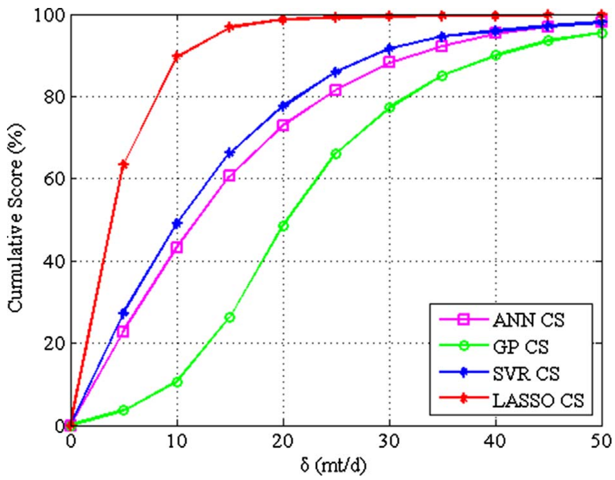

Fig. 4. Comparisons of the CS results of four different methods.

effectiveness of the proposed LASSO-based model.

The RMSD and MAE results of the proposed model and other three models for the same test data sets are shown in Table 4. As expected, the proposed model had the lowest RMSD and MAE which were 7.4 and 4.9 mt/d respectively. The lowest RMSD means that the proposed model has better fitting performance and is able to accurately predict the change of fuel consumption under the different navigational conditions.

Finally, these models were implemented in MATLAB R2011a and ran all the experiments on a 3.5 GHz Intel(R) Xeon(R) CPU E5-1650 v2 processor with 8 GB RAM under Windows 7 (64 bits). The efficiencies of the different models were evaluated according to

Table 4
RMSD, MAE results and Running time for the four models.

Methods	RMSD (mt/d)	MAE (mt/d)	Running Time (s)
ANN	19.5	14.96	34.9
SVR	18.7	13.5	28.2
GP	27.5	23.4	87413.5
LASSO	7.4	4.9	8.5

the training time as shown in Table 4. It is obvious that the time cost of LASSO-based model is less than the time cost of other models for the same training dataset.

5. Conclusion

In this study, we presented a framework of predicting ship fuel consumption based on LASSO regression algorithm. In this framework, both the realistic ship operational dataset and the weather data were considered, and the prediction results were quantitatively compared with the real fuel consumption data. In addition, the proposed LASSO-based method was validated in prediction accuracy and computational performance. The proposed LASSO-based method outperforms other traditional methods, and it also has several good characteristics such as interpretability, generalization ability and numerical stability.

In future work, we will collect much more voyage dataset from different voyages or ships to validate and further improve the model and the framework of predicting fuel consumption.

Acknowledgement

The authors would like to thank the anonymous reviewers and editors for their comments and suggestions. The research is supported by Shanghai Shuguang Plan Project (No:15SG44) and NSFC of China (No: 51379121, and 61304230).

References

BalBesikci, E., Arslan, O., Turan, O., et al., 2016. An artificial neural network based decision support system for energy efficient ship operations. Comput. Oper. Res. 66, 393–401.
Belloni, A., Chernozhukov, V., Wang, L., 2011. Pivotal recovery of sparse signals via conic programming. Biometrika 98 (4), 791–806.
Bocchetti, D., Lepore, A., Palumbo, B., et al., 2015. A statistical approach to ship fuel consumption monitoring. J. Ship Res. 59 (3), 162–171.
Efron, B., Hastle, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression (with discussion). Ann. Statist. 32, 407–499.
Fagerholt, K., Laporte, G., Norstad, I., 2010. Reducing fuel emissions by optimizing speed on shipping routes. J. Oper. Res. Soc. 61 (3), 523–529.
Fukasawa, R., He, Q., Santos, F., et al., 2016. A joint routing and speed optimization problem, <https://arxiv.org/abs/1602.08508v2>.
Holtrop, J., 1984. A statistical re-analysis of resistance and propulsion data. Int. Shipbuilding Prog. 31, 272–276.
Holtrop, J., Mennen, G., 1982. An approximate power prediction method. Int. Shipbuilding Prog. 29, 166–170.
IMO, 2009. Guidelines for the Development of a Ship Energy Efficiency Management Plan (SEEMP), MEPC, 1/Circ. 683.
IMO, 2012. Guideline for Development of a Ship Energy Efficiency Management Plan (SEEMP), MEPC, 213(63) Annex 9.
IMO, 2012. Guideline on Survey and Certification of the Energy Efficiency Design Index (EEDI), MEPC, 214(63) Annex 10.
IMO, 2012. Guideline on the Method of Calculation of the Attained Energy Efficiency Design Index (EEDI) for a New Ship, MEPC, 212 Annex 8.
IMO, 2012. Air Pollution and Greenhouse Gas (GHG) Emission from International Shipping, Marpol Annex VI.
IMO, 2012. Guidelines for Calculation of Reference Lines for Use with the Energy Efficiency Design Index (EEDI), MEPC, 215(63) Annex 11.
IMO, 2014. Third IMO GHG Study (Final Report) on Prevention of Air Pollution from Ships, <http://www.iadc.org/wp-content/uploads/2014/02/MEPC-67-6-INF3-2014-Final-Report-complete.pdf>.
Journe, J., 2003. Review of the 1985 full-scale calm water performance tests onboard m.v. mighty servant 3, Tech. Rep. DUT-SHL Report 1361.
Journe, J., Rijke, R., Verleg, G., 1987. Marine performance surveillance with a personal computer, Technical Report.
Lepore, A., Palumbo, B., dos Reis, M., et al., 2017. A comparison of advanced regression techniques for predicting ship CO_2 emissions. Quality Reliab. Eng. Int. 5, 1–11.
Lu, R., Turan, O., 2015. A semi-empirical ship operational performance prediction model for voyage optimization towards energy efficient shipping. Ocean Eng. 110, 18–28.
Meng, Q., Du, Y.Q., Wang, Y.D., 2016. Shipping log data based container ship fuel efficiency modeling. Transp. Res. Part B 83, 207–229.
Montgomery, D., Peck, E., Vining, G., 2012. Introduction to Linear Regression Analysis, fifth ed. Wiley & Sons, Hoboken, NJ.
Norstad, I., Fagerholt, K., Laporte, G., 2011. Tramp ship routing and scheduling with speed optimization. Transp. Res. Part C 19 (5), 853–865.
Osborne, M.R., Presnell, B., Turlach, B.A., 2000. On the LASSO and its dual. J. Comput. Graphical Statist. 9 (2), 319–337.
Petersen, J., Winther, O., 2011. Mining of Ship Operation Data for Energy Conservation, Kgs. Lyngby, Denmark: Technical University of Denmark (DTU). (IMM-PHD-2011; No. 264).
Petersen, J., Jacobsen D., Winther, O., 2011. A machine-learning approach to predict main energy consumption under realistic operational conditions. In: Proceedings of the 10th International Conference on Computer and IT Applications in the Maritime Industries (COMPIT11), pp. 305–316.
Petersen, J., Jacobsen, D., Winther, O., 2011b. Statistical modelling for ship propulsion efficiency. J. Mar. Sci. Technol. 17, 30–39.
Psaraftis, H., Kontovas, C., 2013. Speed models for energy efficient maritime transportation: a taxonomy and survey. Transp. Res. Part C 26, 331–351.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B 58, 267–288.
Wang, S., Meng, Q., 2012. Sailing speed optimization for container ships in a liner shipping network. Transp. Res. Part E 48 (3), 701–714.
Wang, S., Yang, J., 2015. A probabilistic model for latent least squares regression. Neurocomputing 149, 1155–1161.
Wang, K., Yan, X.P., Yuan, Y.P., et al., 2016a. Real-time optimization of ship energy efficiency based on the prediction technology of working condition. Transp. Res. Part D 46, 81–93.
Wang, S., Peng, J., Liu, W., 2016b. Discriminative separable nonnegative matrix factorization by structured sparse regularization. Signal Process. 120, 620–626.
Zou, H., 2006. The adaptive lasso and its oracle properties. J. Am. Statist. Assoc. 101 (476), 1418–1429.