

Prediction of harbour vessel fuel consumption based on machine learning approach

基于机器学习方法预测港口船舶燃料消耗

期刊作者: Zhong Shuo Chen , Jasmine Siu Lee Lam



主要内容

Main Contents

1

摘要

2

背景

3

方法

4

总结



摘要



- 预测与船舶和气象因素相关的港口船舶的燃料消耗，机器学习模型优于统计线性回归模型（Ridge 回归）。
- 本研究进一步研究了气象因子是否增强了燃油消耗的预测，随机森林模型优于其他模型。
- 气象因子对油耗预测具有综合增值作用，将准确率从 0.7 提高到 38.9%。

用到的算法：

支持向量回归
岭回归
随机森林
人工神经网络



背景



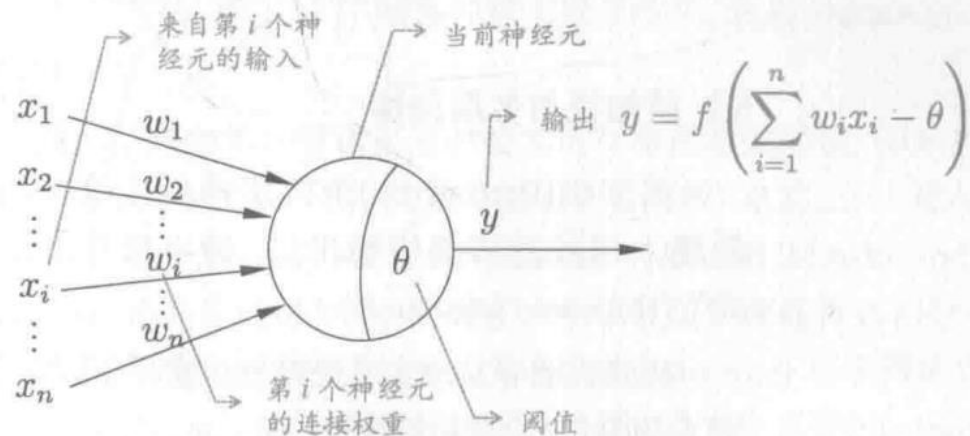


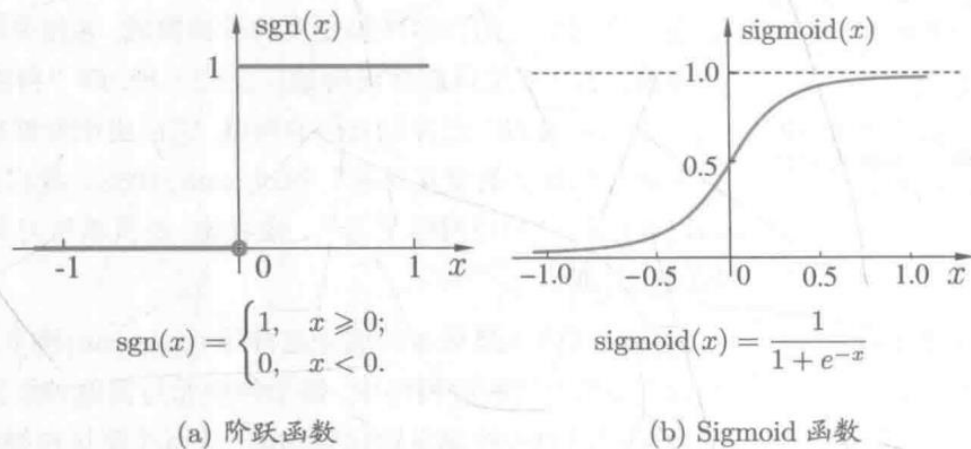
图 5.1 M-P 神经元模型

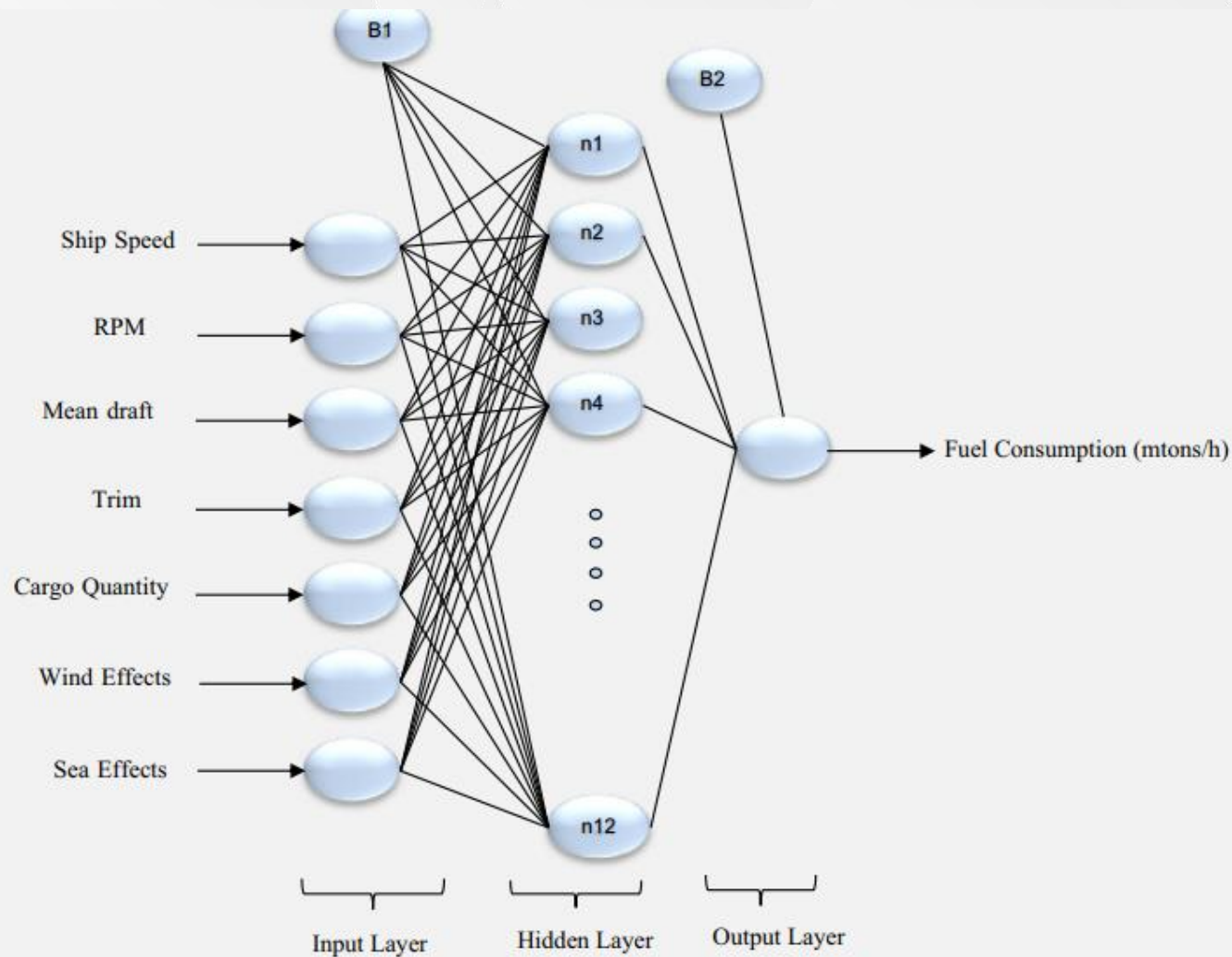
更一般地, 给定训练数据集, 权重 w_i ($i = 1, 2, \dots, n$) 以及阈值 θ 可通过学习得到. 阈值 θ 可看作一个固定输入为 -1.0 的“哑结点” (dummy node) 所对应的连接权重 w_{n+1} , 这样, 权重和阈值的学习就可统一为权重的学习. 感知机学习规则非常简单, 对训练样例 (x, y) , 若当前感知机的输出为 \hat{y} , 则感知机权重将这样调整:

$$w_i \leftarrow w_i + \Delta w_i, \quad (5.1)$$

$$\Delta w_i = \eta(y - \hat{y})x_i, \quad (5.2)$$

其中 $\eta \in (0, 1)$ 称为学习率(learning rate). 从式(5.1) 可看出, 若感知机对训练样例 (x, y) 预测正确, 即 $\hat{y} = y$, 则感知机不发生变化, 否则将根据错误的程度进行权重调整.





$$u_i = \sum_{j=1}^n W_{ij}X_j + b_i$$

$$y_i = f(u_i)$$

$$U_i = C_{1i} \times \text{ship speed} + C_{2i} \times \text{RPM} + C_{3i} \times \text{draft} + C_{4i} \times \text{trim} + C_{5i} \times \text{cargo quantity} + C_{6i} \times \text{wind effect} + C_{7i} \times \text{sea effect} + b_i$$

$$\tanh(U_i) = \frac{1 + e^{-u_i}}{1 - e^{-u_i}} \quad \text{或} \quad g(x) = \frac{1}{1 + e^{-x}}$$

$$K_i = \sum_{l=1}^{n=12} \tanh(U_i) \times lw_i + b_2$$

$$y_i = f(u_i) = \frac{1 + e^{-K_i}}{1 - e^{-K_i}}$$

MR:

$$y = \omega_0 + \omega_1 x_1 + \omega_2 x_2 + \dots + \omega_n x_n$$

该方程通过计算因变量的实际值和因变量的预测值之间的差值，对差值求平方：

$$\hat{\omega} = \underset{(\omega)}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \left(y_i - \omega_0 - \sum_{j=1}^n \omega_j x_{ij} \right)^2 \right\}$$

	MR		ANN	
	Training	Validation	Training	Validation
MSE	0.03	0.038	0.02	0.037
RMSE	0.174	0.196	0.141	0.193

GPR预测由均值函数 $m(x)$ 和协方差函数 $k(x, x)$ 决定，通常定义为： $f(x) \sim GP(m(x), k(x, x))$

考虑到观测到的目标值 y 的噪声，可以建立GPR问题通用模型，即： $y = f(x) + \varepsilon, \varepsilon \sim N(0, \delta_n^2 I_n)$
其中： ε 为噪声， δ_n^2 为噪声方差， I_n 为单位矩阵。

通常，GPR的数据预处理减去平均值，使 $m(x) = 0$ 。通过高斯分布的性质进一步得到观测值 y 和输出样本 y_* 的先验联合分布：

$$\begin{bmatrix} y \\ y_* \end{bmatrix} \sim N \left(0, \begin{bmatrix} k(x, x) + \delta_n^2 I_n & k(x, x_*) \\ k(x_*, x) & k(x_*, x_*) \end{bmatrix} \right)$$

在公式中： y_* 是预测值， $k(x, x_*) = k(x_*, x)$ 是输入样本值 x 与待预测样本输入值 x_* 之间的协方差， $k(x_*, x_*)$ 是要预测的样本输入值的方差。（突出显示的文本是一个方程，它表示两组数据 y 和 y_* 的联合概率分布，假设这两组数据遵循正态分布。假设分布的均值为零。分布的协方差矩阵由块矩阵给出，该方程表示函数值上的先验分布，在观察输入-输出对后，该分布更新为后验分布。然后，可以使用后验分布对新的输入点进行预测。）

从上式中，可以进一步得到预测值 y_* 的后验分布： $y_* | x, y, x_* \sim N(\bar{y}_*, \text{cov}(y_*))$

GPR 的平均预测精度（平均 $R^2 = 0.9887$ ）略高于 BPNN（平均 $R^2 = 0.9817$ ）。然而，与BPNN(平均 $T = 14.7$ s)相比，GPR需要更长的运行时间(平均 $T = 2236.4$ s)。由于运行时间较长，GPR对航路船舶燃料消耗的在线实时预测不太可取。

如何计算:

kernel是给定了，但是kernel里面的超参数(hyperparameters)并没有给定，所谓的learning就是指给定数据之后，我们learning得出这一系列的hyperparameters。

只有当这些hyperparameters都确定了，我们才可以顺利的得到最终的结果。那么具体如何操作有两种方法：

- i) maximum likelihood (ML), 所谓ML就是以给定数据的likelihood最大为[目标函数](#)的[最优化](#)问题，从而估计出kernel里面的hyperparameters.
- ii) Monte Carlo, 这个就是所谓的暴力破解，利用Bayesian formula，结果[高斯先验](#)（也就是我们的多维高斯分布的假设）和likelihood（noise的分布）的知识去强力计算后验的分布。

关于Kernel:它有效的描述了点和点之间的关系，或者说是距离，所以按照其作用kernel原本的名字应该叫“covariance function”

$$k(x, y) = \langle \psi(x), \psi(y) \rangle$$

无论x,y 本身是如何的，维数也好，形式也罢，我们都可以调整这个 ψ 来保证一般的常见的内积定义，而其中这里的 ψ 自然可以看做了一个映射，从一个 R^n 映射到一个一般的希尔伯特空间 V 的映射，而此时内积仍旧是定义在希尔伯特空间 V 的内积。

RR是一种流行的线性回归算法，可以比多元线性回归更好地解决共线性问题。将规范正则化强加到权重可以显著提高线性回归的泛化能力。RR 模型可以描述为：

$$\varepsilon = \|Y - XW\|^2 + \lambda \|W\|^2$$

$$W_{RR} = (X^T X + \lambda I)^{-1} X^T Y$$

$$\widehat{\omega}_{RR} = \underset{(\omega)}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \left(y_i - \omega_0 - \sum_{j=1}^D \omega_j x_{ij} \right)^2 + \lambda \sum_{j=1}^D \omega_j^2 \right\}$$

其中 ε 表示损失函数， λ 表示正则化参数， $\lambda > 0$ 。

Ridge 和 LASSO 回归的主要区别在于惩罚方式。LASSO正则化参数用作绝对值而不是参数的平方:

$$\widehat{\omega}_{LASSO} = \underset{(\omega)}{\operatorname{argmin}} \left\{ \frac{1}{2} \sum_{i=1}^n \left(y_i - \omega_0 - \sum_{j=1}^D \omega_j x_{ij} \right)^2 + \sum_{j=1}^D |\omega_j| \right\}$$

λ 表示正则化参数, $\lambda > 0$

所有的特征都在 RR 中考虑, 而目标方向的子集在 LASSO 中很重要。因此, 在由极高特征组成的数据集中, LASSO 更有效, 因为它提供了稀疏解决方案。然而, 通过RR和导出的估计模型可以进行更成功的估计, Elastic Net是一种源自LASSO和RR的统计方法, 其中进行了绝对正则化和平方正则化。其公式来源于方程式。(4)和(5)如下:

$$\lambda_1 \sum_{j=1}^D |\omega_j| + \lambda_2 \sum_{j=1}^D \omega_j^2$$

超参数 α 和 λ 比率源自 λ_1 和 λ_2 , 用于 Scikit-learn 上的实现, 如下所示:

$$\alpha = \lambda_1 + \lambda_2$$

$$\lambda_{ratio} = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

SVR 使用非线性函数 $\phi(\gamma_i)$ 将输入变量转换为高维空间，并尝试找到一个超平面，可以在容差范围内预测目标。训练样本落入间隔带以内，被认为是正确的。SVR 调整边距以最小化误差 ε ，它可以表示为

$$\varepsilon = C \sum_{i=1}^N (\xi_i + \xi_i^*) + \frac{1}{2} \|\omega\|^2$$

其中约束函数为

$$\psi_i - (\omega * \phi(\gamma_i)) \leq \varepsilon + \xi_i$$

$$(\omega * \phi(\gamma_i)) - \psi_i \leq \varepsilon + \xi_i^*$$

其中 $i = 1, 2, \dots, N$; ξ_i 和 ξ_i^* 是松弛变量; C 表示正则化超参数; ε 表示没有惩罚的区域。该模型通过引入拉格朗日乘子 (Hastie et al., 2009) 转换为

$$f(\gamma, m_i, m_i^*) = \sum_{i=1}^N (m_i - m_i^*) K(\gamma, \gamma_i) + \delta$$

其中 K 表示核函数; m_i 和 m_i^* 表示拉格朗日乘子; 它们可以通过最小化来计算

$$\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (m_i - m_i^*) (m_j - m_j^*) K(\gamma, \gamma_i) - \sum_{i=1}^N \tau_i (m_i - m_i^*) + \varepsilon \sum_{i=1}^N (m_i + m_i^*)$$

在约束 $\sum_{i=1}^N (m_i - m_i^*) = 0$ 下, 其中 $0 \leq m_i, m_i^* \leq C, i = 1, 2, \dots, N$

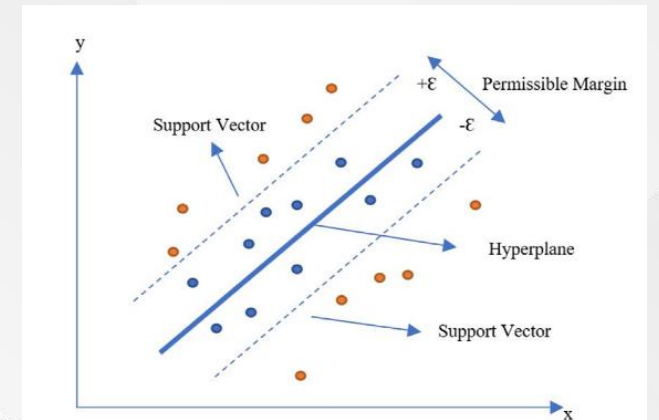


Fig. 3. Support Vector Regression (Smola and Schölkopf, 2004).

随机森林创建多个决策树并将它们组合起来以实现更准确和更稳定的预测。这种结构如图5所示。它以离散的方式给出结果，因为决策是基于树。基于**bagging**方法建立的回归模型的随机森林。在 **bagging** 方法中，通过在样本数据集中再次获取样本并从这些树中提取随机森林来创建新树。

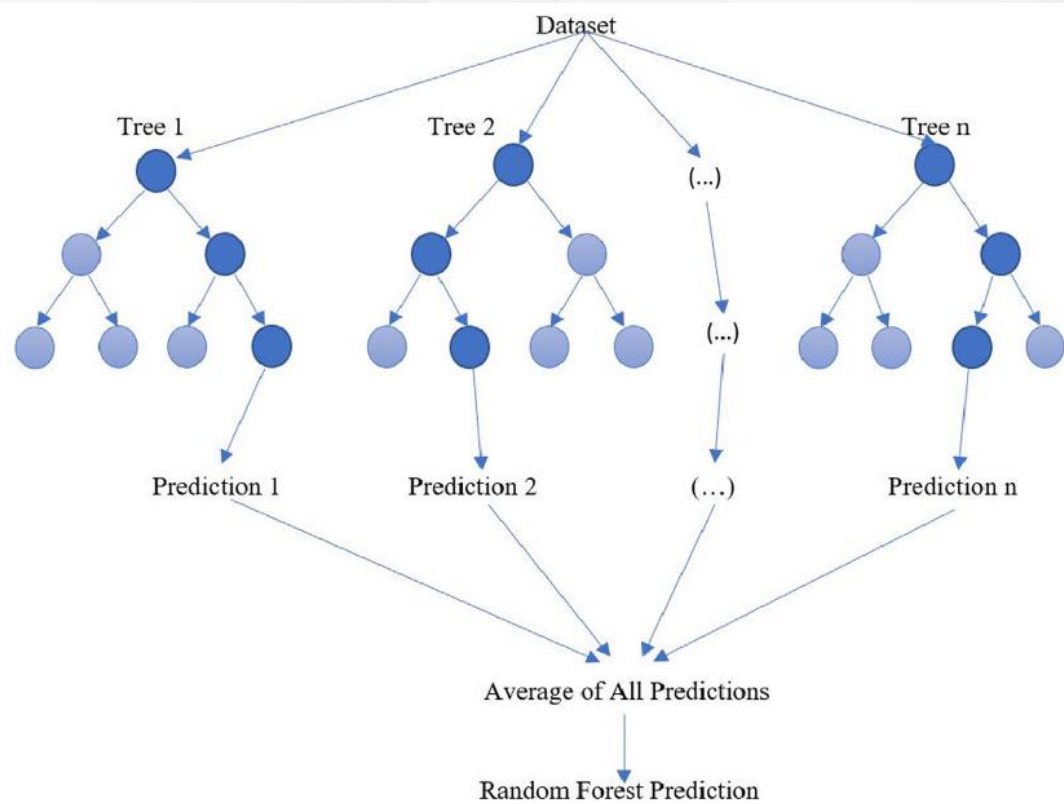


Fig. 5. Random Forest Regression.

- 现有的研究大多集中在海上船舶上，只有少数研究集中在内陆船舶（港口船舶）上
- 本研究调查了气象因素对港口船舶燃料消耗预测的贡献
- 比较了不同类型的机器学习模型和用于燃料消耗预测的统计回归模型



方法

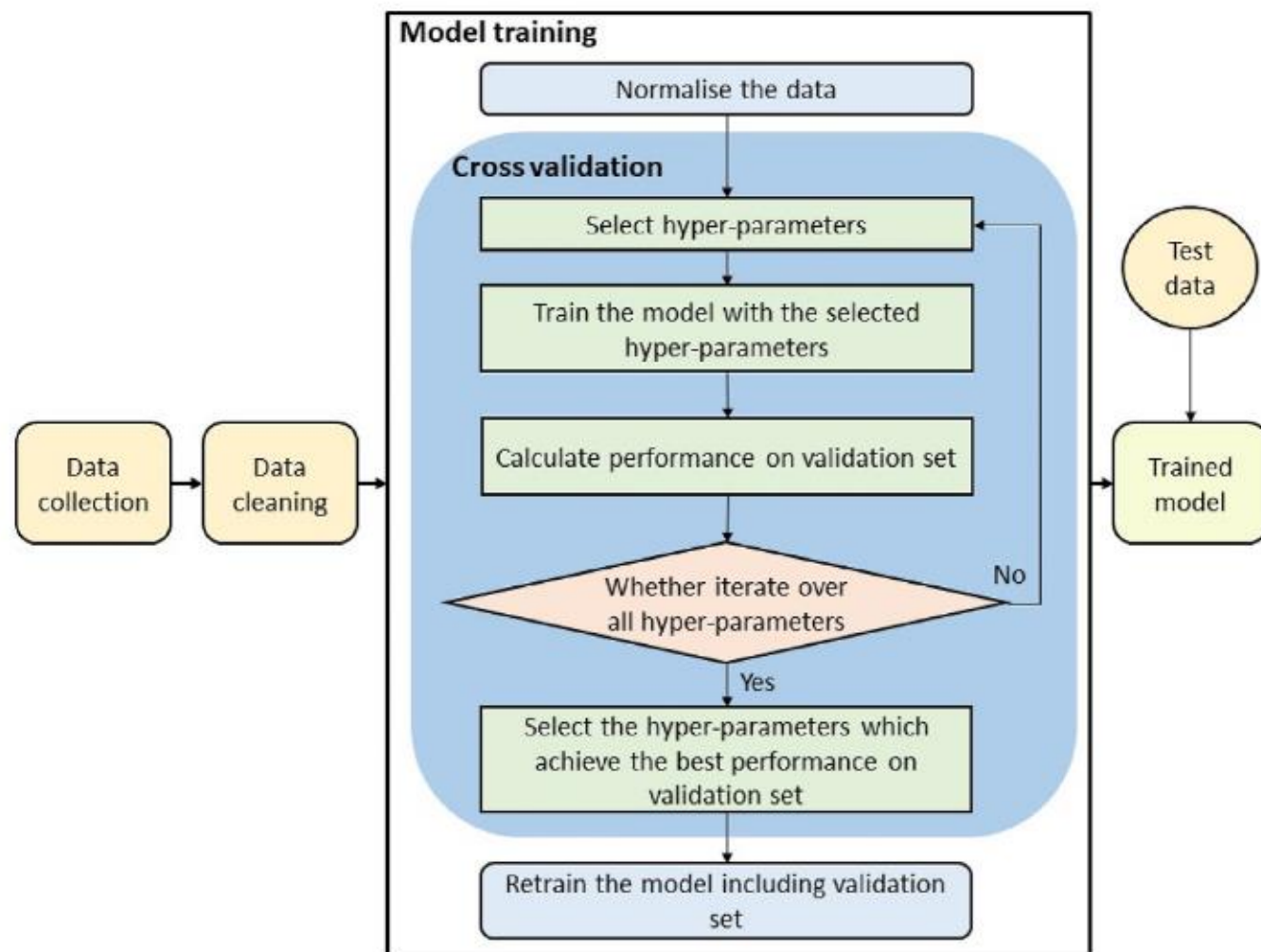


Fig. 1. Model training process.

数据收集:

本研究以新加坡港口水域内的拖船为例，收集自环保局和港务局。AEFC和MEFC通过超声波流量计收集，并通过发动机面板和机载燃油记录进行验证，以确保燃油消耗尽可能准确。

表3给出了分析数据集的一个示例。图3通过对每个数据因子绘制箱形图、his-togram图和小提琴图，给出了整个数据集的分布情况。所有数据在2020年连续14天每小时采集一次。

Table 3
Partial dataset of the analysed tugboat.

Time	10 a.m.	11 a.m.	12 a.m.	...
WD (degree)	226	208	194	...
WS (knots)	6.5	4.1	6	...
TH (metres)	2.1	2.9	3.3	...
TSD (degree)	301	301	121	...
TSR (knots)	1.1	0.5	0.2	...
Temp (°C)	29.8	29.9	29.6	...
RH (%)	66.4	65	65.4	...
Speed (knots)	2.4	0.1	4.5	...
EP (kW)	381.9	0	516.3	...
AEFC (kg)	10.9	10.7	10.6	...
MEFC (kg)	96.8	0	141.4	...

Source: Authors compiled based on collected data.

实际船速 (Speed)、
主机功率 (EP)、
周围温度 (Temp)、
周围相对湿度 (RH)、
辅机油耗 (AEFC)、
主机油耗 (MEFC)、
风向 (WD)、
风速 (WS)、
潮高 (TH)、
潮流方向 (TSD)、
潮流速率 (TSR)。

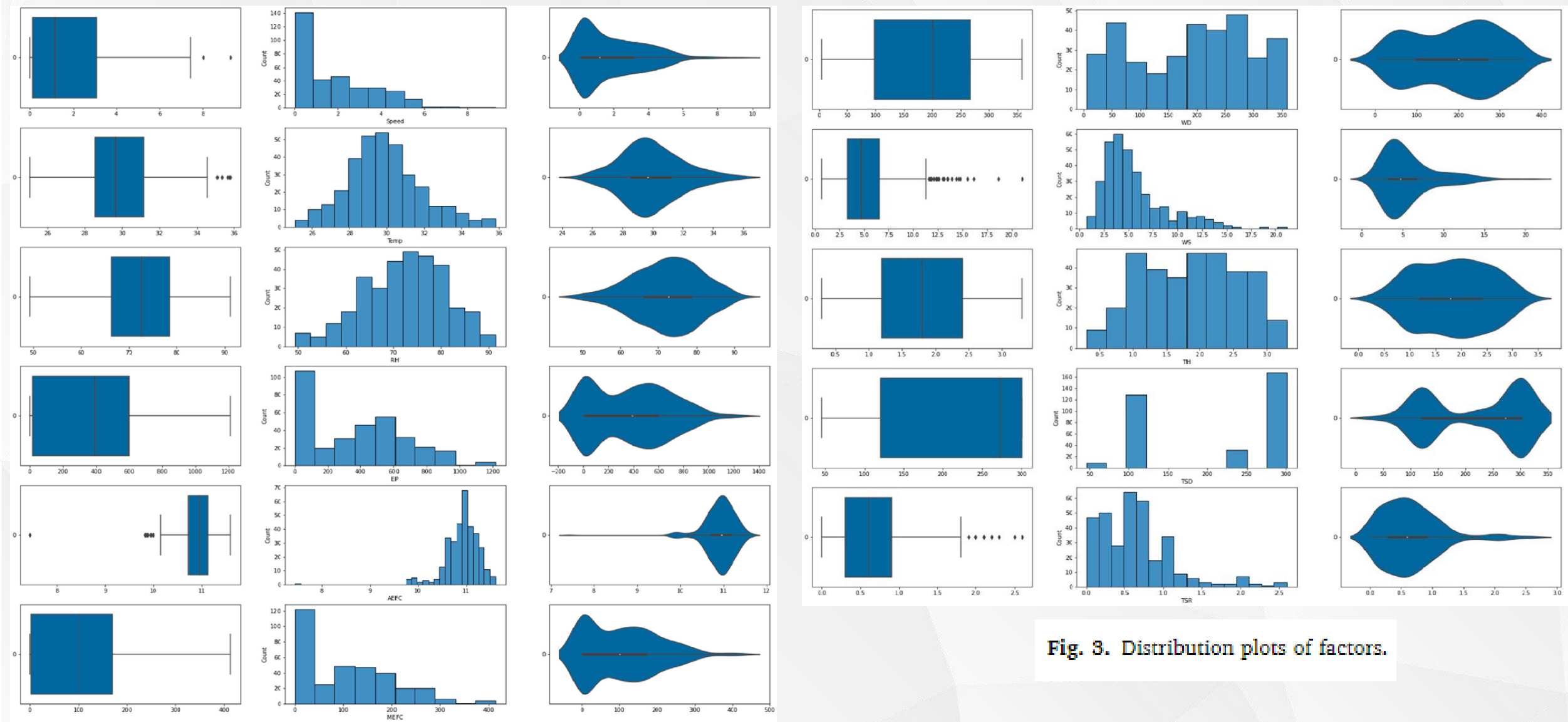


Fig. 8. Distribution plots of factors.

Table 4
Basic statistics of AEFC and MEFC.

Data	Count	Minimum	Maximum	Mean	Standard Deviation
AEFC	334	7.4	11.6	10.9	0.4
MEFC	334	0	413.9	105.0	95.5

辅机的燃油消耗相对稳定，本研究的重点是使用基于机器学习的方法预测主机的燃料消耗。

使用平均绝对误差(MAE)和平均绝对百分比误差(MAPE)来衡量实际值与预测值之间的差异，并评估模型的性能。

$$MAE = \frac{1}{N} \sum_1^N |y_i - y'_i|$$

$$MAPE = \frac{100\%}{N} \sum_1^N \left[\frac{y_i - y'_i}{y_i} \right]$$

r平方(R²)是另一个度量，用于评估因变量的变化有多少可以由回归模型中的自变量解释;取值范围为0 ~ 1。R²的值越接近1，模型越好。调整后的R²可计算为：

$$R^2 = 1 - \left(\frac{N-1}{N-P-1} \right) \left[1 - \frac{\sum_1^N (y'_i - \bar{y})^2}{\sum_1^N (y_i - \bar{y})^2} \right]$$

选择原始数据集的70%作为训练集，10%作为验证集，20%作为测试集。采用交叉验证来评估所有模型中的预测因子，并为每个模型选择最佳超参数。表5总结了超参数的选择范围。

Table 5

Selection ranges of each hyper-parameter.

Hyper-parameter	Selection range
SVR hyper-parameter - C	2^i for i in range $(-5, 0)$
SVR hyper-parameter - ξ	10^i for i in range $(-5, 0)$
SVR hyper-parameter - γ	10^i for i in range $(-5, 0)$
RR hyper-parameter - λ	2^i for i in range $(1, 5)$
RF hyper-parameter - <i>No. of tree estimators</i>	[10, 20, 30, 40]
ANN hyper-parameter - <i>Neurons in the hidden layer</i>	[5, 10, 15, 20, 25, 30]
ANN hyper-parameter - <i>Batch size</i>	[0.1, 0.01, 0.001]
ANN hyper-parameter - <i>Learning rate</i>	[10, 20, 40, 80]
ANN hyper-parameter - <i>Epochs</i>	[16, 32, 64]

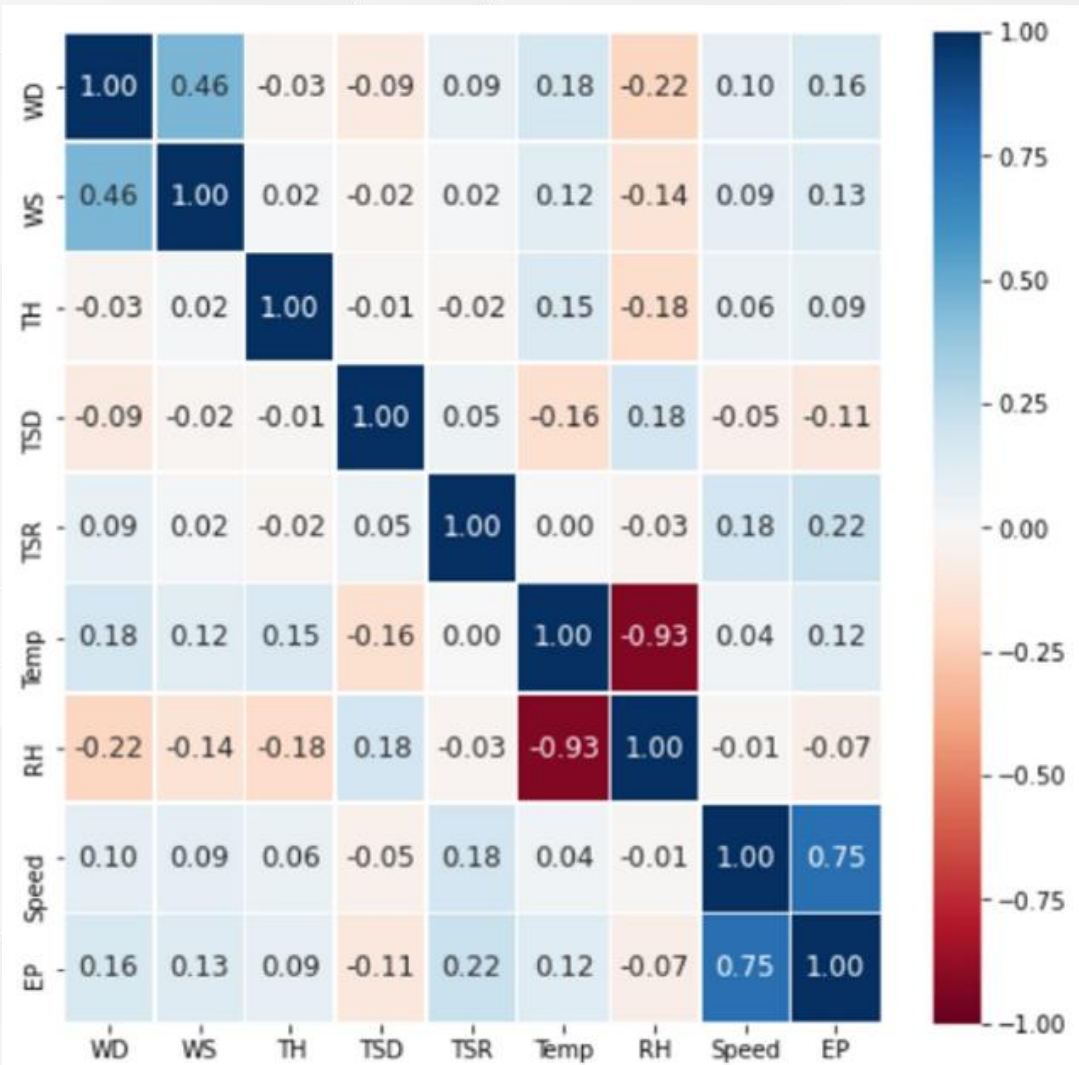


Fig. 4. Correlation heat map of factors.

预测目标为拖船的主机燃油消耗，有两个与船舶相关的因素(航速和主机功率)和七个气象因素可以作为四个模型的输入变量。计算任意两个因素之间的相关性，以提前检验共线性，每对之间的相关性如图4所示。

结果表明，环境温度与相对湿度之间存在很强的相关性。这可以用新加坡的温度和湿度成反比的关系来解释。其他因素之间不存在相关性。

实验结果与拖船的实际情况相一致。当仅使用速度来预测油耗时，精度并不好，这意味着模型不能从单个输入变量(速度)中学习到足够的信息。调整后的R2值从0.442到0.527不等，这表明速度只能代表目标相关变量(油耗)变化的大约一半。当使用EP进行预测时，与仅使用速度相比，这两个误差指标显着降低。调整后的R2值表明，EP更适合表示因变量。

Table 6
Prediction performance using a single ship-related factor.

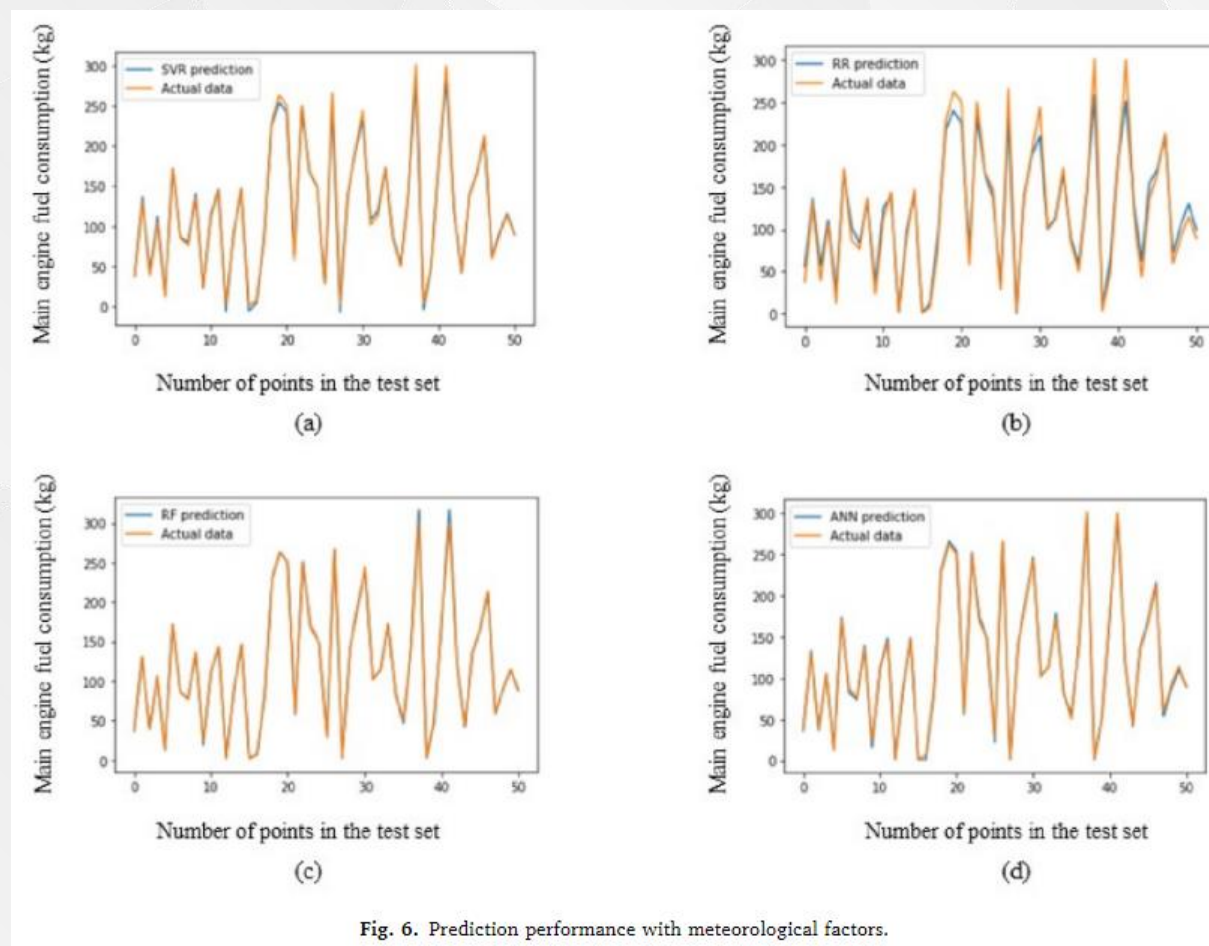
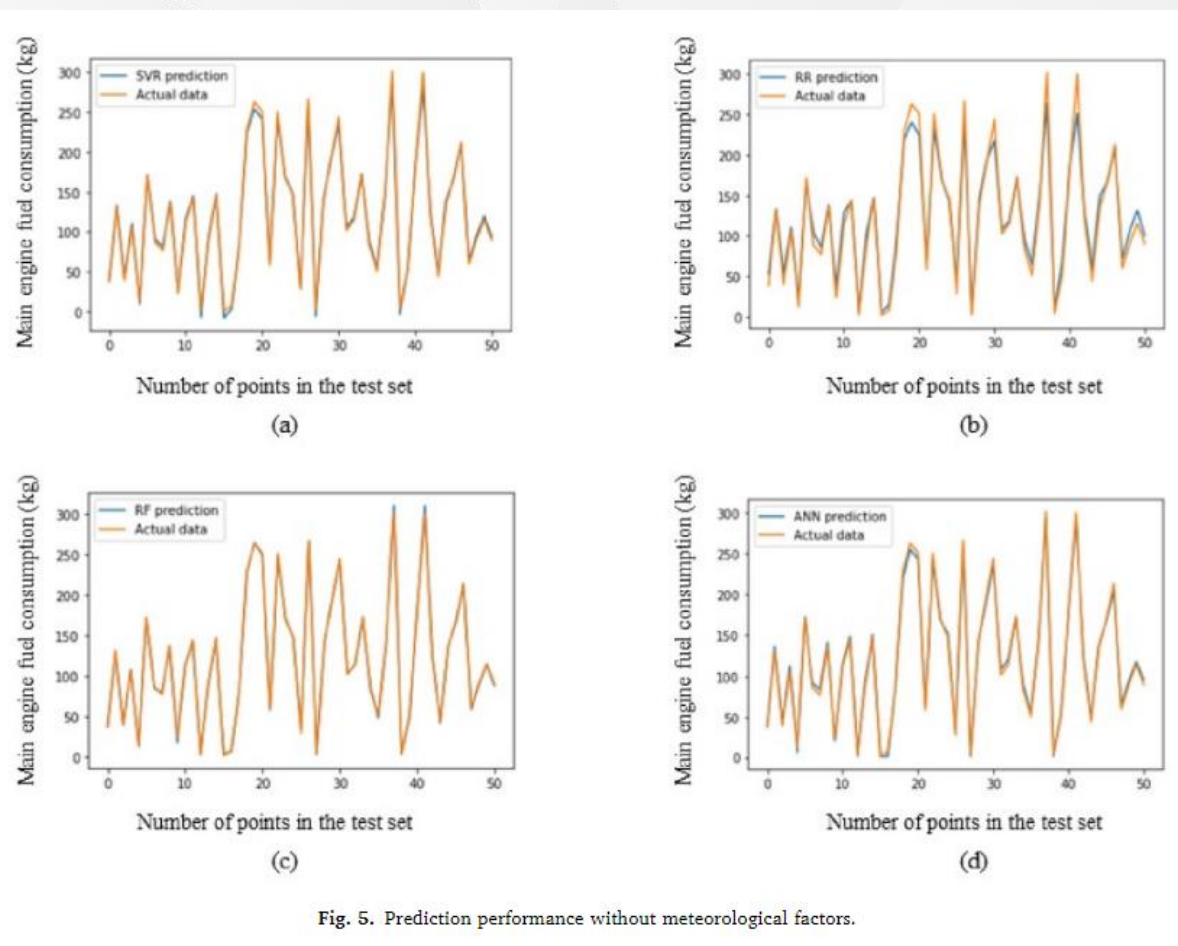
Models	Factors	MAE	MAPE	Adjusted R ²
SVR	Speed	46.896	3.837	0.468
	EP	4.905	0.521	0.990
RR	Speed	49.926	4.660	0.442
	EP	12.301	0.478	0.957
RF	Speed	41.095	0.700	0.521
	EP	1.779	0.083	0.995
ANN	Speed	44.797	2.191	0.527
	EP	3.901	0.102	0.993

通过两组实验考察两组因素对预测的影响。第一组使用两个与船舶相关的因素进行;第二组同时使用所有船舶相关因子和气象因子作为输入变量，如表7和8所示，图5和图6用图形表示。

此外，表6和表7显示，在大多数情况下，使用两个与船舶相关的因素比只使用其中一个因素产生的误差更小，准确性更高。这是因为这两个与船舶相关的因素对于预测拖船的船舶燃料消耗至关重要。

Table 7				
Prediction performance without meteorological factors.				
Models	MAE	MAPE	Adjusted R ²	Best hyper-parameters
SVR	4.469	0.409	0.992	$C = 0.5, \xi = 0.01, \gamma = 0.1$
RR	11.178	0.317	0.960	$\lambda = 2$
RF	1.653	0.078	0.997	No. of tree estimators = 40
ANN	4.336	0.092	0.996	Neurons in the hidden layer = 15, batch size = 16, learning rate = 0.1, epochs = 20

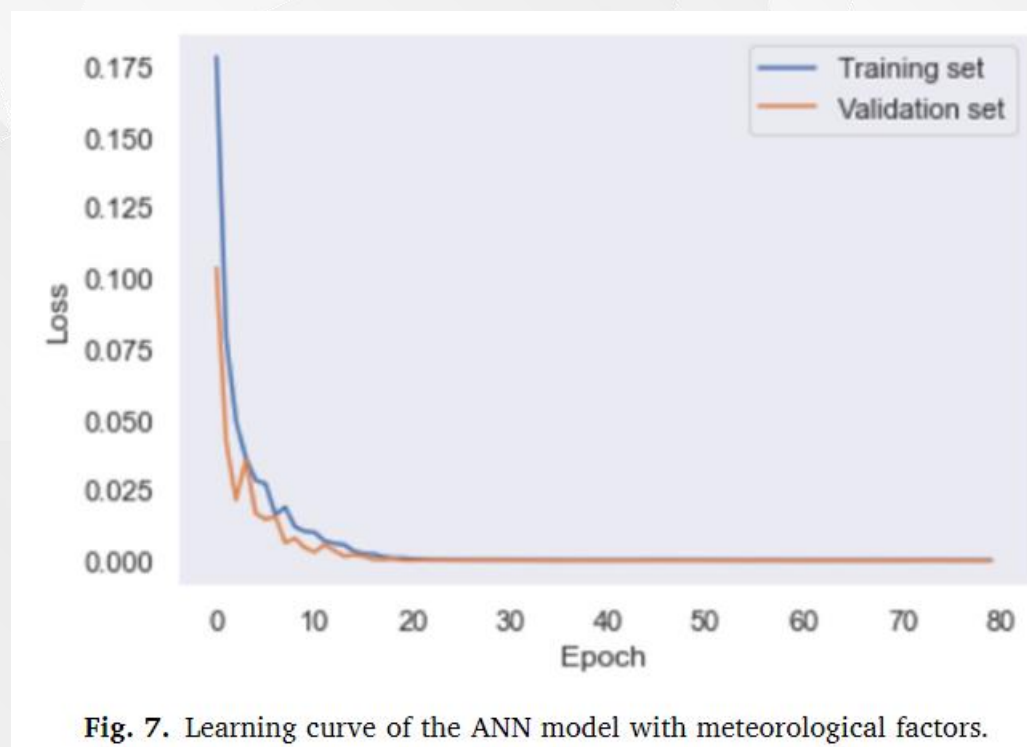
Table 8				
Prediction performance with meteorological factors.				
Models	MAE	MAPE	Adjusted R ²	Best hyper-parameters
SVR	4.145	0.386	0.992	$C = 0.5, \xi = 1e-05, \gamma = 0.1$
RR	11.598	0.194	0.960	$\lambda = 2$
RF	1.642	0.072	0.997	No. of tree estimators = 40
ANN	2.866	0.084	0.997	Neurons in the hidden layer = 20, batch size = 64, learning rate = 0.1, epochs = 80



根据表 7 和表 8 中的误差指标，有或没有气象因素时，从最佳到最差执行的模型都是 RF、ANN、SVR 和 RR。

表 7 和表 8 的比较揭示了气象因素是否有助于提高预测准确性。当添加气象因子作为输入变量时，所有模型都显示出改进。ANN 在四个模型中取得了最显著的改进。在MAE和MAPE方面，ANN的性能分别提高了33.9%和8.7%，这意味着ANN从气象因素中提取重要信息进行预测。

图 7 中的学习曲线表示训练集和验证集的损失减少，这有助于对模型的过拟合问题进行很好的检查。根据图 7 和表 8 中的误差指标，即使是表现最差的 RR 模型也显示出相关目标中超过 96% 的变化。考虑气象变量后人工神经网络的变化覆盖率提高了 0.1%，从而提高了性能。



从图8和图9可以看出，考虑气象变量后，预测值与实际值之间的残差改进也很明显。

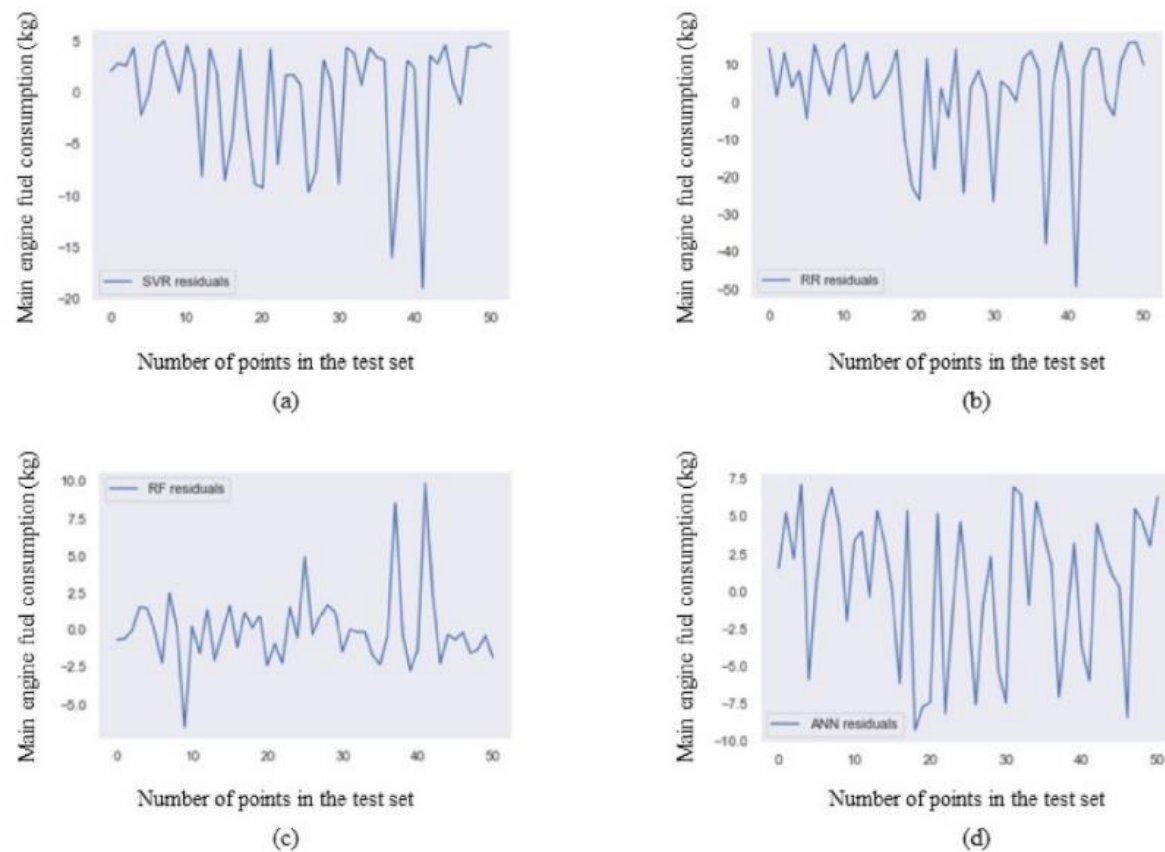


Fig. 8. Prediction residuals without meteorological factors.

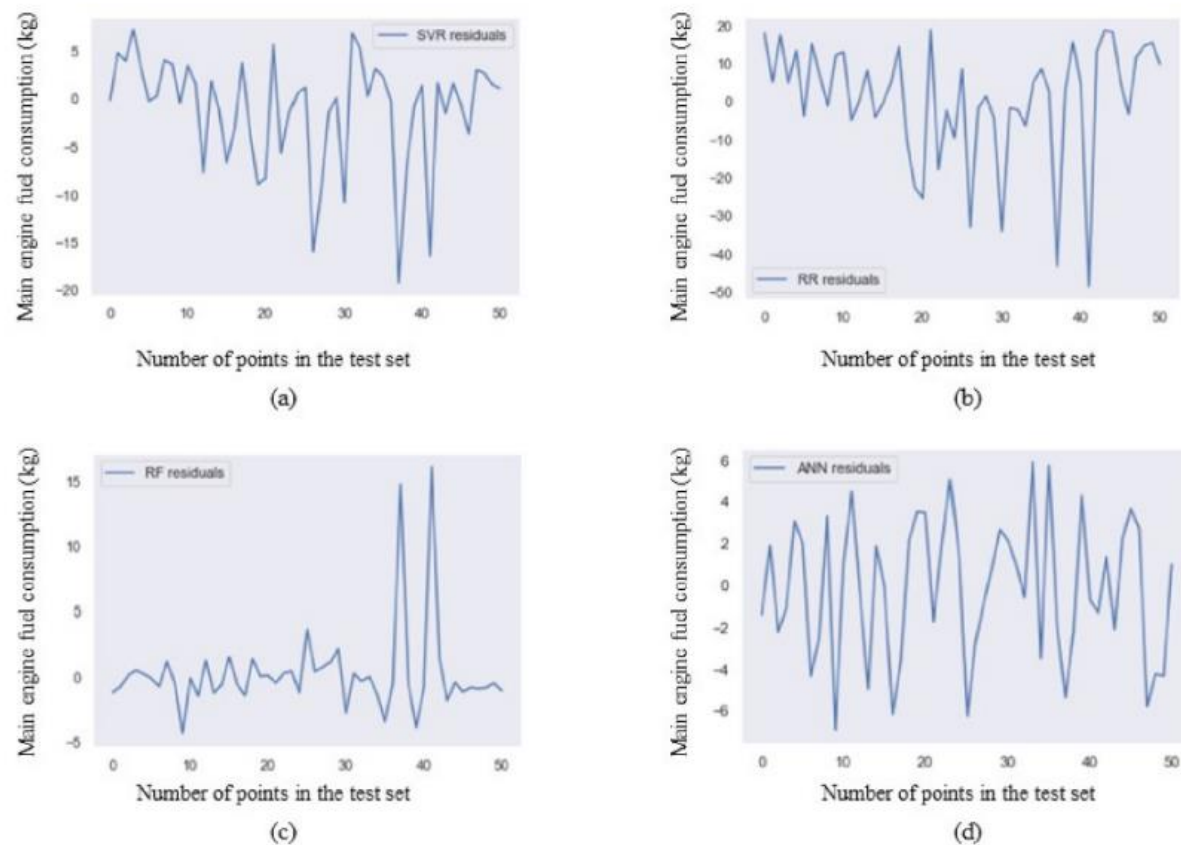


Fig. 9. Prediction residuals with meteorological factors.

Table 9
Prediction performance using a single meteorological factor as an add-on.

Add-on factors	Models	MAE	MAPE
Wind direction (WD)	SVR	4.749	0.362
	RR	11.192	0.265
	RF	1.703	0.083
	ANN	6.910	0.119
Wind speed (WS)	SVR	4.544	0.398
	RR	11.350	0.308
	RF	1.652	0.075
	ANN	7.977	0.537
Tide height (TH)	SVR	4.680	0.412
	RR	11.084	0.262
	RF	1.759	0.081
	ANN	3.376	0.084
Tidal stream direction (TSD)	SVR	4.592	0.399
	RR	11.193	0.329
	RF	1.649	0.081
	ANN	1.300	0.022
Tidal stream rate (TSR)	SVR	3.663	0.435
	RR	10.899	0.279
	RF	1.698	0.082
	ANN	1.561	0.050
Temperature (Temp)	SVR	4.544	0.381
	RR	11.325	0.234
	RF	1.794	0.077
	ANN	2.473	0.070
Relative humidity (RH)	SVR	3.174	0.468
	RR	11.165	0.212
	RF	1.680	0.077
	ANN	1.594	0.052

Table 7
Prediction performance without meteorological factors.

Models	MAE	MAPE	Adjusted R ²	Best hyper-parameters
SVR	4.469	0.409	0.992	$C = 0.5, \xi = 0.01, \gamma = 0.1$
RR	11.178	0.317	0.960	$\lambda = 2$
RF	1.653	0.078	0.997	No. of tree estimators = 40
ANN	4.336	0.092	0.996	Neurons in the hidden layer = 15, batch size = 16, learning rate = 0.1, epochs = 20

研究哪些气象因素有助于改善预测结果，将一个气象因素添加到两个船舶相关因素中，确定每个气象因素对预测的贡献，模型性能结果如表9所示。

表7和表9的比较显示了附加气象因素，有助于提高四个模型的两个误差指标的准确性。换句话说，误差小于没有气象因素的误差。然而，与表 7 中的结果相比，添加 WD 会导致更差的误差指标。

Table 10 Prediction performance after removing a single meteorological factor.			
Removed factors	Models	MAE	MAPE
Wind direction (WD)	SVR	4.218	0.444
	RR	11.569	0.196
	RF	1.797	0.080
	ANN	1.952	0.101
Wind speed (WS)	SVR	4.114	0.402
	RR	11.260	0.181
	RF	1.761	0.076
	ANN	6.274	0.134
Tide height (TH)	SVR	4.676	0.299
	RR	11.812	0.205
	RF	1.981	0.079
	ANN	1.882	0.066
Tidal stream direction (TSD)	SVR	4.061	0.407
	RR	11.536	0.195
	RF	1.689	0.078
	ANN	1.960	0.055
Tidal stream rate (TSR)	SVR	3.913	0.377
	RR	11.570	0.221
	RF	1.685	0.070
	ANN	4.194	0.298
Temperature (Temp)	SVR	4.037	0.461
	RR	11.258	0.192
	RF	1.810	0.079
	ANN	4.366	0.091
Relative humidity (RH)	SVR	4.150	0.463
	RR	11.609	0.197
	RF	1.824	0.075
	ANN	15.846	0.509

Table 8 Prediction performance with meteorological factors.				
Models	MAE	MAPE	Adjusted R ²	Best hyper-parameters
SVR	4.145	0.386	0.992	$C = 0.5, \xi = 1e-05, \gamma = 0.1$
RR	11.598	0.194	0.960	$\lambda = 2$
RF	1.642	0.072	0.997	No. of tree estimators = 40
ANN	2.866	0.084	0.997	Neurons in the hidden layer = 20, batch size = 64, learning rate = 0.1, epochs = 80

我们从所有因素的实验中删除了每次一个气象因素相关因素，结果如表10所示。当比较表8和表10时，当从整个组中删除任何气象因子时，大多数误差都会增加到不同程度的(从0.1%到506.0%)。因此，气象因素可以被认为是共同用于预测的有用因素。

Table 11
Prediction results and performance of partial test set.

Time	1 p.m.	2 p.m.	3 p.m.	4 p.m.	5 p.m.	...
WD (degree)	283	273	267	266	265	...
WS (knots)	11.2	14.4	13.5	11.7	10.2	...
TH (metres)	2.8	2.3	1.5	1.1	0.8	...
TSD (degree)	121	121	119	121	121	...
TSR (knots)	0.8	0.8	0.9	0.6	0.3	...
Temp (°C)	30.2	30.1	31.5	30.2	30.9	...
RH (%)	67.1	69.8	65.1	72.4	72.2	...
Speed (knots)	1.0	4.2	2.9	5.0	0.5	...
EP (kW)	201.5	500.4	586.6	716.4	259.7	...
Actual MEFC (kg)	43.4	135.9	165.9	213.3	59.7	...
Prediction results (RF)	42.6	135.0	165.6	212.8	59.6	...
Prediction errors (RF)	-1.8%	-0.7%	-0.2%	-0.2%	-0.2%	...
Prediction results (RF + meteorological factors)	43.5	136.3	166.3	213.4	59.7	...
Prediction errors (RF + meteorological factors)	2.3%	2.9%	2.4%	0.0%	0.0%	...
Prediction results (ANN + meteorological factors)	42.0	134.3	164.2	211.9	60.8	...
Prediction errors (ANN + meteorological factors)	-3.2%	-1.2%	-1.0%	-0.4%	1.8%	...

总体实验表明，气象因子组提高了每个模型的性能。但是，四个模型的性能顺序没有变化。如果模型不适合预测，则通过添加更多的输入变量来颠覆它，但这种情况下不适用于RR 和 SVR。有了多维输入变量，ANN能更好地训练。当可用气象数据时，模型选择有更大的灵活性。缺乏气象信息时，应谨慎选择合适的机器学习模型。

本研究获得了三个训练有素的模型来预测港口船舶燃料消耗：一个 ANN 模型和两个 RF 模型。表 11 显示了从三个训练有素的模型中获得的预测结果和性能的示例。预测值接近实际燃料消耗。

总结

本文全面构建并讨论了四个模型。RF 比其他三个模型表现更好；本研究发现，将气象因素统称为输入变量可以增强预测结果的准确性。

优点：

- 1)本研究填补了使用机器学习方法来预测港口船舶燃料消耗的文献空白。
- 2)讨论了各种模型之间的差异。
- 3)本研究调查了气象因素对预测的贡献。

缺陷：

- 1)这项研究受到高精度的气象信息的限制。低精度气象信息会导致机器学习模型中的噪声并导致预测结果不准确。未来的研究可以探索更多的影响港口船舶燃油消耗的气象因素。
- 2)机器学习模型的另一个限制是数据集中异常利用点的影响，这可能会导致预测值出现更显著的错误。未来的研究可能会尝试改进模型并解决问题，以实现更好的预测精度。

谢谢！

