

基于加权马尔可夫链修正的 ARIMA 预测模型的研究

郑卓 曹辉 高鹤元 刘如磊

(大连海事大学轮机工程学院 辽宁 大连 116026)

摘要 为改善因频繁事后维修导致的设备使用寿命骤减问题,提出基于加权马尔可夫链修正的差分自回归移动平均(ARIMA)模型来预测设备状态参数,做到事后维修转化为视情维修。考虑到 ARIMA 模型存在一定偏差和不稳定性,引入加权马尔可夫模型对 ARIMA 模型残差序列进行分析。采用状态特征值结合线性插值法将预测的残差状态转化为具体值,残差修正的预测值为最终状态值。以预测船舶海水出口温度为例,对比分析单一 ARIMA 模型和修正的 ARIMA 模型的预测结果。结果表明:修正的模型预测精度较单一 ARIMA 模型显著提高,具备可行性和有效性。

关键词 时间序列 ARIMA 模型 加权马尔可夫模型 马氏检验 残差修正

中图分类号 TP206 文献标志码 A DOI:10.3969/j.issn.1000-386x.2020.12.010

AN ARIMA PREDICTION MODEL BASED ON WEIGHTED MARKOV CHAIN CORRECTION

Zheng Zhuo Cao Hui Gao Heyuan Liu Rulei

(College of Marine Engineering, Dalian Maritime University, Dalian 116026, Liaoning, China)

Abstract In order to improve the rapid reduction of equipment service life caused by frequent post-accident maintenance, a differential autoregressive moving average(ARIMA) model based on weighted Markov chain correction is proposed to predict the equipment state parameters, so that the post-accident maintenance can be transformed into condition-based maintenance. Considering the deviation and instability of ARIMA model, the weighted Markov model was introduced to analyze the residual sequence of ARIMA model. The predicted residual state was converted into a specific value by using the state feature value combined with the linear interpolation method, and the predicted value of residual correction was the final state value. Taking the predicted seawater outlet temperature as an example, the prediction results of single ARIMA model and the modified ARIMA model were compared and analyzed. The results show that the prediction accuracy of modified model is significantly improved compared with the single ARIMA model, and the modified model is feasible and effective.

Keywords Time series ARIMA model Weighted Markov model Markov test Residual correction

0 引言

船舶主海水系统的正常运行,为船舶动力装置提供了充分的冷却条件,为船舶正常航行提供了保障。由于主海水系统直接连通舷外,工作环境恶劣,若能够根据海水系统中某些典型参数的变化进行有效分析预测,做到将事后维修转化为视情维修,对于使主海水冷

却系统时刻处在一个正常运行状态以及延长设备使用寿命,具有重要意义。

目前,在工程上现有的方法主要分为两类:单一模型预测法和组合模型预测法。单一预测模型主要有:神经网络法^[1],灰色预测法^[2],回归预测法^[3],时间序列法^[4]。但是,单一模型在处理非线性、非平稳的时间序列上存在一定的不稳定性和局限性^[5]。因此,目前组合模型预测法更为常见。徐梦茹等^[6]在对地区降雨

量的预测中提出马尔可夫与 ARIMA 组合模型,该模型虽然综合了两个模型的优点,但在模型组合时忽略了组合权重的确定对预测精度的影响。孙堂乐等^[7]在对太阳黑子月均值预测中,提出将原始序列分解为若干子序列,对子序列分别建模预测并叠加重构。该方法由于会因为分解算法的固有缺陷,导致高频分量的非线性和非稳定性仍然较高。

本文提出一种采用加权马尔可夫链修正残差的 ARIMA 预测模型的组合模型。首先建立 ARIMA 模型对原始序列预测。ARIMA 模型虽然能较好地反映序列的随机波动性,但会因为差分的过程导致信息丢失、精度降低的问题,因此,再结合加权马尔可夫链对预测残差进行修正。马尔可夫链适合描述随机波动性大的变量的预测问题,常被用于预测模型的残差修正过程中,在风力发电、降水量等预测中都取得了明显成效^[8]。与单一 ARIMA 模型进行对比实验,结果表明本文提出的预测模型精度更高,实用性更强。

1 模型简介

1.1 ARIMA 模型

ARIMA 模型的主要思想就是将所得的时间序列视为一个随机序列,然后用一个数学模型来近似地描述它。ARIMA 模型通过将非平稳时间序列经过 d 次差分转化为平稳的时间序列进行分析。

ARIMA(p, d, q)模型公式如下:

$$y_t = \theta_0 + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (1)$$

式中: p 为自回归模型阶数; q 为滑动平均模型的阶数; d 为差分次数; $\varphi_i (i=1, 2, \dots, p)$ 、 $\theta_j (j=1, 2, \dots, q)$ 为模型的待定系数; ε_t 为 t 时刻的残差; y_t 为观测值。

1.2 加权马尔可夫链模型

加权马尔可夫链模型的主要思想是将一系列相依的随机序列,经马氏检验后,用相关的各阶(各种步长)确定的加权马尔可夫链来预测未来值。

计算各阶自相关系数 r_k :

$$r_k = \frac{\sum_{l=1}^{n-k} (x_l - \bar{x})(x_{l+k} - \bar{x})}{\sum_{l=1}^n (x_l - \bar{x})^2} \quad (2)$$

式中: r_k 表示第 k 阶(滞时为 k 个时期)的自相关系数; x_l 表示第 l 时段的观测值指数; \bar{x} 表示观测值指数均值; n 表示观测值观测序列的长度。

对各阶自相关系数规范化:

$$w_k = \frac{|r_k|}{\sum_{k=1}^m |r_k|} \quad (3)$$

式中: w_k 为各步长的马尔可夫链的权重; m 为按预测需要计算到的最大阶数。

结合式(2)、式(3)将同一状态的各预测概率加权和作为该状态的预测概率 p_i ,即:

$$p_i = \sum_{k=1}^m w_k p_i^{(k)} \quad i \in I \quad (4)$$

式中: $p_i^{(k)}$ 为结合相应的马尔可夫链转移概率矩阵得出的状态概率。

1.3 加权马尔可夫链修正的 ARIMA 模型

加权马尔可夫链修正的 ARIMA 预测模型的流程如图 1 所示。

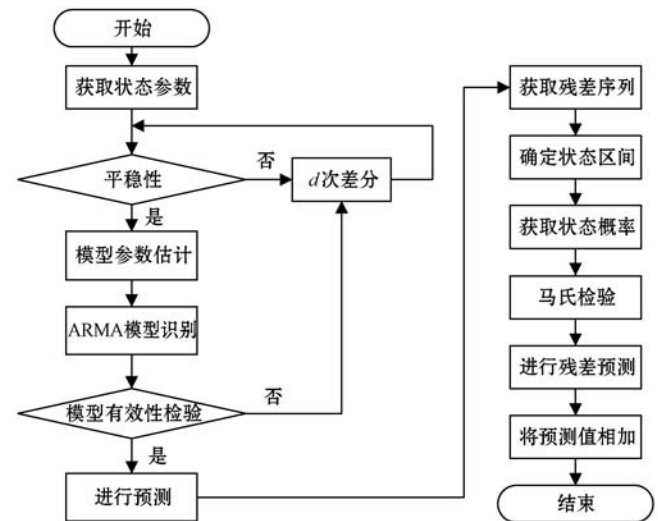


图 1 加权马尔可夫链修正的 ARIMA 模型预测流程图

以 ARIMA 模型作为组合模型预测的第一部分,加权马尔可夫预测模型为第二部分。ARIMA 作为对获取状态参数序列的预测,然后获取预测序列的已知序列部分的残差序列,用加权马尔可夫预测模型预测残差。最后用残差序列修正与其对应的参数序列,得到更为准确的预测值。

2 算例分析

2.1 问题描述与分析

如图 2 所示,在船舶海水系统中,大部分故障都会转移到中央冷却器的换热问题上,如海底门滤器脏堵、海水泵故障、海水管路有渗漏等会导致中央冷却器内的海水量不足,而不变量的低温淡水换热会使中央冷却器内的海水温度升高进而使船舶海水出口温度升高,若中央冷却器换热海水侧结垢,引起换热不足,会使船舶海水出口温度降低至正常温度以下。所以在监

测船舶海水系统运行状态时,海水出口温度是一项重要监测指标。本文以预测“育鲲”轮海水出口温度为具体研究实例展开分析。

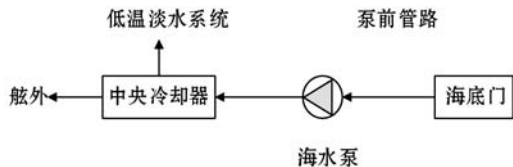


图2 “育鲲”轮海水系统图

2.2 数据获取

本文选取了“育鲲”轮在某航次航行时海水出口温度的数据。具体采集方式为每 2 小时获取一次参数数据,共采集了 6 天的数据即 72 组数据,取前 5 天即 60 组数据作为模型的输入,用于建立预测模型的基础数据,根据这 60 组数据进行建模并预测,后 12 组数据作为预测结果的验证数据^[9]。如果进行长期的故障预测,随着时间的流逝,为了保证模型的准确性,可采取采集的数据顺次更迭的方式,使每次建模时数据总能保持 72 组数据。

2.3 ARIMA 模型预测

(1) 数据预处理。针对这 6 天的海水出口温度历史数据,取前 5 天历史数据作为建模的基本数据,并绘制中央冷却器海水出口温度 T 曲线图,如图 3 所示。

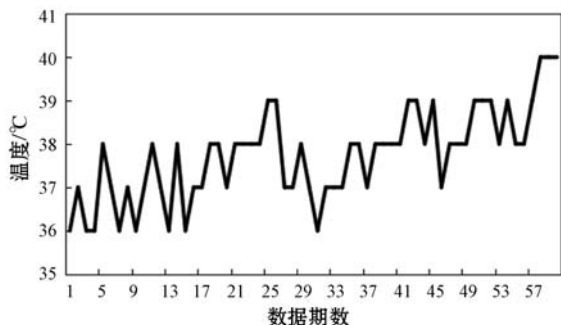


图3 中央冷却器海水出口温度图

可以看出,海水出口温度曲线呈由平稳到逐渐上升趋势。显然这组参数序列是不平稳的,所以应选用 ARIMA 模型进行建模,首先进行平稳化处理,即差分次数的确定。中央冷却器海水出口温度 T 经差分平稳化处理后的图 4 所示。

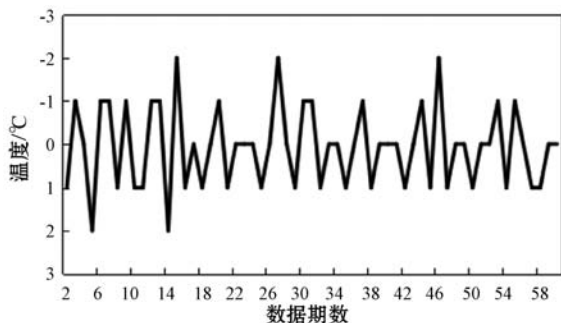


图4 中央冷却器海水出口温度 1 阶差分图

(2) 模型识别。经过时间序列的平稳化处理 and ADF 检验之后,在 MATLAB 中使用 autocorr 和 parcorr 函数得出序列的自相关图和偏自相关图,如图 5、图 6 所示。

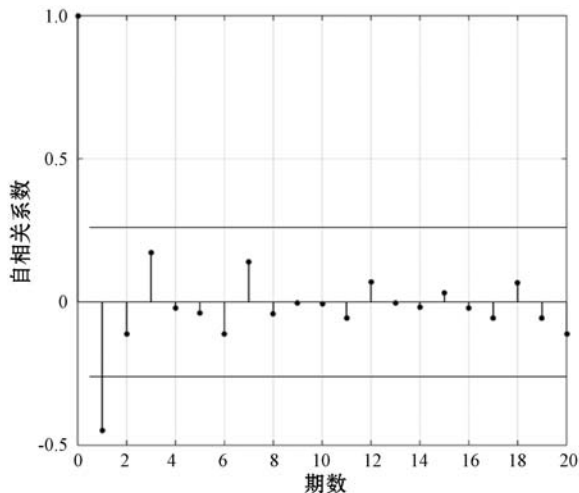


图5 样本自相关函数图

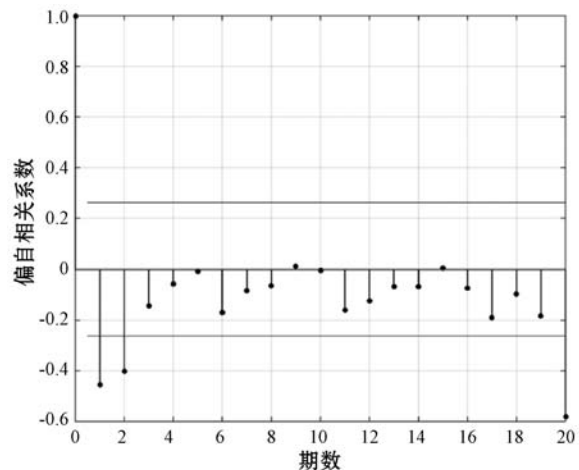


图6 样本偏自相关函数图

图 5、图 6 显示自相关函数和偏自相关函数均拖尾,可以确定该模型应该为 ARIMA(p, q) 模型。在自相关图中,1 阶和 3 阶的函数值可视为显著不为 0,说明 q 可以取到 3 阶,同理在偏自相关图中,1 阶和 2 阶的函数值显著不为 0,说明 p 可以取到 2。

由于仅依靠相关系数图判断模型的阶数具有一定的主观性,本文结合 AIC (Akaike Information Criterion)、SC (Schwarz Criterion) 信息准则进行了误差分析,选出最优模型 ARIMA(2,1,3),如表 1 所示。

表 1 模型检验结果

| (p, d, q) | AIC | SC |
|-------------|---------|---------|
| (1, 1, 1) | 2.461 7 | 2.558 1 |
| (1, 1, 2) | 2.427 8 | 2.556 3 |
| (1, 1, 3) | 2.456 0 | 2.616 6 |
| (2, 1, 1) | 2.463 7 | 2.593 2 |
| (2, 1, 2) | 2.490 7 | 2.652 6 |
| (2, 1, 3) | 2.266 2 | 2.460 4 |

(3) 模型预测。将最优模型在 Eviews 软件中采用最小二乘法进行拟合预测得到接下来的 12 期数据。

2.4 加权马尔可夫模型预测

(1) 马氏检验。检验随机序列是否具有“马氏性”是应用马尔可夫链解决实际条件的重要条件,忽略“马氏性”的检验是不科学的,也是不严谨的。

计算残差序列 Z 的均值近似为 0 和均方差 S 近似为 0.5。根据均值-均方差将序列划分为 6 个状态,如表 2 所示。

表 2 残差状态等级划分

| 状态 i | 划分标准 | 残差区间 |
|--------|--------------------------------|--------------|
| 1 | $[\bar{Z} - 3s, \bar{Z} - 2s)$ | $[-1.5, -1)$ |
| 2 | $[\bar{Z} - 2s, \bar{Z} - s)$ | $[-1, -0.5)$ |
| 3 | $[\bar{Z} - s, \bar{Z})$ | $[0.5, 0)$ |
| 4 | $[\bar{Z}, \bar{Z} + s)$ | $[0, 0.5)$ |
| 5 | $[\bar{Z} + s, \bar{Z} + 2s)$ | $[0.5, 1)$ |
| 6 | $[\bar{Z} + 2s, \bar{Z} + 3s)$ | $[1, 1.5)$ |

因此,可以得到残差序列中从状态 i 经过一步转移达到状态 j 的频数矩阵 f_{ij} 。

$$(f_{ij})_{6 \times 6} = \begin{bmatrix} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 3 & 2 & 4 & 0 \\ 1 & 3 & 1 & 7 & 0 & 0 \\ 2 & 2 & 4 & 6 & 4 & 0 \\ 1 & 3 & 2 & 2 & 2 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

一步转移矩阵为:

$$P_1 = (P_{ij})_{6 \times 6} = \begin{bmatrix} 0.200 & 0.200 & 0.400 & 0.200 & 0 & 0 \\ 0 & 0 & 0.333 & 0.222 & 0.444 & 0 \\ 0.083 & 0.250 & 0.083 & 0.583 & 0 & 0 \\ 0.111 & 0.111 & 0.222 & 0.333 & 0.222 & 0.056 \\ 0.091 & 0.272 & 0.182 & 0.182 & 0.182 & 0.091 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

二步转移矩阵为 $P_2 = P_1^2$, 同理三步、四步、五步、六步转移矩阵分别为: $P_3 = P_1^3$ 、 $P_4 = P_1^4$ 、 $P_5 = P_1^5$ 、 $P_6 = P_1^6$ 。边际概率值为: $P_{\cdot 1} = 0.088$, $P_{\cdot 2} = 0.158$, $P_{\cdot 3} = 0.211$, $P_{\cdot 4} = 0.316$, $P_{\cdot 5} = 0.193$, $P_{\cdot 6} = 0.018$ 根据频数矩阵、一步转移矩阵和边际概率可以得到统计量 $\chi^2 = 2 \sum_{i=1}^5 \sum_{j=1}^5 f_{ij} \left| \ln \frac{P_{ij}}{P_{\cdot j}} \right|$ 计算表,如表 3 所示。

表 3 统计量计算表

| 状态 | $f_{i1} \left \ln \frac{P_{i1}}{P_{\cdot 1}} \right $ | $f_{i2} \left \ln \frac{P_{i2}}{P_{\cdot 2}} \right $ | $f_{i3} \left \ln \frac{P_{i3}}{P_{\cdot 3}} \right $ | $f_{i4} \left \ln \frac{P_{i4}}{P_{\cdot 4}} \right $ | $f_{i5} \left \ln \frac{P_{i5}}{P_{\cdot 5}} \right $ | $f_{i6} \left \ln \frac{P_{i6}}{P_{\cdot 6}} \right $ | 合计 |
|----|--|--|--|--|--|--|--------|
| 1 | 0.824 | 0.236 | 1.284 | 0.457 | 0 | 0 | 2.801 |
| 2 | 0 | 0 | 1.379 | 0.703 | 3.337 | 0 | 5.418 |
| 3 | 0.051 | 1.379 | 0.927 | 4.296 | 0 | 0 | 6.652 |
| 4 | 0.365 | 0.811 | 0 | 0 | 0.348 | 1.099 | 2.622 |
| 5 | 0.036 | 1.640 | 0.293 | 1.104 | 0.119 | 1.645 | 4.837 |
| 6 | 0 | 0 | 0 | 0 | 1.645 | 0 | 1.645 |
| 合计 | 1.276 | 4.066 | 3.882 | 6.559 | 5.449 | 2.844 | 48.051 |

在给定显著性水平 $\alpha = 0.005$ 时,查表可得 $\chi_{\alpha}^2((m-1)^2) = \chi_{0.005}^2(16) = 46.928$, 因为统计量 χ^2 的值为 48.051, $\chi^2 > \chi_{\alpha}^2((m-1)^2)$, 故满足马氏性,接下来可以使用马尔可夫模型预测。

(2) 模型预测。在 MATLAB 中根据式(2)、式(3)可以计算出规范化后的各种步长的马尔可夫链权重,根据第 1 期到第 60 期样本序列的残差序列分别预测第 61 期到第 72 期残差序列,以预测第 61 期为例,如表 4 所示。

表 4 第 61 期残差状态预测

| 初期 | 状态 | 滞时 | 权重 | 状态 1 | 状态 2 | 状态 3 | 状态 4 | 状态 5 | 状态 6 |
|-------------|----|----|-------|------|------|------|------|------|------|
| 55 | 2 | 6 | -1.53 | 0.13 | 0.13 | 0 | 0.25 | 0.50 | 0 |
| 56 | 3 | 5 | -1.15 | 0.09 | 0.18 | 0.09 | 0.45 | 0.18 | 0 |
| 57 | 4 | 4 | -0.95 | 0.18 | 0.29 | 0.12 | 0.24 | 0.18 | 0.06 |
| 58 | 5 | 3 | 1.51 | 0.11 | 0.11 | 0.33 | 0.33 | 0.11 | 0 |
| 59 | 5 | 2 | 2.08 | 0 | 0 | 0.10 | 0.60 | 0.30 | 0 |
| 60 | 5 | 1 | 1.40 | 0.09 | 0.27 | 0.18 | 0.18 | 0.18 | 0.09 |
| P_i (加权和) | | | | 0.60 | 0.98 | 0.82 | 2.05 | 1.45 | 0.15 |

可以看出, $P_i = 2.05$ 对应的状态为 4, 由于预测的需求, 需要将加权马尔可夫预测的状态值转化为具体数值, 夏乐天^[10]认为可以采用随机数学或模糊数学的方法予以推断, 例如采用状态特征值结合线性插值法进行推断。

计算状态特征值公式为:

$$\mu = \sum_{j \in S} j^{\beta} P_j. \quad (5)$$

式中: $\beta > 0$ 为调整因子; S 为状态空间; 通常情况下 $i \leq \mu < i+1$, 令对应状态 i 的区间下限为 \underline{x} , 上限为 \bar{x} , 则预测值 x 可根据下限法求得:

$$x = \bar{x} + (-1)^k (\mu - i) (\bar{x} - \underline{x}) \quad (6)$$

式中:

$$k = \begin{cases} 1 & \text{状态值与指标值增加相反} \\ 0 & \text{其他} \end{cases}$$

本例中, β 取 1.1, 根据式(5)可以算出 μ 为 4.728。

用式(6)可得 61 期残差预测值:

$$x = \bar{x} - (\mu - i)(\bar{x} - x) = -4.7279 \times 0.5 = -0.3639$$

同理剩下的 11 期残差值根据同样的方法预测,但是为了保证实验的准确性与时效性,每预测出新一期的数据就要替换掉一期的样本数据,从第 1 期开始。

2.5 加权马尔可夫模型修正 ARIMA 模型

因为 ARIMA 模型较为简单,只需内生变量而不需要其他外生变量,其实本质上只能预测线性关系,不能预测非线性关系^[11],单一模型预测会存在误差,所以采用马尔可夫模型预测的残差修正 ARIMA 模型:

$$T_i = y_i + x \quad (7)$$

根据式(7)可绘制修正后的预测值、ARIMA 模型的预测值和真实值的曲线拟合如图 7 所示。

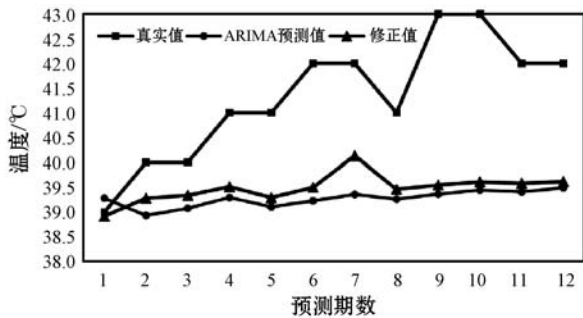


图 7 拟合曲线图

可以看出,被修正的预测值每一期都比 ARIMA 模型预测值更接近真实值。为了更明显地看出修正模型较 ARIMA 模型预测的精度更高,本文分别计算了两个模型预测的平均相对误差,如表 5 所示。

表 5 模型误差对比 %

| 期数 | ARIMA 相对误差 | 修正模型相对误差 | 对比 ARIMA 模型提高 |
|----|------------|----------|---------------|
| 1 | 0.728 | 0.206 | 71.735 |
| 2 | 2.664 | 1.790 | 32.798 |
| 3 | 2.301 | 1.644 | 28.543 |
| 4 | 4.149 | 3.621 | 12.722 |
| 5 | 4.613 | 4.152 | 9.983 |
| 6 | 6.591 | 5.952 | 9.697 |
| 7 | 6.286 | 4.441 | 29.355 |
| 8 | 4.229 | 3.749 | 11.351 |
| 9 | 8.450 | 8.021 | 5.080 |
| 10 | 8.260 | 7.877 | 4.629 |
| 11 | 6.162 | 5.752 | 6.662 |
| 12 | 5.967 | 5.660 | 5.158 |

可以看出,被修正模型的相对误差对比单一 ARIMA 模型相对误差大大降低,平均降低 19%。随着预

测期数的增加,加权马尔可夫对于模型修正的效果也逐渐降低,但是整体模型的预测精度还是达到了一个比较好的效果,所以在预测时并不是预测期数可以无限长,需要选取适当的预测期数,才能达到理想的效果。

3 结 语

本文分别采用加权马尔可夫残差修正的 ARIMA 模型与单一 ARIMA 对“育鲲”轮海水出口温度进行预测,结果显示加权马尔可夫链修正的 ARIMA 模型达到了更好的预测效果,对于 12 期的预测数据,在精度上有显著提升。该模型对海水出口温度预测精度的提高,为早日做到将事后维修转化为视情维修奠定了基础,同样也对船舶其他系统及设备故障预测方面提供很好的借鉴。

参 考 文 献

- [1] 孔国利,王爱菊. 基于改进神经网络的船舶主机温度预测[J]. 舰船科学技术,2018,40(18):85-87.
- [2] 姜日凡,张显库. 基于灰色预测的船舶航向筒捷鲁棒控制[J]. 大连工业大学学报,2018,37(1):63-66.
- [3] 郭卫霞,李红平. 在 VTS 中利用回归分析方法预测研究[J]. 舰船科学技术,2016,38(8):100-102.
- [4] Yang Y, Gao W, Guo C. Aero-engine lubricating oil metal content prediction using non-stationary time series ARIMA model[C]//2017 10th International Symposium on Computational Intelligence and Design,2017.
- [5] 陈静杰,孟琦. 基于两步分解法和 SARIMA 的非饱和机场能耗预测[J]. 计算机应用与软件,2019,36(4):46-50,78.
- [6] 徐梦茹,王学明. 马尔可夫与 ARIMA 组合模型对地区降雨量的预测研究[J]. 计算机应用与软件,2019,36(3):34-37,67.
- [7] 孙堂乐,李国辉. EEMD 与 RBF 神经网络的太阳黑子月均值预测[J]. 计算机工程与应用,2017,53(24):252-256,262.
- [8] 张静,褚晓红,黄学安,等. 一种基于加权马尔可夫链修正的 SVM 光伏出力预测模型[J]. 电力系统保护与控制,2019(19):63-68.
- [9] 孙晓磊,邹永久,张鹏,等. 基于混合预测模型的船舶海水冷却系统状态参数预测[J]. 舰船科学技术,2018,40(8):104-109.
- [10] 夏乐天. 马尔可夫链预测方法及其在水文序列中的应用研究[D]. 江苏:河海大学,2005.
- [11] 吴林川,孙婴婴. 加权马尔可夫链在榆林市降水量预测中的应用[J]. 人民长江,2017,48(S1):82-84,100.