

基于 EMD-ARIMA 组合模型的 长江航运干散货运价指数预测

杨银花¹⁾ 金 雁¹⁾ 汪 敏^{*1)} 张矢宇²⁾

(武汉理工大学船海与能源动力工程学院¹⁾ 武汉 430063) (武汉理工大学交通与物流工程学院²⁾ 武汉 430063)

摘要:针对长江航运干散货运价指数(YBFI)样本点少、周期性不明显、非线性,以及非平稳特性,从分析数据内在波动性出发,提出一种基于经验模态分解(EMD)-差分整合移动平均自回归(ARIMA)组合模型.对比传统 ARIMA 模型、简单季节预测两种方法,EMD 可对 YBFI 序列进行降噪分解,保留数据的内在特性;分解后的序列用 ARIMA 模型、三角函数拟合,效果良好.重组后分析误差,发现该组合预测模型与传统单一模型相比误差较小,预测精度更高.

关键词:长江航运干散货运价指数(YBFI);经验模态分解(EMD);差分整合移动平均自回归(ARIMA);组合预测

中图分类号:U416.4

doi:10.3963/j.issn.2095-3844.2022.05.008

0 引 言

长江航运干散货运价指数(Yangtze River bulk freight index, YBFI)是用来反映我国内河干散货物运输市场货运水平变动情况的重要航运经济指标.由于内河干散货市场的动态性和不稳定性都与沿海干散货市场有较大差异,例如,内河运价指数样本数据点少、周期性不明显且比沿海相关指数更容易受到外界因素的干扰等,沿海干散货相关预测分析方法不适用.现阶段对内河航运干散货物运输市场波动分析及预测的研究较少,国内外学者主要对中国沿海散货运输市场的 CBF、国际(波罗的海)干散货物运输市场的 BDI 和中国集装箱市场变化趋势的 CCFI 等进行了研究.主流的研究方法有:GARCH 模型、小波分析理论、支持向量机(SVM)等. GARCH 模型能够衡量波动的非均衡性,但精度无法保证;小波分析理论缺乏适应性,容易受到基函数选择的约束;SVM 等机器学习方法容易出现过拟合导致预测失效^[1-2]. EMD 是一种经验、直观、自适应的数据处理方法,

能很好地提取非线性和非平稳性数据的时间尺度特征,结合传统预测方法从而克服单一方法的不足.

文中选取 2009 年 5 月—2021 年 4 月 YBFI 月度数据为样本,结合“分解-预测-重构”分析思路,将 EMD 分解后的数据作为样本,从而降低数据噪声和网络复杂性,再对分解序列进行预测,综合得出最终预测值,从而提高 YBFI 预测分析的准确性.

1 模型原理

1.1 YBFI 序列预测思路

YBFI 时间序列具有含噪声、非线性、非平稳和多尺度性.现今对航运指数研究基本是单一算法预测或针对序列整体进行建模^[3],少有学者对数据的内在波动性出发进行预测研究.预测主要思想为“分解-预测-重构”^[4].

采用 EMD 方法对 YBFI 序列进行降噪分解,得到不同频率周期相对稳定的 YBFI 本征模

收稿日期:2022-08-18

第一作者:杨银花(1997—),女,硕士生,主要研究领域为水运系统分析

* 通信作者:汪敏(1974—),女,博士,副教授,主要研究领域为船舶现代化设计方法、水运系统分析、不确定决策

基金项目:绿色智能内河船舶创新专项(42200012)

态分量(intrinsic model function, IMF)和趋势项(residual);运用 ARIMA 模型或者 Holt-Winters 模型对子序列和趋势项进行预测,将各个 IMF 的预测结果进行集成组合即得到最终的 YBFI 预测结果. YBFI 预测模型构建流程见图 1.

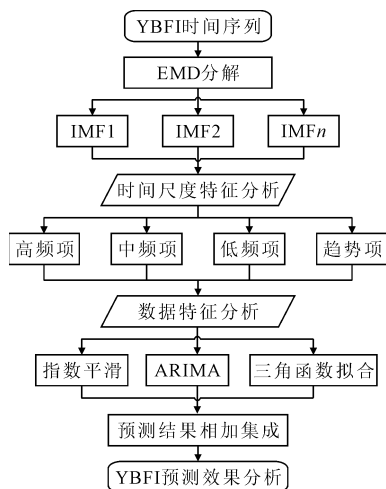


图1 YBFI 预测模型构建流程图

1.2 经验模态分解算法

经验模态分解算法(empirical mode decomposition, EMD)是一种处理非线性、平稳时间序列的前沿时频分析处理方法^[5]. 分解得到不同波动时间尺度的 IMF 需同时满足 2 个条件:①各分量的极值点确定的包络线均值为 0;②各分量通过极值点的个数与通过零点的个数不多于一个. 具体建模步骤见图 2.

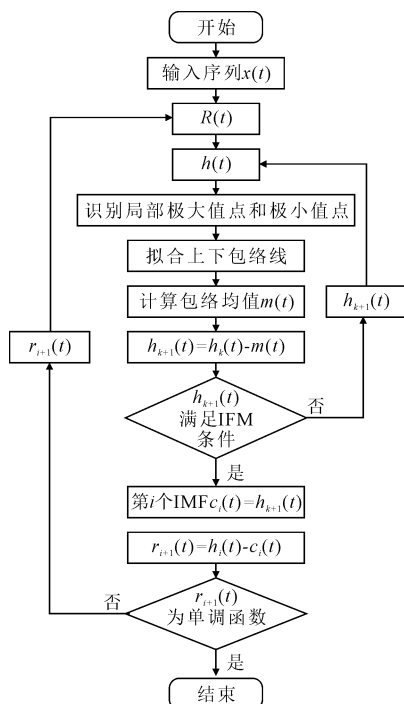


图2 EMD 分解过程

1.3 ARIMA 模型

ARIMA 模型是将自回归过程(AR)与移动平均过程(MA)相结合的时间序列组合模型. 其识别流程图见图 3.

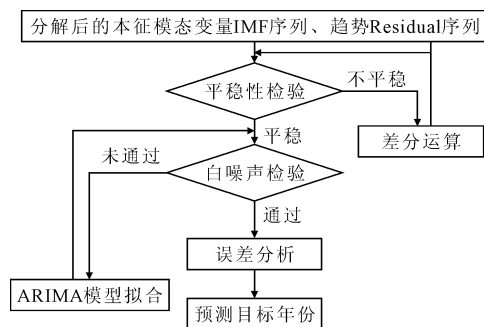


图3 ARIMA 模型识别流程图

该模型的一般表示方式为: $ARIMA(p, d, q) \times (P, D, Q)_S$. 其中: S 为周期步长; d 为非季节部分提取趋势信息所用的差分阶数; D 为季节部分提取趋势信息所用的差分阶数. 乘法模型的实质是通过差分将原始序列化为平稳序列, 再对其进行拟合, 其模型结构为

$${}_d^D \nabla x_t = \frac{\Theta(B)\Theta_S(B)}{\Phi(B)\Phi_S(B)}\epsilon_t$$

式中: $\{\epsilon_t\}$ 为白噪声序列; $\Theta(B)$ 为 q 阶移动平均系数多项式; $\Phi(B)$ 为 p 阶自回归系数多项式; ∇ 为差分算子; B 为滞后算子. 在 ARIMA 建模过程中, 建模的重点在于确定 (p, d, q) 、 (P, D, Q) 的值. 建模的主要思路^[6]为: 分析数据特征, 对数据进行相应变换, 直到通过 ADF 检验获得平稳时间序列为止, 这是 ARIMA 建模的前提条件; 根据 ACF/PACF 图形的峰值个数确定模型的参数, 根据参数建模预测并对预测结果进行评估.

1.4 预测结果误差评估标准

根据统计学误差分析规则, 选择下列 7 个指标作误差评价, 见表 1.

表1 误差评估指标

指标名称	含义	评价准则
R^2	拟合优度, 表征模型预测能力	越接近 1 越好
RMSE	均方根误差, 表征高值预测效果	越小越好
MAPE	平均相对误差绝对值, 表征平稳部分预测效果	一般不超过 10%
MAE	平均绝对误差	越小越好
RE	相对误差, 表征整体误差效果	一般不超过 5%
AIC	$2 \times (\text{模型中参数个数} - \text{模型的最大似然函数值的自然对数})$	越小越好
DW 统计量	durbin watson stat, 检验残差分布是否为正态分布	越接近 2 越好

注: ARIMA 模型在差分过程中会损失部分数据特征, 故其 R^2 一般不会很高, 在 0.3~0.5 则拟合较好.

2 实证过程

2.1 数据来源及样本

由于 2008 年 1 月—2009 年 4 月数据振幅较大,初步预测结果与现实情况不符.为了准确预测长江航运干散货 2021—2030 年的运价指数数据,选取交通运输部长江航务管理局官网公布的从 2009 年 5 月—2021 年 4 月共 144 个样本数据.对数据进行初步调整并绘制时序图,见图 4.

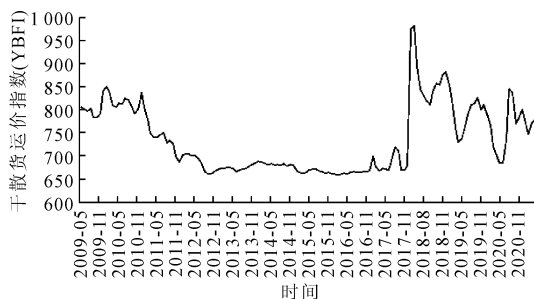


图 4 长江航运干散货运价指数(YBFI)时序图

2.2 YBFI 序列的非线性、非平稳性检验

由图 4 可知:数据呈现出较明显的季节性和波动性,具有明显的非线性特征.在 Eviews 软件中进行 ADF 非平稳性检验,见表 2. t 统计量(-2.893 456)大于不同显著水平下的临界值,在 10%置信水平下不拒绝存在单位根的假设, P 值(0.167 8)大于 0.05,故 YBFI 序列是非平稳序列.

表 2 ADF 非平稳性检验结果

不同水平的临界值	t 统计量	P 值
	-2.893 456	0.167 8
1% level	-4.024 452	
5% level	-3.442 006	
10% level	-3.145 608	

2.3 EMD 分解

EMD 方法相比传统的小波分析、傅里叶分解方法存在一定的优势,EMD 脱离了预定好的基函数的约束,能够最大程度保留时间序列自身的特性. YBFI 序列经 EMD 分解后得到三个 IMF 分量和一个趋势项 Residual,见图 5.

由图 5 可知:IMF1 的波动频率最高,并依次递减,平均振幅也从大到小变动,表现出 YBFI 序列的内在多尺度波动特性.从周期性角度看, IMF1 的周期最小,初步分析是市场基础经济要素对 YBFI 序列的影响; IMF2、IMF3 依次变大,初步分析 IMF2 是政府宏观调控,定期出台的政策规定对 YBFI 的影响,而 IMF3 则可能是航运

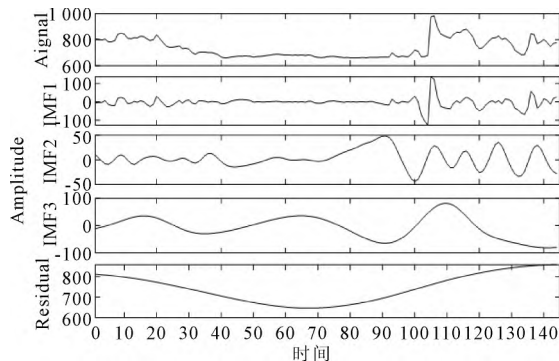


图 5 EMD 分解结果

市场重大事件影响,例如 2008 年金融危机、2020 年新冠疫情等.趋势项自 2008 年(根据官网数据可得)便呈现缓慢下降趋势,到 2014 年末 2015 年初达到谷值,之后开始缓慢增长,查阅有关资料,航运市场中周期为 10~15 年^[7],这与趋势项展示的规律一致.

2.4 预测和重构

通过 EMD 分解将 YBFI 序列分解为相对平稳的 IMF 分量,将三个 IMF 分量运用 ARIMA 模型或 Holt-Winters 指数平滑法进行比较、预测、分析,此类分析主要在 SPSS 和 Eviews 软件中实现.趋势项 Residual 数列采用曲线拟合方法,根据其分布特征和干散货运输市场的周期波动特性,运用 MATLAB 2018a 实验平台进行拟合研究,通过训练,发现三角函数对 Residual 数列的拟合效果较好.

1) IMF1 拟合模型 通过拟合发现 ARIMA(2,0,10)×(2,0,0)₁₂ 能够很好的反映 IMF1 序列的波动特性.其拟合优度 R^2 为 0.483(接近 0.5,拟合较好),均方根误差 RMSE 仅为 21.093; DW 统计量为 2.032; AIC 值为 9.034.

2) IMF2 拟合模型 通过比较 ARIMA 模型、指数平滑模型以及专家建模结果,发现 AR-MA(4,7)能对 IMF2 的周期、波动特性进行很好的预测.其 R^2 达到了一般线性拟合模型 R^2 ,为 0.996,接近 1;均方根误差 RMSE 仅为 1.257;平均绝对误差 MAE 仅为 0.803; DW 值为 1.672; AIC 值为 3.102.

3) IMF3 拟合模型 通过比较拟合发现 AR-MA(6,1)能对 IMF3 的特征数列进行较好的拟合.在该模型拟合过程中, R^2 为 0.998,几乎等于 1;均方根误差 RMSE 与平均绝对误差 MAE 均处于较低区间, DW 值为 1.912,说明拟合精度较优.

4) Residual 拟合模型 趋势项主要反映了

YBFI 序列在发展过程中的整体趋势和变化周期. 经过在 MATLAB 实验平台中反复验算, 调参, 最终确定该序列的较优拟合、预测模型方程式为

$$\begin{aligned} \text{Residual_PRE}_t = & 778.2 + 14.49 \times \\ & \cos(t \times 0.02786) - 95.89 \times \sin(t \times 0.02786) + \\ & 24.64 \times \cos(2 \times t \times 0.02786) + \\ & 22.71 \times \sin(2 \times t \times 0.02786) - 3.695 \times \\ & \cos(3 \times t \times 0.02786) + 0.4971 \times \\ & \sin(3 \times t \times 0.02786) \end{aligned}$$

式中: Residual_PRE_t 为第 t 期趋势项 Residual 的预测值; t 为时间序号, 记 2009 年 5 月为 $t=1$; 依次往后 $t=2, 3, \dots, n$.

图 6 为 Residual 的拟合预测模型效果和残差分析图.

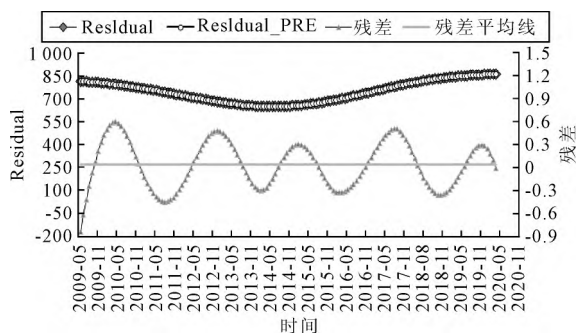


图 6 Residual 拟合与残差分布

从残差分布曲线, 发现拟合效果较好, 残差基本在 $[-0.95, 0.60]$ 之间, 且平均残差仅为 0.036 09, 与 0 无明显差异.

5) YBFI 预测结果 由式(1), 将上述经验模态分量和趋势项求和即可得到相应年份 YBFI 预测结果, 见图 7.

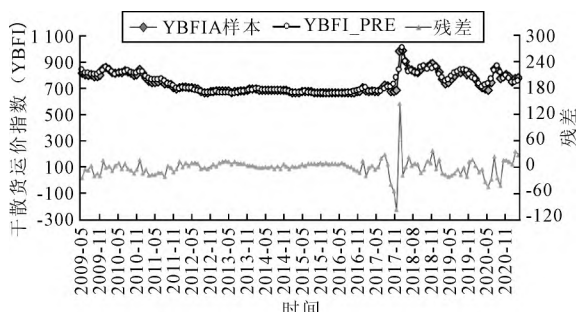


图 7 YBFI 拟合效果图

预测值略有滞后, 但基本能够反映样本波动规律和变化趋势. 通过计算, 得到残差均值仅为 0.312 8 (与 0 无明显差异), 说明总体误差水平较低; 均方根误差 $RMSE$ 为 20.317, 若以基期 1 000 点算, 均方根误差仅为 2%, 说明高值部分预测效果较好; 平均相对误差绝对值 $MAPE$ 为 1.62%, 在 10% 范围内, 表明平稳部分预测效果较好; 平

均绝对误差 $MAE=12.184$; 相对误差 RE 中仅有五组数据超过 5%, 2 组数据超过 10%, 占样本总数的比例分别为 0.034 7、0.013 9, 可忽略不计, 故可判断整体误差较低; R^2 为 0.973 5, 说明该模型对 YBFI 序列的预测效果较准确.

2.5 结果分析

将 EMD-ARIMA 组合模型相关参数与传统、单一的季节性预测方法^[8]进行对比, 见表 3.

表 3 EMD-ARIMA 组合模型及其对比模型预测误差比较

指标	EMD-ARIMA 组合模型	ARIMA(4,1,4)× (2,0,0) ¹²	简单季节 模型
R^2	0.974	0.846	0.824
$RMSE$	20.317	29.804	30.683
$MAPE$	1.620	1.672	2.057
MAE	12.184	12.393	15.882

总体来看, 相比 ARIMA 模型和简单季节模型, EMD-ARIMA 组合模型对 YBFI 序列的预测表现更优, 说明 EMD 分解能有效提取含噪声、非线性、非平稳性数据的时间尺度特征, 通过对本征模态变量 IMF 和趋势项 Residual 的预测结果进行重新组合, 可以得到准确的 YBFI 预测结果.

该模型对未来年份 YBFI 预测结果 (部分) 见表 4.

表 4 未来年份预测结果

预测年	月份	YBFI 预测值
2022	5	808.268
2024	8	799.095
2026	11	764.941
2028	2	760.193
2030	4	754.901

由表 4 可知: 2022—2023 年, YBFI 将达到一个较高水平, 之后又缓慢下跌, 在不考虑“通货膨胀”等因素的情况下, 预计 2030 年的 YBFI 将处于一个较低的水平. 根据我国内河运输的长期发展现状来看, 在未来很长一段时间内, 若没有较大的技术突破, 干散货航运市场将维持原来的周期波动规律, 整体略有上涨趋势.

3 结 论

1) EMD 模型能有效分解 YBFI 序列, 在保留数据自身的内在波动特性的同时对不同波动频率数据进行分解, 为后续预测操作提供周期较稳定的分解序列.

2) EMD-ARIMA 组合模型的相对误差 RE 中超过 5% 的样本仅占样本总数的 0.0347, 其整体误差水平基本在 5% 内.

3) EMD-ARIMA 组合模型预测效果和精度均优于传统单一的时间序列预测方法。在预测精度和拟合优度方面提高了将近 15%, 平稳部分和中高值预测效果均得到较大幅度提升, 其 RMSE、MAPE、MAE 值(相比单一 ARIMA 模型)分别下降了 31.83%、3.11%、1.69%。

本方法可利用有限数量样本, 较好模拟出长江干散货运价指数周期性不明显、非线性、非平稳的波动特性, 为长江干线散货运输市场发展态势预测提供支持。

参考文献

- [1] 贾红雨, 周晨昕, 王宇涵, 等. 中国沿海散货运价指数波动特性研究[J]. 统计与决策, 2021, 37(3): 158-161.
- [2] ZHAI M M, LI W H, TIE P, et al. Research on the predictive effect of a combined model of ARIMA and neural networks on human brucellosis in Shanxi Province, China: a time series predictive analysis[J]. BMC Infectious Diseases, 2021, 21(1): 56-61.
- [3] 贾红雨, 周晨昕, 王宇涵, 等. 基于 EEMD-PSO-LSS-VM 的中国沿海散货运价指数预测[J]. 大连海事大学学报, 2020, 46(1): 107-113.
- [4] 汤霞, 匡海波, 郭媛媛, 等. 基于 VMD 的中国出口集装箱运价指数分析与组合预测[J]. 系统工程理论与实践, 2021, 41(1): 176-187.
- [5] 周晨昕. 基于加权组合模型的沿海散货运价指数预测研究[D]. 大连: 大连海事大学, 2020.
- [6] 徐映梅, 陈尧. 季节 ARIMA 模型与 LSTM 神经网络预测的比较[J]. 统计与决策, 2021, 37(2): 46-50.
- [7] 黄辉. 航运市场的周期理论和发展趋势预测[J]. 水运管理, 2016, 38(2): 8-11.
- [8] 王燕. 时间序列分析: 基于 R[M]. 北京: 中国人民大学出版社, 2015.

Prediction of the Yangtze River Shipping Dry Bulk Freight Price Index Based on the EMD-ARIMA Combined Model

YANG Yinhua¹⁾ JIN Yan¹⁾ WANG Min¹⁾ ZHANG Shiyu²⁾

(School of Energy and Power Engineering, Wuhan University of Technology, Wuhan 430063, China)¹⁾

(School of Transportation and Logistics Engineering, Wuhan University of Technology, Wuhan 430063, China)²⁾

Abstract: According to the characteristics of Yangtze River Shipping Dry Bulk Freight Index (YBFI), such as few sample points, unobvious periodicity, nonlinearity and non-stationarity, a combined model based on empirical mode decomposition (EMD) and differential integration moving average autoregressive (ARIMA) was proposed from the perspective of analyzing the inherent volatility of data. Compared with traditional ARIMA model and simple seasonal prediction, EMD can denoise and decompose YBFI series, and keep the intrinsic characteristics of data. The decomposed sequence is fitted by ARIMA model and trigonometric function, which has a good effect. After the error analysis, it is found that the combined forecasting model has smaller error and higher forecasting accuracy than the traditional single model.

Key words: Yangtze River bulk freight index (YBFI); empirical mode decomposition (EMD); differential integrated moving average autoregressive (ARIMA); combination prediction