

DOI: 10.16783/j.cnki.nwnuz.2020.01.004

一种基于经验模态分解的时间序列预测方法

马宇红^{1,2}, 强亚蓉¹, 杨梅¹

(1. 西北师范大学 数学与统计学院, 甘肃 兰州 730070;

2. 西北师范大学 学报编辑部, 甘肃 兰州 730070)

摘要: 针对来源于实际问题的时间序列非线性、非平稳、多尺度复合的特点建立了一种基于经验模态分解(EMD)的 ARIMA 时间序列预测模型, 即 EMD-ARIMA 模型。首先, 借助经验模态分解将时间序列分解为多个不同时间尺度的内在模函数和一个趋势项, 并确定每个内在模函数的季节性趋势; 其次, 对每个内在模函数使用季节性 ARIMA 模型进行预测, 对趋势项使用趋势移动平均模型进行预测; 最后, 将所有内在模函数和趋势项的预测结果进行复合得到原时间序列的预测结果。数值实验结果表明, EMD-ARIMA 方法能够揭示真实时间序列内在的多尺度复合特征和季节性变化规律; 与经典的 ARIMA 模型和人工神经网络(ANN)模型相比, EMD-ARIMA 模型明显提高了预测精度, 因而是一种可靠的非线性、非平稳时间序列预测方法。

关键词: 时间序列; 经验模态分解; 内在模函数; 季节性 ARIMA 模型

中图分类号: TP 181

文献标志码: A

文章编号: 1001-988X(2020)01-0027-08

A time series prediction method based on empirical mode decomposition

MA Yu-hong^{1,2}, QIANG Ya-rong¹, YANG Mei¹

(1. College of Mathematics and Statistics, Northwest Normal University, Lanzhou 730070, Gansu, China;

2. Editorial Department of the University Journal, Northwest Normal University, Lanzhou 730070, Gansu, China)

Abstract: A novel time series prediction model based on empirical mode decomposition (EMD) and ARIMA model is established according to the characteristics of nonlinear, nonstationary and multiscale composite for a large amount of time series existing in practical problems. Firstly, a time series is decomposed into several intrinsic mode functions (IMF) with different time scales and a trend term based on EMD method, and the periodic of each IMF is determined. Secondly, seasonal autoregressive integrated moving averaging (ARIMA) model is used to predict every IMF, while trend moving average (TMA) model is used to predict the trend term. Finally, the prediction results of all of sub-time series are combined to obtain final prediction results of the original time series. The experimental results show that EMD-ARIMA method not only reveals the characteristics of intrinsic multi-scale composite and laws of seasonal variation of actual time series, but also significantly improves the prediction accuracy of time series compared with the classical ARIMA model and artificial neural network (ANN) model, so it is a reliable prediction method for nonlinear and nonstationary time series.

Key words: time series; empirical mode decomposition (EMD); intrinsic mode function (IMF); seasonal ARIMA model

收稿日期: 2019-07-10; **修改稿收到日期:** 2019-09-16

基金项目: 国家自然科学基金资助项目(51368055)

作者简介: 马宇红 (1971—), 男, 甘肃天水人, 副编审, 博士。主要研究方向为时空序列分析及预测。

E-mail: mayh@nwnu.edu.cn

0 引言

20 世纪以来,科学技术的飞速发展促进了空间观测技术、计算机技术和网络通信技术等取得了重大突破。随着地理信息系统、遥感测绘技术和全球定位系统的发展和广泛应用,社会各领域已经积累了海量的地理时空数据,如经济统计数据、卫星影像数据、环境监测数据、交通流量数据等^[1]。时空数据中蕴含丰富而复杂的时间和空间关系,所以深入分析和挖掘时空数据隐含的内在规律,构建有效的时空数据分析和预测模型,不仅具有重要的学术价值,而且对于揭示事件发展的时空演化规律、有效指导行业生产也具有一定的科学意义和现实的经济效益。

时间序列分析和预测是时空数据分析和预测的基础。传统的数据分析方法基本都基于线性和平稳性假设,如傅里叶变换等,近年来发展起来的小波分析和 Wagner-Ville 分布也只能处理线性非平稳数据,然而源于实际生活的绝大多数时间序列数据均具有明显的非线性、非平稳特征,这类数据的分析与处理是一个非常具有挑战性的问题。

一个时间序列通常由几种变化形式叠加或耦合而成,如长期趋势变动、季节性变动、周期变化和不规则波动等,最简单的时间序列预测方法有移动平均法、指数平滑法、自适应滤波法和趋势外推法等。现代时间序列分析起源于 1927 年英国统计学家 Yule^[2]提出的自回归(auto-regressive, 确 AR)模型,该模型与 1931 年英国统计学家 Walker^[3]提出的移动平均(moving average, MA)模型和 ARMA 模型,构成了时间序列分析的基础。

时间序列模型主要有平稳模型和非平稳模型两类,平稳模型假设过程保持概率特性上的统计均衡性,不随时间而改变,即变化都在一个固定的均值水平上,具有固定的方差^[4]。时间序列的平稳变换是 ARMA 模型预测的基础。Salles 等^[5]的研究表明,合适的非平稳时间序列变换方法对超过 50%的时间序列预测精度提高 30%以上,对 10%的时间序列预测精度提高 95%以上。ARMA 模型属于平稳时间序列模型,而 ARIMA 模型属于非平稳模型,它先对时间序列做差分,使之变为平稳序列,然后再用 ARMA 模型进行预测。

1927 年,Box 等^[4]建立了时间序列分析、预测以及 ARIMA 模型识别、参数估计和诊断检测的

系统方法。模型识别就是确定自相关过程和移动平均过程的阶数,一般根据时间序列自相关函数和偏相关函数的拖尾性和截尾性进行确定,也可以使用 AIC 准则^[6]和 BIC 准则^[7]进行筛选,其他经典的模型识别方法还有 R 和 S 排列法^[8]、Corner 法^[9]、GPAC 法^[10]、EACF 法^[11]以及 SCAN 法^[12]等。2013 年,Marrelli 等^[13]利用极大似然准则,提出了 ARMA 模型的一种递归辨识算法;2019 年,Roberto 等^[14]在附加噪声条件下建立了识别 ARMA 模型的两步变量误差法。

ARIMA 模型的参数估计方法主要有极大似然估计^[15-17]、拟极大似然估计^[18]、最小二乘估计^[19-20]、Bayes 估计^[21]等。模型的诊断检验就是检验当前建立的模型是否适当,主要包括模型整体的拟合优度检验和每个参数的显著性水平检验。

随着机器学习技术的发展,ARIMA 模型与神经网络的深度融合已经成为时间序列分析的热点,神经网络是一种灵活的计算框架和通用逼近器,可应用于多种预测问题,具有较高的预测精度,人工神经网络与 ARIMA 模型^[22]、概率神经网络与 ARIMA 模型^[23]、深度置信网络与 ARIMA 模型^[24]的结合能够克服单一模型的局限性,显著改善预测精度。2019 年,Bahareh 等^[25]结合人工神经网络和 ARIMA 模型研究血小板的需求量,有效减少了供应链中需求的不确定性;2017 年,Qin 等^[26]结合 ARIMA 模型较强的预测能力以及深度信念网络良好的稳定性和强大的学习能力,建立了我国舟山及温州沿海的赤潮预报模型,取得了很好的预测精度。2004 年,Pai 等^[27]将支持向量机强大的非线性回归能力引入 ARIMA 模型进行股票价格预测,显著提高了 ARIMA 模型的预测性能。

一方面,ARIMA 模型要求时间序列经过一次或多次差分变换成为平稳的,然而大量实际的时间序列很难满足此要求;另一方面,实际的时间序列往往是多种周期性或季节性因素综合作用的结果,所以如果能够将时间序列内在的周期性因素逐一剥离出来,则预测其未来发展趋势就简单多了。

受此启发,文中建立了一种新的时间序列预测模型,即 EMD-ARIMA 模型,它的优势是不仅可以直接应用于非线性、非平稳时间序列,而且能够揭示时间序列蕴含的不同时间尺度的周期性变化。首先,利用 EMD 方法将原时间序列分解为多个不同时间尺度的内在模函数和一个趋势项;其次,对

每个内在模函数使用季节性 ARIMA 模型进行预测，对趋势项使用趋势移动平均模型进行预测；最后，把所有子序列的预测结果复合得到原时间序列的预测结果。选取 Husdom Bay 公司的月度貂皮销售数据、IBM 公司普通股的日收盘价数据和黄河兰州站的日含沙量数据作为研究对象，比较 EMD-ARIMA、ARIMA 和 ANN 3 种模型的预测性能，结果表明，EMD-ARIMA 模型比 ARIMA 模型和 ANN 模型的预测精度更高，而且适用于中长期预测，是一种可靠的时间序列预测方法。

1 模型与方法

1.1 EMD 方法

近年来，Huang 等^[28]基于 Hilbert-Huang 变换建立的经验模态分解(EMD)是一种适用于非线性非平稳信号的时频分析方法，其在大气科学、机械损伤检测、图像分析及海洋生态等领域已经取得了很大的成功。EMD 是一种自适应方法，它将任意时间序列分解为多个不同时间尺度的内在模态函数(IMF)和一个长期趋势项，内在模态函数成为表征数据的基函数，因为这个基是自适应的，并且是局部确定的，所以可为潜在过程提供具有物理意义的表现形式。

EMD 方法用时间序列上、下包络的平均值来确定瞬时平均位置，从而分解出内在模函数 $I(t)$ 。内在模函数 $I(t)$ 的分解过程如下：

1) 求上、下包络。首先，找出原时间序列 $X(t)$ 的局部极大值，局部极大值定义为时间序列中的某个时刻值，其前一时刻的值和后一时刻的值都不能比它大。然后，用三阶样条函数对局部极大值进行插值，得到原时间序列 $X(t)$ 的上包络序列 $X_{\max}(t)$ 。同理，基于时间序列 $X(t)$ 的局部极小值得到下包络序列 $X_{\min}(t)$ 。

2) 求瞬时平均值。对 $X_{\max}(t)$ 和 $X_{\min}(t)$ 求平均，得到瞬时平均值

$$m(t) = (X_{\max}(t) + X_{\min}(t)) / 2.$$

3) 求类距平值序列。用原时间序列 $X(t)$ 减去瞬时平均值 $m(t)$ ，得到类距平值序列

$$h(t) = X(t) - m(t).$$

如果 $h(t)$ 中极值点的个数和跨零点的个数相等或者只相差一个，并且瞬时平均值 $m(t)$ 等于零，则 $h(t)$ 就是内在模函数，否则把 $h(t)$ 视为原序列，重复上述做法，直到满足内在模函数的定义为止，

求出内在模函数。

4) 经验模态分解。分离出第一个内在模函数 $I_1(t)$ 以后，用原时间序列 $X(t)$ 减去 $I_1(t)$ ，得到剩余项 $r_1(t)$ ，即

$$r_1(t) = X(t) - I_1(t).$$

这样就完成了第一个内在模函数的分解过程。把 $r_1(t)$ 作为新的时间序列，按照以上步骤，依次分解出第二、第三，直至第 n 个内在模函数 $I_n(t)$ ，直到 $r_n(t)$ 成为一个单调序列或仅含有一个极值点，则结束模态分解， $r_n(t)$ 称为趋势项。

若把分解后的各个分量合并起来，就得到原时间序列

$$X(t) = \sum_{i=1}^n I_i(t) + r_n(t).$$

1.2 ARIMA 模型

ARIMA 模型能够对非平稳时间序列进行建模，是经典的时间序列预测方法之一。ARIMA 模型的建模过程包括数据预处理、平稳性检验、模型定阶、参数估计和模型检验等步骤。记 p 为自回归阶数， q 为移动平均阶数， d 为差分阶数，那么自回归求和移动平均过程记为 ARIMA(p, d, q)。如果不做差分运算，则简记为 ARMA(p, q)，其表达式为

$$X(t) = \varphi_1 X(t-1) + \varphi_2 X(t-2) + \dots + \varphi_p X(t-p) - \theta_1 \varepsilon(t-1) - \theta_2 \varepsilon(t-2) - \dots - \theta_q \varepsilon(t-q) + \varepsilon(t),$$

其中 ε_t 为高斯白噪声； $\varphi_1, \varphi_2, \dots, \varphi_p$ 为自回归参数； $\theta_1, \theta_2, \dots, \theta_q$ 为移动平均参数。

1.3 趋势移动平均法

时间序列没有明显的趋势变化时，使用简单移动平均就能够得到较好的预测效果，但时间序列具有明显的增长或下降趋势时，使用简单移动平均法进行预测就会出现比较大的偏差，趋势移动平均法(trend moving average, TMA)对于同时存在直线趋势与周期波动的时间序列，是一种既能反映趋势变化，又可以有效分离周期变动的的时间序列预测方法^[29]。

TMA 方法的基本思想是：记时间序列 $X(t)$ 自 t 期开始直到 $t+N-1$ 期共 N 期的一次移动平均为

$$M^{(1)}(t) = \frac{1}{N} (X(t) + X(t-1) + \dots + X(t-N+1));$$

对一次移动平均再做平均就是二次移动平均，即

$$M^{(2)}(t) = \frac{1}{N}(M^{(1)}(t) + M^{(1)}(t-1) + \dots + M^{(1)}(t-N+1)).$$

于是时间序列 $X(t)$ 未来的发展趋势可以按照下列模型进行预测:

$$\tilde{X}(t+m) = a_t + b_t m,$$

其中, t 为当前时期; m 为预测期数; a_t 为截距; b_t 为斜率; a_t 和 b_t 称为平滑系数. 经过推导可知

$$\begin{cases} a_t = 2M^{(1)}(t) - M^{(2)}(t), \\ b_t = \frac{2}{N-1}(M^{(1)}(t) - M^{(2)}(t)). \end{cases}$$

显然, TMA 模型对于线性增长的时间序列具有很好的预测效果, 但对于非线性增长(如指数增长)的时间序列则需要做多次移动平均才能取得满意的预测效果.

1.4 EMD-ARIMA 模型

EMD-ARIMA 模型把 EMD 方法与季节性 ARIMA 模型结合起来, 该预测过程首先对时间序列数据进行 EMD 分解, 得到多个内在模函数和一个趋势项, 然后对每个内在模函数计算平均周期, 再使用季节性 ARIMA 模型^[30]进行预测, 对趋势项使用趋势移动平均模型进行预测(如果内在模函数的平均周期较大, 变化趋势明显, 则使用趋势移动平均法效果更好), 最后把所有分量的预测结果进行复合就得到原时间序列的预测结果.

1.5 误差分析

为了比较不同模型的预测性能, 本文引入平均绝对百分比误差(mean absolute percent error, MAPE) E_{MAPE} 、平均绝对误差(mean absolute deviation, MAD) E_{MAD} 和均方根误差(root mean squared error, RMSE) E_{RMS} 3 种误差, 其定义如下:

$$E_{MAPE} = \frac{100}{T} \sum_{t=1}^T \left| \frac{X(t) - \tilde{X}(t)}{X(t)} \right|,$$

$$E_{MAD} = \frac{1}{T} \sum_{t=1}^T |X(t) - \tilde{X}(t)|,$$

$$E_{RMS} = \sqrt{\frac{1}{T} \sum_{t=1}^T (X(t) - \tilde{X}(t))^2},$$

其中, $X(t)$ 为 t 时刻的测量值; $\tilde{X}(t)$ 为 t 时刻的预测值.

2 实证分析

2.1 数据来源与特征

选取 3 组不同规模的时间序列进行预测建模.

数据 1 是 Husdom Bay 公司的年度貂皮销售数据, 时间为 1850—1911 年, 共 62 个数据; 数据 2 是 IBM 公司普通股的交易日收盘价数据, 时间为 1959 年 6 月 29 日—1960 年 6 月 30 日, 共 255 个数据; 数据 3 是黄河兰州站的日平均含沙量数据, 时间从 2009 年 1 月 1 日—2015 年 12 月 31 日, 共 2556 个数据. 数据 1 和 2 来源于 George 等的专著《时间序列分析预测与控制》^[4], 数据 3 根据《中华人民共和国水文年鉴》多年数据整理而来. 数据 1 将前 54 个数据作为训练集, 后 8 个数据作为测试集, 做短期预测; 数据 2 将前 235 个数据作为训练集, 后 20 个数据作为测试集, 做中期预测; 数据 3 将前 2191 个数据(前 6 年)作为训练集, 后 365 个数据(2015 年)作为测试集, 做长期预测.

2.2 经验模态分解与内在模函数

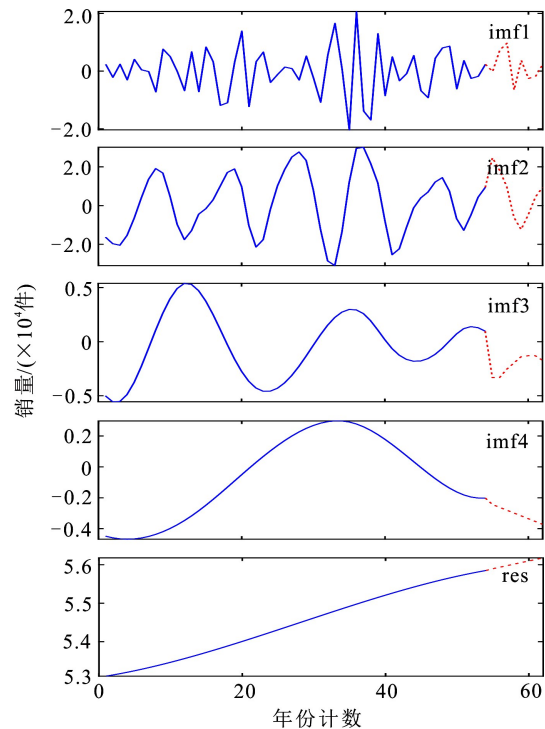


图 1 数据 1 的 EMD 分量预测结果

Fig 1 Prediction results of EMD components for dataset 1

数据 1 的训练集经经验模态分解得到 4 个内在模函数和一个趋势项, 内在模函数的平均周期分别为 4 a, 11 a, 18 a 和 54 a, 说明了 Husdom Bay 公司年度貂皮销售量由 4 个周期变化的销售因素复合而来, 趋势项是一条是单调递增的直线, 说明该公司的貂皮销量一直在缓慢增长(图 1 实线).

数据 2 的训练集经经验模态分解得到 5 个内在模函数和一个趋势项, 5 个内在模函数的平均周期分别为 4 d, 8 d, 17 d, 34 d 和 79 d, 因为我们只分

析了一个年度的交易数据，并且股票市场因为节假日休市或意外停盘等因素，这几个周期很难说明是什么因素造成的；趋势项说明 IBM 公司普通股的价格从 1959 年 7 月开始缓慢下降，10 月最低，此后开始缓慢回升，至来年 5、6 月份达到最高，其股价的长期趋势是持续上升的(图 2 实线)。

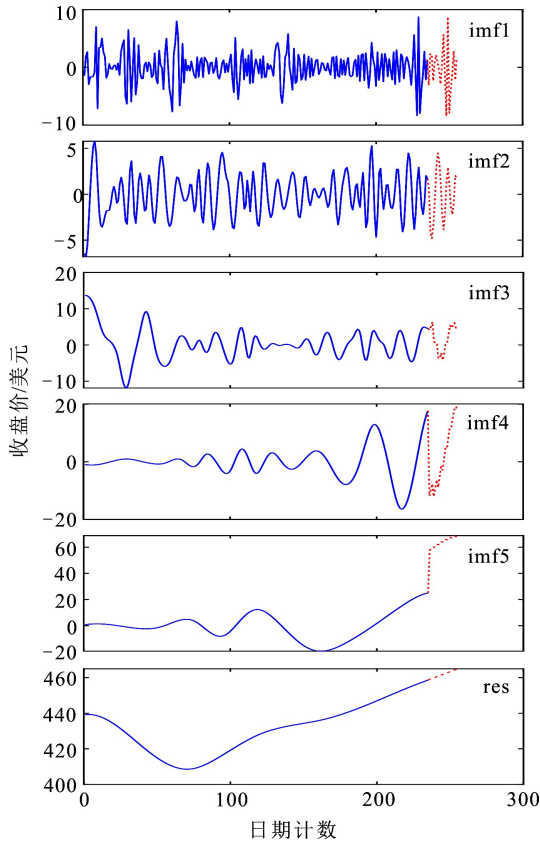


图 2 数据 2 的 EMD 分量预测结果

Fig 2 Prediction results of EMD components for dataset 2

数据 3 经经验模态分解共得到 11 个内在模函数和一个趋势项，内在模函数的平均周期分别为 10 d,13 d,17 d,31 d,38 d,49 d,65 d,100 d,244 d, 439 d 和 548 d，前 9 个 IMF 反映黄河兰州站含沙量的季节性与洪枯季变化，后 2 个 IMF 反映含沙量的年际变化，其具体的物理意义还有待进一步分析；趋势项说明黄河兰州段含沙量的总体变化趋势是：自 2009 年开始黄河日含沙量持续上升，至 2012 年 7 月达到峰值，此后开始缓慢下降，2014 年 6 月最低，此后又有所回升(图 3 实线)。

2.3 EMD-ARIMA 模型预测分析

对经验模态分解得到的内在模函数使用季节性 ARIMA 模型、趋势项(残差)使用 TMA 模型进行预测，图 1~3 中虚线分别是数据 1~3 的每个内在模函数和趋势项的预测结果，从中可以看出，每个

内在模函数的预测结果较好地重现了周期特征，体现了事件季节性因素的影响。当然，因为内在模函数并不是严格周期的，局部具有拉伸或放大现象，

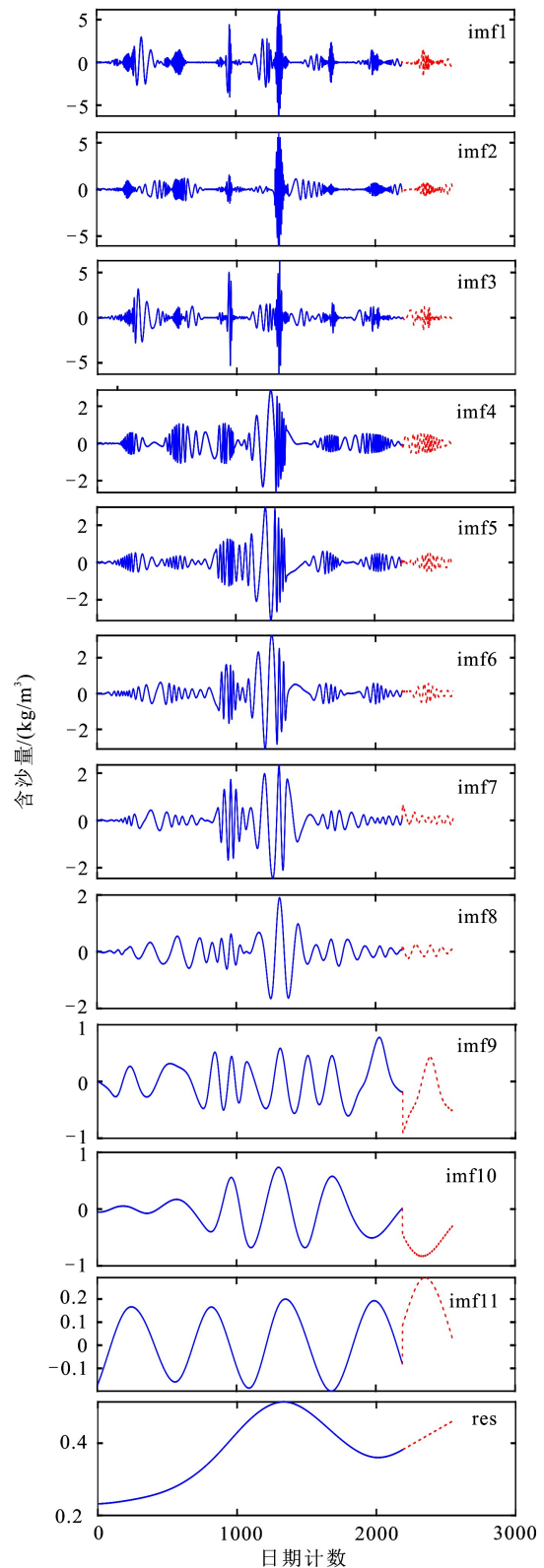


图 3 数据 3 的 EMD 方法分量预测结果

Fig 3 Prediction results of EMD components for dataset 3

所以使用季节性 ARIMA 模型存在不小的误差, 特别对于大周期尺度的内在模函数, 性能更差。

将图 1 中所有内在模函数和趋势项的预测结果复合就得到数据 1 的预测结果(图 4), 其中虚线为预测值, 实线为实际值。数据 2 和数据 3 的预测结果分别见图 5 和图 6。从中可以看出, 数据 1 和数据 2 的预测结果整体偏大, 而数据 3 的预测结果整体偏小, 这很大程度上是因为趋势项的递增或递减造成的。

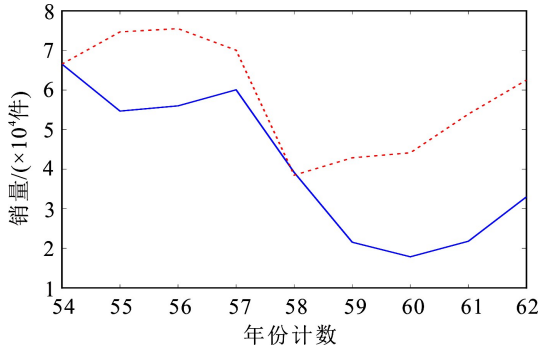


图 4 数据 1 的预测结果

Fig 4 Prediction results for dataset 1

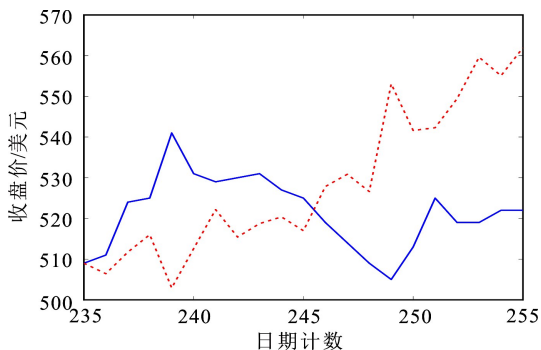


图 5 数据 2 的预测结果

Fig 5 Prediction results for dataset 2

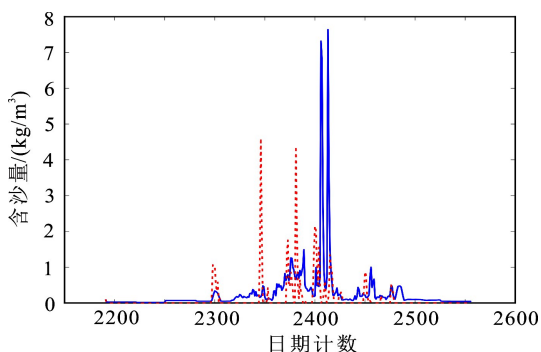


图 6 数据 3 的预测结果

Fig 6 Prediction results for dataset 3

分析测试期事件峰值的发生强度和发生时间可以发现: 数据 1 中, 真实的 4 个销售峰值依次为 54 673, 55 996, 60 053 和 66 549, 分别出现在 55,

56, 57 和 54 年份, 而预测的 4 个销售峰值依次为 66 549, 70 054, 74 685 和 75 507, 依次出现在 54, 57, 55 和 56 年份, 虽然预测峰值明显偏大, 但峰值出现的时间完全一样, 对销售旺季的时机预测令人满意; 数据 2 虽然测试期内股票价格相差较小, 但股价峰值出现的时机却相距甚远; 数据 3 中, 实际测量数据中黄河兰州段含沙量有 5 个比较大的峰值, 分别为 2.69, 3.61, 6.81, 7.32 和 7.64, 依次出现在 2015 年 8 月上旬, 预测的含沙量最大的 5 个峰值分别为 2.03, 2.06, 2.15, 4.37 和 4.60, 依次出现在 2015 年 7 月底, 峰值普遍偏小, 而且时机几乎相差 25 d 左右。这说明近年来黄河兰州段洪峰强度不仅逐年增加, 而且洪峰发生的时机从 7 月底向 8 月初延伸。

2.4 误差分析

数据 1 的 E_{MAP} 为 71.6174%, E_{MAD} 和 E_{RMS} 分别为 19 932 和 22 164, 相对于 40 000 多的平均销售量而言并不大; 数据 2 的 E_{MAP} 为 3.9545%, 精度较高, E_{MAD} 和 E_{RMS} 分别为 20.5967 和 24.4141, 相对于 520 多美元的均价而言也很小; 数据 3 的 E_{MAP} 为 114.8324%, 明显偏大, 这是因为含沙量的峰值远远大于平均含沙量, 而峰值发生的实际时间和预测时间不一致所致; 数据 3 的 E_{MAD} 和 E_{RMS} 分别为 0.2844 和 0.8212, 精度较高。

引入实际曲线和预测曲线包络的积分均值衡量 EMD-ARIMA 模型的平均预测精度。计算表明, Husdom Bay 公司 1904—1911 年貂皮的年平均销售量为 43 490 件, 预测平均销量为 59 691 件, 相对误差为 37.25%; IBM 公司普通股 1960 年 6 月最后 20 个交易日的平均收盘价为 521.8 美元, 预测平均收盘价为 528.5 美元, 相对误差为 0.01%; 2015 年度黄河兰州段平均日含沙量为 $0.51 \text{ kg} \cdot \text{m}^{-1}$, 预测日平均含沙量为 $0.33 \text{ kg} \cdot \text{m}^{-1}$, 相对误差为 35.75%。这表明数据 2 的平均预测精度很高, 而数据 1 和数据 3 的平均预测精度较差。

2.5 性能比较

以数据 1~3 为例, 比较 EMD-ARIMA 模型与经典的 ARIMA 模型、ANN 模型的 E_{MAP} , E_{MAD} 和 E_{RMS} 3 种预测误差, 结果见表 1, 其中 ANN 模型为 3 层 BP 神经网络, 隐含层规模依次设置为 20, 10 和 6, 参数取默认值。

表 1 3 种模型的预测误差
Tab 1 Prediction errors of 3 models

数据 集	数据 规模	预测 期数	IMF 数目	模型	E_{MAP}	改进率	E_{MAD}	改进率	E_{RMS}	改进率
1	62	8	4	ANN	94.3343	24.1%	2.5728×10^4	22.5%	2.8732×10^4	22.9%
				ARIMA	83.9319	14.7%	1.9986×10^4	0.3%	2.4484×10^4	9.5%
				EMD-ARIMA	71.6174		1.9932×10^4		2.2164×10^4	
2	255	20	5	ANN	4.4990	12.1%	23.6730	13.0%	28.3065	13.8%
				ARIMA	2.6420	-49.7%	13.9154	-48.0%	16.0066	-52.5%
				EMD-ARIMA	3.9545		20.5967		24.4141	
3	2556	365	11	ANN	117.4438	2.5%	0.2804	-0.01%	0.8127	-0.01%
				ARIMA	134.5177	14.6%	0.4482	36.5%	0.8782	6.5%
				EMD-ARIMA	114.8324		0.2844		0.8212	

从表 1 可以看出，就短期预测而言(数据 1)，EMD-ARIMA 模型的性能最好，相对于 ANN 模型和 ARIMA 模型， E_{MAP} 分别减小了 24.1% 和 14.7%， E_{MAD} 分别减小了 22.5% 和 0.3%， E_{RMS} 分别减少了 22.9% 和 9.5%；就中期预测而言(数据 2)，ARIMA 模型的性能最好，虽然 EMD-ARIMA 模型和 ANN 模型不及 ARIMA 模型，但 EMD-ARIMA 模型却明显优于 ANN 模型， E_{MAP} ， E_{MAD} 和 E_{RMS} 分别减少了 12.1%，13.0% 和 13.8%；就长期预测而言(数据 3)，EMD-ARIMA 模型性能最好，ANN 模型与之相差不大，但远远高于 ARIMA 模型， E_{MAP} ， E_{MSD} 和 E_{RMS} 分别减小了 14.6%，36.5% 和 6.5%。综合来看，在 3 种预测模型中，EMD-ARIMA 模型的综合预测性能最好。

分析数据特征和 3 种模型的预测误差可以发现，数据 1 是线性平稳的，而且没有明显的周期性变化，所以 3 种模型的预测性能差别不大，EMD-ARIMA 模型略好；数据 2 是线性非平稳的，ARIMA 模型经过差分运算成为线性平稳的，所以预测性能最好，而 ANN 和 EMD-ARIMA 模型均没有使用差分技巧，所以性能较弱，但 EMD-ARIMA 模型仍然好于 ANN 模型；数据 3 是非线性非平稳的，多次差分依然难以保证数据的平稳性，所以 ARIMA 模型的预测性能就下降了，而 ANN 模型没有平稳性要求，EMD-ARIMA 模型经过经验模态分解以后每个内在模函数具有季节性数据，再经过周期差分后能够保证数据的平稳性，所以预测性能显著改善。综合而言，EMD-ARIMA 模型特别适用于非线性、非平稳且没有明显周期性变化的时间序列，而且对于中长期预测效果更好。

3 结论

结合 EMD 方法建立了一种新的 ARIMA 时间序列预测模型，即 EMD-ARIMA 模型。与传统的预测模型相比，其优势在于适用于非线性非平稳时间序列，尤其是没有明显周期性变化的时间序列。因为从实际中测量得到的时间序列很多都是非线性非平稳的，所以文中提出的模型具有很强的实用价值。

为了比较 EMD-ARIMA 模型与 ANN 和 ARIMA 模型的预测性能，选取 3 组不同规模的时间序列做短期、中长期和长期预测，比较 E_{MAP} ， E_{MAD} 和 E_{RMS} 3 种误差。结果表明，EMD-ARIMA 模型对于没有明显周期性变化的非线性、非平稳时间序列的中长期预测具有很好的预测精度，是一种可靠的时间序列预测方法。

参考文献：

[1] 王佳璆, 邓敏, 陈涛, 等. 时空序列数据分析和建模[M]. 北京: 科学出版社, 2012.

[2] YULE G U. On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers[J]. *Philos Trans Roy Soc*, 1927, **A226**: 267.

[3] WALKER G. On periodicity in series of related terms[J]. *Proc Roy Soc*, 1931, **A131**: 518.

[4] GORGE EP B, GWILYM M J, GREGORY C R. 时间序列分析: 预测与控制[M]. 王成章, 尤梅芳, 郝杨, 译. 北京: 机械工业出版社, 2011.

[5] SALLESR, BELLOZE K, PORTO F, et al. Nonstationary time series transformation methods: an experimental review [J]. *Knowledge-Based Systems*, 2019, **164**: 274.

[6] AKAIKE H. A new look at the statistical model

- identification[J]. *IEEE Trans Automatic Control*, 1974, AC-19: 716.
- [7] SCHWARZ G. Estimating the dimension of a model [J]. *Ann Statist*, 1978, **6**: 461.
- [8] GRAY H L, KELLEY G D, MCINTIRE D D. A new approach to ARMA modelling [J]. *Commun Statist*, 1978, **B7**: 1.
- [9] DE GOOIJER J G, HEUTS R M J. The corner method: an investigation of an order determination procedure for general ARMA processes[J]. *J Oper Res Soc*, 1981, **32**: 1039.
- [10] WOODWARD W A, GRAY H L. On the relationship between the S array and the Box-Jenkins method of ARMA model identification[J]. *J Amer Statist Assoc*, 1981, **76**: 579.
- [11] TSAY R S, TIAO G C. Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models[J]. *J Amer Statist Assoc*, 1984, **79**: 84.
- [12] TSAY R S, TIAO G C. Use of canonical analysis in time series model identification [J]. *Biometrika*, 1985, **72**: 299.
- [13] MARELLI D, YOU Ke-you, FU Min-yue. Identification of ARMA models using intermittent and quantized output observations[J]. *Automatica*, 2013, **49**(2): 360.
- [14] ROBERTO D, ERIC G, FERNANDO M. ARMA model identification from noisy observations based on a two-step errors-in-variables approach [J]. *IFAC-Papers Online*, 2017, **50**(1): 14143.
- [15] WINCEK M A, REINSEL G C. An exact maximum likelihood estimation procedure for regression-ARMA time series models with possibly nonconsecutive data [J]. *J Roy Statist Soc*, 1986, **B48**: 303.
- [16] MAURICIO J A. Exact maximum likelihood estimation of partially nonstationary vector ARMA models [J]. *Computational Statistics & Data Analysis*, 2006, **50**(12): 3644.
- [17] MEITZ M, SAIKKONERN P. Maximum likelihood estimation of a noninvertible ARMA model with autoregressive conditional heteroskedasticity [J]. *Journal of Multivariate Analysis*, 2013, **114**: 227.
- [18] LING Shi-qing. Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models [J]. *Journal of Econometrics*, 2007, **140**(2): 849.
- [19] GORGE K. A note on an iterative least-squares estimation method for ARMA and VARMA models [J]. *Economics Letters*, 2003, **79**(3): 305.
- [20] LIU Li-juan Liu, WANG Yan, WANG Cheng, et al. Maximum likelihood recursive least squares estimation for multivariate equation-error ARMA systems [J]. *Journal of the Franklin Institute*, 2018, **355**(15): 7609.
- [21] BOX G E P, DRAPER N R. The Bayes estimation of common parameters from several responses[J]. *Biometrika*, 1965, **52**: 355.
- [22] MEHDI K, MEHDI B. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting [J]. *Applied Soft Computing*, 2011, **11**(2): 2664.
- [23] MEHDI K, MEHDI B, GHOLAM A R A. Hybridization of autoregressive integrated moving average (ARIMA) with probabilistic neural networks (PNNs)[J]. *Computers & Industrial Engineering*, 2012, **63**(1): 37.
- [24] XU Wen-quan, PENG Hui, ZENG Xiao-yong, et al. Deep belief network-based AR model for nonlinear time series forecasting [J]. *Applied Soft Computing*, 2019, **77**: 605.
- [25] BAHAREHF, BEHNAM M, FARZAD F J, et al. Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models [J]. *Computers in Biology and Medicine*, 2019, **113**: Article 103415.
- [26] QIN Meng-jiao, LI Zhi-hang, DU Zhen-hong. Red tide time series forecasting by combining ARIMA and deep belief network [J]. *Knowledge-Based Systems*, 2017, **125**(1): 39.
- [27] PAI Ping-feng, LIN Chih-sheng. A hybrid ARIMA and support vector machines model in stock price forecasting [J]. *Omega*, 2005, **33**(6): 497.
- [28] HUANG N E, SHEN S, LONG S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis [J]. *Proc Roy Soc London*, 1998, **A454**: 903.
- [29] MAJEROVÁ I, PRAŽÁK T. Estimation of economic development in Papua New Guinea: linear trend analysis or moving average model? [J]. *Procedia: Social and Behavioral Sciences*, 2014, **110**: 450.
- [30] CHANG Xing-hu, GAO Meng, WANG Yan, et al. Seasonal autoregressive integrated moving average model for precipitation time series [J]. *Journal of Mathematics and Statistics*, 2012, **8**(4): 500.

(责任编辑 马宇鸿)