

Prediction of Temperature Based on EEMD-ARIMA Model

Yujie Jia, Penglei Chen, Li Zhu

Ningbo University of Technology, Ningbo Zhejiang
Email: zhulwhu@163.com

Received: Apr. 9th, 2020; accepted: Apr. 22nd, 2020; published: Apr. 29th, 2020

Abstract

The average annual temperature data are modeled and predicted by ensemble empirical mode decomposition (EEMD) and autoregressive integrated moving (ARIMA) model. First, this paper performs the noise processing on the average annual temperature, and performs empirical mode decomposition (EMD) to make the components smoothed. Then, each component is predicted by ARIMA model. Finally, the component prediction results are added to obtain the final predicted value of the average annual temperature. By comparison, it is found that the prediction results of the EEMD-ARIMA model are more accurate than the prediction results of the EMD-ARIMA model and the ARIMA model.

Keywords

Empirical Mode Decomposition (EMD), Ensemble Empirical Mode Decomposition (EEMD), ARIMA Model, Average Annual Temperature, Forecast

基于EEMD-ARIMA模型的气温预测研究

贾雨杰, 陈鹏蕾, 朱 莉

宁波工程学院, 浙江 宁波
Email: zhulwhu@163.com

收稿日期: 2020年4月9日; 录用日期: 2020年4月22日; 发布日期: 2020年4月29日

摘 要

采用集合经验模态分解(EEMD)与差分整合移动平均自回归(ARIMA)模型相结合的方法对年均气温数据进行建模预测。首先对年均气温做加噪处理后进行经验模态分解(EMD),使其分量平稳化,然后再对各分量采用ARIMA模型预测,最后将各预测结果叠加得到年均气温的预测值。通过比较发现,EEMD-ARIMA模型比EMD-ARIMA模型及ARIMA模型的预测结果具有更高的精确度。

关键词

经验模态分解(EMD), 集合经验模态分解(EEMD), ARIMA模型, 年均气温, 预测

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

气温是描述某一地区气候特征的一个主要因素, 通过对气温进行预测, 可以在极端天气来临前起到预警作用, 并及时采取预防措施。传统的预测方法主要有多元线性回归法[1]、自回归移动平均法[2]和灰色预测法[3]等, 这些方法的预测结果比较趋于平均值, 不仅难以应付海量的数据和天气的动态变化, 而且用于预测的气象数据多为时间序列, 传统预测方法没有将数据的时间相关性考虑在内, 导致出现预测准确率低的问题, 因此正逐渐被淘汰[4]。目前, 应用较广泛的是以支持向量机、神经网络等为代表的人工智能算法。如小波分析和非线性自回归(NVR)神经网络结合技术[5]、多元时间序列局部支持向量回归方法[6], 这些研究都能考虑到气象要素的时序多元性, 却没实现精细化的气温预测, 只能对相对稳定的平均气温进行预测, 限于研究的手段限制, 最终得到的精度有限。

1998年, N.E. Huang 提出了一种新的信号分析方法——经验模态分解(EMD) [7]。该方法在处理非平稳非线性的数据上具有很大的优势, 一经提出后, 就在不同的工程领域得到了迅速有效的应用, 例如在海洋、大气、天体观测资料与地震记录分析、机械故障诊断[8]、密频动力系统的阻尼识别以及大型土木工程结构的模态参数识别等方面。但是, 如果数据出现模态混叠现象, 该方法的性能会出现大幅度下降的情况, 因此 N.E. Huang 为了改进 EMD 的不足, 提出了 EEMD 方法[9], 用以解决 EMD 中的模态混叠问题, 该方法在气候、工程等领域有较多的应用。

气温数据具有非平稳非线性、时间相关性的特征。为了提高气温预测的准确性, 本文提出 EEMD 和 ARIMA 模型相结合的方法对气温数据进行预测。EEMD 能对非平稳数据进行平稳化处理, 并且能有效解决数据中存在的模态混叠问题, 使其达到时间序列数据分析的要求, ARIMA 时间序列分析又充分考虑到了气温数据的时间相关性, 将两种方法结合起来对气温数据进行处理及预测, 将会有效提高预测的准确性。

2. EEMD 基本原理

1998年 N.E. Huang [7]提出了 EMD 方法分析非线性、非平稳序列, 此方法的基本思想是把一个频率不规则的波转化为多个单一频率的波(本征模态函数 IMF)和残波的形式。但有时 EMD 会出现不同程度的模态混叠现象, 为了解决这个现象, WU 等[9]提出了集合经验模态分解方法(EEMD)。EEMD 是一种基于 EMD, 用噪声进行辅助的数据分析方法, EEMD 的核心思想是将白噪声加入原始信号中进行多次分解, 把分解多次的本征模态函数(IMF)和趋势项取均值, 该值为最终的 IMF 分量和 rs 余项。利用 EEMD 分解原始信号的具体步骤如下:

- 1) 添加白噪声的次数(即试验总次数) n_e 和幅值系数 ε 。
- 2) 将白噪声加入原始信号 $x(t)$ 中, $w(t)$ 是第 i 次加入的白噪声序列, 可得到信号 $X'(t)$:

$$X'(t) = x(t) + \varepsilon * w(t) \quad (1)$$

3) 找出要分解的时间序列 $X'(t)$ 所有的局部最大值点和最小值点, 利用三次样条函数的方法拟合 $X'(t)$ 所有的最大值点和最小值点, 构成上包络线和下包络线。

4) $b_1(t)$ 为上下包络线的平均值。将信号 $X'(t)$ 与 $b_1(t)$ 相减, 得到新序列:

$$X_1''(t) = X'(t) - b_1(t) \tag{2}$$

5) 判断 $X_1''(t)$ 是否满足 IMF 分量的条件, 如果满足, 则 $X_1''(t)$ 是筛选出的第一个 IMF 分量为 IMF_1 ; 如果不满足, $X_1''(t)$ 将作为新的原始序列重新回到步骤(3)和步骤(4)继续进行筛选, 直到满足 IMF 分量的条件。

6) 把 IMF_1 从信号 $X'(t)$ 中减去得到 x_1 , 如式(3)所示。把 x_1 作为要分解的新信号, 重复步骤(3)至步骤(6) n 次, 直到 x_n 或者 IMF_n 小于给定的数值或 x_n 为单调函数, 即可结束分解。

$$x_1 = X'(t) - IMF_1 \tag{3}$$

7) 用 SD (限值标准差)判断筛选是否终止。当 SD 小于阈值 α 时, 筛选结束。 α 的一般取值范围是 0.2~0.3。 $X'(t)$ 经步骤(3)至步骤(7)处理后即可得到不同尺度的 IMF 和余项。

$$SD = \frac{\sum_{t=0}^T (X_{k-1}''(t) - X_k''(t))^2}{\sum_{t=0}^T X_{k-1}''(t)} \tag{4}$$

8) 利用白噪声频谱的均值为零, 将 n_e 次分解得到的各 IMF 进行均值处理, 得到 EEMD 分解后最终的 IMF 分量 $c_n(t)$ 和趋势项 $r(t)$:

$$c_s(t) = \frac{1}{n_e} \sum_{i=1}^{n_e} IMF_{is}(t) \tag{5}$$

$$r(t) = \frac{1}{n_e} \sum_{i=1}^{n_e} r_i(t) \tag{6}$$

9) 最后 $x(t)$ 的分解结果为:

$$x(t) = \sum_{s=1}^n c_s(t) + r(t) \tag{7}$$

3. EEMD-ARIMA 预测模型基本原理

3.1. ARIMA 模型

ARIMA(p, d, q)是最常用的时间序列模型[10] [11], 其中 AR 是自回归过程, 参数 p 是其自回归项数; MA 是移动平均过程, 参数 q 是其移动平均项数; 参数 d 是使研究对象的时间序列达到平稳状态的差分次数。ARIMA(p, d, q)模型结构为:

$$\begin{cases} \Phi(B)\nabla^d x_t = \Theta(B)\varepsilon_t \\ E(\varepsilon_t) = 0, Var(\varepsilon_t) = \sigma_t^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_s \varepsilon_t) = 0, \forall s < t \end{cases} \tag{8}$$

式中, $\nabla^d = (1-B)^d$; $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, 为平稳可逆 ARMA(p, q)模型的自回归系数多项式;

$\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$, 为平稳可逆 ARMA(p, q)模型的移动平滑系数多项式。ARIMA(p, d, q)模型可简记为: $\nabla^d x_t = \frac{\Theta(B)}{\Phi(B)} \varepsilon_t$, 式中 $\{\varepsilon_t\}$ 为零均值白噪声序列。

3.2. EEMD-ARIMA 模型

气温是一个非平稳序列, 所以在预测之前要先对序列进行平稳化处理。因此, 本文建立基于 EEMD 分解的 EEMD-ARIMA 预测模型。模型的建模流程如图 1 所示。

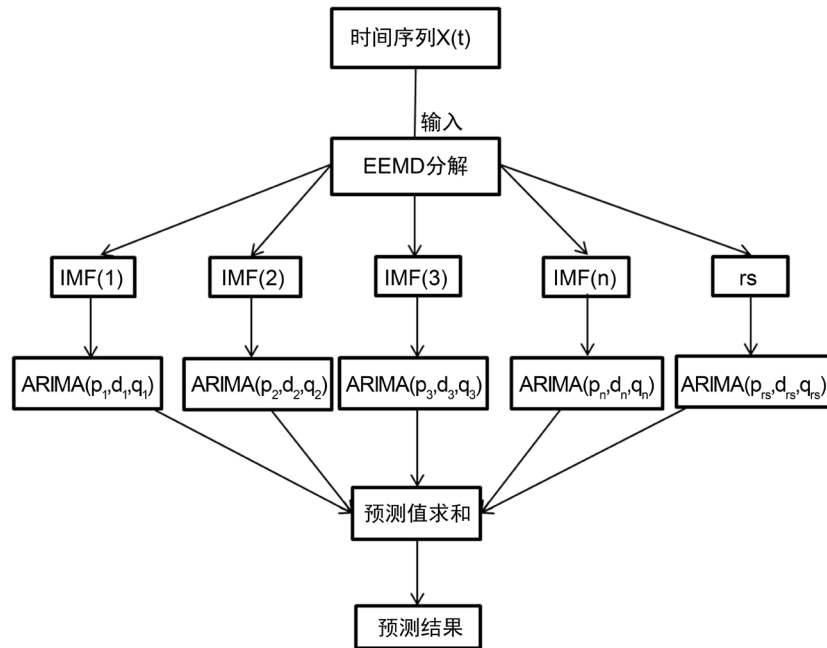


Figure 1. Flow chart of EEMD-ARIMA prediction model
图 1. EEMD-ARIMA 预测模型流程图

EEMD-ARIMA 预测模型的具体步骤如下:

- 1) 首先, 将原始信号 $X(t)$ 输入 EEMD 中进行分解, $X(t)$ 会被分解成若干个 IMF 和余项 rs 。
- 2) 其次, 因为通过 EEMD 分解得到的若干个 IMF 和 rs 之间相互独立, 所以, 对每一个 IMF 和 rs 进行 ARIMA 建模, 并求出相应的预测值。
- 3) 最后, 将所求得的分量的预测值相加, 其和为最终的气温预测数据。

4. 模型应用

本节利用 EEMD-ARIMA 模型对沈阳市年平均气温数据建模分析。采用 1951 至 2013 年的年平均气温数据作为训练样本, 并用 2014 年至 2018 年的年均气温作为测试集。将 EEMD-ARIMA 的结果与 EMD-ARIMA 和 ARIMA 的结果对比分析。

4.1. EMD 与 EEMD 去噪效果分析

用 EMD 和 EEMD 分别对沈阳 1951 至 2013 年的年平均气温数据进行处理。对原始信号进行 EMD 处理后得到 4 个 IMF 分量和 1 个残余分量, 如图 2 所示。在图 2 中, 分量 IMF1 代表着信号高频成分, 含有的噪声成分最多; 分量 IMF2 和分量 IMF3 尺度区分不明显, 存在模态混叠现象。

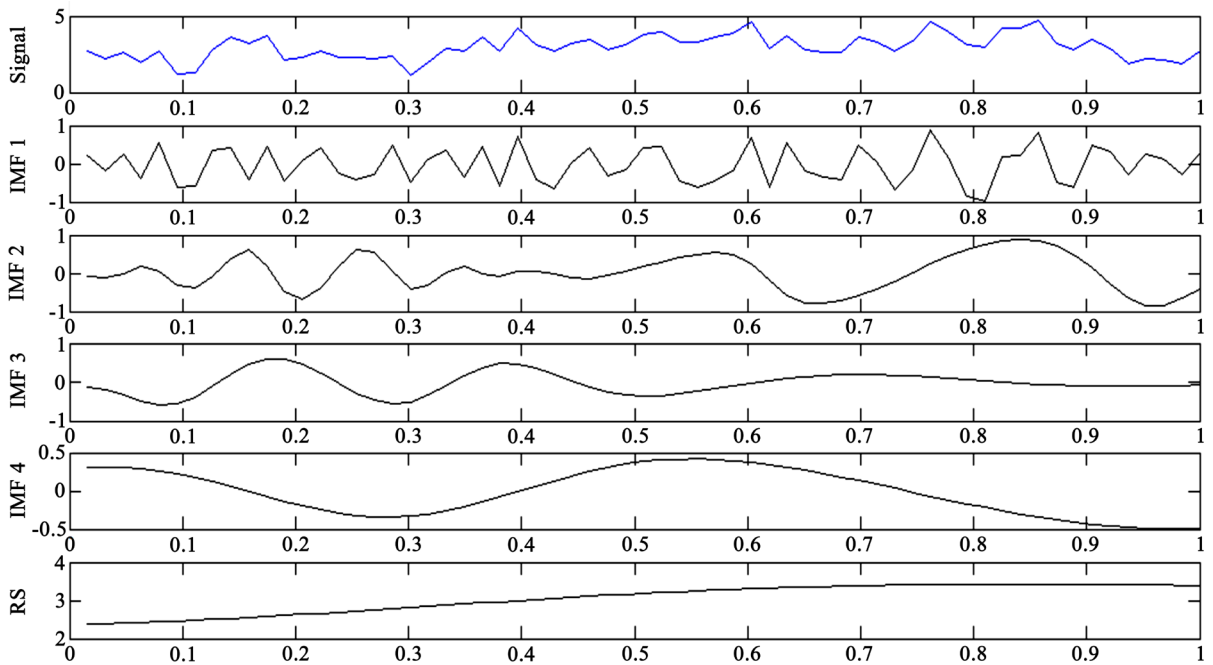


Figure 2. EMD component of annual average temperature in Shenyang
图 2. 沈阳年平均气温 EMD 分解分量

对原始信号进行 EEMD 处理，其中 EEMD 加入了一个 20dB 的白噪声且总体白噪声集成次数为 25 次。处理后得到 4 个 IMF 分量和 1 个残余分量，如图 3 所示。其中 IMF1 为信号高频成分，频率沿时间

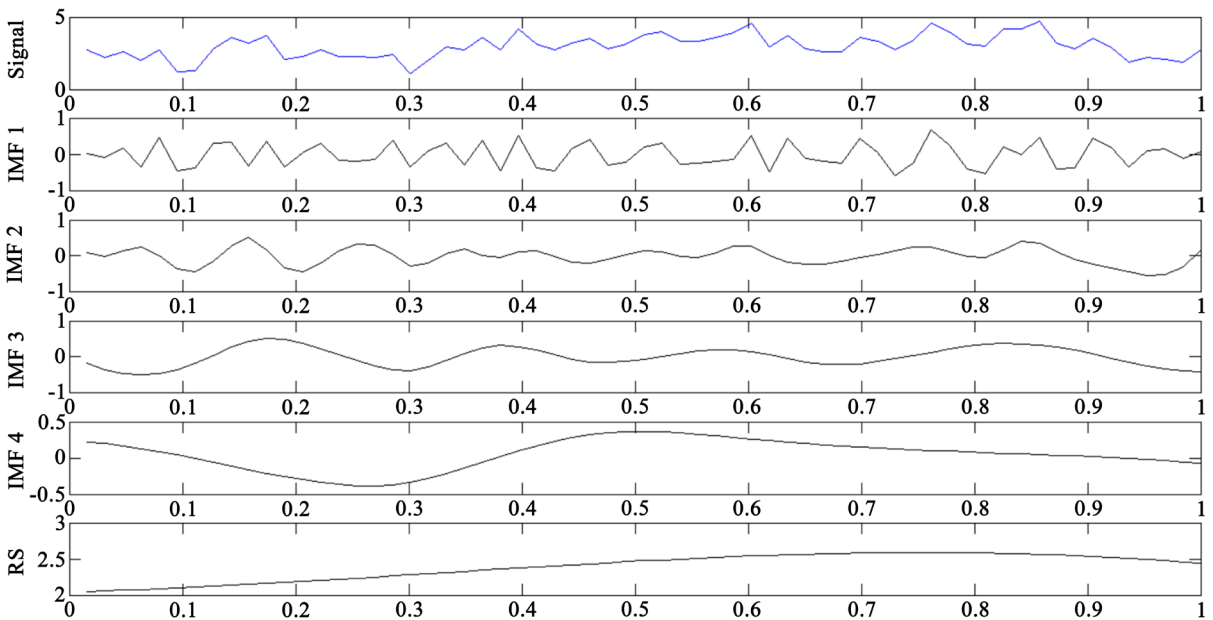


Figure 3. EEMD component of annual average temperature in Shenyang
图 3. 沈阳年平均气温 EEMD 分解分量

轴基本不发生较大的变化，且幅值很小，可以视为信号噪声。各个 IMF 分量具有不同的时间尺度且时间尺度区分较为明显，说明 EEMD 在一定程度上解决了 EMD 存在的模态混叠问题。其中 IMF1 和 IMF2

不具有规律性,为随机分量, IMF3 和 IMF4 为周期分量, rs 为趋势分量。为了查看 EEMD 分解原始信号的完备性水平,对沈阳年平均气温和用 EEMD 分解得到的 IMF 分量与 rs (残余分量)进行逐步拟合,得到不同平滑程度的重构序列,如图 4 所示。随着依次添加 IMF 分量,重构序列与原始信号越来越接近,直至最后与原始信号差不多完全拟合,说明 EEMD 分解该原始信号的完备性很好。

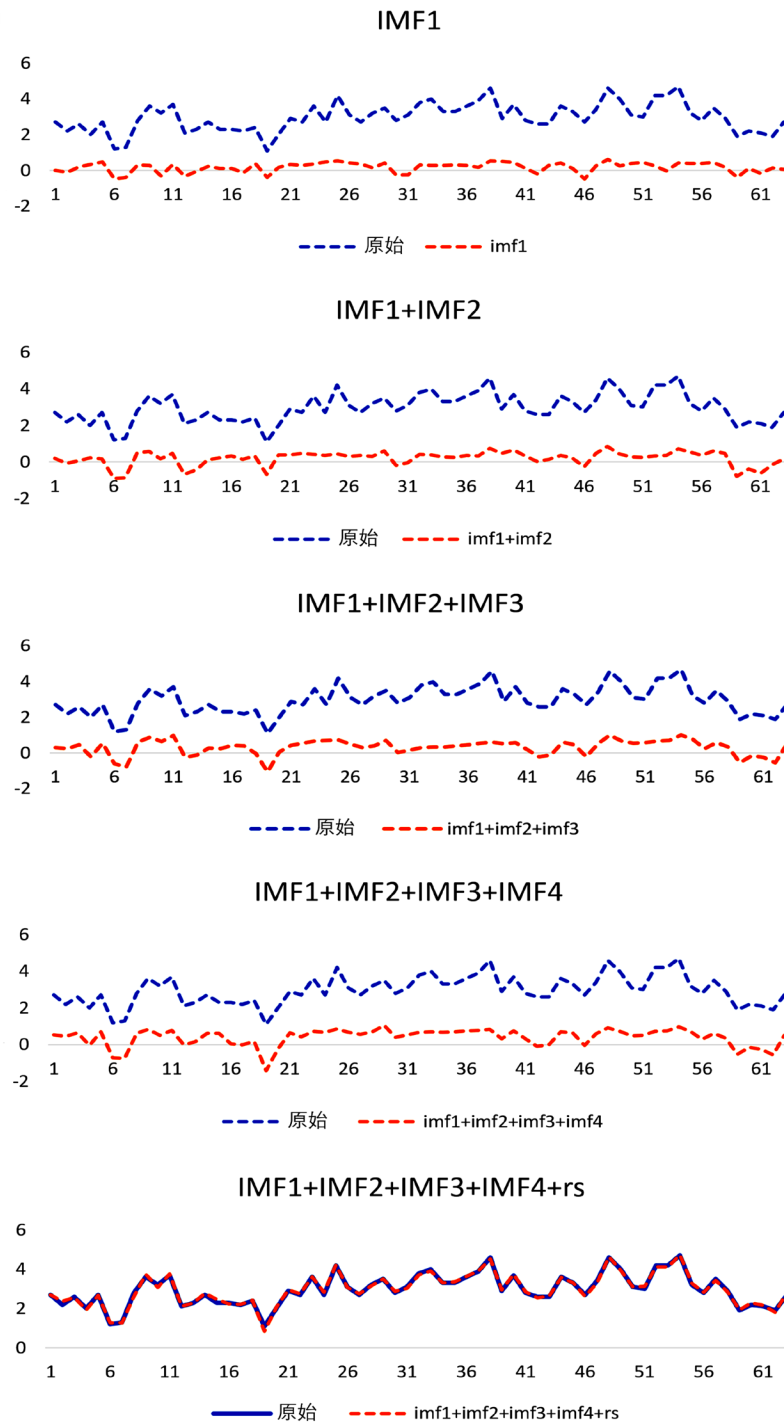


Figure 4. Sequence reconstruction of EEMD decomposition components
图 4. EEMD 分解分量的序列重构

利用均方误差(MSE)对 EMD 和 EEMD 的去噪结果进行量化评价。均方误差是反映估计量与被估计量之间差异程度的一种度量,所以 MSE 越小,表示估计量的可靠性越大,去噪效果会更好。其定义式如式(9)所示:

$$MSE = \frac{\sum_{t=1}^n |x(t) - x'(t)|^2}{n} \tag{9}$$

式中: $x(t)$ 为原始时间序列, $x'(t)$ 为去噪后的时间序列。

计算 MSE 得出的结果如表 1 所示。EEMD 的均方误差更小,其去噪效果优于 EMD。

Table 1. MSE calculation value
表 1. 均方误差(MSE)计算值

方法	MSE
EMD	9.4838
EEMD	8.9211

4.2. 年均气温预测结果及分析

如上,用 EMD、EEMD 对原始信号进行分解,得到两组不同的 IMF,对两组中不同的 IMF 选择合适的 ARIMA 模型建模预测。EMD-ARIMA 模型、EEMD-ARIMA 模型和 ARIMA 模型的预测结果如图 5 所示,可以看出,EEMD-ARIMA 预测模型的预测结果最接近原始值。

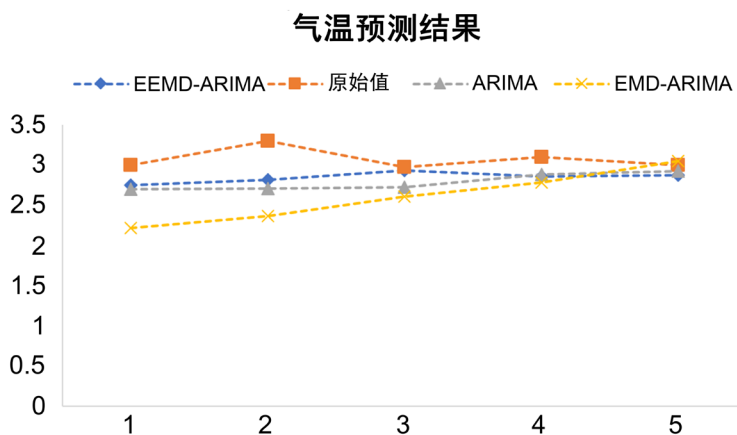


Figure 5. Temperature prediction chart of three models
图 5. 三种模型的气温预测图

利用平均绝对误差(MAE)评价三种模型的预测效果。平均绝对误差是所有单个观测值与算术平均值的偏差的绝对值的平均,其中 MAE 越小,说明预测值与原始值的偏差越小,预测模型的准确性越高。MAE 的定义式如式(10)所示:

$$MAE = \frac{1}{n} \sum_{t=1}^n |x(t) - x'(t)| \tag{10}$$

式中 $x(t)$ 为原始值, $x'(t)$ 为预测值。

三种模型 MAE 的计算结果如表 2 所示。从表 2 中可以看出 EEMD-ARIMA 预测模型的 MAE 最小,其次是 ARIMA 模型,EEMD-ARIMA 预测模型的 MAE 最大,故 EEMD-ARIMA 预测模型的预测效果最好。

Table 2. Calculated value of mean absolute error (MAE)**表 2.** 平均绝对误差(MAE)计算值

模型	MAE
EEMD-ARIMA 模型	0.227
EMD-ARIMA 模型	0.489
ARIMA 模型	0.286

5. 结论

气温时间序列数据具有非线性、非平稳的问题,故本文用 EMD 和 EEMD 对原始信号进行分解,对分量进行分析,然后对各分量分别建立 ARIMA 预测模型。通过比较 EEMD-ARIMA 预测模型、EMD-ARIMA 预测模型和 ARIMA 模型的预测结果,发现针对沈阳地区的年均气温预测,EEMD-ARIMA 模型预测结果均优于 EMD-ARIMA 模型和 ARIMA 模型。

基金项目

浙江省大学生科技创新活动计划暨新苗人才计划项目(2019R428014);宁波市自然科学基金(2018A610195)。

参考文献

- [1] Abuella, M. and Chowdhury, B. (2015) Solar Power Probabilistic Forecasting by Using Multiple Linear Regression Analysis. *SoutheastCon 2015*, Fort Lauderdale, 9-12 April 2015, 1-5. <https://doi.org/10.1109/SECON.2015.7132869>
- [2] Pratiher, S., Mukhopadhyay, S., Barman, R., et al. (2016) Recurrence Quantification & ARIMA Based Forecasting of Rainfall-Temperature Dynamics. 2016 *International Conference on Signal Processing and Communication (ICSC)*, Noida, 26-28 December 2016, 490-495. <https://doi.org/10.1109/ICSPCom.2016.7980630>
- [3] Huang, C.Y., Liu, Y.W., Tzeng, W.C., et al. (2011) Short Term Wind Speed Predictions by Using the Grey Prediction Model Based Forecast Method. 2011 *IEEE Green Technologies Conference (IEEE-Green)*, Baton Rouge, 14-15 April 2011, 1-5. <https://doi.org/10.1109/GREEN.2011.5754856>
- [4] 陶晔, 杜景林. 基于随机森林的长短期记忆网络气温预测[J]. *计算机工程与设计*, 2019, 40(3): 737-743.
- [5] 黄建风, 陆文聪. 基于小波-NAR神经网络的气象要素时间序列预测与天气指数彩虹期权估值[J]. *系统工程理论与实践*, 2016, 36(5): 1146-1155.
- [6] 王定成, 曹智丽, 陈北京, 等. 日气温多元时间序列局部支持向量回归预测[J]. *系统仿真学报*, 2016, 28(3): 654-660.
- [7] Huang, N.E., Shen, Z., Long, S.R., et al. (1988) The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **454**, 903-995. <https://doi.org/10.1098/rspa.1998.0193>
- [8] 葛兴来, 邹丹. 基于多层降噪技术及 Hilbert 变换的轴承故障诊断方法[J/OL]. *电机与控报*, 1-12. <http://kns.cnki.net/kcms/detail/23.1408.TM.20191226.1452.006.html>, 2020-03-06.
- [9] Wu, Z.H. and Huang, N.E. (2008) Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method. *Advances in Adaptive Data Analysis*, **1**, 1-41. <https://doi.org/10.1142/S1793536909000047>
- [10] 麦鸿坤, 肖坚红, 吴熙辰, 等. 基于 R 语言的负荷预测 ARIMA 模型并行化研究[J]. *电网技术*, 2015, 39(11): 3216-3220.
- [11] Huang, S.J. and Shih, K.R. (2003) Short-Term Load Forecasting via ARMA Model Identification Including Non-Gaussian Process Considerations. *IEEE Transactions on Power Systems*, **18**, 673-679. <https://doi.org/10.1109/TPWRS.2003.811010>