

基于因素分解的月最高气温预测

王文炯, 苏俊钊, 王忠雯

临沂大学数学与统计学院, 山东 临沂

Email: 1157858359@qq.com, 8734711@qq.com, 3327374410@qq.com

收稿日期: 2021年4月11日; 录用日期: 2021年4月25日; 发布日期: 2021年5月12日

摘要

利用因素分解方法中的Holt-Winters三参数指数平滑法对月最高气温预测进行了研究。该类方法主要考察序列外部会影响序列波动的确定性因素, 查看序列有无各明显特征并按其固定特征进行因素分解。根据Holt-Winters三参数指数平滑法并利用所获得的黑龙江克山气象站1952年到2017年各月的最高气温作为支撑数据构建相应模型进行预测, 最终将预测出的未来的12个月即2017年每个月的最高温度与2017年实际每月最高温度进行对比且计算出平均绝对百分误差值。

关键词

因素分解, 加法模型, 指数平滑, 最高气温预测

The Prediction of Monthly Maximum Temperature Based on Factor Decomposition

Wenjiong Wang, Junfan Su, Zhongwen Wang

School of Mathematics and Statistics, Linyi University, Linyi Shandong

Email: 1157858359@qq.com, 8734711@qq.com, 3327374410@qq.com

Received: Apr. 11th, 2021; accepted: Apr. 25th, 2021; published: May 12th, 2021

Abstract

The Holt-Winters three parameter exponential smoothing method in factor decomposition methods is used to carry on research about the prediction of the monthly maximum temperature. This kind of method mainly investigates the certainty factors outside the series that will affect the fluctuation of the series, checks whether there are obvious characteristics of the series, and

conducts the factor decomposition according to the fixed characteristics. According to the Holt-Winters three parameter exponential smoothing method, and using the obtained monthly maximum temperature of Keshan meteorological station in Heilongjiang Province from 1952 to 2017 as the supporting data to build the corresponding model for conducting prediction, and finally use the values after predicting for the next 12 months, which are the maximum temperature of each month in 2017, compared with the actual maximum temperature of each month in 2017 and calculated the value of average absolute percentage error.

Keywords

Factor Decomposition, Additive Model, Exponential Smoothing, Forecast of Maximum Temperature

Copyright © 2021 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

气温是指大气层中气体的温度，而月最高气温则是指一个月中出现过的最高温度，在直接受日射所影响的同时，气温也是受位置、大气、地形、洋流、植被和厄尔尼诺现象等自然条件及地球温室效应等人为活动影响的复杂变量，被多种不可确定成分所控制，从而使利用传统的建模方法搭建模型十分困难，尤其是在预测月最高气温这种变化较频繁的数据时，模型的构筑更为困难。所以对月最高气温的变化规律有一定的认识，不止需要传统理论分析工具的协助，还有必要通过新技术、新方法，从不同的角度，借助不同的理论对月最高气温变化规律进行更加深入的分析，以期寻找到合适且准确的描述月最高气温变化的工具，并提高月最高气温的预测精度。

时间序列分析是统计学科的一个重要分支，它通过对离散指标的随机过程的统计分析，研究数据序列的互相依赖关系，最后进行对未知序列的定量预测，包括一般的统计分析、统计模型的建立与推断及数据序列的最优预测和控制等内容[1]。时间序列分析在日常生产生活中的研究和应用十分活跃，并在经济、金融、天文、气象、海洋、物理、化学、医学、质量控制等诸多领域都有广泛的应用，并且由于其模型建立便捷、预测结果有效等诸多原因，目前已经成为众多行业经常使用的统计方法。近些年，计算技术和信号处理技术的迅速发展使得时间序列分析的理论及方法愈发完善，尤其是在参数估计、模型定阶及架构等方面，更是取得了显著的成果[2]。与此同时，基于时间序列分析的应用也日益广泛，且所得出的成果水平不断提高。本文就是利用了时间序列分析方法中一种重要的因素分解方法(Holt-Winters 三参数指数平滑法)来建立合适的模型，以对某地区的月最高气温进行预测。

2. Holt-Winters 三参数指数平滑加法模型

时间序列的种类是多种多样的，其中最基础的可将时间序列分为平稳时间序列、非平稳时间序列和白噪声序列，不同种类的时间序列都有自己独特的规律和性质。在众多性质规律不同的时间序列中，有一部分带有季节效应，且呈现出周期波动规律。根据统计学家 100 多年来对于时间序列中季节性、周期性信息的提取方法的研究，目前针对有季节效应的序列的分析方法主要分为基于因素分解方法和基于自回归移动平均模型方法(Autoregressive Integrated Moving Average Model, 简记 ARIMA)两大类。由于本文中的序列使用 ARIMA 加法模型所拟合出来的模型预测效果并不理想，且温度会出现零的情况故不能使

用 ARIMA 乘法模型, 所以最终选择使用因素分解的方法对序列进行时序分析而不是 ARIMA 模型。使用基于因素分解的思想进行确定性时序分析, 一是可以克服其他因素的干扰, 单纯测度出某个确定性因素(诸如季节、趋势、交易日)对序列的影响; 二是能够根据序列呈现的确定性特征选择适当的方法对序列进行综合预测[3]。而通过对序列长期趋势和季节效应的判断, 可以将序列主要分为三类: 没有长期趋势也没有季节效应的序列可选择简单指数平滑模型进行预测; 有长期趋势但没有季节效应的序列可通过 Holt 两参数指数平滑模型进行对序列未来发展的预测; 而对长期趋势不做要求但必须有季节效应时, 预测模型则应选择 Holt-Winters 三参数指数平滑模型(分为加法模型和乘法模型)。这里主要介绍季节加法模型的构造原理。

对于季节加法模型, 序列可用如下式子表示:

$$x_t = a_0 + b_t + c_t + \varepsilon_t$$

其中 a_0 是截距; b 是斜率; ε_t 是随机波动, 且 $\varepsilon_t \sim N(0, \sigma^2)$; c_t 是 t 时刻由季节效应导致的序列偏差。

假设每个季节周期长度为 n , 每期的季节指数分别为 S_1, S_2, \dots, S_n , 如若设 t 时刻为季节周期的第 j 期 ($1 \leq j \leq n$), 此时 c_t 可以写为:

$$c_t = S_j + e_t, e_t \sim N(0, \sigma_e^2)$$

则 x_t 可以表示为如下公式:

$$\begin{aligned} x_t &= a_0 + b(t-1) + b + c_t + \varepsilon_t \\ &= (x_{t-1} - c_{t-1} - \varepsilon_{t-1}) + (b + \varepsilon_t) + (S_j + e_t) \end{aligned}$$

记

$$\begin{aligned} a(t-1) &= x_{t-1} - c_{t-1} - \varepsilon_{t-1} \\ b(t) &= b + \varepsilon_t \\ c(t) &= S_j + e_t \end{aligned}$$

易知, $a(t-1)$ 为 $t-1$ 时刻消除季节效应的序列截距项的无偏估计值, $b(t)$ 为 t 时刻斜率 b 的无偏估计值, $c(t)$ 为 t 时刻季节指数 S_j 的无偏估计值。

此时 x_t 可以表示为:

$$x_t = a(t-1) + b(t) + c(t)$$

Holt-Winters 三参数指数平滑法就是指分别使用指数平滑的方法, 不断迭代递推参数 $\hat{a}(t)$, $\hat{b}(t)$ 和 $\hat{c}(t)$ 的值, 递推公式如下所示:

$$\begin{aligned} \hat{a}(t) &= \alpha[x_t - c(t-n)] + (1-\alpha)[\hat{a}(t-1) + \hat{b}(t-1)] \\ \hat{b}(t) &= \beta[\hat{a}(t) - \hat{a}(t-1)] + (1-\beta)\hat{b}(t-1) \\ \hat{c}(t) &= \gamma[x_t - \hat{a}(t)] + (1-\gamma)c(t-n) \end{aligned}$$

式子中, x_t 是序列在 t 时刻得到的最新观察值; n 是季节效应的周期长度; α, β, γ 均为平滑系数, 且满足 $0 < \alpha, \beta, \gamma < 1$ 。

使用 Holt-Winters 三参数指数平滑加法公式, 向前 k 期的预测值为:

$$\hat{x}_{t+k} = \hat{a}(t) + \hat{b}(t)k + \hat{c}(t+k), \forall k \geq 1$$

若设 $t+k$ 期为季节周期的第 j 期, 则 $\hat{c}(t+k) = \hat{S}_j (j=1,2,\dots,n)$ 。

3. 实例分析

本文采用黑龙江克山气象站站点(区站号为 50658, 经度为 $125^{\circ}88'$, 纬度为 $48^{\circ}05'$, 高度为 236.5 m) 的每月最高气温为例进行数据的分析与预测。

选取黑龙江克山站 1952 年到 2017 年各月的最高气温作为该模型的支撑数据。在处理数据过程中, 使用以下步骤对数据进行分析预测:

第 1 步: 数据输入

将该站点 1952 年到 2016 年各月的最高气温这 780 个数据写入 Excel 表格, 并导入 R 软件, 生成时间序列。

第 2 步: 判断平稳性

使用 R 软件对数据进行分析, 得出时间序列的时序图、自相关图和偏相关图, 结果分别见图 1、图 2 和图 3。分析图 1 可以看出该序列为有季节性的非平稳时间序列, 分析图 2 和图 3 可以明显看出序列的自相关系数和偏相关系数呈正弦波动趋势, 可以得出该序列是非平稳的时间序列。

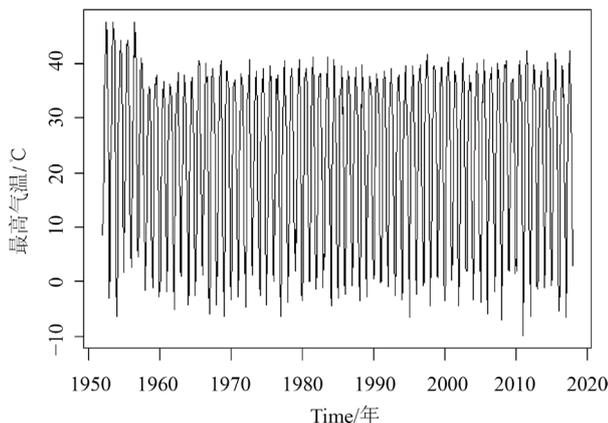


Figure 1. Timing sequence diagram

图 1. 时序图

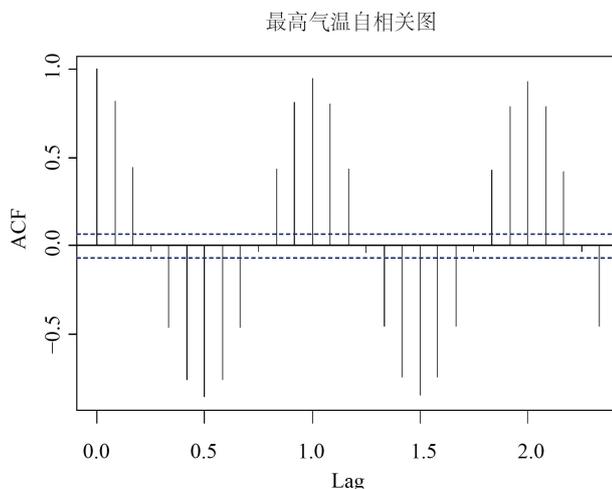


Figure 2. Autocorrelogram

图 2. 自相关图

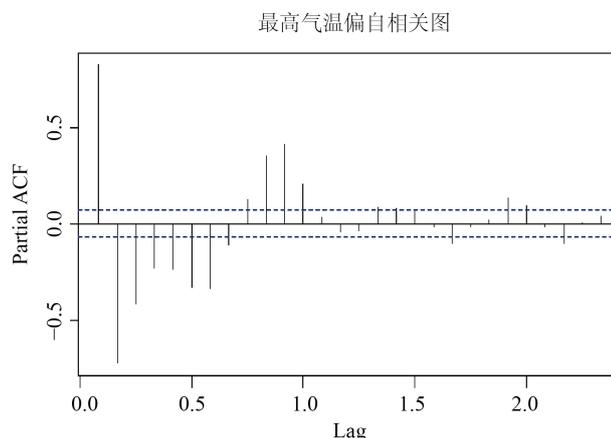


Figure 3. Partial correlogram
图 3. 偏相关图

第 3 步：建立模型

从图 1 中可以看出，该序列具有以年为周期的季节效应，但既没有线性趋势，也没有较大的循环波动特征，所以判断这个序列受到两个因素的影响：季节效应和随机波动。

图 1 显示，每个季节的振幅相对来说是比较稳定的，所以选择加法模型。现使用 Holt-Winters 三参数指数平滑模型对该序列进行适当的拟合与预测。

使用 R 软件可以直接得出模型的参数，结果如表 1 所示：

Table 1. Model coefficient
表 1. 模型系数

Holt-Winters	α	β	γ
系数	0.07212809	0.02342573	0.193682

通过 Holt-Winters 三参数指数平滑加法迭代公式，可以得到三参数最后的迭代值为：

$$a(t) = 23.331407111, b(t) = -0.009288751$$

参数 $c(t)$ 的最后 12 个估计值对应的是 12 个月的季节指数，如表 2 所示。

Table 2. Month and corresponding season index
表 2. 月份及对应季节指数

月份 j	季节指数 S_j	月份 j	季节指数 S_j
1	-27.130056098	7	16.175846602
2	-21.321119384	8	13.635027148
3	-4.592140179	9	10.164572359
4	5.697052829	10	0.410807994
5	11.238295315	11	-9.484287431
6	14.468918890	12	-21.266541650

因此，此序列向前任意 k 期的预测值为

$$\hat{x}_{t+k} = 23.331407111 - 0.009288751k + S_j, \forall k \geq 1$$

式中, j 为 $t+k$ 期对应的季节。

第 4 步: 模型检验

基于 R 软件对所建立的模型进行白噪声检验, 检验 P 值为 0.1401, 大于 0.05, 不能拒绝原假设, 残差序列可视为白噪声序列, 所以模型显著性成立, 检验通过, 可以认为使用 Holt-Winters 三参数指数平滑法所建立的模型是合理有效的。

第 5 步: 结果预测

可以利用所建立的模型对未来的 12 个月即 2017 年每个月的最大温度进行预测, 继而与 2017 年实际每月最高温度进行对比, 来检验模型的准确性。预测及对比结果见表 3 与图 4, 其中图 4 中虚线为序列的实际观察值, 实线为序列的指数平滑估计预测值, 深色阴影部分为预测值的 80%置信区间, 浅色阴影部分为预测值的 95%置信区间, 可以看出预测值与实际观察值的趋势大致相似, 经计算可得实际数据与预测数据的平均绝对误差值为 29.5%。

Table 3. Comparison table of predicted value and actual value

表 3. 预测值与实际值对照表

月份	预测值	实际值	月份	预测值	实际值
1	-3.8079	-4.8	7	39.442	38.6
2	1.9917	4.0	8	36.892	38.5
3	18.711	22.7	9	33.412	30.8
4	28.991	29.1	10	23.649	28.3
5	34.523	32.5	11	13.745	9.4
6	37.745	36.6	12	1.9534	0.7

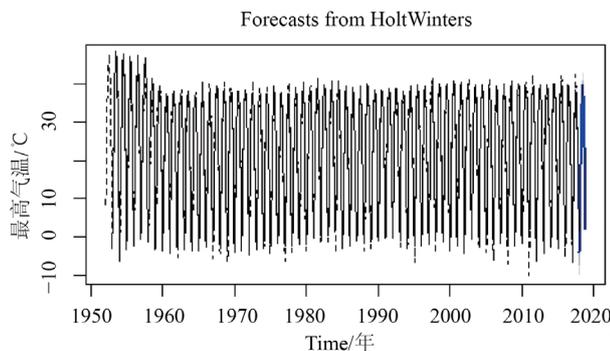


Figure 4. Forecast effect picture

图 4. 预测效果图

通过模型的建立, 参数估计, 模型检验三个步骤, 能够得到一个完整的 Holt-Winters 三参数指数平滑模型, 模型虽然合理有效, 但经计算的平均绝对百分误差 MAPE 值为 29.5%, 误差还是相对较大, 实际预测结果并不理想。说明该季节加法模型虽能大体预测出月最高气温的变化趋势, 但所预测的数值还是有所偏差, 且因月最高气温受较多因素影响, 仅凭因素分解的方法不能十分准确的预测月最高气温。

4. 结语

在对黑龙江克山站点的月最高气温序列进行分析预测的过程中, 首先通过对时间序列固定特征的判

断, 确定了选用 Holt-Winters 三参数指数平滑法中的季节加法模型, 其次借助了 R 软件拟合建立了合理的具体模型, 最后经过白噪声检验, 认为所建立的最高气温预测模型是合理有效的, 最终可对相应的时间序列做出未来发展情况的预测。

由于影响一个地区的月最高气温变化的因素是多种多样的, 例如: 降水量变化、其他地区的洋流影响、生态环境的改变、厄尔尼诺现象、以及人为因素的影响等, 只用时间序列分析方法来预测月最高气温的效果较差, 需考虑多方面的因素, 综合现有的预测模型, 建立适合当地的月最高气温预测模型, 从而使得月最高气温预测更为可靠[4]。

参考文献

- [1] 孙红杰. 芦苇床生态稳定污泥的效果与作用机理研究[D]: [博士学位论文]. 沈阳: 东北大学, 2013.
- [2] 王勇, 顾海燕, 徐文科. 基于 ARMA 模型的河川年径流量预测[J]. 哈尔滨商业大学学报(自然科学版), 2011(3): 320-323.
- [3] 王燕. 应用时间序列分析[M]. 第 2 版. 北京: 中国人民大学出版社, 2008.
- [4] 程敏, 张耀文, 姜纪沂, 等. 基于时间序列模型的降雨量预测分析[J]. 水科学与工程技术, 2019, 213(1): 5-9.