2021-05-25 T17: 40

尊敬的《大气科学》编辑部:

我们将附上题为"基于 EOF 的自动气象站气温观测资料修复方法"的论文, 希望能被考虑在贵杂志"大气科学"上发表。投稿时所有作者均同意投稿。在此, 我谨代表我的合著者声明,我所描述的研究是从未发表过的原始研究,也没有考 虑在其他地方全部或部分发表。所有列出的作者都认可了随信附上的手稿。

我国已经建成了高时空分辨率的地面自动观测站,能够对边界层的一些常见 天气现象以及爆发性气象灾害进行实时的监测,因此,地面观测资料在天气预报 预警中具有重要作用。但是由于地面自动站资料容易出现质量问题,错误的自动 站观测资料会对天气预测和气候变化研究带来严重的影响。而在资料同化应用 中,对错误资料进行有效订正的前提下,通过对订正后资料进行合理的、定量化 的观测误差调整,将使得这些资料在模式中的应用变为可能;在气候应用中,提 供一个在时间和空间上更加完整的观测资料数据集,以确保历史观测资料的连续 性,有益于真实可靠地评估历史气候趋势和变率,尤其是对于气候态和极端事件 的研究非常重要。因此,如何高精度的修复错误的地面观测资料,更好发挥现有 区域自动站网的效益,具有很好的研究价值。

本文在有效识别自动站错误资料的前提下,提出了一种基于 EOF (Empirical Orthogonal Function)分析法的错误资料修复方法。该方法基于 EOF 分解的各模态之间的独立性,可以利用 EOF 方法提取该区域内不同空间尺度下的观测信息,在气温日变化条件的约束下,充分利用气温观测的空间连续性和典型分布特征,可以逐步恢复目标位置在不同空间尺度下的观测信息。对重构数据的精度分析以及实际的修复结果显示,EOF 迭代法不受地形等因素影响,能够较高精度地恢复自动站的错误气温观测值,重构结果很好的保持了修复数据的空间连续性和日变化特征。因此,新修复方法可以为地面观测数据的应用提供重要的数据支撑。

我希望这篇文章适合"大气科学"。

下面是推荐的三个审稿人专家,供您参考:

1)姓名:李娟,邮箱:lj@cma.gov.cn,联系电话:13671329159,地址:北京 市中关村南大街 46 号中国气象局

2)姓名: 孙丞虎, 邮箱: sunch@cma.gov.cn, 联系电话: 13522273371, 地址: 北京市中关村南大街 46 号中国气象局 3)姓名:郝增周,邮箱: hzyx80@sio.org.cn,联系电话: 13588165270 地址: 浙江省杭州市西湖区保俶北路 36 号自然资源部第二海洋研究所

我们非常感谢您对我们手稿的考虑,并期待早日收到审稿人的意见。如果您 有任何疑问,请及时与我联系,以下是我的地址与联系方式:

投稿人: 沈王彬 研究方向: 气象资料的质量控制与同化 地址: 江苏省南京市浦口区宁六路南京信息工程大学 联系电话: 18351806901 邮箱: wangbinshen@nuist.edu.cn

基于 EOF 的自动气象站气温观测资料修复方法

沈王彬 1.2、李昕 3、秦正坤 1.2、张冰 3

1.南京信息工程大学大气科学学院,江苏,南京 210044;

2.南京信息工程大学 大气科学学院 资料同化研究与应用联合中心,江苏,南京 210044:

3.南京气象科技创新研究院/江苏省气象科学研究所,江苏,南京 210009;

摘要 全国目前已经建成了近7万个自动气象观测站点,然而自动气象站观测资料一直存在资料 质量较低的问题,大量错误资料的存在极大影响了其在气象研究中的应用,因此对错误观测的数 据进行准确的修复是一项重要工作。本文利用 2019 年 12 月 1 日 00 时至 7 日 23 时,共 168 个时 次的地面自动站温度观测资料,在利用 EOF (Empirical Orthogonal Function)质量控制方法识别 异常观测资料的基础上,提出了一种基于迭代 EOF 分析方法的错误资料修复方法。通过理想修 复试验的精度分析表明,新修复方法能够很好地修复错误的地面自动站观测气温,修复方法的误 差约为 0.48℃。而基于 Cressman 插值等这一类依赖单点观测信息进行修复的方法更容易受到小 尺度信号干扰而引入非自然观测信息,对地面温度的修复误差可以达到 1.55℃。实际的修复结果 分析也证明新修复方法充分利用了 EOF 分析方法的时空分离作用和模态正交性特点,通过迭代 方法逐步消除错误资料的影响,从而获得了与周边观测资料有更好时空连续性的修复结果。

关键词: 自动气象站; 地面温度; 经验正交函数分解

Restoration method for automatic station temperature observation

data based on EOF

Shen Wangbin^{1,2}, Li Xin³, Qin ZhengKun^{1,2}, Zhang Bin³

¹School of Atmospheric Science, Nanjing University of Information Science & technology, Nanjing210044, China;

²Joint Center of data assimilation research and applications, school of atmospheric science,Nanjing University of Information Science & technology,Nanjing 210044,China;

³ Nanjing Joint Institute for Atmospheric Sciences, Nanjing 210009, China





作者简介 沈王彬, 男, 1995年6月出生, 博士生, 主要从事气象资料的质量控制与同化研究。E-mail: wangbinshen@nuist.edu.cn

资助项目 国家重点研发计划资助(2018YFC1507302),国家自然科学基金青年项目(41805076),中国气象科学研究院基本科研业务费专项资金(2019Z006)

Funded by National Key R&D Program of China(2018YFC1507302), National Natural Science Foundation of China(41805076), Fundamental Research Funds of the Chinese Academy of Meteorological Sciences(2019Z006)

Abstract: With the construction of about 70,000 automatic weather stations across China, comprehensively automatic meteorological observation has been realized. However, the real application of this kind of observations always surfers from their low quality. Large number of error data seriously affect the practical application of observation. Therefore, it is a particularly important work to repair those abnormal observations. Using a total of 168 times of hourly surface temperature observations of automatic weather station during December 1-7, 2019 provided by Jiangsu meteorological bureau, a restoration method based on the Empirical Orthogonal Function method is proposed. The accuracy analysis of the ideal restoration experiments shows that the new restoration method can well repair the wrong observations, and the error of the restoration method is about 0.48 °C. The methods based on Cressman interpolation, which rely on single point observation information, are more vulnerable to small-scale signal interference and introduce unnatural observation information, and the surface temperature repair error can reach 1.55 °C. The analysis of the actual repair results also proves that the new repair method makes full use of the time-space separation and modal orthogonality of EOF analysis method, and gradually eliminates the influence of wrong data through iterative method, so as to obtain better space-time continuity repair results with the surrounding observation data. Key words: Ground observation automatic station; Surface temperature; Empirical Orthogonal Function

0 引言



随着观测技术的发展,目前我国已经建立了巨大且密集的地面气象观测网络,全国已建成约 七万个自动气象观测站,自动气象观测站能够定时、定点的观测和记录地面气象信息,并且能够 在复杂的地形下进行无人观测(张慧,2012)。但是随着自动化气象观测事业的发展同时也带了 些许问题,张颖超等(2017)、徐浩然(2019)指出随着观测自动化的实现,原有的业务模式发 生改变,但是地面观测技术的不稳定导致地面观测资料更容易出现质量问题,一方面众多台站的 观测资料往往由于采集、传输或仪器自身的系统问题,不可能达到完全正确无误,另一方面,区 域自动站资料的质量控制和评估工作较为滞后,直接影响了资料在预报预测业务中的使用。

为了提高区域自动气象站观测的数据质量,需要通过质量控制来保证数据质量(Lorenc等, 1988)。近十年来,我国气象工作者在实时观测资料自动质量控制方面做了大量的研究工作。国 内外有关地面气象资料的常规质量控制方法主要有气候学界限值检查、台站或区域极值检查、要 素间内部一致性检查、时间一致性检查以及空间一致性检查等方法。王海军等(2007)、陶士伟 等(2009)、陈兴旺(2011)、赵煜飞等(2011)都利用常规质量控制方法对我国的地面观测资 料进行了质量控制试验。李良富等(2006)和叶小岭等(2014)还尝试了利用人工智能的黑板模型和基因表达式编程方法进行了质量控制。基于空间分析方法 EOF(Empirical Orthogonal Function)分解方法,Qin 等(2010)和邵宇行等(2021)针对全国高时间和空间分辨率的地面自动站观测资料,对前人提出的基于 EOF分析的质量控制方法进行适应性的调整,并利用我国中东部的自动站观测资料进行实际质量控制试验,发现该方法可以很好适用于高时空分辨率的地面自动站观测资料。

错误的自动站观测资料会对天气预报和气候变化研究带来严重的影响,根据研究目的不同, 气象学家对错误资料的处理方法也不尽相同。气候研究往往是对资料的稳定性提出较高的要求, 如何把误差的大小控制在气象业务所允许的范围内,提供一个在时间和空间上更加完整的观测资 料数据集,以确保历史观测资料的连续性是自动站观测资料应用上面临的问题(王晓春等,1994; 江志红等,2008;曹丽娟等,2011)。而在资料同化应用中,为了保证同化增量的合理性,往 往倾向于直接剔除可疑的观测资料(薛纪善,2009),但错误资料也并非完全不可利用,在进行 有效订正的前提下,通过对订正后资料进行合理的、定量化的观测误差调整,将使得这些资料在 模式中的应用变为可能。

资料序列的订正方法研究是国内外学者所共同关注的问题之一,并已有不少研究。根据方法 的原理不同,目前的错误资料或者缺测资料的修复方法大致可以分为以下三类:一是利用观测资 料本身的规律或者不同气象变量之间的物理关系来进行资料修复,多用于气候资料(Annan et al., 2013; 李天琪等,2015;周方成等,2020),或者规律性显著的资料(Göttsche and Olesen et al., 2001; Duan et al., 2012);二是利用随机模型方法,利用扰动的随机样本之间的相似性来修复数 据,比如神经网络方法(Wu,2019)、自组织映射法和期望最大化法等(Sorjamaa,2007; Zhao et al., 2020);三是根据变量的时空特征,利用最优化和多维统计方法来修复资料,比如最优插 值方法(朱江等,1995; Everson et al., 1997)、经验正交函数(EOF)分解方法(Alvera-azcarate et al, 2005;盛峥等,2009; Shen et al., 2019)。

同时,研究也发现没有绝对最优的空间内插方法,必须对所研究的数据进行空间探索分析, 根据数据的特点,选择相应方法,对某一种气象要素合适的插值方法未必对另一种气象要素合适 (蔡彦枫等,2015)。我国目前已经建成了六万多个地面自动观测站,高时空分辨率的地面自动 站资料最大的特征就是小尺度特征强,时空变率大,容易出现极值情况,尤其是对于天气变化尺 度的研究,仅依赖于温度的日变化等变量固有的规律性特征进行修复容易出现较大的误差,而较 多的极值情况也导致了基于随机性的修复方法存在一定的不足,所以依赖于周边资料来修复地面 站温度资料是一种比较可行的方法。 最优化和多维分析方法都可以依据周边资料来修复单个错误资料,但是最优化方法往往需要 资料的误差等先验信息,所以基于 EOF 等多维分析方法的错误资料修复方法的就成为更为简单 易行的方法。基于 EOF 分析的时空分离的特征,Alvera-azcarate et al. (2009)和 Shen et al. (2019) 都提出了基于 EOF 空间分析方法的错误资料修复方法。EOF 方法的优点在于利用空间降维技术 将原始变量场的时空结构进行正交分解,得到为数很少的不相关典型模态来代替原始变量场。 EOF 方法提取的典型模态存在两个重要特征,一是模态是在考虑了变量场时间和空间变化特征 后确定的,不容易受到少数极值资料的影响,这种特征对于提取小尺度信号强的地表信息时尤为 重要;二是 EOF 方法提取的各个模态是相互正交的,各个模态及其对应的系数在处理过程中不 会相互影响,所以可以考虑通过迭代的方法逐步确定各个模态的系数,不需要考虑不同模态系数 之间的相互影响。利用 EOF 分析方法的这两个特征,我们就可以建立一个基于迭代的错误资料 修复方法。首先利用 EOF 分析方法将研究区域的大气温度分离为不同尺度的空间模态,然后分 别针对代表不同时空尺度的 EOF 模态,通过迭代的方法消除错误资料对 EOF 分析模态的影响。 由于个别错误资料不会影响各个模态所对应的时间系数,结合迭代修正的 EOF 分析模态和稳定 的时间系数,我们就可以具有很好的时空连续性的修复结果。

本文将针对全国自动气象站观测资料高时空分辨率的特征,在有效识别自动气象站异常观测 资料的基础上,利用 EOF 分析方法,建立一个新的基于 EOF 迭代方法的错误资料修复方法,并 对中国东部自动站气温观测资料进行了错误观测资料的修复试验,希望为自动气象站错误观测数 据的修复提供科学依据。

1.资料和方法

1.1 数据说明

本文选取的资料为江苏省气象局通过 CIMISS 数据库实时接收的 2019 年 12 月 1 日 00 时至 7 日 23 时的逐小时地面自动站观测资料,共 168 个时次的地面观测站温度观测资料。选取的研 究区域为113°E-123°E,29°N-39°N。高水平分辨率是自动站观测资料的主要特点,图 1 给出了自动站站点的空间分布特征,其中站点分布最密的主要是在江苏、山东和安徽三个省份。



Figure 1. Spatial distribution of ground automatic observation stations (black points)

由于观测资料中存在大量缺测资料,为了保证研究结果的普遍性,我们需要对某些时刻缺测 值进行时间三次样条插值,但是由于某些站缺测时刻过多导致时间插值后存在极端异常值。所以 为了尽可能的消除时间插值对检测结果的影响,我们首先对观测资料进行筛选,筛选的条件如下: 1)前五个时次无缺测,2)后五个时刻无缺测,3)168个时次至少有1/10资料无缺测。最终该区 域包含4983个观测站。

1.2 质量控制方法

对自动气象站气温观测资料进行质量控制,准确的识别出其中错误的资料是修复错误资料的前提。为此,邵宇行等人(2021)提出了基于 EOF 的高时空分辨率自动站温度观测资料质量控制方法。按照该方法首先利用传统质量控制方法对资料进行了预处理,去除了数据中明显存在的误差。在此基础上,以ECWMF(European Center for Medium-Range Weather Forecasts)的ERA5 2m 温度再分析资料作为背景场信息,使用 EOF 质量控制方法对筛选出的 4983 个站点进行了质量控制。

图 2 给出了基于该质量控制方法识别出的 168 个时次的错误资料个数曲线。平均剔除资料为每日 48 个,剔除率约为 0.96%,大部分时间错误资料量都是稳定在 50 个左右,只是在第 136 时次和第 146 次错误资料个数异常偏多。



图2 2019年12月1-7日质量控制方法剔除资料量的时间变化曲线 Figure 2. The hourly data count of abnormal data detected by the quality control method during December 1-7, 2019

为了明确剔除资料的正确性,图3给出了2019年12月1日11时四个离群观测站点及其周围站点的气温分布情况,黑色阴影表示地形高度。该时次一共识别出89个离群站点,主要分布在安徽省的中部及南部。根据四个错误站点以及周围站点的温度分布可以看出,错误站点在第36个时次相较于周围站点观测温度均为异常高值,因此可以认定基于EOF的质量控制方法识别的错误资料是合理的。虽然质量控制方法能够剔除绝大部分温度异常的站点,但是依然可能存在漏检现象,上图中的区域中部(116.45°E,30.72°N)可以看到存在一个明显低于周边站点温度的资料,这可能还需要更长时间的资料来评估研究区域地面温度的空间连续性特征,后续研究中我们将引入更长时间资料来完善目前的质量控制方法。



图3 2019年12月1日11时(第36时次)局部温度数值(单位: ℃)。图中黑色阴影表示地形高度 (单位: m),红色为剔除站点

Figure 3. Spatial distribution of observed temperature (unit: °C) at 11:00 December 1, 2019 (36th time, Beijing time). The shading indicates the terrain(unit: m), red dots represent abnormal stations

从错误站点的识别结果可以看出,自动站观测资料在每个时刻均存在50个左右的错误资料, 在数值天气预报需求日益精细化的背景下,错误资料会对天气预报带来很多消极的影响。因此, 在对错误站点进行准确识别的基础上,本研究的主要的目的是对错误气温资料进行准确的修复。

1.3 地面自动站错误资料修复方法

在利用前期质量控制方法有效识别错误的气温观测资料基础上,我们将尝试对错误资料进行修复研究。为了不失一般性,我们将尝试修复其中任选的一个错误观测站 p 的气温观测资料。

高时空分辨率的地面自动站温度观测资料也具有时空变率强的特点,所以基于变量规律性特征或相似样本的修复方法存在一定的不足,因此利用周边观测信息修复某个站点的错误资料最为可行。基于最优插值方法和基于多维分析方法的修复方法是最常用的基于周边观测信息修复单个观测资料的方法。这里我们分别选用了两类修复方法,一是基于 Cressman 空间插值方法的修复方法,二是利用 Shen 等(Shen et al., 2019)建立的基于 EOF 分析方法的迭代修复方法。

1.3.1 基于 Cressman 空间插值修复方法

Cressman插值方法是一种比较经典的非规则格点空间插值方法(Cressman, 1959),该方法的基本原理是基于相邻越近的站点在空间分布上具有越相近特征的假设,由于地面自动站高空间分辨率的特点,因此利用错误站点周围正确的气温观测资料,通过空间插值方法可以获得该点的观测信息。该方法的基本形式如下:

$$F = \frac{\sum W(D_i) \cdot f_i}{\sum W(D_i)} \tag{1}$$

其中 f_i 是第 i 个需要被修复的观测站点的气温值, $W(D_i)$ 是 Cressman 权重系数,其形式为:

$$W(D_i) = \begin{cases} \frac{R^2 - D^2}{R^2 + D^2}, D_i \le R\\ 0, D_i > R \end{cases}$$
(2)

式中R为影响半径,这里经验的设定为 0.75° , D_i 为错误站点与周围正确观测站点的距离。 由权重函数可以看出,距离被修复的点越近的点权重函数越大,符合 Cressman 插值基本原理。

1.3.2 基于 EOF 迭代修复方法

基于 EOF 的迭代修复方法首先是建立用于 EOF 分析的数据矩阵。假定地面站 p 点被检测为错误观测站,再选取空间内所有观测站的观测资料,剔除其中已经检测的所有错误气温观测资料,剩余观测站数记为 n。考虑到地面温度的日变化,对于已选的 n 个观测点,选择每个观测点上当日 24 个时次的逐小时观测数据,假定修复某观测站第 12 时的气温观测资料,那么这里的 24 个时次选为:当日 00 时至 23 时连续 24 个时次的观测气温,记为*T_{i,j}*,其中

i=1,2,...,24表示一共选取 24个时次的观测资料, j=n表示一共选取 n个站点进行重构。

最终用于修复的数据矩阵可以表示为 A_{m×n}:

$$A_{m \times n} = \begin{bmatrix} T_{1,1} & T_{1,2} & \cdots & T_{1,p} & \cdots & T_{1,n} \\ T_{2,1} & T_{1,2} & \cdots & T_{2,p} & \cdots & T_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{x,1} & T_{x,2} & \cdots & T_{x,p} & \cdots & T_{x,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{23,1} & T_{23,2} & \cdots & T_{23,p} & \cdots & T_{23,n} \\ T_{24,1} & T_{24,1} & \cdots & T_{24,p} & \cdots & T_{24,n} \end{bmatrix}$$
(3)

其中,m=24,代表所选的24个时次,假设其中第 x 个时次,第 p 观测点 $T_{x,p}$ 为错误资料, 也是用于重构的观测点。为了增强错误资料与其周边资料点之间的小尺度异常特征,可以设 $T_{x,p} = 0$ 。对上述矩阵A进行 EOF 分解,我们可以得到:

$$A_{m \times n} = V_{m \times n} Z_{m \times n} = \sum_{k=1}^{m} V_{k}^{T} \overline{Z_{k}} = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{m} v_{i,k} Z_{k,j}$$
(4)

这里的 $\vec{v_k}$ 一般称为模态向量,而 $\vec{z_k}$ 则是系数向量,k代表第k个模态,因为所选的观测站数肯定是大于时次数,所以EOF分析获得的模态数最多为m个。

由于 EOF 第一模态提取的是资料的平均特征,所以有理由认为第一模态不会包含异常点 0 值的信息,那么利用第一模态进行重构就可以得到:

$$T_{x,p}^{1} = v_{1,1} z_{1,p}$$
(5)

获得准确的第一模态信息是保证方法修复效果的前提条件。虽然 0 值的异常信息不会被包含到第一模态中,但是 0 值仍然可能使得应有的第一模态信息被分散到其他模态中。因此需要通 过迭代的方式逐步完善第一模态所代表的平均特征。

在获得 $T^1_{x,p}$ 后,我们将该值带入到原来的数据矩阵,组成新的数据矩阵 A^1 ,即:

$$A^{1} = \begin{bmatrix} T_{1,1} & T_{1,2} & \cdots & T_{1,p} & \cdots & T_{1,n} \\ T_{2,1} & T_{1,2} & \cdots & T_{2,p} & \cdots & T_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{x,1} & T_{x,2} & \cdots & T_{x,p}^{1} & \cdots & T_{x,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{23,1} & T_{23,2} & \cdots & T_{23,p} & \cdots & T_{23,n} \\ T_{24,1} & T_{24,1} & \cdots & T_{24,p} & \cdots & T_{24,n} \end{bmatrix}$$
(6)

对矩阵 A^1 再次进行 EOF 分解可得:

$$A^{1} = V^{1}Z^{1} = \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{m} v_{i,k}^{1} z_{k,j}^{1}$$
(7)

同样获得基于第一模态重构获得的 p 点气温, 即:

$$T_{x,p}^2 = v_{1,1}^1 z_{1,p}^1 \tag{8}$$

在 A^1 矩阵中利用第一模态的重构资料替换了A矩阵中的 0 值,从而减小了异常资料对第 一模态完整性的影响,所以 A^1 的 EOF 第一模态更为接近第一模态的真值,即 $T_{x,p}$ 未出现错误观 测的条件下获得的 EOF 第一模态。由于真值的唯一性,这就使得迭代过程中 EOF 的第一模态会 逐渐稳定,因此每次迭代过程中 EOF 第一模态重构值的变化量可以作为 EOF 模态稳定性的指标。 所以我们计算前后两次重构的第 x 个时次 p 点的差值绝对值, $\left|T_{x,p}^1 - T_{x,p}^2\right|$ 。如果该差值大于指 定阈值,则认为还有平均特征信息遗留在第一模态之外,需要重复上述过程,继续构造新的数据 矩阵A,并进行 EOF 分解,建立第一模态重构气温。直到前后两次重构的第 x 个时次 p 点气温 的差值小于指定阈值,假定已经进行了 t_1 次迭代后满足要求,即

$$\left|T_{x,p}^{t_1-1}-T_{x,p}^{t_1}\right|\leq \varepsilon$$

(9)

可以认为第一模态所代表的平均特征已经能够完全保留,这也意味着第 x 个时次 p 点气温的 平均特征也得到了很好的再现。文中 ε 我们经验性的指定为 0.01。

为了研究迭代过程,随机选取了 2019 年 12 月 4 日 14 时(第 87 个时次)中的某个站点观测 为例。根据以上描述,用于重构的时次为 12 月 4 日 00 时至 23 时共 24 个时次的气温观测,剔除 其中所有错误的站点后,图 4(a)为剩余的站点分布,图中五角星标注的位置为被重构的站点,图 4(b)为第一模态迭代重构过程中 $\left|T_{x,p}^{t_1-1} - T_{x,p}^{t_1}\right|$ 的变化曲线。可以看到大约经过 6 次迭代后,该值 逐渐趋近于 0,说明迭代趋于稳定,第一模态所代表的平均特征也已经被完全保留。



图 4 (a)用来重构的气温分布以及(b) $\left|T_{x,p}^{t_1-1} - T_{x,p}^{t_1}\right|$ 随迭代次数的变化曲线(单位: °C) Figure 4. (a) Spatial distribution of temperature for reconstruction and (b) $\left|T_{x,p}^{t_1-1} - T_{x,p}^{t_1}\right|$ varying with the number of iterations for station P (unit: °C)

实现了第 x 个时次 p 点气温第一模态特征完全再现的基础上,我们继续提取第二模态的信息。经过 t₁次迭代后,构造的新的数据矩阵为 A^{t₁} 为:

$$A^{t_{1}} = \begin{bmatrix} T_{1,1} & T_{1,2} & \cdots & T_{1,p} & \cdots & T_{1,n} \\ T_{2,1} & T_{1,2} & \cdots & T_{2,p} & \cdots & T_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{x,1} & T_{x,2} & \cdots & T_{x,p}^{t_{1}} & \cdots & T_{x,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ T_{23,1} & T_{23,2} & \cdots & T_{23,p} & \cdots & T_{23,n} \\ T_{24,1} & T_{24,1} & \cdots & T_{24,p} & \cdots & T_{24,n} \end{bmatrix}$$
(10)

对矩阵 A^{t_1} 进行EOF分解,并利用前两个模态重构 p 点地面温度:

$$T_{x,p}^{t_1+1} = \sum_{k=1}^{2} v_{1,k}^{t_1} z_{k,p}^{t_1}$$
(11)

从前面的分析可以知道,我们已经获得了 $T_{x,p}$ 稳定的第一模态信息,并构造了新的数据矩阵,这等同于将除第1模态外的其他模态分量设为0,简单的重复上述迭代过程,我们可以获得具有前两个模态准确信息的 $T_{x,p}^{t_2}$ 。同样通过对更多模态进行上述迭代过程,我们可以依据 EOF 提供的周边观测点信息和其它未受干扰的通道观测信息的约束条件下,获取包含尽可能多正确观测信息。

理论上可以对所有模态进行提取,所以图 5 也给出了不同模态迭代过程中误差的变化曲线。 可以看到,前七个主要模态迭代结束后,重构值趋于稳定,此时误差在 0.5 ℃左右。继续对剩余 的模态进行提取,可以看到,由于剩余模态包含一些小尺度的气温场信息,因此重构误差曲线出 现了一些波动,但是当所有模态均被提取后,重构误差小于 0.1 ℃。



图 5 重构过程中观测值与重构值的差随迭代次数的变化曲线(单位: ℃)

Figure 5. Temperature differences between the observed and the reconstructed temperature varying with the number of iterations (unit: $^{\circ}$ C)

2. 结果与分析

2.1 重构方法精度分析

在新的修复方法适用之前,应该对该方法的修复精度进行评估。首先选取单个时刻的气温观 测资料中的正确资料作为验证数据。图 6 给出了 2019 年 12 月 4 日 14 时(第 87 个时次)用 EOF 质量控制方法识别出的错误站点的分布。可以看到,该时次一共出现 19 个错误站点。



图 6 2019 年 12 月 4 日 14 时(第 87个时次) EOF 质量控制方法识别出的错误站点分布 Figure 6. Spatial distribution of abnormal data (red points) at 14:00 PM December 4, 2019 (87th time, BT)

对该时刻所有正确资料逐个进行重构,将重构结果与观测进行对比,并统计其均方根误差 可以有效的验证重构方法的精度。均方根的计算公式如下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{N} \left(T_i^{ob} - T_i^{re}\right)^2}$$
(12)

式中,n表示选取的站点数, T_i^{ob} 为站点的观测气温值, T_i^{re} 为重构得到的站点气温。若均方根误差越小,则表明重构结果精度更高。

图 7 给出了 2019 年 12 月 4 日 14 时(第 87 个时次)正确观测气温、两种方法重构的气温 分布以及观测与重构的偏差(观测-重构)。从自动站正确观测资料的分布可以看到,自动站空 间分辨率较高,因此能很好分辨更多小尺度的温度变化特征。空间分布形态也是评价插值方法的 重要方面,对比基于 EOF 迭代法和 Cressman 插值法得到的气温分布,可以看到,EOF 迭代法重 构的气温与观测非常接近,重构场不仅能准确的描述气温观测的空间特征,对于观测中存在的一 些极值也能很好的还原,这是由于 EOF 方法考虑了变量场时间和空间变化特征,不容易受到少 数极值资料的干扰,也不易受观测站点分布的疏密程度和不均匀性的影响。而 Cressman 插值法 得到的重构场只能大致反映观测气温的特征,但是在极值和极值附近区域重构值与观测存在较大 差异,原因是 Cressman 方法只考虑周围站点的距离来确定其权重,而不考虑它们之间的空间相 关性,即是站点距离越近则变量值越大,分析值就越靠近,从而在极值及其附近区域出现偏向的 特征。统计结果表明 EOF 迭代法的重构误差为 0.41℃,而 Cressman 插值的重构误差为 1.25℃, 远大于 EOF 迭代法,从重构误差的分布上可以看到,EOF 迭代法的重构误差基本在 0.5℃以内, 并且误差的分布比较随机,与气温本身高低无关。Cressman 插值重构的误差较高,尤其在地形 变化强烈的区域附近,如安徽省的大别山以及黄山附近。由于地形影响,气温在大地形附近出现 了较大温度梯度,但是 Cressman 插值仅考虑测站之间的距离,不能结合观测站之间的空间相关 性,因此插值结果误差较大。



图 7 2019 年 12 月 4 日 14 时(第 87 个时次)观测气温、重构气温的分布以及偏差(单位: ℃) (a)观测气温的分布; (b) EOF 迭代法重构气温的分布; (c) EOF 迭代法重构气温的误差; (d) Cressman 插值法重构得到的气温分布; (e) Cressman 插值法重构气温的误差 Figure 7. Spatial distribution of observed and restored temperature and their difference distribution at 14:00. PM December 4, 2019 (87th time, BT)(unit: ℃) (a) Observed temperature; (b) restored temperature of EOF method; (c) restoration error of EOF method; (d) reconstructed temperature of Cressman interpolation method; (e) reconstruction error of Cressman interpolation method 为了增大样本数来提高统计可信度,我们又选择了 2019 年 12 月 1 日-7 日共 168 个时次, 共计 808260 个正确的气温观测,分别使用 EOF 迭代法和 Cressman 插值法进行重构。经统计, EOF 迭代法和的重构误差分别为 0.48 和 1.55 ℃,Cressman 插值的误差远大于 EOF 迭代法,这 与图 7 的结论一致。图 8 给出了所有正确观测气温与重构气温的散点图,黑色点表示 EOF 迭代 法对应的散点,灰色点表示 Cressman 插值法对应的散点。可以明确看到,EOF 迭代法重构的气 温与观测值较为接近,重构误差较小,而 Cressman 插值法由于受多方面的干扰重构效果较差, 重构后的气温表现出来的特征不明显,其偏差明显大于 EOF 迭代法。



图 8 观测气温与 EOF 迭代法(黑色点)以及 Cressman 插值法重构气温(灰色点)散点图(单位: ℃) Figure 8. Scatter plot between observed and reconstructed temperature of the EOF iteration method (black spots) and the Cressman interpolation method (gray spots) (unit: ℃)

进一步分析重构误差的分布情况,图9给出的是观测值与重构值偏差的概率密度函数分布曲线。可以看到,EOF迭代法重构误差分布的峰度系数为210,误差绝对值基本集中在0.5℃以内, 只有极少数误差绝对值大于1℃,但是极值也在2℃以内。而Cressman 插值重构误差峰度系数为 83.67,很大一部分站点的重构误差大于1℃。因此,相对于Cressman 插值而言,EOF迭代法具 有更高的稳定性和更高的精度。



图 9 EOF 迭代法(蓝色实线)以及 Cressman 插值(红色虚线)的重构误差概率密度曲线(单位: ℃) Figure 9. The probability density function of reconstruction error for EOF iteration method (the blue solid line) and Cressman interpolation method(the red dotted line) during 1-7 December 2019(unit: ℃)

2.2 实际重构结果分析

在根据正确数据对 EOF 迭代法的修复精度进行理想评估后,我们将修复方法应用到实际错 误资料的修复中进行检验。分别使用 EOF 迭代法和 Cressman 插值法对基于 EOF 的自动站质量 控制方法识别出的错误站点进行修复。图 10(a)给出了 2019 年 12 月 4 日 14 时(第 87 个时次) 局部温度数值(红色为错误站点温度数值),图中的阴影表示地形高度。两个站点在第 87 时次 被判定为离群点,根据两个站点第 87 时次的温度及其周围站点温度分布可以看出,两个站点在 第 87 时次相对于周围站点温度为异常高值,并且该错误站点的地面温度观测并未受地形的影响, 因此可以认定质量控制方法判定的错误资料是合理的。使用两种方法对该错误站点进行重构,图 10(b)、(c)分别给出了基于 EOF 迭代法和 Cressman 插值法修复的地面气温,可以看到,EOF 迭代法重构的气温值明显低于错误的观测气温,并且修复值与周围观测值更加接近,而 Cressman 插值结果与周围观测相比则异常偏低,这是由于错误站点南部出现了地形高度增加的情况,随着 地形高度的增加地面温度也出现了明显的降低,因此区域南部气温普遍低于区域北部,由于 Cressman 插值只考虑测站空间距离因素,南部的低温资料会使得最终插值结果偏低。EOF 迭代 法则能综合考虑气温的空间典型分布特征,使得重构结果不受地形影响。因此,EOF 迭代法修 复后的气温分布连续性更佳,其平滑性和规律性更好,修复效果也更好。



图 10 2019 年 12 月 4 日 14 时(第 87个时次)(a)局部观测气温、(b) EOF 迭代法重构气温、 (c) Cressman 插值气温数值(单位: ℃)。图中黑色阴影表示地形高度(单位: m) Figure 10. Spatial distribution of observed (a) and restored temperature by (b) the EOF iteration method and (c) the Cressman interpolation method at 14:00 PM on December 4, 2019 (87th time, BT) (unit: ℃). The shading indicates the terrain(unit: m)

重构的气温分布已经验证了修复气温在空间分布上的合理性。另一方面,自动气象站逐小时 高分辨率的气温观测资料为精细化研究气温的日变化差异、季节及气候特征带来了可能性,因此, 修复后的气温随时间变化是否合理也是评估修复气温质量的一项重要指标。图11给出了图10中两 个识别出的错误站点的观测以及修复后气温的日变化曲线,可以看到,与周围站观测气温相比, 两个存在错误观测资料的站点均在第15时次出现了异常偏高的观测气温(红色线),在该时刻远 高于周围所有观测站的气温(灰色线)。对北部站点而言,经过EOF迭代法修复后(蓝色线), 错误站点气温的时间变化趋势与周围站点基本一致,符合气温日变化的规律,而Cressman插值得 到的气温(黑色线)在该时刻出现异常偏低的不合理变化趋势。这是由于EOF迭代法不仅能考虑 到气温的典型空间分布特征,同时其在气温的时间变化趋势约束下对错误观测进行重构,最终得 到的重构结果能同时满足气温的时空特征,而Cressman插值结果则受到观测站南部地形影响,插值结果异常偏低。



图 11 观测气温(红色)、重构后气温(蓝色)、Cressman 插值后(黑色)气温以及周围站点的 气温(灰色)序列(单位:℃)

Figure 11. Time series of (a) observed and restored temperature by (b) the EOF iteration method and (c) the Cressman interpolation method during 10:00-20:00 PM on December 4, 2019 (BT) (unit: $^{\circ}$ C)

3. 结论与讨论

错误的自动站观测资料对天气预报和气候变化研究带来很大的限制,地面气候研究以及各个 领域需要的专业气象服务往往需要完整的自动站观测数据,因此对异常观测的资料序列进行订 正,确保观测资料的准确性和连续性至关重要。

本文利用江苏省气象局提供的 2019 年 12 月 1 日 00 时至 7 日 23 时,共 168 个时次的地面自 动站温度观测资料,在 EOF 的质量控制方法有效识别异常观测资料的基础上,提出一种新的基 于 EOF 的迭代方法用于修复错误的自动气象站地面气温观测。

结果表明,利用本文提出的 EOF 迭代修复法可以较高精度的对地面气温观测进行重构,其 重构误差为 0.48℃,远小于 Crssman 插值法,并且 EOF 迭代法修复后的气温在时空特征上都具 有合理性。这是因为 EOF 迭代法不仅考虑了地面气温的典型空间分布,还利用了气温随时间变 化的特征对重构进行约束,从而从整体上考虑了插补资料的效果。而 Cressman 一类的插值方法 只选取了错误点周边的部分资料,通过资料的连续性进行插值,但是其插值过程中不能考虑其他 周边资料的约束,所以容易受到个别极端资料的影响,这对于小尺度信息较强的地表变量来说, 更容易出现不稳定现象。

由于资料长度的限制,本文没有对长时间的气温观测进行修复试验,在后续的研究中我们 将针对更长时间的观测资料进行重构检验。另一方面,目前的工作只针对地面气温展开,在以后 研究工作中还应考虑利用EOF迭代法对其他一些地面观测变量进行修复,例如相对湿度、降水等。



参考文献(References)

- [1] Alvera-Azcárate A, Barth A, Sirjacobs D, et al. 2009. Enhancing temporal correlations in EOF expansions for the reconstruction of missing data using DINEOF [J]. Ocean Science, 5(4): 475-485.
- [2] Annan J D, Hargreaves J C. 2013. A new global reconstruction of temperature changes at the Last Glacial Maximum [J]. Climate of the Past, 9(1): 367-376.
- [3] 蔡彦枫, 张灿亨, 黄勇. 2015. 广东沿海地区极值风速空间插值方法的对比研究 [J]. 南方 能源建设, 2(S1): 187-192.
 - Cai Yanfeng, Zhang Canheng, Huang Yong. 2015. Study on the comparison of spatial interpolation methods used to estimate extreme wind speed spatial distribution using irregularly distributed station data in guangdong coastal area [J]. SOUTHERN ENERGY CONSTRUCTION, 2(S1): 187-192.
- [4] 曹丽娟. 2011. 地面气候资料均一性研究进展 [J]. 气候变化研究进展, 7(2):129-135.
- Cao Lijuan. 2011. Progresses in Research of Homogenization of Climate Data [J]. Adv. Clim. Change Res., 7(2): 129-135.
- [5] 陈兴旺. 2011. 自动气象站数据质量控制方法及误差分析 [J]. 科技资讯, (23): 129-131.
 Chen Xingwang. 2011. Data quality control method and error analysis of automatic meteorological station [J], Science & Technology Information, (23): 129-131.
- [6] Cressman G. P. 1959. An operational objective analysis system [J]. Mon. Wea. Rev., 87: 365-374.
- [7] Duan S B, Li Z L, Wang N, et al. 2012. Evaluation of six land-surface diurnal temperature cycle models using clear-sky in situ and satellite data [J]. Remote Sensing of Environment, 124: 15-25.
- [8] Everson, R., Cornillon, P., Sirovich, L., & Webber, A. 1997. An empirical eigenfunction analysis sea surface temperatures in the western North Atlantic [J]. Journal of Physical Oceanography, 27(3): 468-479.
- [9] Frank-M GÖttsche, Folke S Olesen. 2001. Modelling of diurnal cycles of brightness temperature extracted from METEOSAT data [J]. Remote Sensing of Environment, 76(3): 337–348.
- [10] 江志红, 黄群, 李庆祥. 2008. 近50年中国降水序列均一性检验与订正研究 [J]. 气候与环境 研究, 13(1): 67-74. doi: 10.3878/j.issn.1006-9585.2008.01.09
 - Jiang Zhihong, Huang Qun, Li Qingxiang. 2008. Study of Precipitation Series Homogeneous Adjustment and Their Correction over China in the Last 50 Years [J]. Climatic and Environmental Research, 13(1): 67-74. doi: 10.3878/j.issn.1006-9585.2008.01.09
- [11] 李良富, 王汉杰, 刘金玉, 宋帅. 2006. 基于黑板模型的地面气象数据质量控制 [J]. 气象科技, 34(2): 199-204.
 - Li Liangfu, Wang Hanjie, Liu Jinyu, Song Shuai. 2006. Surface Meteorological Data Quality Control Based on Blackboard Model [J]. Meteorological Science and Technology (in Chinese), 34(2): 199-204.
- [12] 李天祺,朱秀芳,潘耀忠,刘宪锋. 2015. MODIS陆地表面温度数据重构方法研究 [J]. 北京师范大学学报(自然科学版), 51(S1): 70-76.
 - Li Tianqi, Zhu Xiufang, Pan Yaozhong, Liu Xianfeng. 2015. Study on reconstruction methods of MODIS LST product [J]. Journal of Beijing Normal University(Natural Science), 51(S1): 70-76.
- [13] Lorenc A c, Hammon o. 1988. Objective quality control of observations using Bayesianmethods. Theory, and a practical implementation [J]. QuarJ RoyMeteor Soc, 114(480): 515-543.
- [14] Qin Z K, Zou X, Li G, et al. 2010. Quality control of surface station temperature data with non-Gaussian observation-minus-background distributions [J]. Journal of Geophysical Research: Atmospheres, 115(D16).
- [15] 邵宇行,秦正坤,李昕. 2021. 基于EOF的高时空分辨率自动站温度观测资料质量控制 [J]. 大气科学学报, accepted.
 - Shao Yuhang, Qin Zhengkun, Li Xin. 2021. High spatial and temporal resolution automatic station temperature observation data quality control based on EOF [J]. Trans Atmos Sci, accepted
- [16] Shen W, Qin Z, Lin Z. 2019. A New Restoration Method for Radio Frequency Interference Effects on AMSR-2 over North America [J]. Remote Sensing, 11(24): 2917.

- [17] 盛峥, 石汉青, 丁又专. 2009. 利用DINEOF方法重构缺测的卫星遥感海温数据 [J]. 海 洋科学进展, 27(2): 243-249.
 - Sheng Zheng, Shi Hanqing, Ding Youzhuan. 2009. Missing Satellite-Based Sea Surface Temperature Data Reconstructed by DINEOF Method. Advances in Marine Science (in Chinese), 27(2): 243-249.
- [18] Sorjamaa A, Lendasse A. 2007. Time series prediction as a problem of missing values: Application to estsp2007 and nn3 competition benchmarks[C]//2007 International Joint Conference on Neural Networks. IEEE: 2948-2953.
- [19] 陶士伟, 仲跻芹, 徐枝芳, 郝民. 2009. 地面自动站资料质量控制方案及应用 [J]. 高原 气象, 28(5): 1202-1209.
 - Tao Shiwei, Zhong Qiqin, Xu Zhifang, Hao Min. 2009. Quality control schemes and its application to automatic surface weather observation system [J]. Plateau Meteorology (in Chinese), 28(5): 1202-1209.
- [20] 王海军,杨志彪,杨代才,龚贤创. 2007. 自动气象站实时资料自动质量控制方法及其应用. 气象, 33(10): 102-109.
 - Wang Haijun, Yang Zhibiao, Yang Daicai, Gong Xianchuang. 2007. The Method and Application of Automatic Quality Control for Real Time Data from Automatic Weather Stations [J]. Meteorological Monthly (in Chinese), 33(10): 102-109.
- [21] 王晓春, 蔡雅萍, 梁幼林. 1994. 用Bivariate-Test方法分析温度资料的不均一性 [J]. 大 气科学, 18(z1): 856-867. doi: 10.3878/j.issn.1006-9895.1994.z1.10
 - Wang Xiaochun, Cai Yaping, Liang Youlin. 1994. To Detection of the Inhomogeneity of Yearly-mean Surface Temperature Data Using the Bivariate-Test Method [J]. Chinese Journal of Atmospheric Sciences, 18(z1): 856-867. doi: 10.3878/j.issn.1006-9895.1994.z1.10
- [22] Wu P, Yin Z, Yang H, et al. 2019. Reconstructing geostationary satellite land surface temperature imagery based on a multiscale feature connected convolutional neural network [J]. Remote Sensing, 11(3): 300.
- [23] 徐浩然. 2019. 地面气象观测自动化存在的弊端及其对策 [J].河南科技, (22): 151-152.
 Xu Haoran. 2019. Disadvantages and countermeasures of surface meteorological observation automation [J]. Henan Science and Technology, (22): 151-152.
- [24] 薛纪善. 2009. 气象卫星资料同化的科学问题与前景 [J]. 气象学报, 67(6): 903-911.
 Xue Jishan. 2009. Scientific issues and perspective of assimilation of meteorological satellite data [J]. Acta Meteorologica Sinica (in Chinese), 67(6): 903-911.
- [25] 叶小岭, 施林红, 熊雄, 王璐. 2016. AI方法在地面气温观测资料质量控制中的应用 [J]. 气候与环境研究, 21(1): 1-7. doi: 10.3878/j.issn.1006-9585.2015.15073
 - Ye Xiaoling, Shi Linhong, Xiong Xiong, Wang Lu. 2016. Application of AI Method to Quality Control in Surface Temperature Observation Data [J]. Climatic and Environmental Research, 21(1): 1-7. doi: 10.3878/j.issn.1006-9585.2015.15073
- [26] 张慧. 2012. 分析自动站资料对气象预报的重要性 [J]. 农业与技术, 32(12): 124.
 - Zhang Hui. 2012. The importance of analyzing data from automatic stations for weather Forecasting [J]. Agriculture and Technology, 32(12): 124.
- [27] 张颖超,姚润进,熊雄,沈云培. 2017. PSO-PSR-ELM集成学习算法在地面气温观测资料 质量 控制中的应用 [J]. 气候 与环境研究, 22(1): 59-70. doi: 10.3878/j.issn.1006-9585.2016.16013
 - Zhang Yingchao, Yao Runjin , Xiong Xiong, Shen Yunpei. 2017. Application of PSO-PSR-ELM-Based Ensemble Learning Algorithm in Quality Control of Surface Temperature Observations [J]. Climatic and Environmental Research, 22(1): 59-70. doi: 10.3878/j.issn.1006-9585.2016.16013
- [28] Zhao J, Lange H, Meissner H. 2020. Gap-filling continuously-measured soil respiration data: A highlight of time- series-based methods [J]. Agricultural and Forest Meteorology, 285-286.
- [29] 赵煜飞, 任芝花, 张强. 2011. 适用于全国气象自动站正点相对湿度资料的质量控制方法 [J].气象科学, 31(6): 687-693.
 - Zhao Yufei, Ren Zhihua, Zhang Qiang. 2011. Quality control procedures for hourly relative humidity data from national automatic weather stations [J]. Scientia Meteorologica Sinica (in Chinese), 31(6): 687-693.
- [30] 周芳成, 唐世浩, 韩秀珍, 等. 2021. 云下遥感地表温度重构方法研究 [J]. 国土资源遥感, 33(01): 78-85.

Zhou Fangcheng, Tang Shihao, Han Xiuzhen, et al. 2021. Research of reconstructing missing remotely sensed land surface temperature data in cloudy sky [J]. Remote Sensing for Land and Resources, 33(01): 78-85.

- [31] 朱江, 徐启春, 王赐震, 李许花. 1995. 海温数值预报资料同化试验I. 客观分析的最优插 值法试验 [J]. 海洋学报(中文版), (06): 9-20.
 - Zhu Jiang, Xu Qichun, Wang Cizhen, Li Xuhua. 1995. Data assimilation experiment of SST numerical forecast Experiment of optimal interpolation method for objective analysis [J]. Acta Oceanologica Sinica, (6): 9-20.