

张颖超, 姚润进, 熊雄, 等. 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. Zhang Yingchao, Yao Runjin, Xiong Xiong, et al. 2017. Application of PSO-PSR-ELM-based ensemble learning algorithm in quality control of surface temperature observations [J]. Climatic and Environmental Research (in Chinese), 22 (1): 59–70, doi:10.3878/j.issn.1006-9585.2016.16013.

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

张颖超^{1,2} 姚润进¹ 熊雄³ 沈云培¹

¹ 南京信息工程大学信息与控制学院, 南京 210044

² 南京信息工程大学气象灾害预报预警与评估协同创新中心, 南京 210044

³ 南京信息工程大学气象灾害教育部重点实验室, 南京 210044

摘 要 针对台站密度低、新建台站以及特种单要素站等无法获得有效邻站或内部参考资料情况下的质量控制问题, 从气温时间序列的混沌特性出发, 考虑气温在短时间内的连续性和稳定性, 提出一种基于粒子群 (Particle Swarm Optimization, PSO) 改进的相空间重构法 (Phase Space Reconstruction, PSR) 和极限学习机 (Extreme Learning Machine, ELM) 的集成学习算法的地面逐时气温观测资料的单站质量控制方法, 实现气温资料的质量控制。为检验该方法的适用性, 运用该方法对江苏省八市 2007~2009 年的地面气温观测资料进行质量控制, 并与传统单站方法及切比雪夫多项式内插法 (Tshebyshev Polynomial Interpolation, TPI) 进行对比。实验结果表明, 该方法相比于 TPI 和传统方法可以更有效地标记出可疑数据, 具有检错率高、地区和气候适应性、可控性强等优点。

关键词 质量控制 气温 混沌性 粒子群 相空间重构 极限学习机

文章编号 1006-9585 (2017) 01-0059-12

中图分类号 P413

文献标识码 A

doi:10.3878/j.issn.1006-9585.2016.16013

Application of PSO-PSR-ELM-Based Ensemble Learning Algorithm in Quality Control of Surface Temperature Observations

ZHANG Yingchao^{1,2}, YAO Runjin¹, XIONG Xiong³, and SHEN Yunpei¹

¹ School of Information and Control, Nanjing University of Information Science and Technology, Nanjing 210044

² Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing 210044

³ Key Laboratory of Meteorological Disaster of Ministry of Education, Nanjing University of Information Science and Technology, Nanjing 210044

Abstract In order to overcome the quality control problem over areas where the station density is low, or for some stations that have no adjacent stations and lack effective internal reference data, for example those newly deployed stations and special single factor stations, a new quality control method for surface temperature observations based on the ensemble learning algorithm of Phase Space Reconstruction (PSR) and Extreme Learning Machine (ELM) that was improved by Particle Swarm Optimization (PSO) was introduced in detail in this paper. This method considers the chaotic

收稿日期 2016-01-13; 网络预出版日期 2016-07-06

作者简介 张颖超, 男, 1960 年出生, 教授, 主要从事气象资料质量控制、气象灾害评估与损失预测、气象服务效益等研究。Email: yc.nim@163.com

通讯作者 姚润进, Email: runjin_1991@163.com

资助项目 国家自然科学基金项目 41675156, 江苏省六大人才高峰项目 WLW-021, 江苏省高校优势学科建设工程资助项目

Funded by National Natural Science Foundation of China (Grant 41675156), Six Talent Peaks Project in Jiangsu Province (Grant WLW-021), A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions

characteristics of the time series of temperature. In order to assess the feasibility and applicability of the proposed method, it was applied to hourly temperature observations from 2007 to 2009 in eight cities of Jiangsu Province. Results were examined and compared against that of conventional single-station quality control method and Tshebyshev Polynomial Interpolation (TPI) method. It was found that the method introduced in this study can flag suspicious data more effectively, and it also has the advantages of high identification accuracy and good adaptability and controllability over different regions with various climate backgrounds.

Keywords Quality control, Temperature, Chaos, Particle swarm optimization, Phase space reconstruction, Extreme learning machine

1 引言

地面气象观测资料是一切气象、气候研究的基本前提,同时也是短期数值天气预报的基础 (Ingleby and Lorenc, 1993; Feng et al., 2004)。我国自 20 世纪 90 年代末开始在部分台站布设地面自动气象观测系统,至 2011 年,全国有 2500 多个地面自动气象站应用到气象观测业务 (韩海涛和李仲龙, 2012)。这一方面使观测资料的获取更加全面,但同时也给许多观测资料带来质量上的问题。因此,寻求以提高数据质量为目的的质量控制方法具有一定的必要性和迫切性。

地面气象观测资料质量控制方法可以分为单站质量控制和多站联网质量控制方法。对于多站联网质量控制,近年来得到了国内外一些学者的重视,提出了一些有效的方法。最佳内插法 (Lorenc, 1981)、反距离加权法 (Wade, 1987)、空间回归检查法 (Hubbard and You, 2005) 通过观测值与真值之间的残差来标记可疑数据实现质量控制;贝叶斯质量控制法 (Lorenc and Hammon, 1988) 引入贝叶斯理论,从观测资料出错概率的角度进行质量控制;熊安元 (2003) 描述了北欧气象观测资料的控制方法,并与任芝花和熊安元 (2007) 提出了我国地面观测资料的三级质量控制方案;尹嫦娥等 (2010) 引入误差稳定性理论来进行质量控制;王海军和刘莹 (2012) 参考邻站观测资料,提出了基于线性回归数据估计方法的质量控制法。然而,这些方法都严重依赖邻站的参考资料,对于我国西部一些地区,由于经济、环境等原因,观测网络密度较小,甚至存在一些没有邻站的观测站。在这些地区想获得高质量的气象观测资料,探索和研究有效的单站质量控制方法显得尤为重要。

单站质量控制法根据气象要素的物理限制、相互关系以及时间一致性对单站数据质量进行检查。Kubecka (2001) 在定义日最高气温后对日最高气温

进行了单站质量控制;Reek et al. (1992) 在给定的 8 个约束条件的基础上对降水和气温进行了内部一致性检验;Lanzante (1996) 用双权重平均法和双权重标准差法对观测资料进行时间一致性检验来减少奇异值的影响。我国地面观测资料三级质量控制业务系统中,单站级质量控制方法采用的是传统的单站质量控制法,即包括格式检查、范围检查、极值检查以及内部一致性检查相结合的综合法 (Meek and Hatfield, 1994),是被检站无有效邻站参考信息时资料有效性、精确性的基本保障。近年来单站质量控制方法有了新的发展,李永等 (2013) 试验了用 TPI 方法对连续气象观测资料进行单站实时控制,提高了实时质量控制业务应用的灵敏度;叶小岭等 (2014) 引入基因表达式编程,利用相对湿度与气温的耦合关系,实现了温度的单站级质量控制,拓宽了基于智能方程拟合的质量控制新思路;周青等 (2015) 基于自动气象站历史观测资料研究了特殊天气发生时气候要素的变化规律及一致性,提出了高湿事件和单站突变事件的判定条件并应用到质量控制算法中,降低了数据误判率。这些方法具有较传统方法更有效地标记常规台站分布的观测资料中的疑误数据的能力。为进一步满足单站、单要素质量控制的要求,本文从气温的混沌特性出发,根据气温在短期内的连续性和稳定性,提出一种基于粒子群 (Particle Swarm Optimization, PSO) 改进的相空间重构法 (Phase Space Reconstruction, PSR) 和极限学习机 (Extreme Learning Machine, ELM) 的集成学习算法并在验证其优越性后引入到质量控制中,形成一种检错率高、适应性强、适应于无法获得外部及内部参考资料的台站分布情况下的单站气温质量控制法——PSO-PSR-ELM 质量控制法。

2 资料选取

2.1 资料

本文选取江苏省南京 (NJ)、徐州 (XZ)、连

云港 (LYG)、淮安 (HA)、盐城 (YC)、扬州 (YZ)、南通 (NT) 和无锡 (WX) 八个城市观测站 2007~2009 年逐时气温资料作为基准资料, 这些数据源于中国气象局气象信息中心, 均经过了一般的质量控制流程, 已经剔除了明显的粗大误差, 可以正确反映气温的分布。

为验证本文所提方法的可操作性, 需人为地向被检站的基准资料中加入随机误差, 模拟被检站的观测值即模拟观测资料; 通过识别随机误差的能力, 即检错率来验证 PSO-PSR-ELM 质量控制方法的可行性。检错率定义为实际识别的误差个数占插入误差个数的百分比。随机添加人为误差的方法参照 Hubbard et al. (2007) 在多站质量控制中的人为误差插入法, 即在被检数据中随机插入占总数 3% 的错误值。插入随机值的大小 K_x 通过如下公式:

$$K_x = s_x p_x, \quad (1)$$

其中, p 为服从区间 $[-3.5, 3.5]$ 、均值为 0 的均匀分布的随机数, s 是目标站气温的标准差, x 为误差插入的位置。

为进一步评估本文方法的可靠性和优越性, 选用欧洲中期天气预报中心 (ECMWF) 的 ERA-Interim 再分析资料作为独立观测样本资料与 PSO-PSR-ELM 计算的预测资料进行对比 (赵天保和符淙斌, 2006)。ERA-Interim 资料选择同期逐日 6 小时 2 m 气温、分辨率为 $0.75^\circ \times 0.75^\circ$ 的网格资料, 即格点再分析资料, 并通过插值方法将网格资料插值到对应观测站, 即台站再分析资料 (赵天保等, 2004)。本文选择反距离加权算法将观测站点周围 4 个 ERA-Interim 格点资料插值到台站, 得到台站 ERA-Interim 再分析资料。台站再分析资料值 r_{tep} 按照如下公式计算:

$$r_{\text{tep}}(i, t) = \sum_{j=1}^4 \frac{1}{d(t, i, j)} r_{\text{tepg}}(t, i, j), \quad (2)$$

其中, $r_{\text{tep}}(i, t)$ 表示 t 时刻的台站 i 的再分析资料, $r_{\text{tepg}}(t, i, j)$ 、 $d(t, i, j)$ 分别为 t 时刻观测站 i 周围第 j 个格点再分析资料值和观测站与格点之间的距离。

2.2 混沌性识别

丁一汇 (2009)、段晚锁等 (2013) 指出, 大气系统本质上是一个混沌系统, 具有混沌性和短期可预测性。为验证气温时间序列的混沌特性, 本文通过时间序列的最大 Lyapunov 指数方法 (Rosenstein et al., 1993) 进行分析。Lyapunov 指数作为沿轨道长期运动的平均结果, 是系统整体特征

的表现, 总以实数的形式存在。当最大 Lyapunov 指数 $\lambda < 0$, 系统对初始条件的敏感程度较差, 系统在此方向上的运动相对稳定; 当 $\lambda > 0$ 时, 在此方向上的相连各点分离较快, 即轨道运动变得不稳定, 呈现出一种混沌特性; 当 $\lambda = 0$, 此时系统处于稳定与混沌的临界状态 (Rosenstein et al., 1993)。因此, 当一个系统的 $\lambda > 0$ 时, 即可推断出该系统一定是混沌系统。

本文最大 Lyapunov 指数采用小数据量法 (Rosenstein et al., 1993) 计算, 其中嵌入维数和时间延迟参数通过虚假近邻法 (Kennel and Abarbanel, 2002) 和互信息法 (肖方红等, 2005) 的相空间重构后获得。考虑本文混沌识别序列即气温观测资料的逐时性, 时间延迟以小时为单位。表 1 是八市观测站 2007~2009 年观测气温时间序列相应的嵌入维数、时间延迟和最大 Lyapunov 指数。从表 1 可以看出, 八市气温时间序列的最大 Lyapunov 指数均大于 0, 表明其均有明显的混沌性。

表 1 2007~2009 年八市观测站气温时间序列的最大 Lyapunov 指数

Table 1 The maximum Lyapunov exponents of the temperature time series collected from observation stations in eight cities from 2007 to 2009

城市	嵌入维数	延迟时间/h	最大 Lyapunov 指数
南京	15	2	0.0708
徐州	16	3	0.0191
连云港	25	2	0.0112
淮安	26	3	0.0376
盐城	30	4	0.0051
扬州	19	4	0.0076
南通	21	5	0.0043
无锡	24	3	0.0233

3 方法

基于上述分析, 气温时间序列呈现出较明显的混沌特性。本文从这点出发, 结合 PSR 和 ELM, 并从预测性能指标的角度智能选择模型参数, 提出一种 PSO 改进的 PSR 和 ELM 的集成学习算法。在验证其优越性后, 将其引入到质量控制领域, 形成 PSO-PSR-ELM 质量控制法。

3.1 PSR 和 ELM 组合算法及改进

3.1.1 PSR 和 ELM 算法

PSR 以 Takens 延迟嵌入定理 (Takens, 1981) 为基础, 该定理为一维混沌时间序列可以重构一个

与原动力系统在拓扑意义下等价的相空间提供了理论上的保障。G-P 算法(Grassberger and Procaccia, 1983)的提出实现了混沌时间序列的研究从已知混沌系统到任何一个未知混沌系统的跨越,为混沌时间序列研究方法的实际应用提供了现实基础。目前,基于 PSR 的混沌时间序列分析方法已在网络流量分析、保密通信、生物医学、大气科学和金融等领域得到了广泛的应用。

ELM 是 Huang et al. (2006) 依据摩尔—彭罗斯广义逆矩阵理论提出的一种机器学习算法。该算法与神经网络和支持向量机相比,由于随机初始化输入权值和隐含层偏置且仅需一步即可解析出学习网络的输出权值,所以具有更简单的网络结构、较强的泛化性能和更快的学习收敛速度,在非线性拟合预测领域得到了较好的应用效果。

基于 PSR 的混沌时间序列预测的重点是:如何根据 t 时刻的序列 $\{x(t)\}$ 得到下一时刻的 $x(t+1)=f(X(t))$, 即 $f(*)$ 的求取, 其中 $X(t)$ 为 $\{x(t)\}$ 的重构相空间。考虑 ELM 的非线性学习能力, 本文结合 PSR 和 ELM。首先通过 PSR 将混沌时间序列分解成恰当的能够反映原系统动力特性的高维矢量作为 ELM 学习的样本, 然后运用 ELM 的自适应和自学习特点, 建立时间序列的预测模型。

3.1.2 改进的 PSR 和 ELM 集成学习算法

PSR 是时间序列研究的关键, 即怎样重构一个恰当的高维相空间以还原出原时间序列的动力特性。Packard et al. (1980) 提出了两种有效的混沌时间序列重构相空间方法, 即导数重构法和坐标延迟重构法。在实际应用中, 由于大多不知混沌时间序列的先验信息, 且考虑数值微分对误差的敏感性, 所以本文选取坐标延迟相空间重构法。

基于坐标延迟法的 PSR 的核心是两个参数的确定, 即嵌入维数 m 和时间延迟 τ 。有关 m 和 τ 的确定方法已有很多, 大体可以分为两类, 即独立估计法和联合估计法(陈帝伊, 2012)。其中, 独立估计法包括针对 m 的虚假近邻法、Cao 法、G-P 关联维数法等; 针对 τ 的自相关法、复相关法和互信息法等。联合估计法有时间窗口法、C-C 法等, 且随着研究的深入和技术的突破, 在更多的实际应用中发现, 联合估计法能够更好地描述出系统的特性。但这些方法均未结合预测性能统筹考虑, 因此, 本文从预测性能指标的角度出发, 提出一种新的基于

最优预测性能的双参数联合智能估计法, 并与传统的基于虚假近邻法和互信息法的独立估计法进行对比, 即对改进前后的 PSR-ELM 预测性能进行评估。

考虑 PSO 算法(Eberhart and Kennedy, 1995)具有较强的全局收敛能力和鲁棒性, 本文通过 PSO 选取具有最佳预测性能的 m 和 τ 。PSO 算法假设在一个 D 维的目标搜索空间中, m 个粒子组成的种群 $\mathbf{x}=(x_1, x_2, \dots, x_m)^T$, 即所求问题的 m 个潜在解, 第 i 个粒子的信息可以表示为位置 $\mathbf{x}_i=(x_{i1}, x_{i2}, \dots, x_{iD})^T$ 和速度 $\mathbf{v}_i=(v_{i1}, v_{i2}, \dots, v_{iD})^T$, 粒子之间通过跟踪个体极值和群体极值来交换传递信息, 即第 i 个体极值 $\mathbf{p}_i=(p_{i1}, p_{i2}, \dots, p_{iD})$, 种群极值 $\mathbf{g}_i=(g_1, g_2, \dots, g_D)^T$, 每个粒子按照如下公式更新自身的位置和速度:

$$\mathbf{x}_{id}^{k+1} = \mathbf{x}_{id}^k + \mathbf{v}_{id}^{k+1}, \quad (3)$$

$$\mathbf{v}_{id}^{k+1} = w\mathbf{v}_{id}^k + c_1r_1(\mathbf{p}_{id}^k - \mathbf{x}_{id}^k) + c_2r_2(\mathbf{g}_d^k - \mathbf{x}_{id}^k), \quad (4)$$

式中, w 为惯性因子, 用于控制粒子的搜索能力; c_1 和 c_2 为学习因子, 用于控制粒子的学习认知能力和粒子间信息共享能力; r_1 和 r_2 是 $(0, 1)$ 区间随机数; $\mathbf{p}_{id}^k$ 为粒子 i 在第 k 次迭代过程中第 d 维的最优位置; $\mathbf{g}_d^k$ 为第 k 次迭代过程中种群第 d 维 $(1 \leq d \leq D)$ 的最优位置。同时为防止粒子的盲目搜索, 一般将粒子位置和速度控制在一定区间 $[-x_{\max}, x_{\max}]$, $[-v_{\max}, v_{\max}]$ 。

改进算法的主要思想是: 把 PSR 的两个参数 m 和 τ 看作 PSO 的两个粒子, 选择 ELM 测试均方根误差(Root Mean Square Error, RMSE)作为粒子的适应度值:

$$F_{\text{fitness}}(t) = \sqrt{\frac{\sum_{i=1}^m (X_{\text{test}} - \bar{X})^2}{n}}, \quad t=1, 2, \dots, n, \quad (5)$$

式中, $\bar{X}$ 为测试样本平均值, X_{test} 为根据 ELM 测试样本估计值, n 是时间序列的长度, m 为测试样本个数。算法选择全局最优值的过程就是获取最佳 m 和 τ 的过程。PSO-PSR-ELM 集成学习算法的流程如图 1 所示, 具体在 3.2 节 PSO-PSR-ELM 单站质量控制法的步骤中有详细说明。

为了说明改进后的算法具有更好地描述混沌时间序列动态特性的能力, 也即具有更好的预测性能, 对改进前后的 PSR-ELM 算法进行对比性实验。实验资料选取南京站 2009 年地面气温观测资料, 以月为单位运用两种方法进行单步预测。考虑到 PSO 算法的依概率收敛性, 采取多次实验取平均的

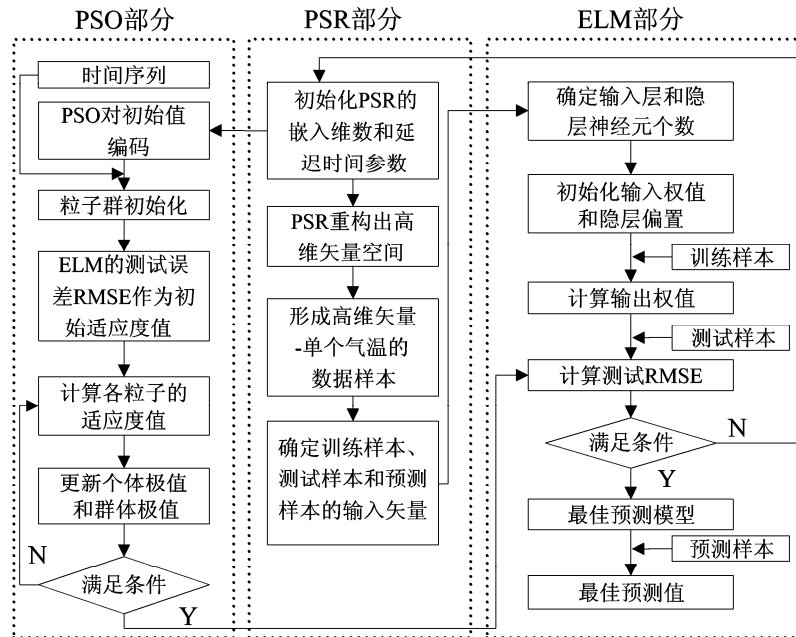


图1 PSO算法改进PSR和ELM组合算法流程图

Fig. 1 The flow chart of combination algorithm based on PSR (Phase Space Reconstruction) and ELM (Extreme Learning Machine) improved by PSO (Particle Swarm Optimization)

策略, 本实验设为 100 次。PSO-PSR-ELM 算法运行参数设置: 粒子位置的整数规划方法采用最终取整法(杨荣华和刘建华, 2011), 粒子位置参照表 1, 设置在区间[10, 30]和[2, 6], 即 m 和 τ 的范围; 惯性权值采用 Shi and Eberhart (1998) 提出的线性递减惯性权重, 初始和终止惯性权值分别设为 0.9 和 0.4, 粒子最大进化代数设为 40, 加速因子 c_1 、 c_2 均设为 1.49; 适应度值具体按照公式 (5) 计算; ELM 的输入层和隐含层神经元个数均取 PSR 的嵌入维数, 避免了人为选择的主观性和繁复性; 训练样本和测试样本按照 3:1 比例随机分配。

改进前、后 PSR-ELM 算法的预测指标可以用 RMSE, 平均绝对误差 (Mean Absolute Error, MAE) 和纳什效率系数 (Nash-Sutcliffe model efficiency Coefficient, NSC) 评估。其中, NSC 的值介于 0 到 1, 数值越接近 1, 模型的性能越好; RMSE 和 MAE 为大于 0 的数, 越接近 0, 模型的性能越好。图 2 是改进前、后 PSR-ELM 算法对南京站 2009 年气温时间序列三种预测性能的对比。从图 2 可以看出, PSO 优化后的 PSR-ELM 组合算法具有更好的预测效果。

3.2 PSO-PSR-ELM 质量控制法

上述讨论表明 PSO 改进的 PSR-ELM 算法对地面气温时间序列具有更好的预测性能。所以, 将其

引入到气温的质量控制领域, 形成一种 PSO-PSR-ELM 单站质量控制方法。

PSO-PSR-ELM 单站质量控制法的主要思想是: 首先通过 PSO 改进的 PSR 算法将历史气温时间序列重构为恰当的高维矢量空间, 即样本空间; 然后通过 ELM 对此样本空间进行学习; 最后对学习网络输入对应于下一时刻 (或多个时刻) 时间序列的高维矢量空间, 即可求得此时对应的气温的估计值。若被检站该时刻的观测值是由非气象因素造成的, 观测值与估计值之间的差值一定会高于平均水平, 即认为该时刻的观测值可疑, 需进行可疑性标记。PSO-PSR-ELM 质量控制法具体步骤如下, 其中算法运行参数按照 3.1.2 节讨论设置。

(1) 取被检站被检 t 时刻前 20 天逐时气温参考资料, 长为 480 小时的历史气温时间序列, 即 $\{tep(t-1), tep(t-2), \dots, tep[t-(n-1)]\}$, 其中 $n = 480$ 。通过 PSO 得到最佳 m 和 τ 进行 PSR, 获得 M 个 ($M = 480 - (m-1) \cdot \tau$) 维数为 m 的高维矢量空间 TEP, 并与对应的气温时间序列构成 ELM 的样本集, 即 (TEP, tep)。其中, 包括 ($M-l$) 个训练样本集和测试样本集, 即 (TEP _{l} (j), tep _{l} ($j+1$)), l 为可控步长, j 依次取 1 到 $M-l$ 之间的整数, 记作 $j \in \text{Abel}([1, M-l])$, 下同; l 个待估样本, 输

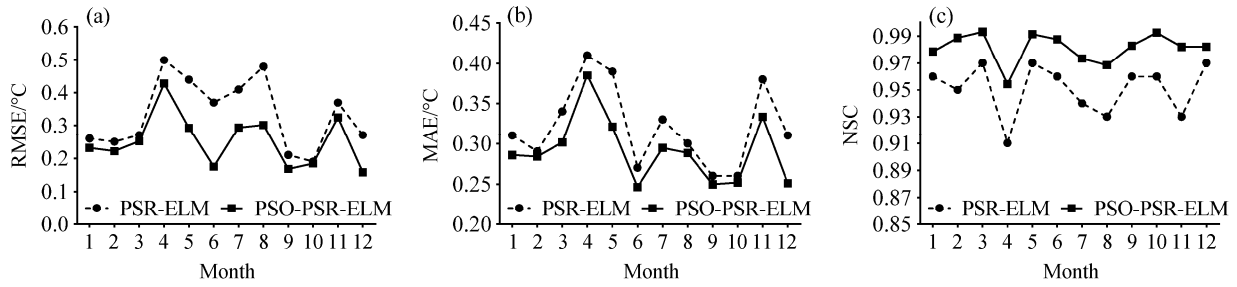


图2 PSO改进前、后PSR-ELM算法对2009年南京站气温序列三种预测指标 (a) RMSE、(b) MAE、(c) NSC的对比

Fig. 2 Comparisons of the three predictive indices (a) RMSE, (b) Mean Absolute Error (MAE), and (c) Nash-Sutcliffe model efficiency Coefficient (NSC) from PSR-ELM algorithm before and after improved by PSO for temperature at Nanjing Station in 2009

入为 $TEP_i (j \in \text{Abel}([M-l+1, M]))$ ，输出即为 $t \in \text{Abel}([t, t+l-1])$ 时刻待估气温值。

(2) 根据 $(M-l)$ 个训练、测试样本集，运用 ELM 算法进行训练学习，得到气温序列在高维空间的模型，即 $\text{tep}(t) = \text{elm}[TEP_i(j), \text{tep}_i(j+1)]$ ，其中 $j \in \text{Abel}([1, M-l])$ 。

(3) 取被检站被检 $t \in \text{Abel}([t, t+l-1])$ 时刻气温 $\text{tep}_{\text{obs}}(t)$ 的高维矢量 TEP_i ，并将其输入步骤(2)训练得到的 elm 模型中，可得到当前 $t \in \text{Abel}([t, t+l-1])$ 时刻气温的预测值，即 $\text{tep}_{\text{est}}(t) = \text{elm}(TEP_i(j \in \text{Abel}([M-l+1, M])))$ 。

(4) 将被检站当前 $t \in \text{Abel}([t, t+l-1])$ 时刻的气温观测值 $\text{tep}_{\text{obs}}(t)$ 与步骤(3)中的估计值进行对比：

$$|\text{tep}_{\text{obs}}(t) - \text{tep}_{\text{est}}(t)| \leq f \cdot S, \quad (6)$$

其中， S 为观测气温序列 $\text{tep}_{\text{obs}}(t)$ 的均方根误差， f 是可调参数，这里称为质量控制参数。若上式成立，则认为 $\text{tep}_{\text{obs}}(t)$ 正确；反之，则认为 $\text{tep}_{\text{obs}}(t)$ 疑误，需做标记。

(5) 重复步骤(1)~(4)，对被检站 $(t+l)$ 时刻的气温时间序列进行检验。

4 方法可靠性分析及不确定性评估

4.1 可靠性分析

从 3.2 节可知，本文质量控制法存在两部分，即被检时刻资料的估计及其与实际观测资料的对比。这两个部分相互关联，共同影响质量控制的效果，具体表现为：预测资料的估计是基于被检站的基准资料，根据一定的模型计算得到的，而本文认为基准资料可以代表气温的实际变化规律，所以预测资料 tep_{est} 在一定的误差范围内可以代表气温的实际变化情况，可以作为观测资料 tep_{obs} 即被检站

数据正确与否判断的一个参考依据；且模型的预测精度越高即预测资料 tep_{est} 越接近气温的实际变化规律，对观测值 tep_{obs} 的质量控制结果也就越具说服力和正确性（叶小岭等，2014）。所以，下文从模型预测结果的精度、预测资料的优越性及检错效果对 PSO-PSR-ELM 质控法的可靠性和优越性进行分析。

4.1.1 预测结果对比

本文将预测资料与基准资料进行对比，完成对预测模型的精度评估；同时，运用独立观测样本资料即台站 ERA-Interim 再分析资料与基准资料的对比结果与预测资料的对比结果进行对比分析，进行独立性验证和优越性分析。

选取南京等 8 市 2007~2009 年的逐日 6 小时基准资料、PSO-PSR-ELM 集成学习算法求的预测资料的逐日 6 小时资料以及台站 ERA-Interim 逐日 6 小时再分析资料，进行时间维度和空间维度的对比，比较结果运用偏差、均方根误差和相关系数三个指标进行评估，指标计算公式如下：

$$V = a_{\text{tep}} - b_{\text{tep}}, \quad (7)$$

$$R_{\text{RMSE}} = \sqrt{\frac{\sum_{i=1}^N (a_{\text{tep}} - b_{\text{tep}})^2}{N}}, \quad (8)$$

$$R = \frac{\sum_{i=1}^N [a_{\text{tep}} - M(a_{\text{tep}})][b_{\text{tep}} - M(b_{\text{tep}})]}{\sqrt{\sum_{i=1}^N [a_{\text{tep}} - M(a_{\text{tep}})]^2 \sum_{i=1}^N [b_{\text{tep}} - M(b_{\text{tep}})]^2}}, \quad (9)$$

其中， V 、 R_{RMSE} 、 R 分别代表偏差、均方根误差和相关系数， a_{tep} 代表分析资料，此处即预测资料和台站 ERA-Interim 再分析资料， b_{tep} 为基准资料， N 为样本的个数， $M()$ 为求平均函数。图 3 为南京站 2007~2009 年两种资料与基准资料的月均方根误差和相关系数曲线图。从图 3 可以看出，两种资料

与基准资料的均方根误差和相关系数月变化趋势整体一致; 预测资料较台站 ERA-Interim 再分析资料具有更小的均方根误差, 更高的相关系数。其他七个台站在时间 (月份) 上的变化趋势与南京基本一致。

图 4 是南京等八市 2007~2009 年两种资料的年均偏差分布情况。从图 4 可以看出, 预测资料具有较台站 ERA-Interim 再分析资料更小的偏差, 且基本呈现冷偏差, 冷偏差低于 0.2°C ; 而台站 ERA-Interim 再分析资料总体呈现暖偏差, 暖偏差低于 0.3°C , 这与高路和郝璐 (2014) 和高志刚等 (2015) 对 ERA-Interim 再分析资料的分析吻合。

通过两种资料与基准资料的对比, 可以得出 PSO-PSR-ELM 计算值即预测资料具有较小的误差和较高的相关性, 表现出更高的描述真实气温变化规律的能力; 且在不同台站总体呈现出的冷偏差也表明了本文方法较强的一致性和稳定性。

4.1.2 检错效果分析

为了验证 PSO-PSR-ELM 质量控制法的实际可行性, 运用该方法对南京等八市观测站 2007~2009

年的地面气温观测资料进行质量控制 (4.2 节有具体讨论), 表 2 是不同地区、不同年份的年平均检错率。从表 2 可以看出, 八市整体而言年平均检错率基本都达到 80% 以上, 说明该方法对误差的识别率较高, 地区和气候适应性较强; 同一地区、不同年份的检错率基本保持一致, 说明该方法对同一气候特征的地区具有一定的稳定性; 不同地区的检错率基本一致, 但仍存在一定的差异性, 微小的差异可以从地区地貌和气候特征进行分析和解释。南京地处中纬地区, 长江贯城而出, 属北亚热带季风气候区, 四季分明; 徐州以平原为主, 属暖温带湿润半湿润气候, 受东南季风影响较大; 连云港和南通受季风性气候较明显, 前者处鲁中南丘陵与淮北平原的结合地带为湿润型季风, 后者三面环水, 地处中纬度地带和海陆相过渡带, 为东亚冬季风; 淮安地势总体呈西南部略高、东北部略低状, 属于南温带和北亚热带的过渡地区, 兼具南北气候特征; 盐城为平原地区, 海拔差异较小, 东临黄海, 海洋调节作用较明显, 处北亚热带向暖温带过渡地带, 雨水充沛; 扬州地处淮河下游, 地势西高东低, 属

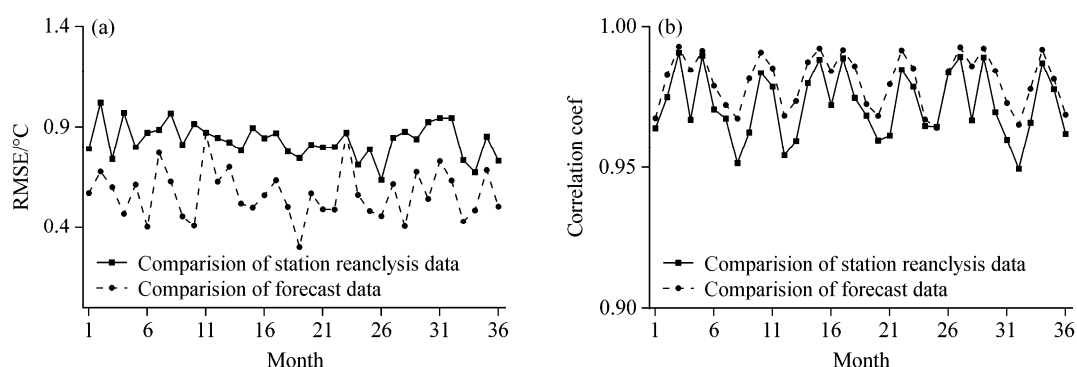


图 3 2007~2009 年南京站两种资料 (预测资料和台站 ERA-Interim 再分析资料) 与基准资料对比的 (a) 均方根误差和 (b) 相关系数曲线

Fig. 3 (a) RMSE and (b) correlation coefficients between forecast data, station ERA-Interim reanalysis data and standard data from 2007 to 2009 at Nanjing station

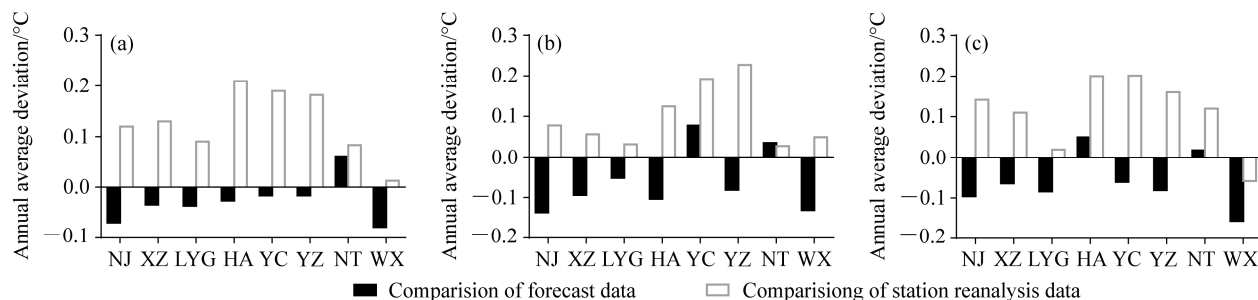


图 4 (a) 2007 年、(b) 2008 年、(c) 2009 年南京等八市台站两种资料的年均偏差分布

Fig. 4 Distributions of annual average deviations of forecast data and station ERA-Interim reanalysis data from (a) 2007, (b) 2008, and (c) 2009 at observation stations of Nanjing (NJ), Xuzhou (XZ), Lianyungang (LYG), Huai'an (HA), Yancheng (YC), Yangzhou (YZ), Nantong (NT), and Wuxi (WX)

于热带湿润气候, 季风显著, 冬夏温差较大; 无锡地势复杂, 西为宜南丘陵山区, 北依长江, 南濒太湖, 属北亚热带湿润区, 小气候特征多种多样。

表 2 八市 2007~2009 年 PSO-PSR-ELM 质量控制法的检错效果

Table 2 Testing results with the quality control method of PSO-PSR-ELM in eight cities from 2007 to 2009

城市	检错率		
	2007	2008	2009
南京	81.02%	83.12%	84.02%
徐州	82.17%	80.68%	82.45%
连云港	80.02%	79.30%	81.28%
淮安	83.15%	82.52%	85.19%
盐城	80.16%	83.13%	82.06%
扬州	84.53%	81.24%	85.19%
南通	81.09%	83.73%	84.02%
无锡	83.71%	82.16%	83.23%

为了说明 PSO-PSR-ELM 法较 TPI 方法及传统单站方法的优越性, 分别使用三种方法对南京站 2007~2009 年地面气温观测资料进行质量控制。本文采用的传统单站质量控制方法 (王海军等, 2007; 窦以文等, 2008) 包括格式检查、范围检查、极值

检查以及内部一致性检查相结合的综合法。其中, 范围检查与极值检查综合成为气候极限值范围检查, 要素置信区间参照窦以文等 (2008) 选取; 内部一致性检查参照王海军等 (2007) 设置进行。TPI 方法选择 4 阶切比雪夫多项式进行插值 (史丽萍等, 2003)。图 5 是南京站 2007~2009 年 TPI 方法及传统方法与 PSO-PSR-ELM 方法的检错效果对比。从图 5 可以看出, 三种方法在南京站不同年份的整体检错效果基本一致, 但 PSO-PSR-ELM 方法检错效果均好于 TPI 方法及传统单站质量控制方法。

为了进一步比较 PSO-PSR-ELM 方法与 TPI 方法及传统方法在不同地区的检错效果, 运用三种方法分别对南京等八市 2007~2009 年的气温资料进行检错。图 6 是三种方法在不同地区 2007~2009 年的检错效果。从图 6 可以明显看出, 不同地区 PSO-PSR-ELM 方法的检错效果整体上均明显优于 TPI 方法及传统单站质量控制法, 较 TPI 方法在检错能力上有一定的提升。

4.2 不确定性评估

4.2.1 质量控制参数 f

在概率统计学中存在两类错误, “第一类错误”

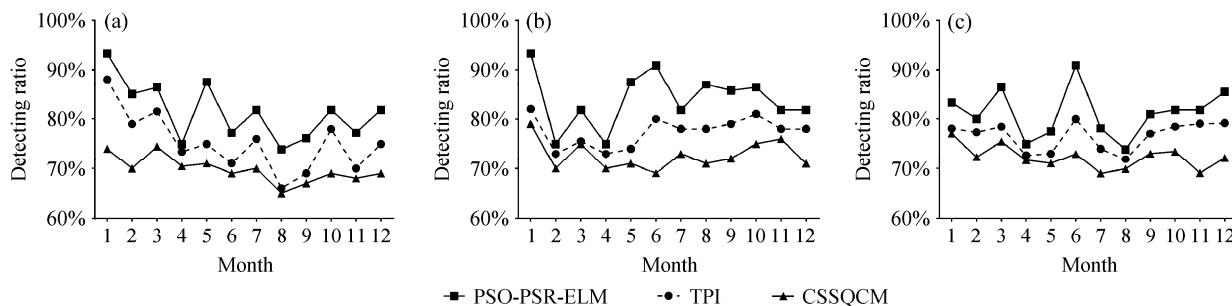


图 5 (a) 2007 年、(b) 2008 年、(c) 2009 年 PSO-PSR-ELM 方法、TPI 方法、传统方法在南京站检错效果对比

Fig. 5 Comparisons of detecting ratios of PSO-PSR-ELM, TPI (Tshebyshev Polynomial Interpolation), and conventional single-station quality control method (CSSQCM) for observations from (a) 2007, (b) 2008, and (c) 2009 at Nanjing station

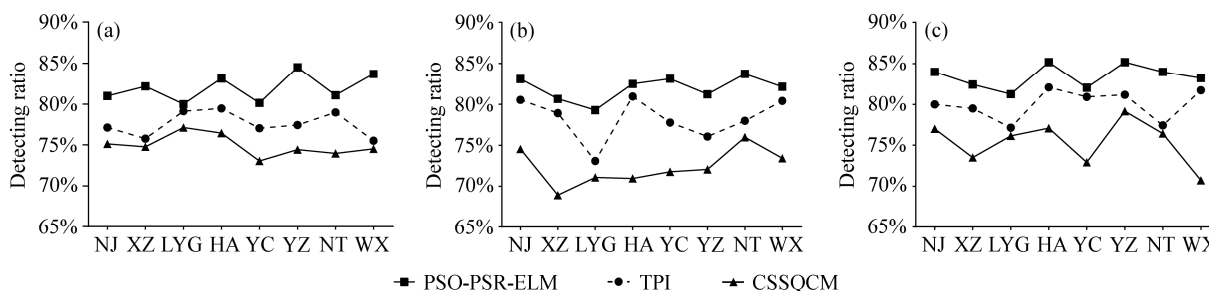


图 6 南京等八市 (a) 2007 年、(b) 2008 年、(c) 2009 年气温的 PSO-PSR-ELM 方法与 TPI 方法及传统方法检错效果对比

Fig. 6 Comparisons of detecting ratios of PSO-PSR-ELM, TPI, and conventional single-station quality control method for temperature from (a) 2007, (b) 2008, and (c) 2009 at observation stations of Nanjing (NJ), Xuzhou (XZ), Lianyungang (LYG), Huai'an (HA), Yancheng (YC), Yangzhou (YZ), Nantong (NT), and Wuxi (WX)

是对正确的假设予以拒绝,即“舍真”;“第二类错误”是对错误的假设进行接收,即“存伪”。屠其璞等(1984)指出,对质量控制而言,在样本容量一定的情况之下,需要在控制“第一类错误”的前提下,使“第二类错误”达到最小。据此,本文选择当两类错误差距最小,即满足“最小”原则时的 f 作为最佳的质控参数,此时对应的检错率即为最佳检错率。

为进一步研究 f 在不同地区、气候的取值差异,对南京、徐州、无锡和淮安四市 2007~2009 年气温观测资料以季度为单位进行质量控制。表 3 是不同站、不同年份、不同季节的最佳 f 值及对应的检错率。从表 3 可以看出,不同地区、不同季节的最佳 f 取值有所差异,但同一地区、同一季节的 f 取值基本一致。基于“最小”原则的 f 选择方法的引入,使 f 的取值在同地区、同气候下表现出一致性

表 3 2007~2009 年四市在 4 个季度下最佳 f 及季度平均检错率

Table 3 The optimum f (quality control parameter) values and corresponding detecting ratios in the four seasons from 2007 to 2009 at observation stations in four cities Nangjing, Xuzhou, Wuxi, and Huai'an

年份	季 度	南京		徐州		无锡		淮安	
		f	检错率	f	检错率	f	检错率	f	检错率
2007	1	1.5	80.41%	1.5	85.64%	1.5	84.79%	1.5	84.56%
	2	1.0	81.37%	1.5	80.18%	1.0	80.06%	1.5	83.64%
	3	1.5	78.93%	1.5	80.71%	2.0	78.98%	2.0	81.79%
	4	1.5	82.49%	1.5	81.36%	1.5	82.17%	1.5	80.93%
2008	1	2.2	81.45%	1.5	85.89%	1.5	82.86%	1.5	85.57%
	2	1.0	79.69%	1.5	78.35%	1.5	79.98%	1.5	81.85%
	3	2.0	80.02%	2.0	78.03%	2.0	78.79%	2.0	84.86%
	4	1.5	83.05%	1.5	83.15%	1.5	81.30%	1.5	80.05%
2009	1	1.5	79.90%	1.5	88.23%	1.5	85.15%	1.5	83.38%
	2	1.0	81.14%	1.5	79.92%	1.5	77.42%	1.5	84.47%
	3	2.0	75.01%	2.0	77.31%	2.0	80.41%	2.0	84.82%
	4	1.5	86.36%	1.5	80.30%	1.5	80.30%	1.5	83.33%

及稳定性,为最佳 f 的选取提供了实际可行性;在不同地区、不同气候下表现出的差异性增强了算法的实际可控性。

4.2.2 可控步长 l

为进一步提高该质控算法的效率以满足实际质控业务的需求,对算法的可控步长进行讨论,即 PSO-PSR-ELM 在 20 天历史气温序列参考下,在允许误差范围内可以达到的最大质量控制步长(李军和张友鹏, 2011)。考虑本文质量控制算法的策略,即根据历史资料通过 PSO-PSR-ELM 得到气温估计值,再比较实际观测值与估计值进行检错,其中估计值的精度直接关系到检错率的高低。下文即从预测误差、检错率及算法耗时三个方面对 PSO-PSR-ELM 质量控制法的可控步长进行评估。

选取南京、徐州、无锡和淮安四市 2007~2009 年气温观测资料进行 PSO-PSR-ELM 质量控制,其中可控步长 l 在多次试验后设置从 1 到 24。图 7 为 PSO-PSR-ELM 质量控制法针对南京站资料预测误差、检错率、100 个逐时气温的平均质控耗时与可控步长之间的关系。从图 7 可以看出,随着可控步长 l 的增大,预测误差在增大,检错效果在降低,检错效率在提高;具体而言,当 $l \in [1, 10]$ 时,预测误差和检错率变化较平缓,检错效率大幅提高; $l \geq 10$ 时 RMSE 和检错率均发生突变,预测误差增大和检错率减低幅度较大,检错效率提高较小。所以,就南京站 2007~2009 年气温资料,PSO-PSR-ELM 法的最佳可控步长为 1,最大可控步长为 10。徐州、无锡和淮安站预测误差、检错率、耗时与可控步长关系和南京站(图 7)走势基本一致,但最大可控步长有所差异,分别为 8、8 和 10,最佳可控步长不变仍为 1。可控步长的讨论,为 PSO-PSR-ELM 法的效率及检错效果提供了定量的分析,用户可根据实际要求,在允许检错误差范围

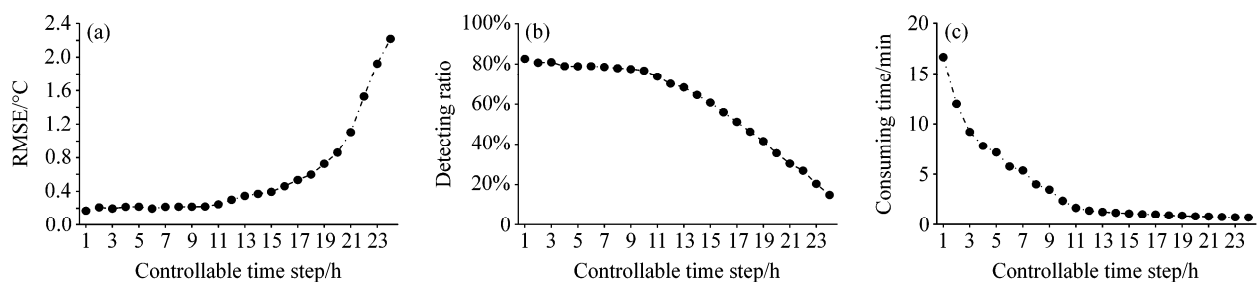


图 7 南京站 2007~2009 年 PSO-PSR-ELM 预测 (a) 均方根误差、(b) 检错率、(c) 质控耗时与可控步长的关系

Fig. 7 Relationships between (a) RMSE, (b) detecting ratio, (c) consuming time of the PSO-PSR-ELM method and controllable time step for observations from 2007 to 2009 at Nanjing station

内、最大可控步长下进行选择,增强了算法的高效性和灵活性。

5 结论

(1) 气温时间序列具有混沌特性。本文采用最大 Lyapunov 指数法分析了南京等八市 2007~2009 年地面气温观测资料,均表现出较明显的混沌特性。

(2) 本文从气温时间序列的混沌特性出发,结合相空间重构技术(PSR)与极限学习机(ELM),并针对 PSR 中嵌入维数和时间延迟的选取,提出一种基于预测性能的优化选择方法,即 PSO-PSR-ELM 集成学习算法,并运用该学习算法对南京站 2007~2009 年地面气温时间序列进行预测,实验结果表明 PSO-PSR-ELM 算法较 PSR-ELM 具有更好的预测性能。

(3) PSO-PSR-ELM 集成学习算法适用于单站质量控制,可以解决台站密度较低、新建台站以及特种单要素站等无法获取邻站和内部有效参考资料的台站的质量控制。PSO-PSR-ELM 质量控制法相比于 TPI 法及传统的单站质量控制法,对不同气候和地区的观测站的资料均具有更高的误差识别能力。

(4) PSO-PSR-ELM 质量控制法实质上是一种机器学习算法,决定了该方法具有检错率高、地区和气候适应性较强的特点,以及理论上的高可移植性。PSO-PSR-ELM 法在南京等八市不同地区、不同气候下均表现出一定的差异性,在同一地区和年份表现出一定的稳定性。

(5) PSO-SPR-ELM 质量控制法的可靠性分析表明:PSO-SPR-ELM 计算的预测资料较台站 ERA-Interim 再分析资料具有更好的描述气温实际变化规律的能力,不同站点总体表现出的冷偏差表明本文方法具有较强的稳定性。

(6) PSO-SPR-ELM 质量控制法的不确定性分析表明:参照“最小”原则的选取方法,质控参数对不同地区和气候具有一定的差异性,增强了方法的实际可行性和可控性;可控步长的定量分析增强了算法的实际应用效率及灵活性。

本文仅是 PSO-PSR-ELM 在气温质量控制上的简单应用,对本文方法的移植性,即其他要素如湿度等要素的适用性还未进行分析,也未考虑地势等因素对气温的影响分布,以及怎样进一步提高算法

的泛化性能。在接下来的研究中,可以从这些方面综合考虑,建立更符合气候要素分布的模型,获得更好的质量控制效果。

参考文献 (References)

- 陈帝伊,柳桦,马孝义. 2012. 基于径向基函数神经网络的混沌时间序列相空间重构双参数联合估计 [J]. 物理学报, 61 (10): 100501. Chen Diyi, Liu Ye, Ma Xiaoyi. 2012. Parameter joint estimation of phase space reconstruction in chaotic time series based on radial basis function neural networks [J]. Acta Physica Sinica (in Chinese), 61 (10): 100501.
- 丁一汇. 2009. 全球气候变化中的物理问题 [J]. 物理, 38 (2): 71-83. Ding Yihui. 2009. Physical problems in the global climate change [J]. Physics (in Chinese), 38 (2): 71-83, doi:10.3321/j.issn:0379-4148.2009.02.001.
- 段晚锁,丁瑞强,周非凡. 2013. 数值天气预报和气候预测可预报性研究的若干动力学方法 [J]. 气候与环境研究, 18 (4): 524-538. Duan Wansuo, Ding Ruiqiang, Zhou Feifan. 2013. Several dynamical methods used in predictability studies for numerical weather forecasts and climate prediction [J]. Climatic and Environmental Research (in Chinese), 18 (4): 524-538, doi:10.3878/j.issn.1006-9585.2012.12009.
- 窦以文,屈玉贵,陶士伟,等. 2008. 北京自动气象站实时数据质量控制应用 [J]. 气象, 34 (8): 77-81. Dou Yiwen, Qu Yugui, Tao Shiwei, et al. 2008. The application of quality control procedures for real-time data from automatic weather stations [J]. Meteorological Monthly (in Chinese), 34(8): 77-81.
- Eberhart R C, Kennedy J. 1995. A new optimizer using particle swarm theory [C]//Proceedings of the Sixth International Symposium on Micro Machine and Human Science. Nagoya, Japan: IEEE, 1: 39-43, doi:10.1109/MHS.1995.494215.
- Feng S, Hu Q, Qian W H. 2004. Quality control of daily meteorological data in China, 1951-2000: A new dataset [J]. International Journal of Climatology, 24 (7): 853-870, doi:10.1002/joc.1047.
- 高路,郝璐. 2014. ERA-Interim 气温数据在中国区域的适用性评估 [J]. 亚热带资源与环境学, 9 (2): 75-81. Gao Lu, Hao Lu. 2014. Verification of ERA-Interim reanalysis data over China [J]. Journal of Subtropical Resources and Environment (in Chinese), 9 (2): 75-81, doi:10.3969/j.issn.1673-7105.2014.02.012.
- 高志刚,骆敬新,刘克修,等. 2015. ERA-Interim 再分析数据在中国沿海的质量评估 [J]. 海洋科学, 39 (5): 92-105. Gao Zhigang, Luo Jingxin, Liu Kexiu, et al. 2015. Evaluation of ERA-Interim reanalysis data along coast of China [J]. Marine Sciences (in Chinese), 39 (5): 92-105, doi: 10.11759/hyxx20140522001.
- Grassberger P, Procaccia I. 1983. Measuring the strangeness of strange attractors [J]. Physica D, 9 (1-2): 189-208, doi:10.1016/0167-2789(83)90298-1.
- 韩海涛,李仲龙. 2012. 地面实时气象数据质量控制方法研究进展 [J]. 干旱气象, 30 (2): 261-265. Han Haitao, Li Zhonglong. 2012. Review of quality control methods of surface real time meteorological data [J]. Journal of Arid Meteorology (in Chinese), 30 (2): 261-265, doi:10.3969/j.issn.1006-7639.2012.02.018.
- Huang G B, Zhu Q Y, Siew C K. 2006. Extreme learning machine: Theory

- and applications [J]. *Neurocomputing*, 70 (1–3): 489–501, doi:10.1016/j.neucom.2005.12.126.
- Hubbard K G, You J S. 2005. Sensitivity analysis of quality assurance using the spatial regression approach—A case study of the maximum/minimum air temperature [J]. *J. Atmos. Oceanic Technol.*, 22 (10): 1520–1530, doi:10.1175/JTECH1790.1.
- Hubbard K G, Guttman N B, You J S, et al. 2007. An improved QC process for temperature in the daily cooperative weather observations [J]. *J. Atmos. Oceanic Technol.*, 24 (2): 206–213, doi:10.1175/JTECH1963.1.
- Ingleby N B, Lorenc A C. 1993. Bayesian quality control using multivariate normal distributions [J]. *Quart. J. Roy. Meteor. Soc.*, 119 (513): 1195–1225, doi:10.1002/qj.49711951316.
- Kennel M B, Abarbanel H D I. 2002. False neighbors and false strands: A reliable minimum embedding dimension algorithm [J]. *Physical Review E*, 66 (2): 026209, doi:10.1103/PhysRevE.66.026209.
- Kubecka P. 2001. A possible world record maximum natural ground surface temperature [J]. *Weather*, 56 (7): 218–221, doi:10.1002/j.1477-8696.2001.tb06577.x.
- Lanzante J R. 1996. Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data [J]. *International Journal of Climatology*, 16 (11): 1197–1226, doi:10.1002/(SICI)1097-0088(199611)16:11<1197:AID-JOC89>3.0.CO;2-L.
- 李军, 张友鹏. 2011. 基于高斯过程的混沌时间序列单步与多步预测 [J]. *物理学报*, 60 (7): 070513. Li Jun, Zhang Youpeng. 2011. Single-step and multiple-step prediction of chaotic time series using Gaussian process model [J]. *Acta Physica Sinica*, 60 (7): 070513.
- 李永, 张美玲, 张军, 等. 2013. 单站实时地面观测数据质量控制方法的试验 [J]. *北京农业*, 9: 146–147. Li Yong, Zhang Meiling, Zhang Jun, et al. 2013. Test of quality control method of single-station and real-time surface observations [J]. *Beijing Agriculture (in Chinese)*, 9: 146–147.
- Lorenc A C. 1981. A global three-dimensional multivariate statistical interpolation scheme [J]. *Mon. Wea. Rev.*, 109 (4): 701–721, doi:10.1175/1520-0493(1981)109<0701:AGTDSM>2.0.CO;2.
- Lorenc A C, Hammon O. 1988. Objective quality control of observations using Bayesian methods—Theory, and a practical implementation [J]. *Quart. J. Roy. Meteor. Soc.*, 114 (480): 515–543, doi:10.1002/qj.49711448012.
- Meek D W, Hatfield J L. 1994. Data quality checking for single station meteorological databases [J]. *Agricultural and Forest Meteorology*, 69 (1–2): 85–109, doi:10.1016/0168-1923(94)90083-3.
- Packard N H, Crutchfield J P, Farmer J D, et al. 1980. Geometry from a time series [J]. *Physical Review Letters*, 45 (9): 712–716, doi:10.1103/PhysRevLett.45.712.
- Reek T, Doty S R, Owen T W. 1992. A deterministic approach to the validation of historical daily temperature and precipitation data from the cooperative network [J]. *Bull. Amer. Meteor. Soc.*, 73 (6): 753–762, doi:10.1175/1520-0477(1992)073<0753:ADATTV>2.0.CO;2.
- 任芝花, 熊安元. 2007. 地面自动站观测资料三级质量控制业务系统的研制 [J]. *气象*, 33 (1): 19–24. Ren Zhihua, Xiong Anyuan. 2007. Operational system development on three-step quality control of observations from AWS [J]. *Meteorological Monthly (in Chinese)*, 33 (1): 19–24, doi:10.3969/j.issn.1000-0526.2007.01.003.
- Rosenstein M T, Collins J J, De Luca C J. 1993. A practical method for calculating largest Lyapunov exponents from small data sets [J]. *Physica D*, 65 (1–2): 117–134, doi:10.1016/0167-2789(93)90009-P.
- 史丽萍, 孙宝元, 于浩洋. 2003. 切比雪夫多项式回归分析方法在测量数据处理中的应用 [J]. *黑龙江大学自然科学学报*, 20 (2): 78–81. Shi Liping, Sun Baoyuan, Yu Haoyang. 2003. An application of Tchebyshev polynomials recursive analysis method on metrical data processing [J]. *Journal of Natural Science of Heilongjiang University (in Chinese)*, 20 (2): 78–81, doi:10.3969/j.issn.1001-7011.2003.02.021.
- Shi Y H, Eberhart R C. 1998. Parameter selection in particle swarm optimization [M]//Porto V W, Saravanan N, Waagen D, et al. *Evolutionary Programming VII*. Berlin Heidelberg: Springer, 591–600, doi:10.1007/BFb0040810.
- Takens F. 1981. Detecting strange attractors in turbulence [M]//Rand D, Young L S. *Dynamical Systems and Turbulence*, Warwick 1980: Lecture Notes in Mathematics. Berlin Heidelberg: Springer, 361–381, doi:10.1007/BFb0091924.
- 屠其璞, 王俊德, 丁裕国, 等. 1984. 气象应用概率统计学 [M]. 北京: 气象出版社, 183–186. Tu Qipu, Wang Junde, Ding Yuguo, et al. 1984. *Probability Statistics of Meteorology Application (in Chinese)* [M]. Beijing: China Meteorological Press, 183–186.
- Wade C G. 1987. A quality control program for surface mesometeorological data [J]. *J. Atmos. Oceanic Technol.*, 4 (3): 435–453, doi:10.1175/1520-0426(1987)004<0435:AQCPFS>2.0.CO;2.
- 王海军, 刘莹. 2012. 综合一致性质量控制方法及其在气温中的应用 [J]. *应用气象学报*, 23 (1): 69–76. Wang Haijun, Liu Ying. 2012. Comprehensive consistency method of data quality controlling with its application to daily temperature [J]. *Journal of Applied Meteorological Science (in Chinese)*, 23 (1): 69–76, doi:10.3969/j.issn.1001-7313.2012.01.008.
- 王海军, 杨志彪, 杨代才, 等. 2007. 自动气象站实时资料自动质量控制方法及其应用 [J]. *气象*, 33 (10): 102–109. Wang Haijun, Yang Zhibiao, Yang Daicai, et al. 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, doi:10.3969/j.issn.1000-0526.2007.10.015.
- 肖方红, 阎桂荣, 韩宇航. 2005. 混沌时序相空间重构参数确定的信息论方法 [J]. *物理学报*, 54 (2): 550–556. Xiao Fanghong, Yan Guirong, Han Yuhang. 2005. Information theory approach to determine embedding parameters for phase space reconstruction of chaotic time series [J]. *Acta Physica Sinica (in Chinese)*, 54 (2): 550–556, doi:10.3321/j.issn:1000-3290.2005.02.012.
- 熊安元. 2003. 北欧气象观测资料的质量控制 [J]. *气象科技*, 31 (5): 314–320. Xiong Anyuan. 2003. Quality control of meteorological observational data in Nordic countries [J]. *Meteorological Science and Technology (in Chinese)*, 31 (5): 314–320, doi:10.3969/j.issn.1671-6345.2003.05.013.
- 杨荣华, 刘建华. 2011. 量子粒子群算法求解整数规划的方法 [J]. *科学技术与工程*, 11 (33): 8195–8198. Yang Ronghua, Liu Jianhua. 2011. Solving integer programming based on quantum particle swarm optimization [J]. *Science Technology and Engineering (in Chinese)*, 11

- (33): 8195–8198, doi:10.3969/j.issn.1671-1815.2011.33.021.
- 叶小岭, 周建华, 熊雄. 2014. 一种基于 GEP 的地面气温观测资料的质量控制方法 [J]. 热带气象学报, 30 (6): 1196–1200. Ye Xiaoling, Zhou Jianhua, Xiong Xiong. 2014. A GEP-based method for quality control of surface temperature observations [J]. Journal of Tropical Meteorology (in Chinese), 30 (6): 1196–1200, doi:10.3969/j.issn.1004-4965.2014.06.021.
- 尹嫦姣, 江志红, 吴息, 等. 2010. 空间差值检验方法在地面气象资料质量控制中的应用 [J]. 气候与环境研究, 15 (3): 229–236. Yin Changjiao, Jiang Zhihong, Wu Xi, et al. 2010. A research on the application of spatial difference method in quality control of surface meteorological data [J]. Climatic and Environmental Research (in Chinese), 15 (3): 229–236, doi: 10.3878/j.issn.1006-9585.2010.03.02.
- 赵天保, 艾丽坤, 冯锦明. 2004. NCEP 再分析资料和中国站点观测资料的分析与比较 [J]. 气候与环境研究, 9 (2): 278–294. Zhao Tianbao, Ai Likun, Feng Jinming. 2004. An intercomparison between NCEP reanalysis and observed data over China [J]. Climatic and Environmental Research (in Chinese), 9 (2): 278–294, doi:10.3969/j.issn.1006-9585.2004.02.005.
- 赵天保, 符淙斌. 2006. 中国区域 ERA-40, NCEP-2 再分析资料与观测资料的初步比较与分析 [J]. 气候与环境研究, 11 (1): 14–32. Zhao Tianbao, Fu Congbin. 2006. Preliminary comparison and analysis between ERA-40, NCEP-2 reanalysis and observations over China [J]. Climatic and Environmental Research (in Chinese), 11 (1): 14–32, doi:10.3969/j.issn.1006-9585.2006.01.002.
- 周青, 张乐坚, 李峰, 等. 2015. 自动站实时数据质量分析及质控算法改进 [J]. 气象科技, 43 (5): 814–822. Zhou Qing, Zhang Lejian, Li Feng, et al. 2015. Quality analysis of real-time AWS data and algorithm improvement of quality control [J]. Meteorological Science and Technology (in Chinese), 43 (5): 814–822, doi:10.3969/j.issn.1671-6345.2015.05.007.