

1976-2018 年上海气候年代际变率的统计学分析

危国锐 (516021910080)

(上海交通大学电子信息与电气工程学院, 上海 200240)

摘要: 本文对 1976-2018 年上海气候数据作了统计学分析. 对一月和七月的逐月平均气温和降水量数据应用非参数检验、总体参数的区间估计和假设检验等方法, 发现近年来上海的一月和七月的逐月平均气温已显著升高, 而一月和七月的逐月降水量无明显变化.

关键词: 气温; 降水量; 气候; 统计学

Statistical Analysis of Shanghai's Climate Interdecadal Variability from 1976 to 2018

Wei Guo-ru

(School of Electronics, Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: This article makes a statistical analysis of Shanghai climate data from 1976 to 2018. The monthly average temperature and precipitation data of January and July were applied to non-parametric testing, interval estimation of overall parameters, and hypothesis testing. It was found that the monthly average temperature of January and July in Shanghai has increased significantly in recent years. , While the monthly precipitation in January and July did not change significantly.

Key words: temperature; precipitation; climate; statistics;

引言

近年来, 气候变化、全球变暖等概念已为人们所熟知. 在全球变暖背景下, 上海的逐月平均气温、逐月降水量等主要气候指标是否已发生显著变化? 本文将利用公开的数据集, 尝试从统计学的角度, 对上述问题作出初步分析.

1 数据集和数据处理方法

本研究主要使用逐月降水量数据集[1]的 1976-2017 年数据, 逐月平均气温数据集[2]的 1976-2017 年数据, 以及地面气象要素数据集[3]的 1979-2018 年数据.

1.1 中国 1km 分辨率逐月降水量数据集 (1901-2017)

该数据来源于“国家青藏高原科学数据中心”(http://data.tpdc.ac.cn), 为中国逐月降水量数据, 空间分辨率为 0.0083333° (约 1km), 时间为 1901.1-2017.12. 数据格式为 NetCDF, 即.nc 格式^[1].

该数据集是根据 CRU 发布的全球 0.5° 气候数据集以及 WorldClim 发布的全球高分辨率气候数据集, 通过 Delta 空间降尺度方案在中国地区降尺度生成的. 并且, 使用 496 个独立气象观测点数据进行验证, 验证结果可信. 本数据集包含的地理

空间范围是全国主要陆地 (包含港澳台地区), 不含南海岛礁等区域. 为了便于存储, 数据均为 int16 型存于 nc 文件中, 降水单位分别为 0.1mm ^{[4][5][6][7]}.

1.2 中国 1km 分辨率逐月平均气温数据集 (1901-2017)

该数据来源于“国家青藏高原科学数据中心”(http://data.tpdc.ac.cn), 为中国逐月平均温度数据, 空间分辨率为 0.0083333° (约 1km), 时间为 1901.1-2017.12. 数据格式为 NetCDF, 即.nc 格式. 数据单位为 0.1°C ^[2].

该数据集是根据 CRU 发布的全球 0.5° 气候数据集以及 WorldClim 发布的全球高分辨率气候数据集, 通过 Delta 空间降尺度方案在中国地区降尺度生成的. 并且, 使用 496 个独立气象观测点数据进行验证, 验证结果可信. 本数据集包含的地理空间范围是全国主要陆地 (包含港澳台地区), 不含南海岛礁等区域^{[4][5][6][7]}.

1.3 中国区域地面气象要素驱动数据集 (1979-2018)

该数据来源于“国家青藏高原科学数据中心”(http://data.tpdc.ac.cn), 包括近地面气温、近地面气压、近地面空气比湿、近地面全风速、地面向下短

完成日期: 2021-06-02

课程名称: 本-(2020-2021-2)-MA1207-2-概率统计

波辐射、地面向下长波辐射、地面降水率共 7 个要素。数据为 NetCDF 格式，时间分辨率为 3 小时，水平空间分辨率为 0.1° 。可为中国区陆面过程模拟提供驱动数据^[3]。

该数据集是以国际上现有的 Princeton 再分析资料、GLDAS 资料、GEWEX-SRB 辐射资料，以及 TRMM 降水资料为背景场，融合了中国气象局常规气象观测数据制作而成。详细过程请参阅参考文献^{[8][9]}。原始资料来自于气象局观测数据、再分析资料和卫星遥感数据。已去除非物理范围的值，采用 ANU-Spline 统计插值。精度介于气象局观测数据和卫星遥感数据之间，好于国际上已有再分析数据的精度^{[8][9]}。

1.4 数据处理方法

从“国家青藏高原科学数据中心”(<http://data.tpdc.ac.cn>)网站申请数据，然后将数据集下载到本地，解压出 NetCDF (*.nc) 数据文件。使用 Panoply 软件^[11]预览 NetCDF 数据。使用 MATLAB 软件^[12]提供的 ncread 等函数，从*.nc 数据文件中提取最接近本地坐标（北纬 31.0250° ，东经 121.4328° ）的位置上的气象要素时间序列。借助 MATLAB 提供的 Statistics and Machine Learning Toolbox，对提取出来的本地气候数据做统计分析，并绘制相关图表。

本研究使用的本地气候数据示于图 1，可见不同数据集的数据存在差异，这是由于这些数据集并非直接取自观测数据。另附 1981-2010 年本地部分气候观测数据于附录。

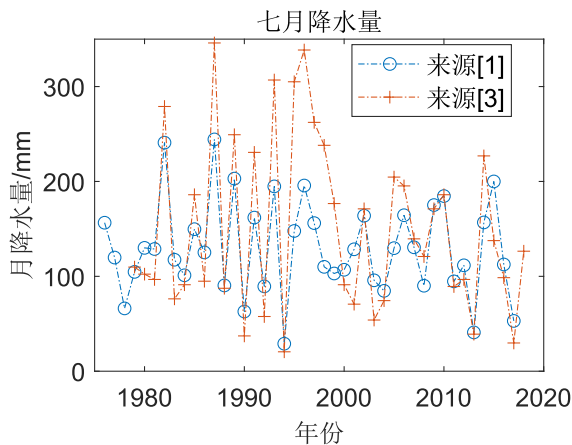


图 1 (a) 七月降水量

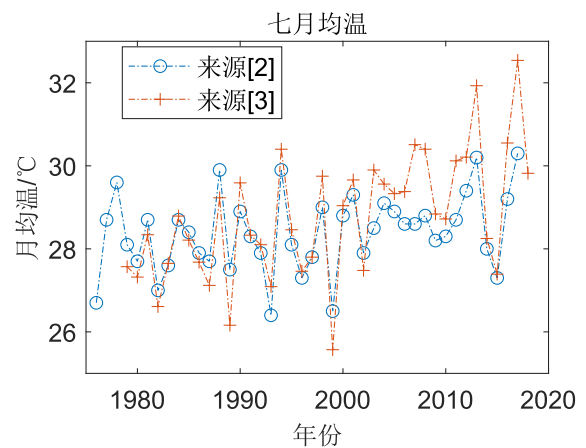


图 1 (b) 七月近地面平均气温

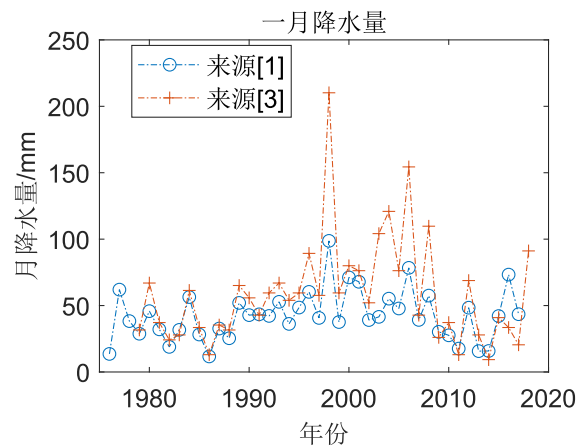


图 1 (c) 一月降水量

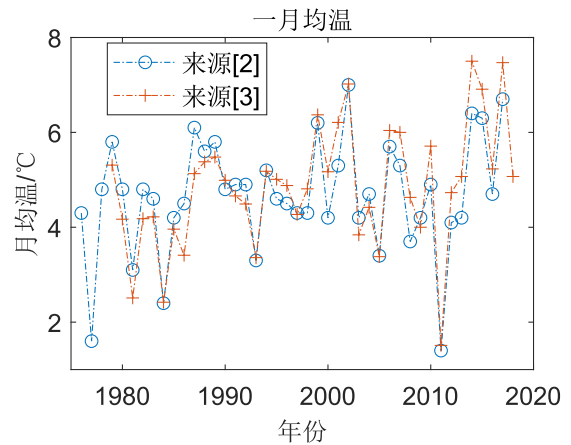


图 1 (d) 一月近地面平均气温

图 1 本研究使用的本地气候数据图示

2 绘制频率直方图并作分布拟合——初步观察本地气候要素的分布特征

提取数据集[1][2]的 1976-2017 年本地一月和七月的月降水量和月均温数据，利用 MATLAB 提

提供的 histfit 等函数, 作出频率分布直方图并作正态拟合, 结果示于图 2. 观察图 2, 可见一月和七月的月降水量和月均温的分布形似正态曲线.

为更直观地考察上述分布和正态分布的接近程度, 可借助 MATLAB 提供的 normplot 等函数, 绘出上述数据的正态概率图, 示于图 3. 如果样本数据具有正态分布, 则数据点 (经验分布) 沿参考线出现. 非正态分布会在正态概率图中引入曲率. 观察图 3, 可见一月降水量、七月降水量、七月均温三个样本比较接近正态分布, 而一月均温似乎偏离正态分布.

从频数分布直方图 2 和正态概率图 3, 已定性地知道月均温和降水量的分布形似正态分布. 下面, 将使用定量方法, 检验这些样本是否来自正态总体.

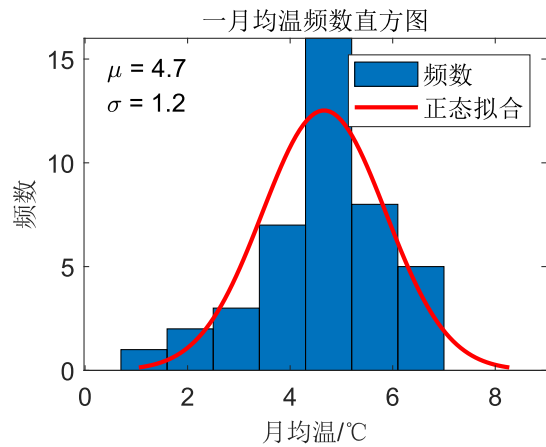


图 2 (c) 一月均温

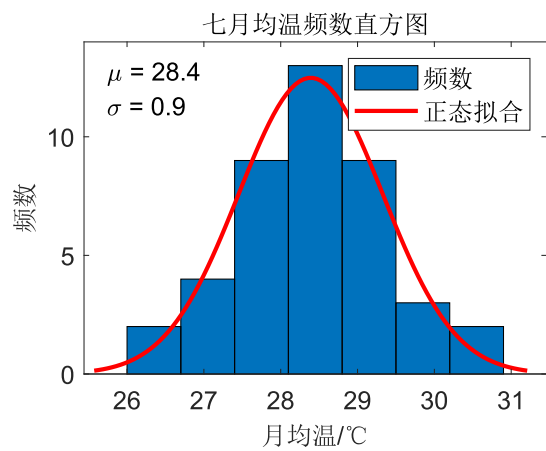


图 2 (d) 七月均温

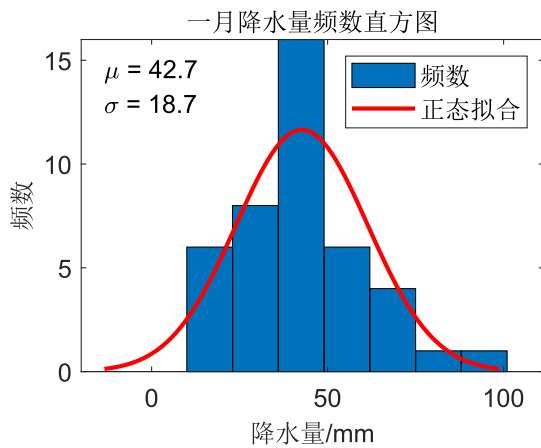


图 2 (a) 一月降水量

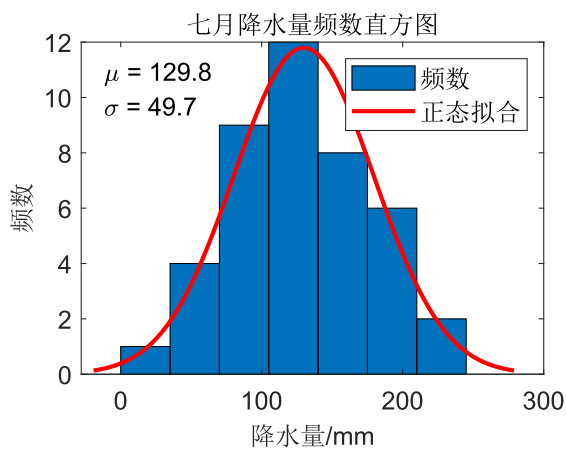


图 2 (b) 七月降水量

图 2 1976-2017 年一月和七月的降水量和均温的频数分布直方图及正态拟合曲线. 数据来源: 数据集[1][2]中的东经 121.4335°、北纬 31.0213° 位置.

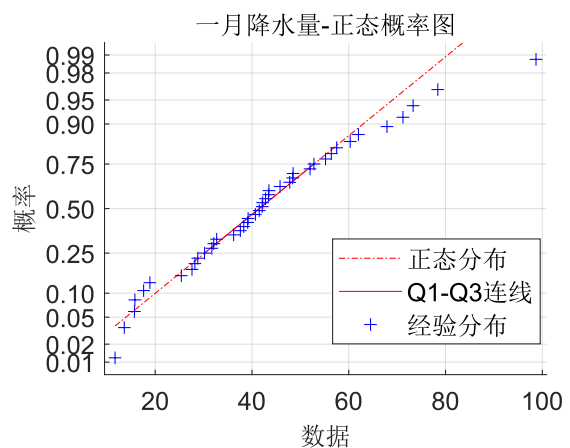


图 3 (a) 一月降水量

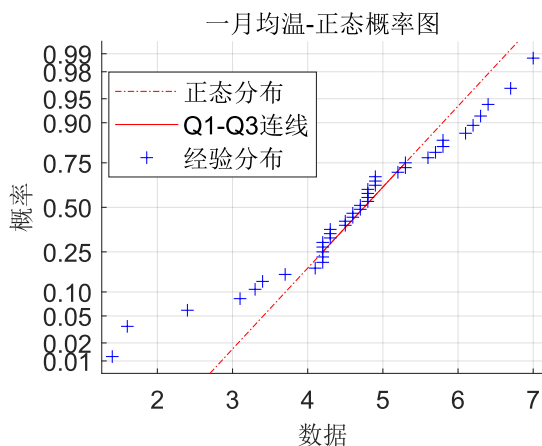


图 3 (b) 一月均温

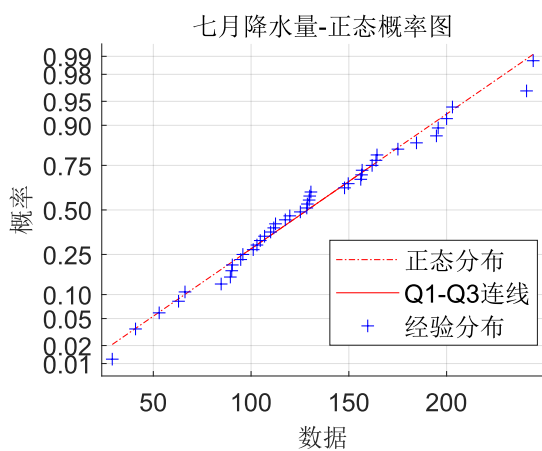


图 3 (c) 七月降水量

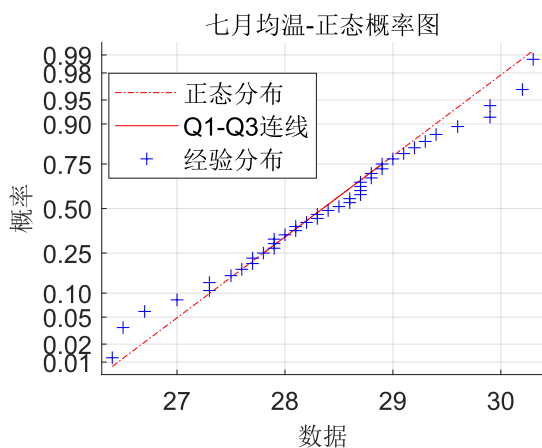


图 3 (d) 七月均温

图 3 1976-2017 年一月和七月的降水量和均温的正态概率图。数据来源数据集[1][2]。如果样本数据具有正态分布，则数据点（经验分布函数）沿参考线出现。非正态分布会在正态概率图中引入曲率。数据来源：数据集[1][2]中的东经 121.4335°、北纬 31.0213° 位置。

3 非参数检验——定量推断本地气候要素

分布类型

非参数检验，又称为分布拟合检验，是根据样本对总体的分布或分布类型提出假设并进行检验的方法^[10]。文献[10]给出了利用 Pearson 定理进行非参数检验的例子。

本研究利用 MATLAB 的 Statistics and Machine Learning Toolbox 中提供的函数 `lillietest`，对本地一月和七月的月降水量和月均温样本数据作 Lilliefors 检验。这种检验方法是以“样本来自正态类型的总体”为原假设 H_0 ，“样本来自非正态类型的总体”为备择假设 H_1 。对上述数据作显著性水平为 5% 的 Lilliefors 检验的结果示于表 1。从表 1 可见，采取不同的样本数据来源，可能得到不同的检验结果。本研究认为，可能的原因包括：数据集质量有差异；样本容量较小；总体确实偏离正态分布（例如气候发生变化）等。

从前面的研究可以看出，用正态分布近似月降水量和月均温的分布是存在偏差的。但为了研究方便，本文仍然在正态分布假设下对月降水量和月平均气温的样本作统计学分析，而且作者认为这种假设不会影响本研究的主要结论。

表 1 Lilliefors 检验结果

样本内容	数据源 ^a	检验结果 ^b
一月降水量	[1]	接受 H_0
	[3]	拒绝 H_0
七月降水量	[1]	接受 H_0
	[3]	拒绝 H_0
一月均温	[2]	拒绝 H_0
	[3]	接受 H_0
七月均温	[2]	接受 H_0
	[3]	接受 H_0

^a 选取了数据集[1][2]中东经 121.4335°，北纬 31.0213° 上 1976-2017 年的数据；数据来源：数据集[3]中东经 121.45°、北纬 31.05° 上 1979-2018 年的数据。

^b 原假设 H_0 ：样本来自正态类型的总体；备择假设 H_1 ：样本来自非正态类型的总体。

4 两个正态总体参数的区间估计和假设

检验——推断两个二十年间的气候变率

为了定量地推断在 1979-1998、1999-2018 这两个二十年间本地气候是否已发生变化，本研究假定两个二十年间的气候状况分别服从正态分布，

利用 MATLAB 的 Statistics and Machine Learning Toolbox 中提供的 ttest, vartest2 等函数, 对样本数据做正态总体的参数估计和假设检验.

首先分别对两个时期的气候样本做非参数检验和单个正态总体参数的区间估计, 结果示于表 2. 从表 2 可见, 两个时期的一月和七月降水量和月平均气温似乎存在差异, 然而这种差异可能是由样本的随机性造成的, 也可能是由于气候变化造成的.

为了从统计学角度定量地推断两个二十年间的的气候是否确实存在差异, 本研究再对两个时期的气候样本数据做两个正态总体参数的区间估计和显著性水平 $\alpha = 0.05$ 的假设检验, 结果示于表 3. 检验结果认为, 相比于 1979-1998 年, 1999-2018

年间的一月、七月的平均气温显著升高, 而一月、七月的降水量无明显变化.

上述区间估计和假设检验, 已定量地给出了对两个总体的均值差的估计. 利用缺口箱线图, 还可以直观地对两个总体的中位数做假设检验. 本研究利用 MATLAB 的 Statistics and Machine Learning Toolbox 中提供的 boxplot 等函数, 绘出了两个时期的一月和七月的月平均气温和月降水量样本的箱线图, 示于图 4. 通过考察箱线图的缺口部分是否有重合部分, 可以直观地得出: 与 1979-1998 年相比, 有 95% 的置信度认为 1999-2018 年的七月平均气温的中位数显著升高, 而其它指标无明显差异.

表2 分别对两个时期的气候样本做非参数检验和单个正态总体参数的区间估计的结果

样本内容	年代	Lilliefors 检验结果	总体均值 μ *	总体标准差 σ †
一月降水量/mm	1979-1998	拒绝 H_0	56.1728 [37.0976, 75.248]	40.7577 [30.9958, 59.5295]
	1999-2018	接受 H_0	62.2178 [43.8111, 80.6245]	39.3293 [29.9096, 57.4433]
七月降水量/mm	1979-1998	拒绝 H_0	175.771 [124.888, 226.654]	108.721 [82.6816, 158.796]
	1999-2018	接受 H_0	124.993 [97.9067, 152.079]	57.8747 [44.0131, 84.5301]
一月均温/°C	1979-1998	接受 H_0	4.39099 [3.97256, 4.80942]	0.894056 [0.679921, 1.30583]
	1999-2018	接受 H_0	5.31399 [4.61963, 6.00835]	1.48363 [1.12829, 2.16695]
七月均温/°C	1979-1998	接受 H_0	28.083 [27.5798, 28.5862]	1.07517 [0.81766, 1.57037]
	1999-2018	接受 H_0	29.46 [28.7313, 30.1887]	1.55709 [1.18415, 2.27424]

*方括号内是总体均值的一个置信度为 95% 的置信区间

†方括号内是总体标准差的一个置信度为 95% 的置信区间

表3 对两个时期的气候样本做两个正态总体参数的区间估计和显著性水平 $\alpha = 0.05$ 的假设检验的结果

样本内容	原假设 H_0 *	总体均值差 $\mu_1 - \mu_2$ †	总体方差比 σ_1^2 / σ_2^2 §	检验结果 ¶
一月降水量/mm	$\mu_1 = \mu_2$ ($\sigma_1^2 = \sigma_2^2$)	[-33.4667, 21.3767]	[0.4251, 2.7133]	接受 H_0 (接受 H_0)
七月降水量/mm		[-11.5092, 113.0652]	[1.3968, 8.9159]	接受 H_0 (拒绝 H_0)
一月均温/°C		[-1.7163, -0.1297]	[0.1437, 0.9175]	拒绝 H_0 (拒绝 H_0)
七月均温/°C		[-2.2817, -0.4723]	[0.1887, 1.2046]	拒绝 H_0 (接受 H_0)

*括号内是对总体方差的原假设

†方括号内是总体均值差的一个置信度为 95% 的置信区间, 采取配对抽样^[10]方法得到

§方括号内是总体方差比的一个置信度为 95% 的置信区间

¶括号内是对总体方差的检验结果

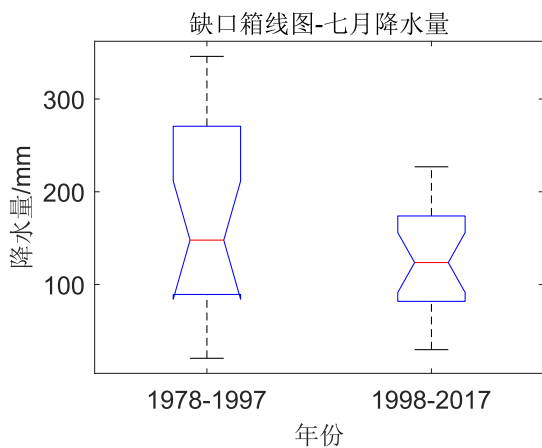


图 4 (a) 七月降水量

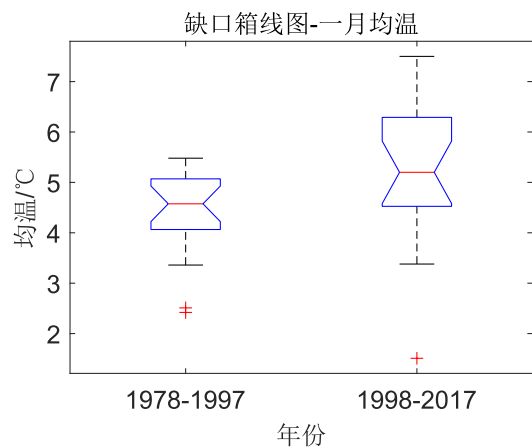


图 4 (d) 一月均温

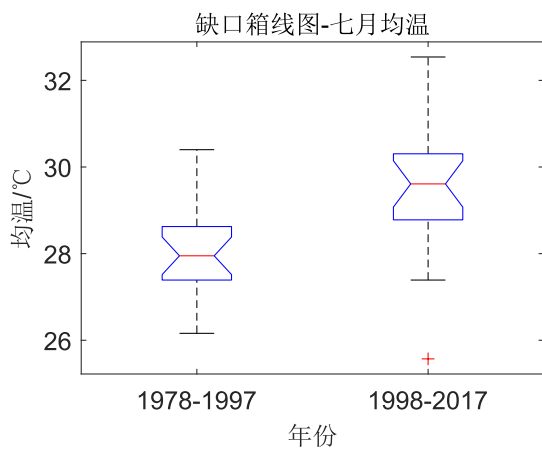


图 4 (b) 七月均温

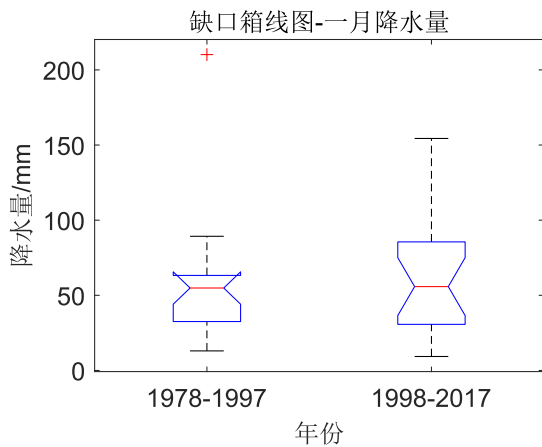


图 4 (c) 一月降水量

图 4 1978-1997 年与 1998-2017 年一月和七月降水量和平均近地面气温缺口箱线图. 若两个样本的箱线图上的缺口不重叠, 则可以有 95% 的置信度认为相应的两个总体的中位数不同. 数据来源: 数据集[3]中东经 121.45°、北纬 31.05° 上 1979-2018 年的数据.

5 结论与讨论

本文利用频率直方图、区间估计、假设检验等统计学方法, 对本地的一月和七月平均气温和逐月降水量数据做了初步分析. 结果表明, 近二十年来, 一月和七月的月平均气温已显著升高, 而一月和七月的逐月降水量无明显变化.

以上结论的准确性可能受到以下限制: 样本数据与真实情况存在差异, 因为它们不是直接取自观测数据; 对样本作出了正态假设, 而非参数检验结果提示该假设值得怀疑; 支撑结论的统计学证据还不够充分, 因为本研究采取的统计学分析方法有限, 且样本比较单一 (仅一月和七月的平均气温和逐月降水量)、样本容量偏小 (小于 50).

致 谢

本研究利用的气候数据[1][2][3]来源于“国家青藏高原科学数据中心” (<http://data.tpdc.ac.cn>).

本研究利用了 UCAR/Unidata 开发的 netCDF 软件 (<http://doi.org/10.5065/D6H70CW6>).

本研究得到了上海交通大学数学科学学院王桂兰老师的指导.

附 录

表 4 的数据来自国家气象信息中心-中国气象数据网 (<http://data.cma.cn/>) 提供的“中国地面累年值月值数据集 (1981-2010 年)”^[13]. 该数据集由各省上报的全国地面月报信息化文件, 基于《气候资料统计整编方法 (1981-2010) (发布版)》, 进行

整编统计而得. 数据集为中国基本、基准和一般地面气象观测站 1981-2010 年, 数据集包括气压、气温、降水、风要素的日气候标准值数据. 地面基础资料专项工作对历史信息文件重新进行了质量检测, 并对存在问题和分歧的站点文件进行了修订. 该数据集没有对资料进行均一性订正, 只是对

明显存在非均一性的台站进行了分段处理, 处理方法如下: 对于 1981-2010 年间位置发生变化的台站, 当迁站前后海拔高度相差超过 100 米或水平距离超过 50 公里, 记录不能合并统计, 进行分段统计处理.

表4 [58361] 闵行台站部分观测数据. 来自“中国地面累年值月值数据集 (1981-2010年)”^[13].

区站号(字符)	月序	累年月平均气温 (摄氏度)	累年 20-20 时平均月降水量 (毫米)	累年 08-08 时平均月降水量 (毫米)
58361	1	4.3	60.2	59.3
58361	2	5.9	62.1	61.5
58361	3	9.5	102.4	103.8
58361	4	14.9	84.7	83.9
58361	5	20.2	103.5	103.2
58361	6	24.1	174.2	171.6
58361	7	28.2	154.2	152.9
58361	8	27.9	165.3	164.7
58361	9	24	122.3	127.1
58361	10	18.8	57.2	57.5
58361	11	12.9	55.5	55.6
58361	12	6.7	39.3	39.6

参 考 文 献

- [1] 彭守璋. 中国 1km 分辨率逐月降水量数据集 (1901-2017). 国家青藏高原科学数据中心, DOI:10.5281/zenodo.3185722, CSTR:, 2020. [PENG Shouzhang. 1-km monthly precipitation dataset for China (1901-2017). National Tibetan Plateau Data Center, DOI:10.5281/zenodo.3185722, CSTR:, 2020]
- [2] 彭守璋. 中国 1km 分辨率逐月平均气温数据集 (1901-2017). 国家青藏高原科学数据中心, DOI:10.11888/Meteoro.tpdc.270961, CSTR:18406.11.Meteoro.tpdc.270961, 2019. [PENG Shouzhang. 1-km monthly mean temperature dataset for china (1901-2017). National Tibetan Plateau Data Center, DOI:10.11888/Meteoro.tpdc.270961, CSTR:18406.11.Meteoro.tpdc.270961, 2019]
- [3] 阳坤, 何杰. 中国区域地面气象要素驱动数据集 (1979-2018). 国家青藏高原科学数据中心, DOI:10.11888/AtmosphericPhysics.tpe.249369.file, CSTR:18406.11.AtmosphericPhysics.tpe.249369.file, 2019. [YANG Kun, HE Jie. China meteorological forcing dataset (1979-2018). National Tibetan Plateau Data Center, DOI:10.11888/AtmosphericPhysics.tpe.249369.file, CSTR:18406.11.AtmosphericPhysics.tpe.249369.file, 2019]
- [4] Peng, S.Z., Ding, Y.X., Liu, W.Z., & Li, Z. (2019). 1 km monthly temperature and precipitation dataset for China from 1901 to 2017. Earth System Science Data, 11, 1931-1946. <https://doi.org/10.5194/essd-11-1931-2019>
- [5] Peng, S.Z., Ding, Y.X., Wen, Z.M., Chen, Y.M., Cao, Y., & Ren, J.Y. (2017). Spatiotemporal change and trend analysis of potential evapotranspiration over the Loess Plateau of China during 2011-2100. Agricultural and Forest Meteorology, 233, 183-194.
- [6] Peng, S., Gang, C., Cao, Y., & Chen, Y. (2017). Assessment of climate change trends over the loess plateau in china from 1901 to 2100. International Journal of Climatology.
- [7] Ding, Y.X., & Peng, S.Z. (2020). Spatiotemporal trends and attribution of drought across China from 1901-2100. Sustainability, 12(2), 477.
- [8] Yang, K., He, J., Tang, W.J., Qin, J., Cheng, C.C.K. (2010). On downward shortwave and long wave radiations over high altitude regions: 0

- bservation and modeling in the Tibetan Plateau. Agricultural and Forest Meteorology, 150 (1), 38-46.
- [9] He, J., Yang, K., Tang, W. Lu, H., Qin, J., Chen, Y.Y., Li, X. (2020). The first high-resolution meteorological forcing dataset for land process studies over China. Scientific Data, 7, 25, <https://doi.org/10.1038/s41597-020-0369-y>.
- [10] 卫淑之, 熊德文, 皮玲. 大学数学概率论与数理统计: 基于案例分析[M]. 北京: 高等教育出版社, 2020.
- [11] NASA GISS: Panoply 4 netCDF, HDF and GRIB Data Viewer[EB/OL]. <https://www.giss.nasa.gov/tools/panoply/>
- [12] MATLAB - 技术计算语言 产品信息 - MATLAB & Simulink (mathworks.cn)[EB/OL]. <https://ww2.mathworks.cn/products/matlab.html>
- [13] 国家气象信息中心-中国气象数据网 (cma.cn). 中国地面累年值月值数据集 (1981-2010年). <http://data.cma.cn/data/detail/dataCode/A.0029.0004.html>

危国锐 男, 1998年生. 学科兴趣: 大气科学, 计算机科学, 电子工程.

E-mail: 313017602@qq.com