1976-2018年上海气候年代际变率的统计学分析

危国锐 120034910021

(上海交通大学海洋学院，上海 200030)

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. [[1]](#footnote-1)

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-rui

(*School of Oceanography*, *Shanghai Jiao Tong University*, *Shanghai* 200030, *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.

# 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.

## 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].

## 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℃[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].

## 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].

## 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.

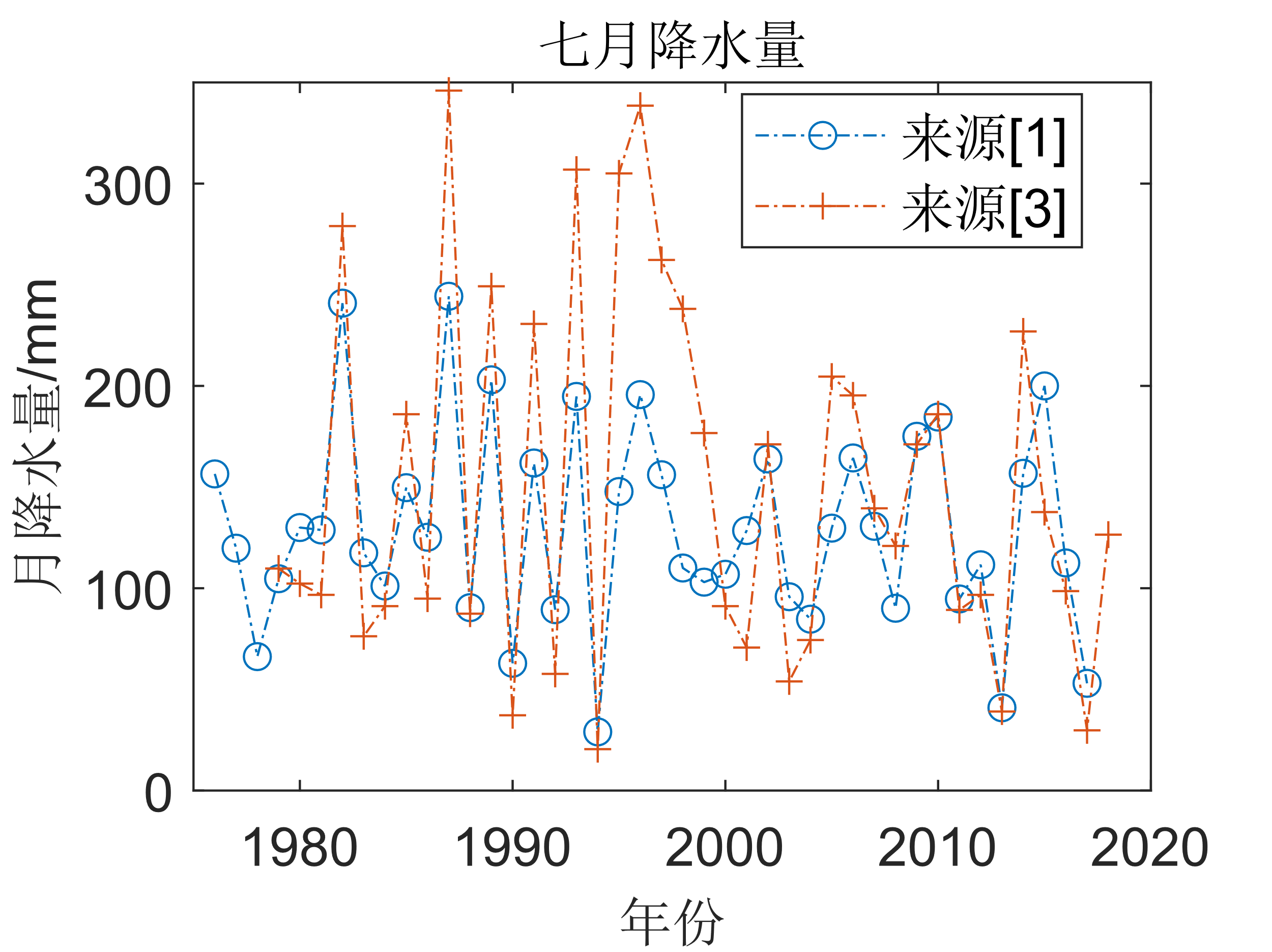


图1（a） 七月降水量

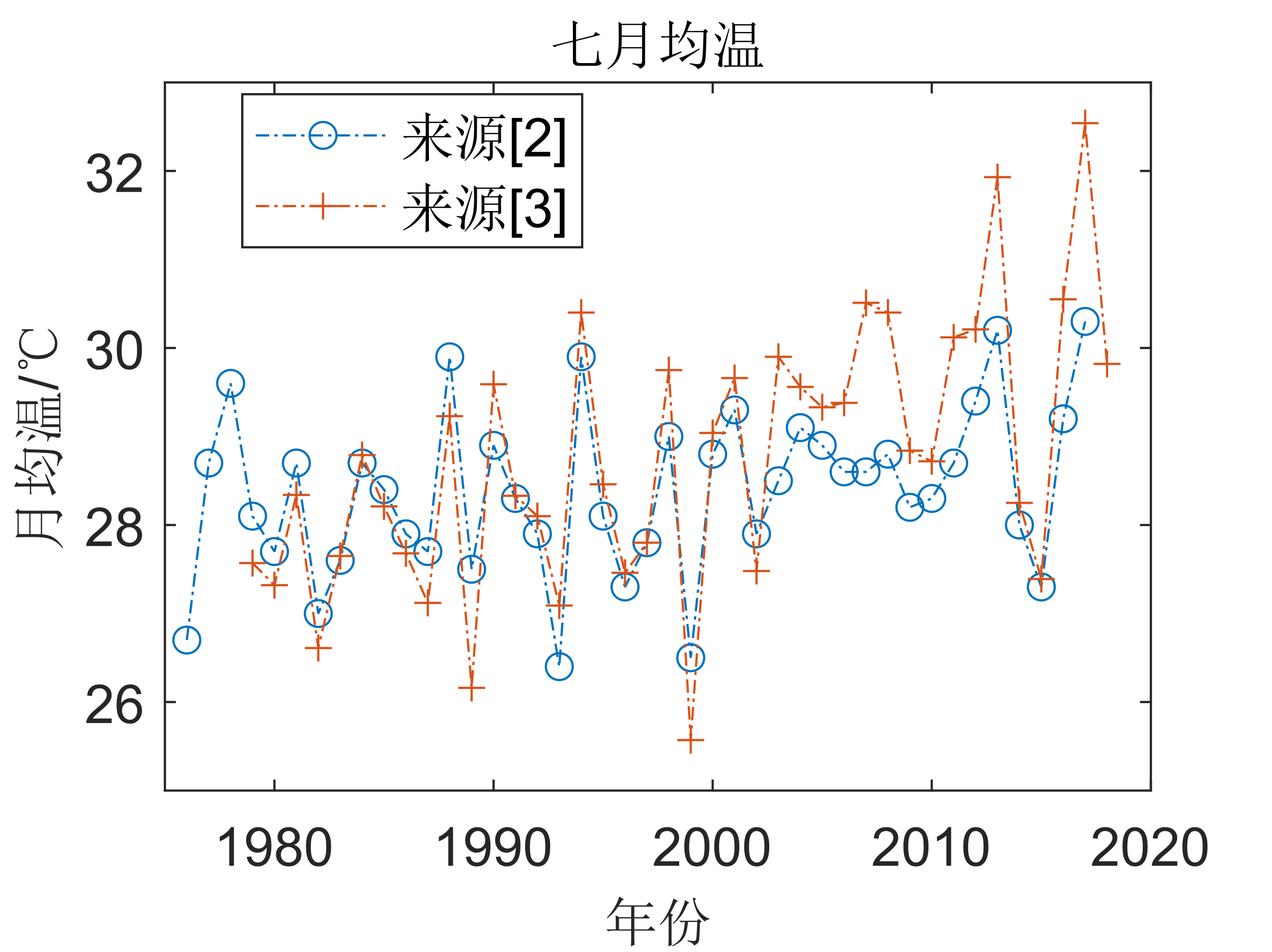


图1（b） 七月近地面平均气温

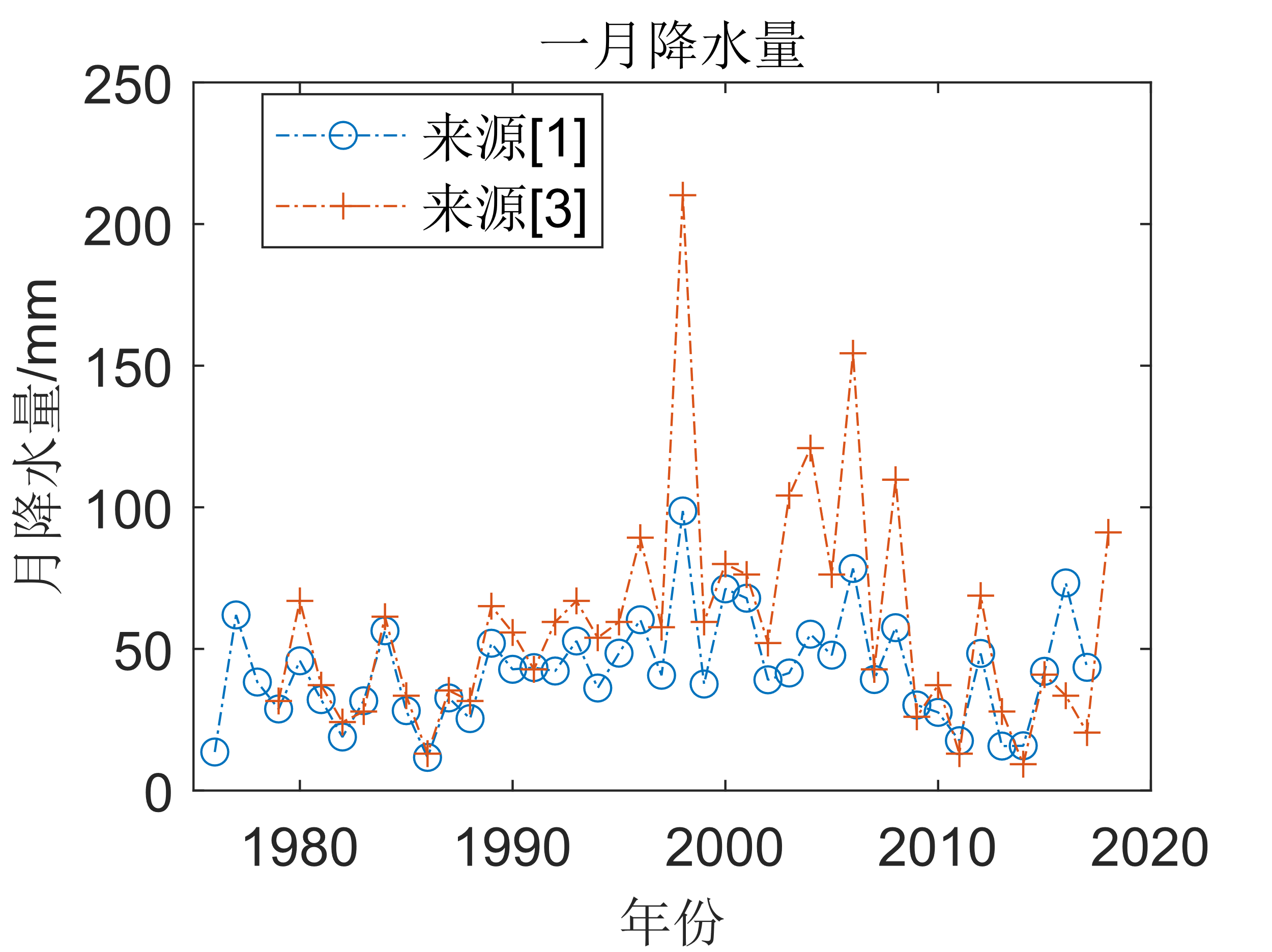


图1（c） 一月降水量

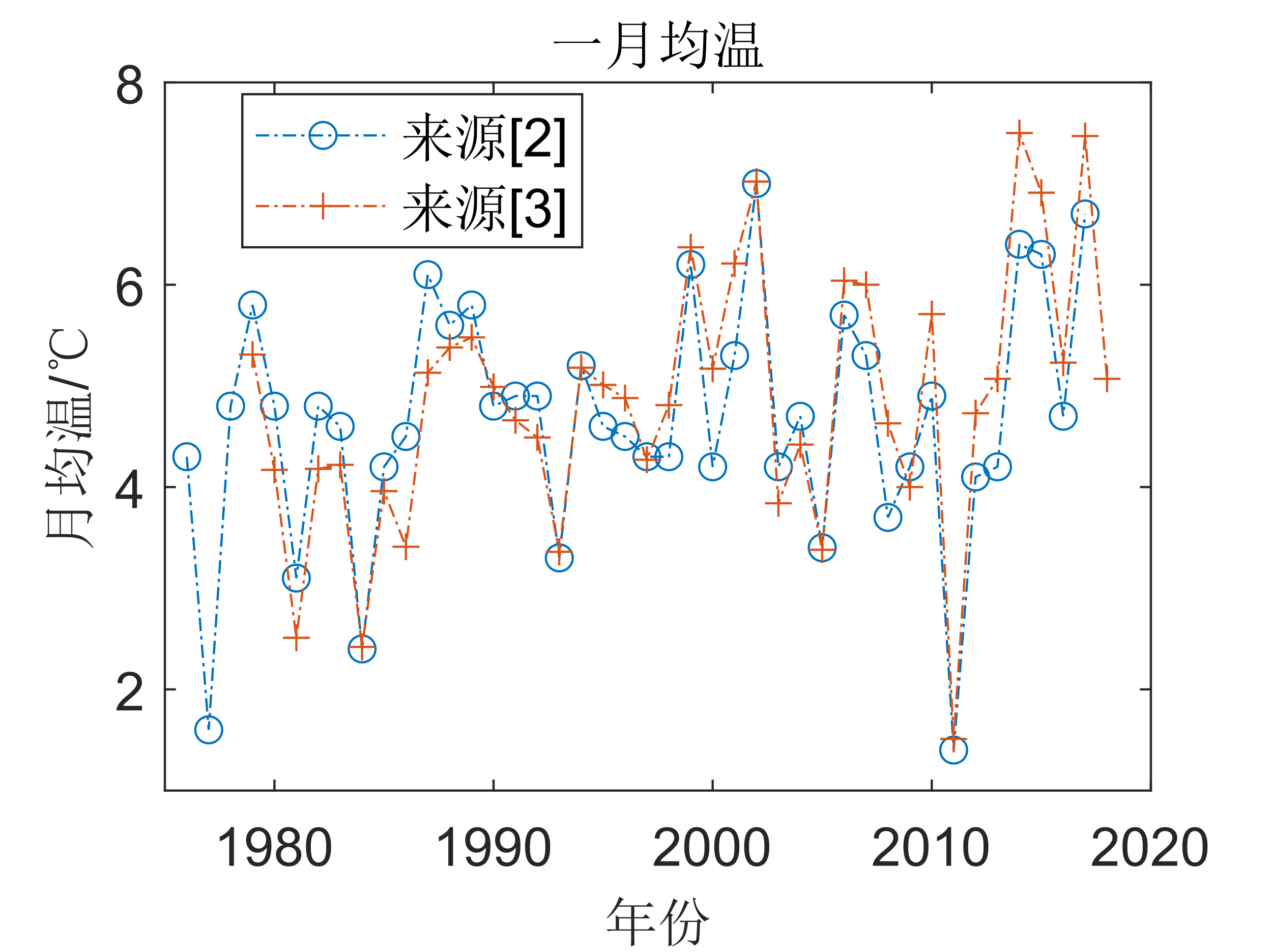


图1（d） 一月近地面平均气温

图1 本研究使用的本地气候数据图示

# 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.

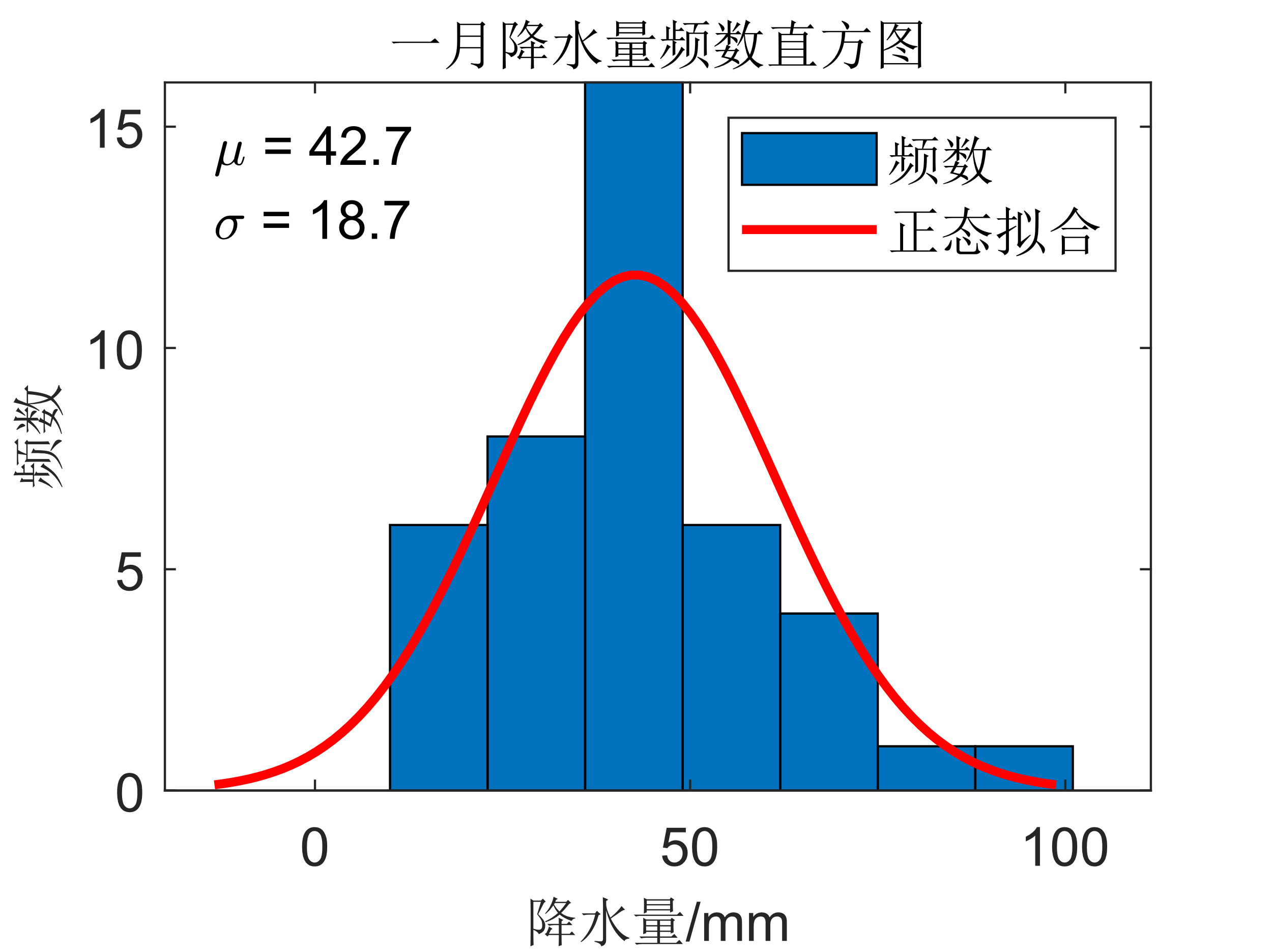


图2（a） 一月降水量

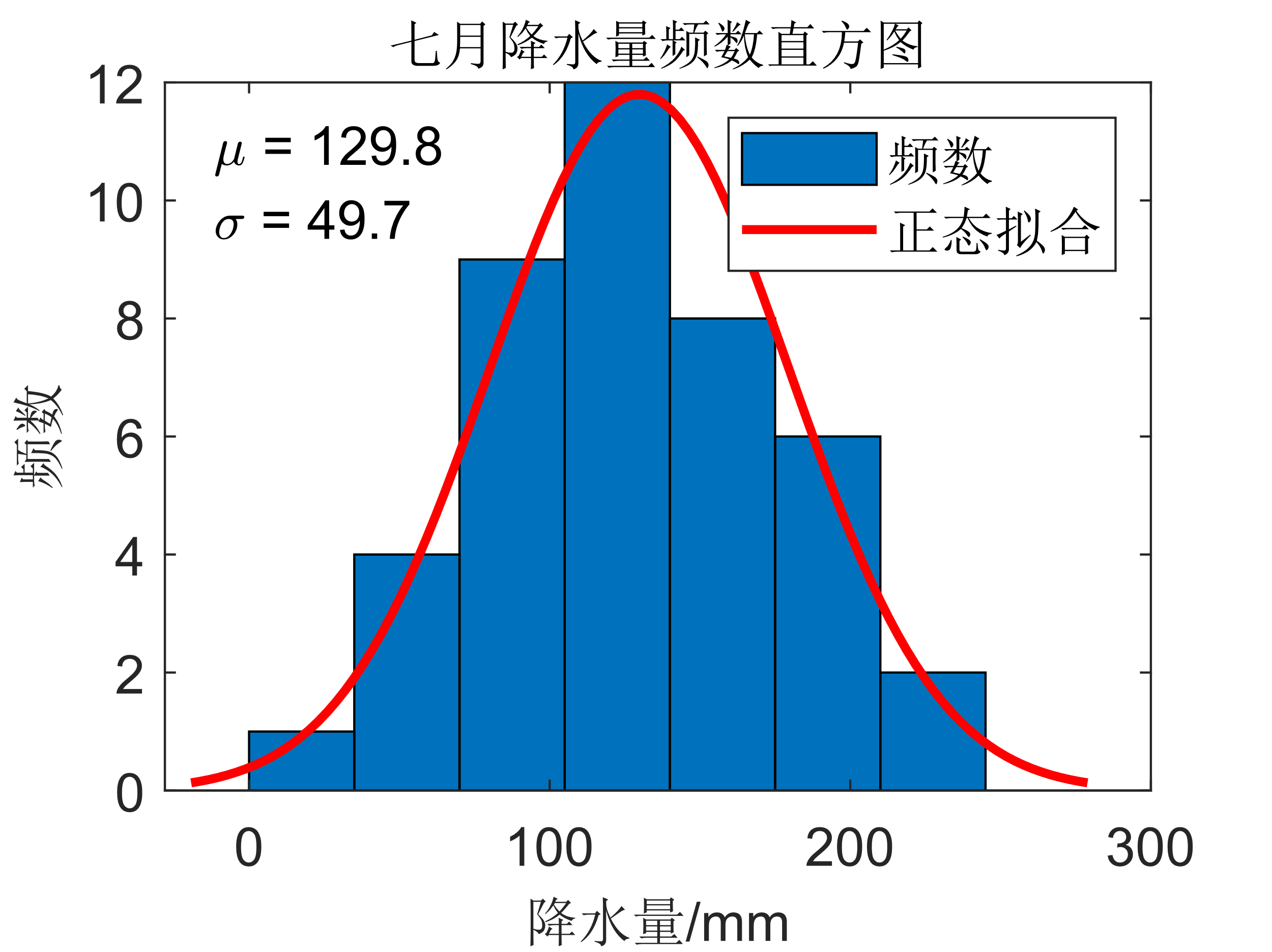


图2（b） 七月降水量

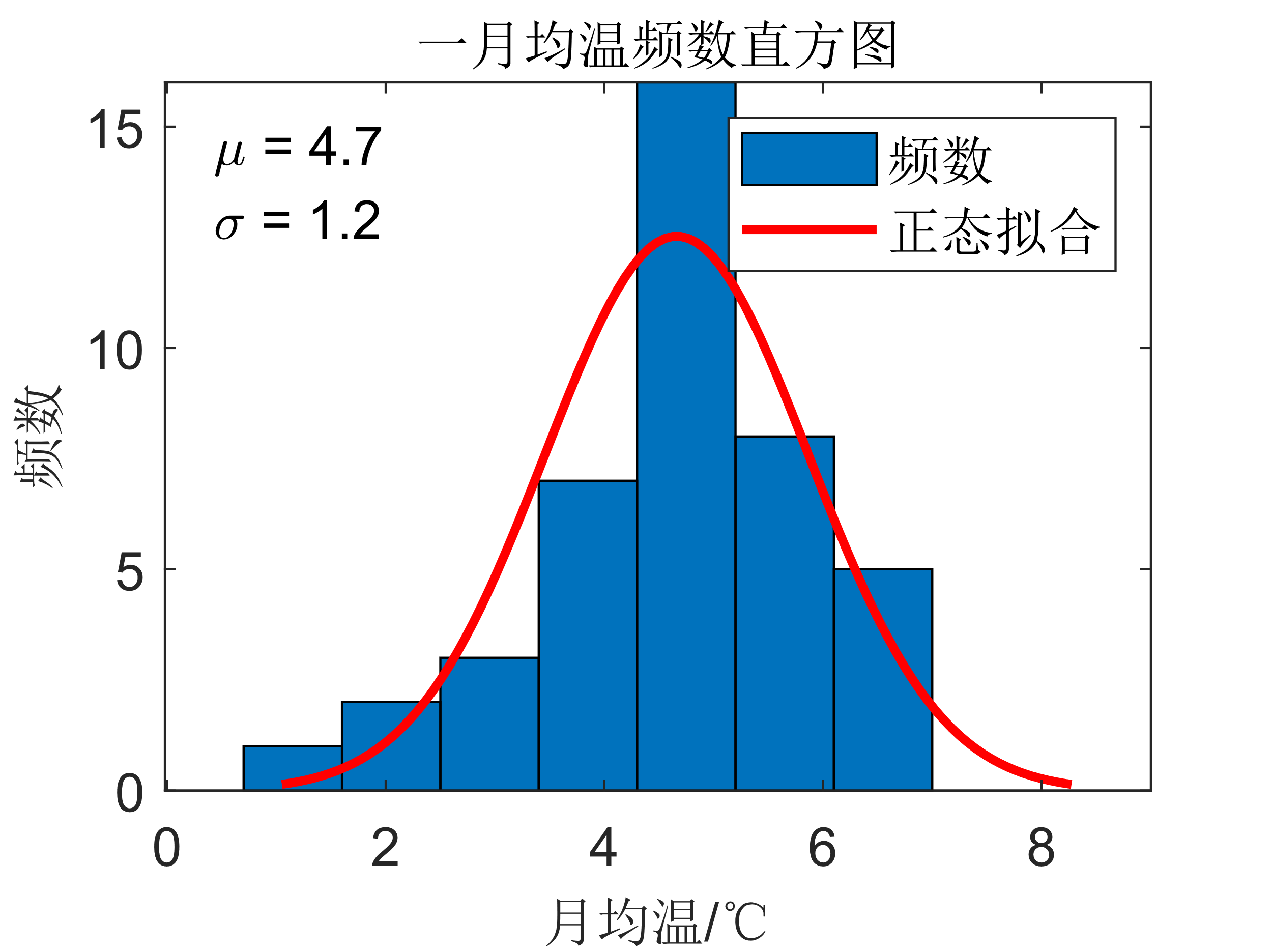


图2（c） 一月均温

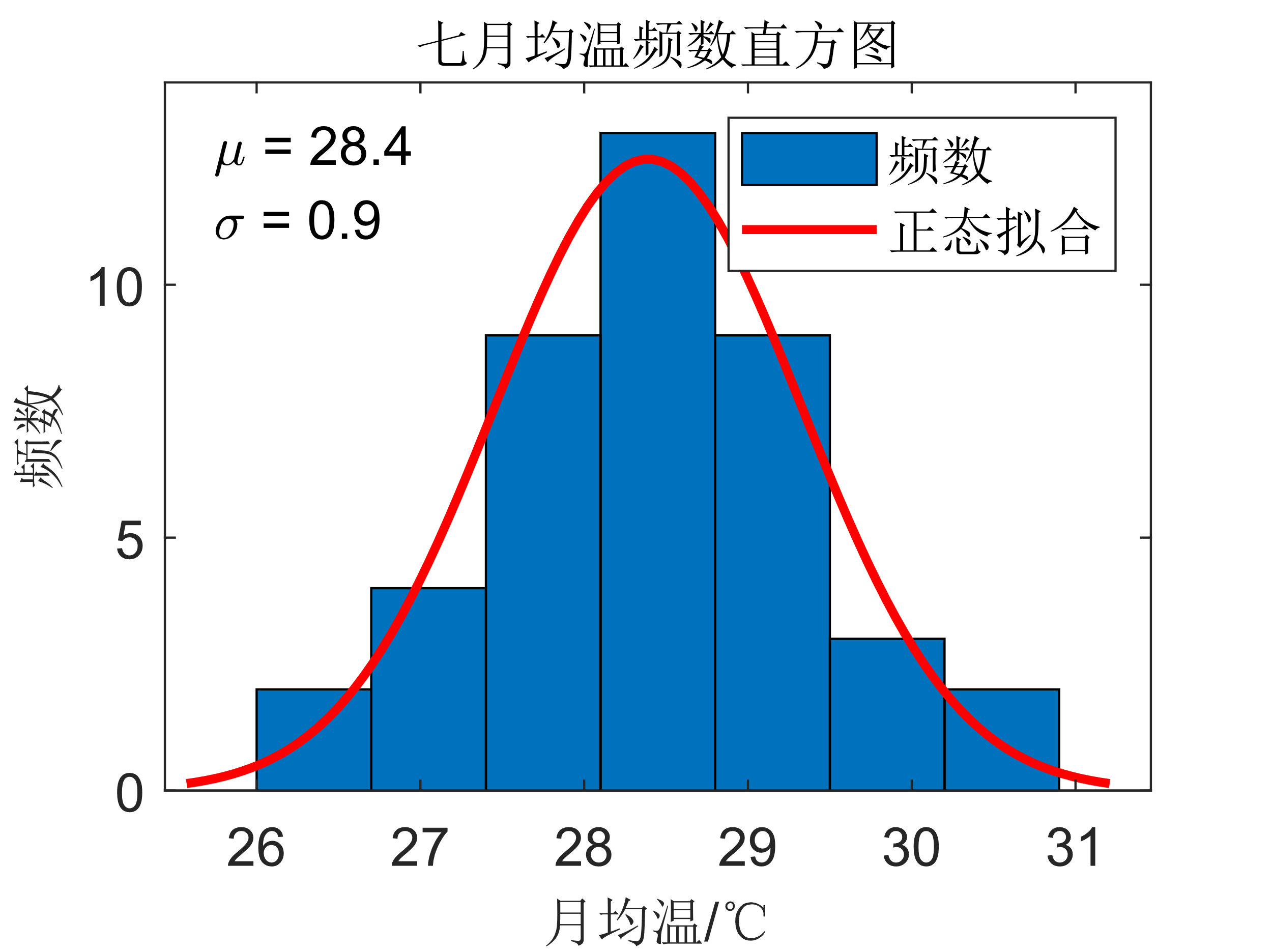


图2（d） 七月均温

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.

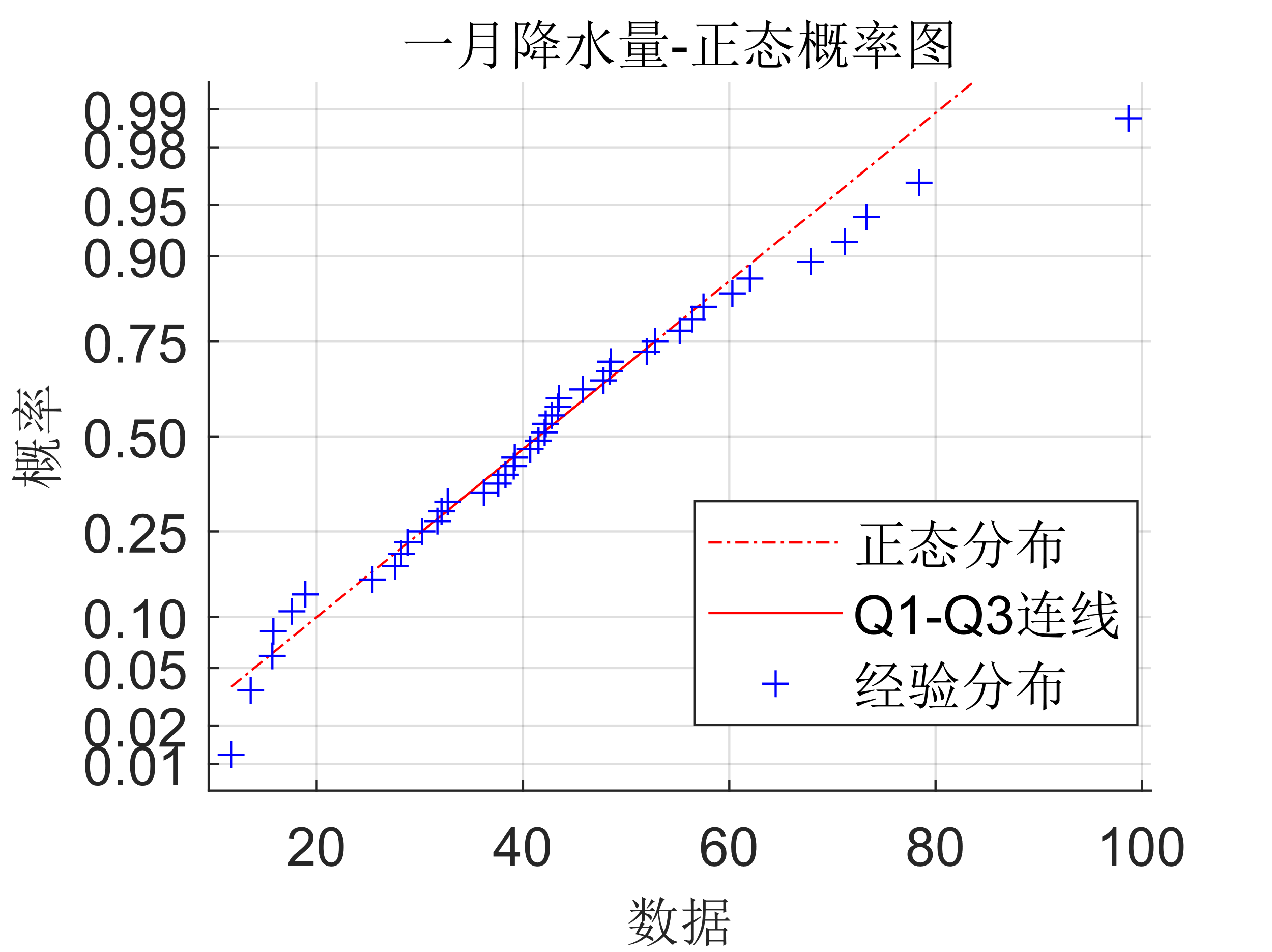


图3（a） 一月降水量

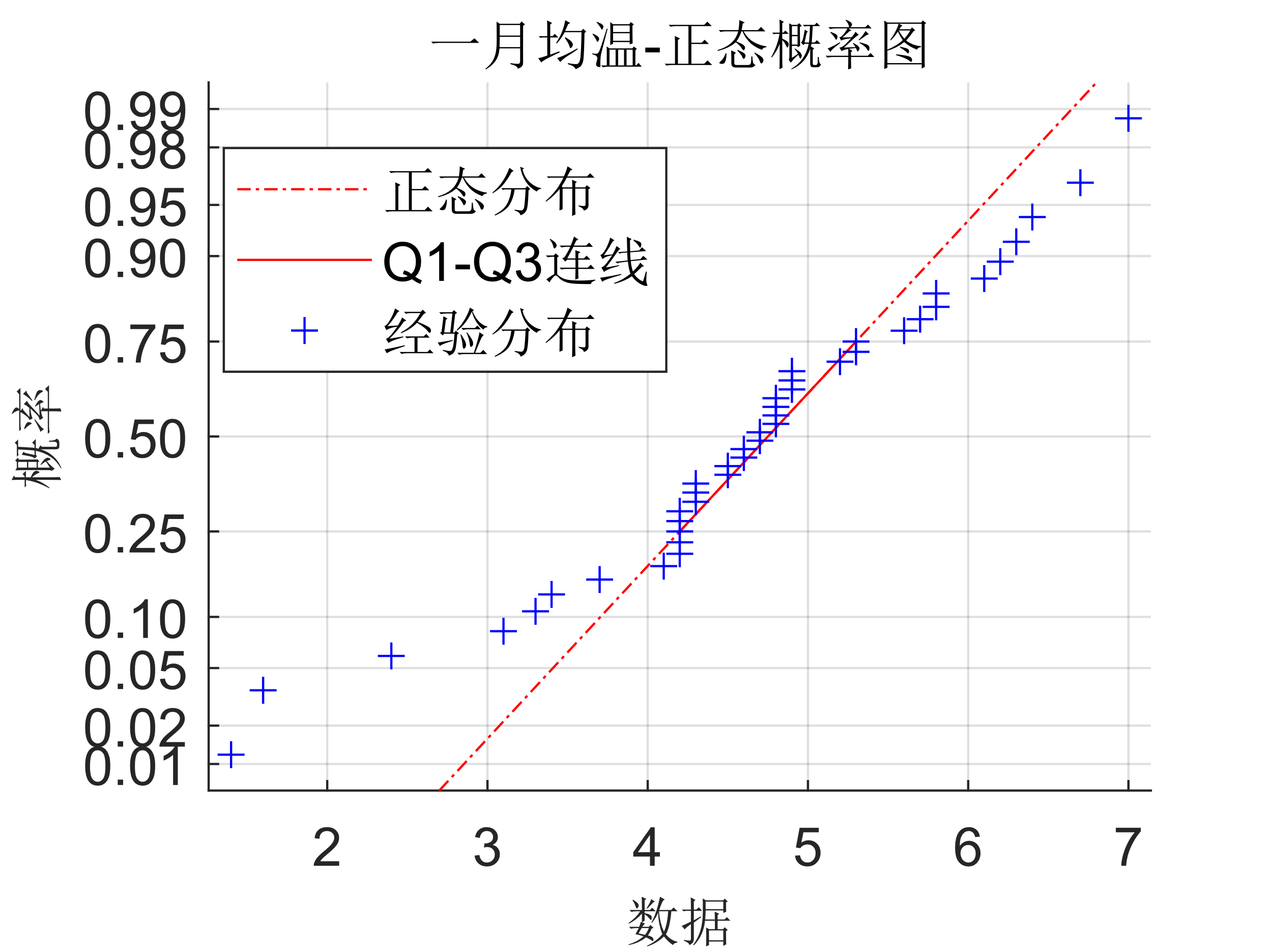


图3（b） 一月均温

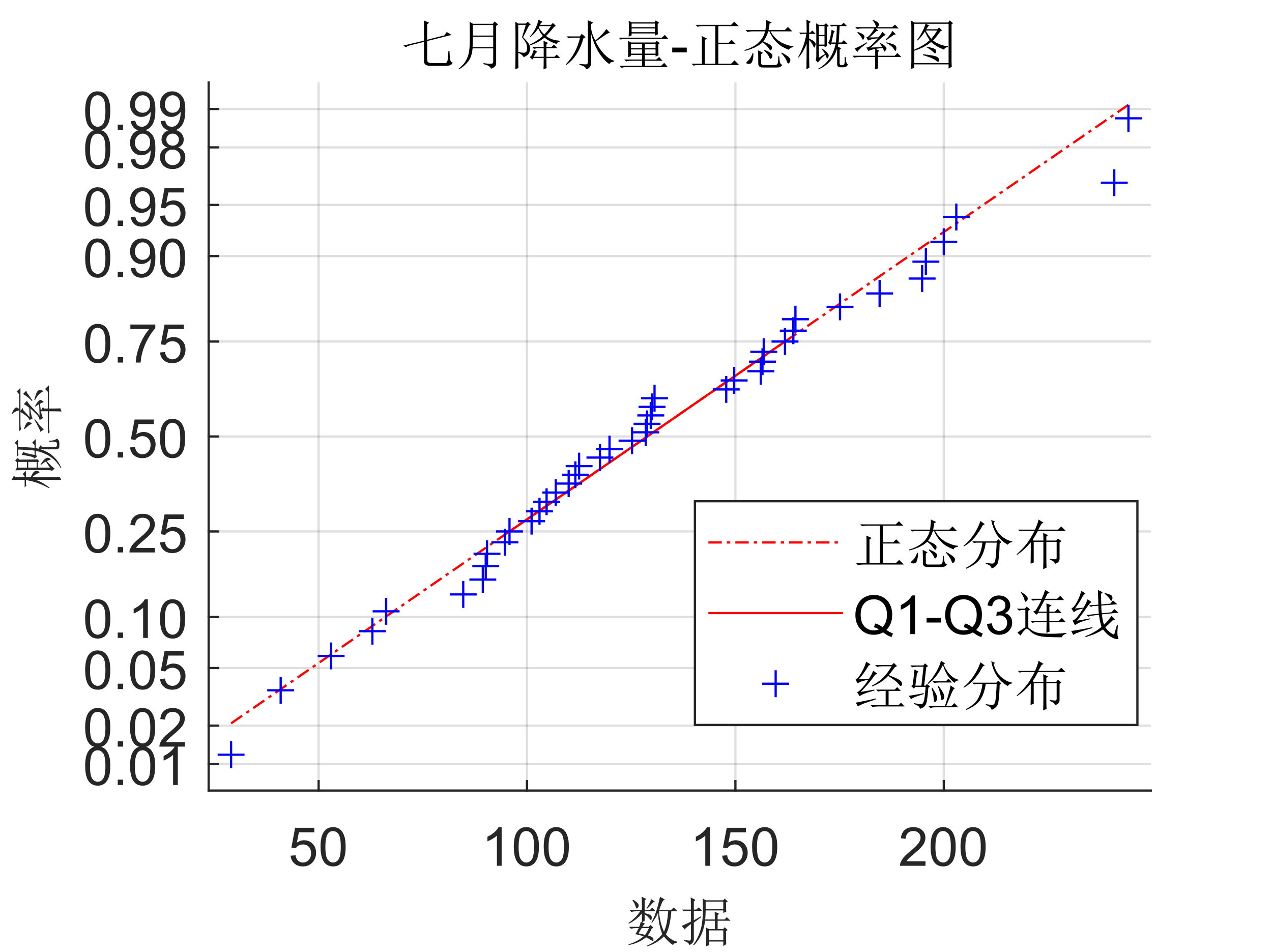


图3（c） 七月降水量

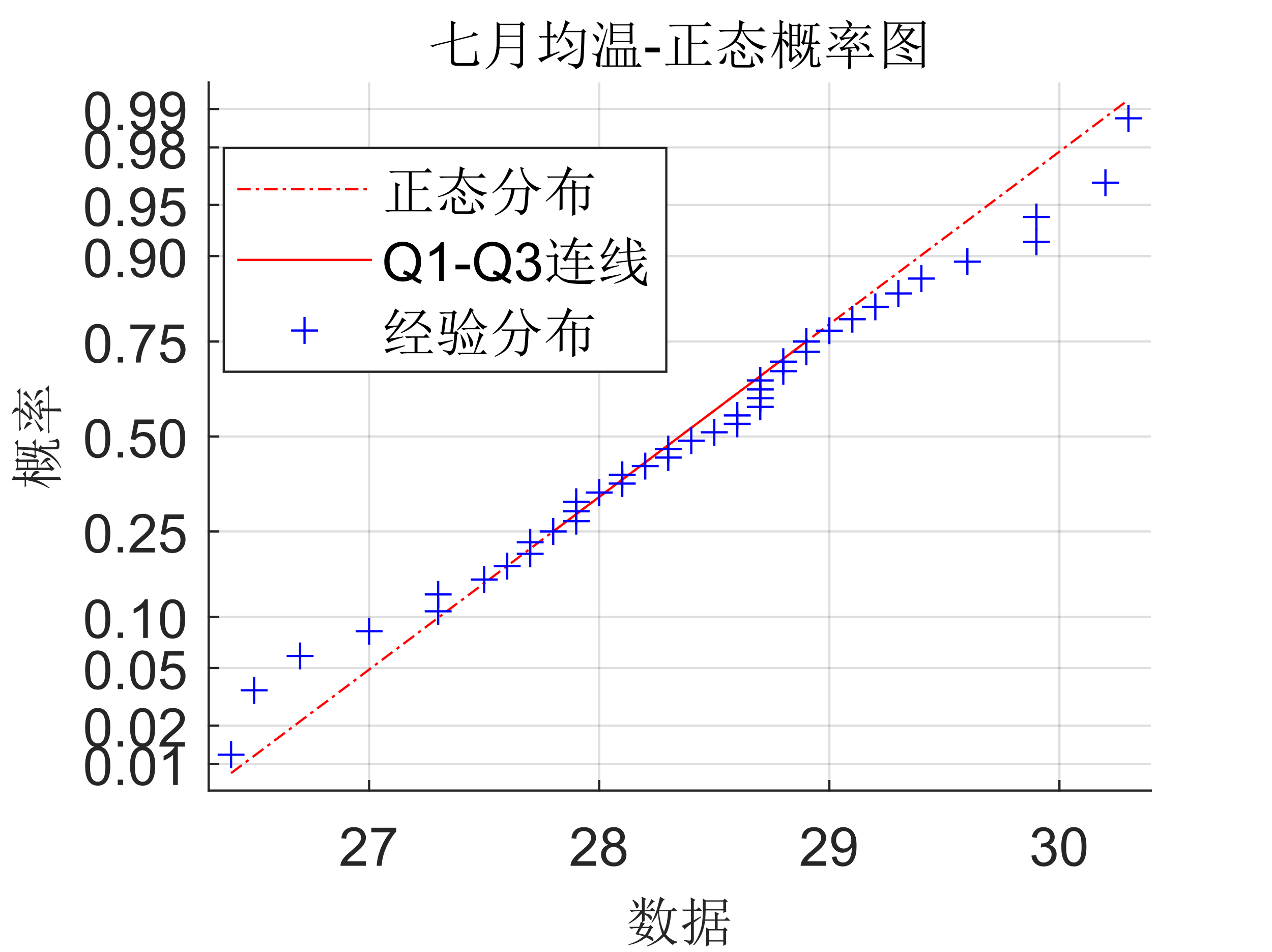


图3（d） 七月均温

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.

# 非参数检验——定量推断本地气候要素的分布类型

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.

# 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= 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 \* | 总体标准差† |
| 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/℃ | 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/℃ | 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= 0.05的假设检验的结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 样本内容 | 原假设*H*0 \* | 总体均值差 † | 总体方差比 § | 检验结果¶ |
| 一月降水量/mm |  | [-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/℃ | [-1.7163, -0.1297] | [0.1437, 0.9175] | 拒绝*H*0（拒绝*H*0） |
| 七月均温/℃ | [-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

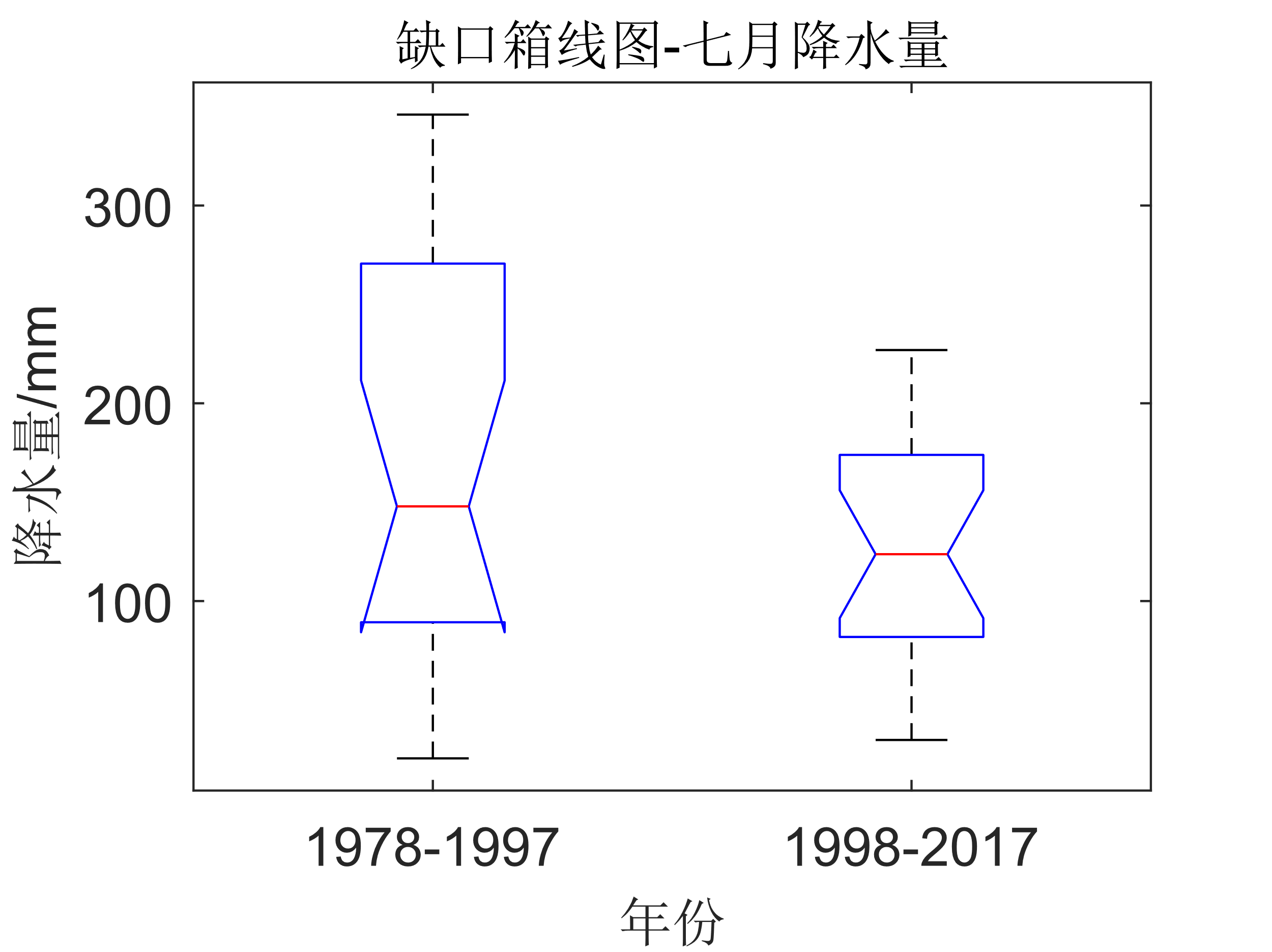


图4（a） 七月降水量

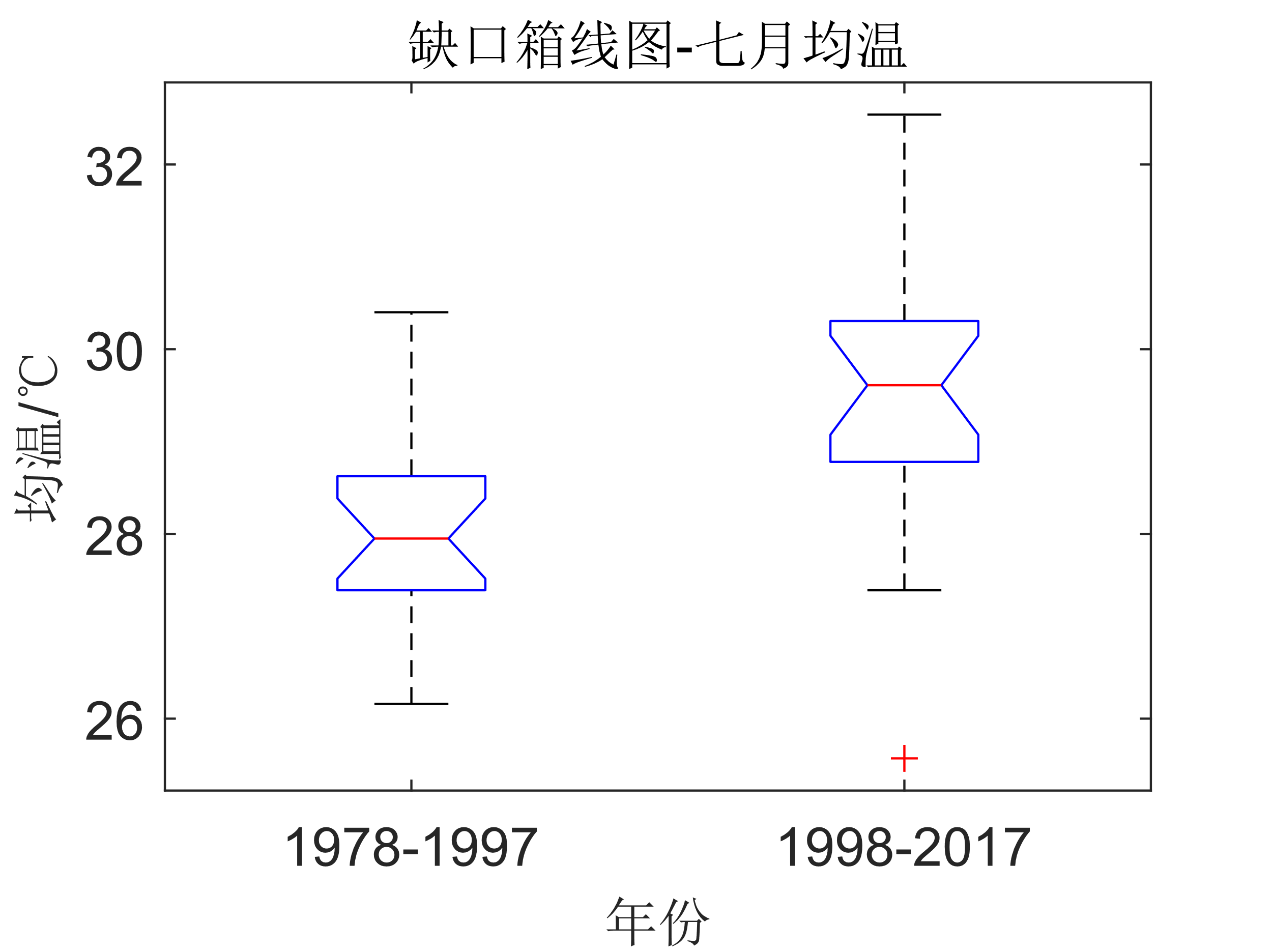


图4（b） 七月均温

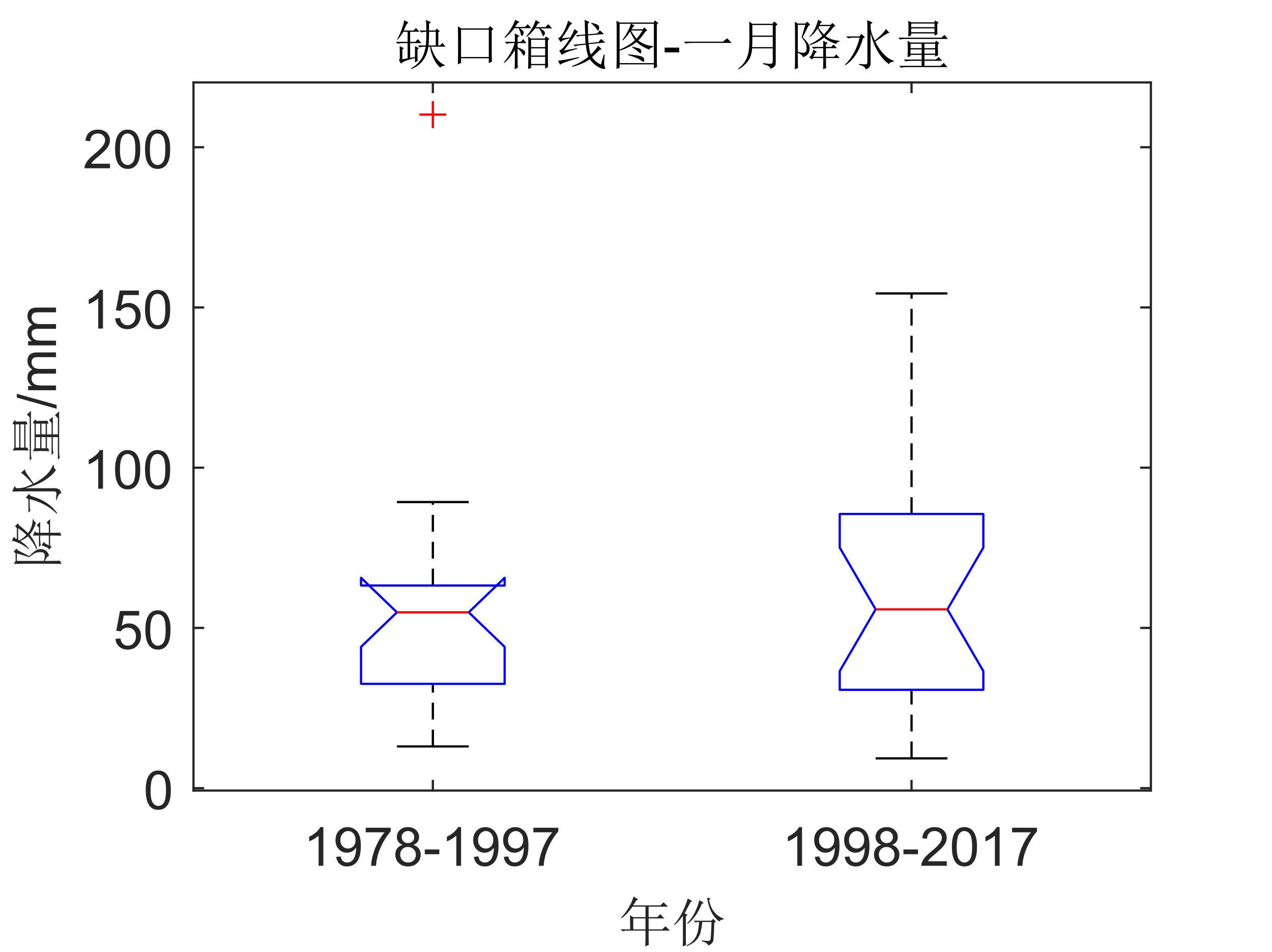


图4（c） 一月降水量

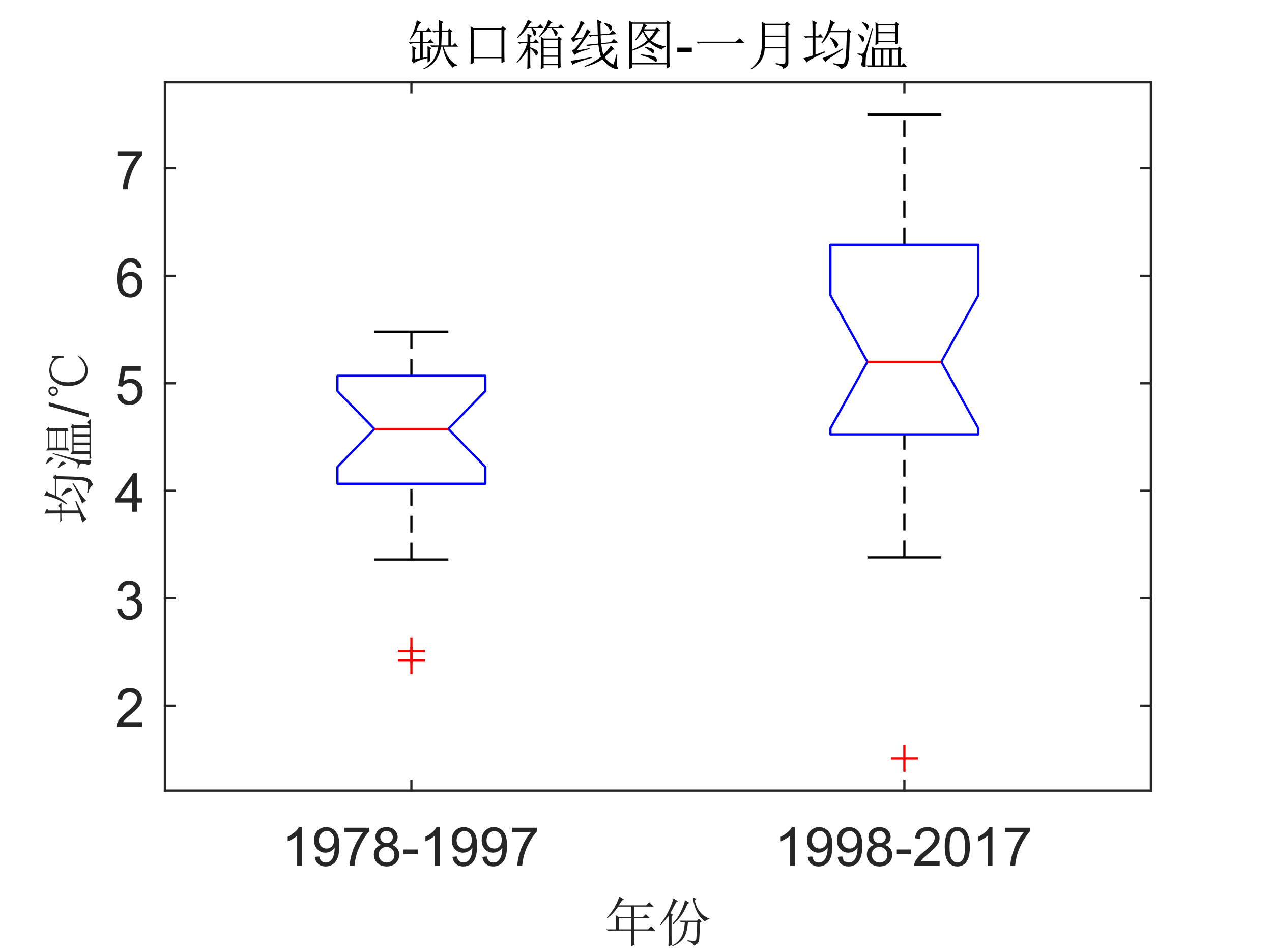


图4（d） 一月均温

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中的缺口不重叠，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.

# 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，然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）.

# 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化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，当迁站前后拔海高度相差超过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 longwave radiations over high altitude regions: Observation 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,硕士研究生. 学科兴趣：大气与海洋科学.

E-mail: weiguorui@sjtu.edu.com

1. 完成日期：2022-06-14

   课程名称：研-STAT6001-M01-基础数理统计 [↑](#footnote-ref-1)