

武汉理工大学

数学建模暑期培训论文

第 1 题

基于经验正交函数分解和多模式误差订正 对中国极端降水的分析及预测

第 1 组

姓名	方向
阮滨（组长）	编程
王宇	建模
王家柯	写作

2020 年 7 月 22 日

摘要

本文基于极端降水特征的多角度指标,通过经验正交函数分解(EOF)提取其时空特征,并运用了 Manner-Kendall 趋势检验、Pearson 相关性系数分析了指标与响应结果之间的关系,之后运用线性变换(LS)、分位点投影变换(QM)等方法对模式误差进行了订正,分别使用时空技巧评分对订正效果进行了评估,最后基于分析结果,订正了未来极端降水的预测结果,并基于此给出了相关建议。

针对问题一需要构建极端降水综合特征,并对数十年来极端降水的时空特征及演变趋势进行分析。通过查阅相关文献,本文首先根据百分位数 P95 确定了极端降水阈值,并从强度、频率、持续性三个方面确定了七个衡量极端降水情况的指标。在此基础上使用了经验正交函数分解法(EOF),对时间和空间两个角度的模态特征进行提取,并通过 Manner-Kendall 趋势分析法对其时间系数进行了趋势性检验,结果显示极端降水强度等指标的增加趋势较为显著。为保证模型结果可靠,本文还对经验正交函数分解法的结果进行了 North 检验,确定了各模块之间的相互独立性。证明了我国极端降水特征在东南地区和西北地区的差异较大,并有随时间逐渐增强的趋势。

针对问题二需要分析全球变暖、ENSO 等因素对全国范围极端降水特征及其概率分布的影响。本文首先通过查阅资料确定了包括太平洋厄尔尼诺系数(ONI)、太平洋年代际震荡指数(PDO)等五个指标作为研究的影响因素,并通过 Savitzky-Golay 滤波对其进行降噪处理。进而选取了极端降水强度(SDII)和基于 Gamma 分布拟合的极端降水概率分布参数作为响应指标,通过 Pearson 相关系数分析了指标之间的关系,发现全球变暖、太阳黑子相对活动数对全国范围极端降水的出现具有正面影响。

针对问题三需要对于国家气候中心的模拟数据进行评估并进行误差订正,并在评估订正效果。本文首先采用了时间技巧评分(SS)和空间技巧评分(M2)对国家气候中心 BCC-CSM1.1-m 全球气候模式模拟的逐日降水模式进行了多角度评估,发现该模式的模拟效果较差。并选用了线性变换(LS)、分位点投影变换(QM)对原始数据进行误差订正,针对订正后的结果再次运用了时空技巧评分对其进行效果评估,发现分位点投影变换的效果较好,SS 和 M2 指标分别提升了 14.40%和 77.82%。

针对问题四需要在误差订正后的模式的基础之上对 21 世纪中期和末期的极端降水情况进行分析。本文结合前文中的数学模型及分析结果,对误差订正后的预测数据进行了统计,并拟合了相应的概率密度函数。结合数据可视化结果及其密度函数的相关特征,发现在未来我国的极端降水强度呈现增加趋势。基于此,本文提出了针对预防极端降水天气和面对极端降水天气的一些建议。

本文的优点在于:1. 使用了经验正交函数分解法,得到了时空模态特征,分析角度全面且符合实际情况;2. 采用了多种误差订正的方案,在一定程度上保证了误差订正效果的优度。

关键词: 经验正交函数分解法 M-K 趋势分析法 分位点投影变换 S-G 平滑滤波

目录

一、 问题重述	3
1.1 问题背景	3
1.2 问题重述	3
二、 模型的假设	3
三、 符号说明	3
四、 问题一的模型建立与求解	4
4.1 问题分析	4
4.2 预备工作	4
4.3 模型的建立	6
4.3.1 建立经验正交函数修改模型 (EOF)	6
4.3.2 Manner-Kendall(M-K) 非参数检验法	7
4.4 模型求解	7
4.5 结果分析	8
4.6 模型检验	8
4.6.1 North 检验	8
五、 问题二模型的建立与求解	10
5.1 问题分析	10
5.2 预备工作	11
5.3 模型建立	12
5.3.1 Pearson 相关性分析	12
5.4 模型求解	13
5.5 结果分析	13
六、 问题三模型建立与求解	13
6.1 问题分析	13
6.2 模型的建立	14
6.2.1 线性变化模型	14
6.2.2 分位点投影变换模型	14
6.3 模型的求解	15
6.4 结果分析	16
6.4.1 时空检验指标	16
七、 问题四的求解与分析	17
7.1 问题分析	17
7.2 问题四求解	17
7.3 结果分析	18
7.3.1 给相关部门的政策与建议	18
八、 模型总结与评价	19
8.1 模型优点	19
8.2 模型缺点	19
8.3 模型改进	19

附录 A 代码	21
A.1 (EOF) matlab 源程序	21
A.2 problem1.m matlab 源程序	21
A.3 (problem2.m 源程序	23
A.4 North 检验 –matlab 源程序	24
A.5 QM 订正 –matlab 源程序	25
A.6 SS,M2–matlab 源程序	25

一、问题重述

1.1 问题背景

近百年来，全球变暖成为全球气候变化的主要趋势，与此同时全球极端天气也处于一个频发的阶段，频繁的极端天气给人们的生产经营活动产生了巨大的影响。因此如何正确地把握极端天气的趋势十分有必要，把握极端降水特征变化，研究其与全球变暖、ENSO 等的关系，预测极端降水未来变化。可以增强我国应对极端天气的能力，并能够给抗灾救灾提供重要理论依据^[2]。

1.2 问题重述

问题一：利用附件所给的中国区域 1961-2005 年夏季逐日降水观测数据，构建表征极端降水综合特征的指标体系，通过建立数学模型分析某一区域几十年来夏季极端降雨的时空特征和演变趋势。

问题二：收集全球变暖和 ENSO 正负位相（即厄尔尼诺和拉尼娜）等数据，分析其对极端降水综合特征及其概率分布等影响

问题三：通过建立数学模型，评价订正后国家气候中心 BCC-CSM1.1-m 全球气候模式模拟的逐日降水数据，从多角度分析可能出现的偏差，并且给出订正方案。

问题四：通过问题三订正后的数据，求出其在概率以及数量上的特征。并根据其数字特征的变化趋势给出相应的气候变、化应对、防灾减灾措施制定方案以及建议。

二、模型的假设

- 假设收集到的数据不存在较大的系统性误差；
- 附件中选取的 553 个观测站点收集的数据对其所在地的降水特征具有较好的代表性。
- 年度数据如二氧化碳指标对全年各季度的影响相同

三、符号说明

符号	意义
$X_{m \times n}$	数据矩阵
$V_{m \times m}$	特征向量
Y_{j+i}	像原始值
C_i	第 i 个值滤波时的系数
$P_{obs,ctr}(d)$	建模时期观测的逐日降水数据
$P_{mod,ctr}(d)$	建模时期模式逐日降水数据
$F_0(x)$	观测时降水地域或等于 x 的概率
$F_m(x)$	预测时降水地域或等于 x 的概率

注：其他未列出的符号以第一次出现时的解释为准

四、问题一的模型建立与求解

4.1 问题分析

针对问题一，利用第 95 百分位数确定极端降水的阈值，分为三个指标：强度指标、连续性指标、频率指标。其中强度指标用极端降水期日均降水量（SDII）、最大日降水量（RX1day）、最大连续五日降水量（RX5day）综合表示，持续性指标用湿日总降水量（PRCPTOT）、连续湿日（CWD）、连续干日（CDD）综合表示，频率指标利用基于广义极值分布计算极端降水重现期来表示。并利用经验正交函数分析法（EOF）来分析极端降雨特征在时空特征。如果通过 North 检验则利用 Mann-Kendall 趋势检验提取演变趋势^[1]。

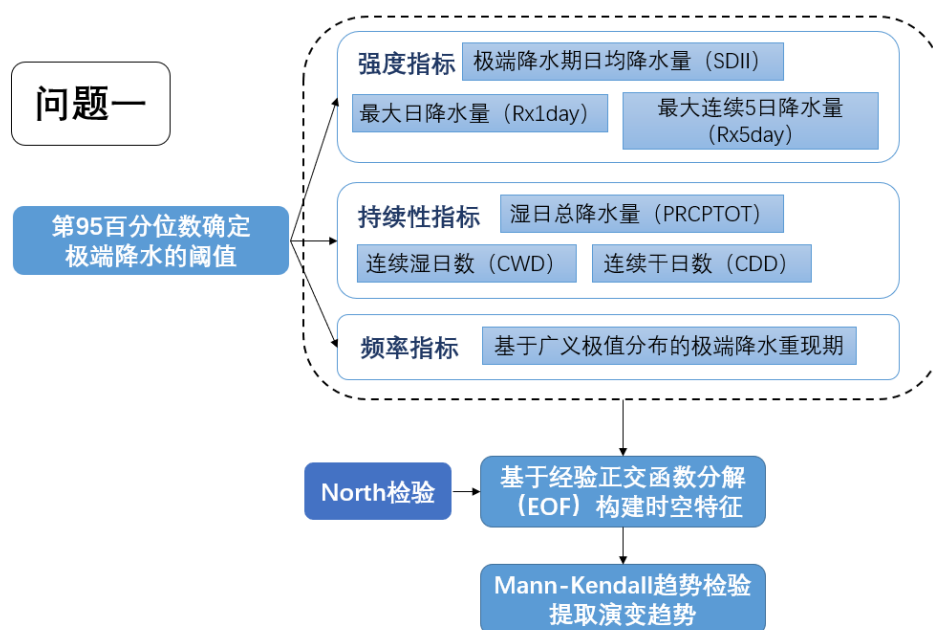


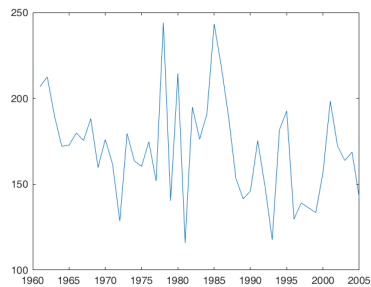
图 1 问题一流程图

4.2 预备工作

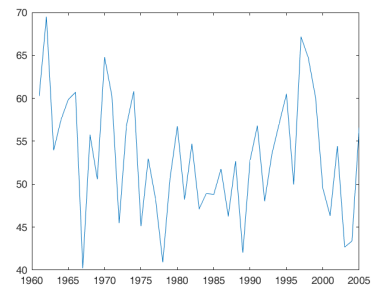
利用问题所给的附件中 553 个站点每日降水量的数据，通过文献所给的统计方法分别求出极端降水期日均降水量（SDII）、最大日降水量（RX1day）、最大连续五日降水量（RX5day）、湿日总降水量（PRCPTOT）、连续湿日（CWD）、连续干日（CDD），基于广义极值分布计算极端降水重现期等指标的参数。

表 1 指标选取

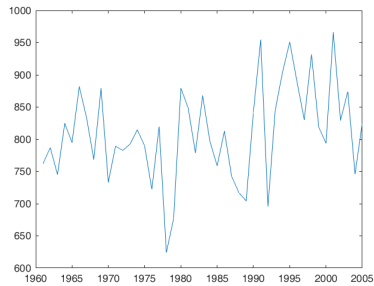
指标	缩写	定义
湿日总降水量	PRCPTOT	每年大于等于 1mm 的降水量总和
连续干日数	CDD	每年最长连续无降水日数
连续湿日数	CWD	每年最长连续降水日数
极端强降水量	R99	大于基准点内 99% 分位点的日降水量总和
降水强度	SDII	降水量与降水日数的比值
1 日最大降水量	Rx1day	最大的日降水量
5 日最大降水量	Rx5day	最大的连续五日降水量



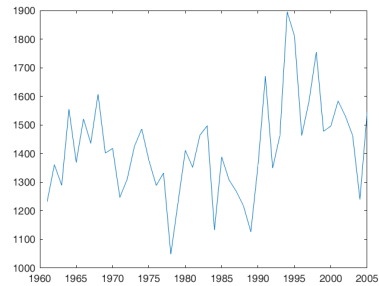
(a) 连续干日数 (CDD)



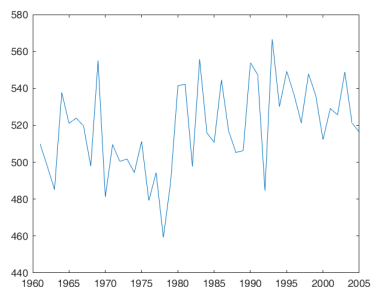
(b) 连续湿日数 (CWD)



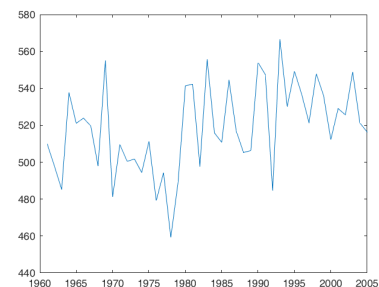
(c) 最大日降水量 (Rx1day)



(d) 最大连续五日降水量 (Rx5day)



(e) 极端降水期日均降水量



(f) 湿日总降水量

图 2 指数收集结果

4.3 模型的建立

4.3.1 建立经验正交函数分解模型（EOF）

为获得极端降水指标体系的时空特征，应采取合适的模型将气候分解至时间维度和空间维度。本文认为气候场分别在时间维度和空间维度上都有相应的模态，并且估计具体到每个站点的极端降水特征都是根据时空维度上的模态线性组合构成的。因此本文选用了正交函数经验分解模型，拟分别提取出典型的时空模态。另外，针对各指标的变化趋势，考虑到数据的分布不一定为正态分布，故采用 Mann-Kendall 趋势检验。此方法不依赖数据的分布，且能够在一定程度上避免异常数据的干扰。

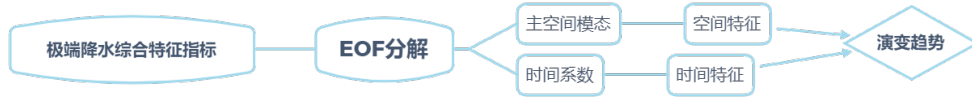


图3 经验正交函数分解模型流程图

step1: 选定要分析的数据，进行数据的预处理，通常处理成距平的形式得到一个数据矩阵 $X_{m \times n}$

step2 计算 X 与其转置矩阵 X^T 的交叉积，得到方阵

$$C_{m \times m} = \frac{1}{n} X \times X^T. \quad (1)$$

如果 X 是已经处理成距平的话，则称 C 为协方差阵；如果 X 已经标准化（即 C 中每一行数据的平均值为 0 标准差为 1）则称为相关系数矩阵。

step3 计算方阵 C 的特征根 ($\lambda_1, \lambda_2, \dots, \lambda_m$) 和特征向量 $V_{m \times m}$ ，二者满足

$$C_{m \times m} \times V_{m \times m} = V_{m \times m} \times E_{m \times m} \quad (2)$$

其中 E 是 $m \times m$ 维对角阵，即

$$E = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{bmatrix}$$

一般将特征根 λ 按从大到小排列。因为数据 X 是真实的观测值，所以 λ 应该大于等于 0。每一个非 0 的特征根对应一列特征向量值，也称 EOF。

step4: 计算主成分。将 EOF 投影到原始资料矩阵上，就得到所有空间特征向量对应的时间系数（即主成分），即

$$PC_{m \times n} = V_{m \times m}^T \times X_{m \times n} \quad (3)$$

其中 PC 中的每一行数据就是对应每个特征向量的时间系数。

step5 计算贡献率。矩阵 X 的方差大小可以简单的用特征根的大小来表示。 λ 越高说明对应的模态越重要，对总方差的贡献越大。第 k 个模态对总的方差解释率为

$$\frac{\lambda_k}{\sum_{i=1}^m \lambda_i} \times 100\% \quad (4)$$

step6 显著性检验。实际资料分析中得到的空间模态是否是随机的, 需要进行统计检验。North 等 (1982) 的研究提出, 在 95% 置信度水平下的特征根的误差

$$\delta\lambda = \lambda\sqrt{\frac{2}{N^*}} \quad (5)$$

其中 λ 是特征根, N^* 是数据的有效自由度。将 λ 按顺序依次检查, 标上误差范围。如果前后两个 λ 之间误差范围有重叠, 则没有通过显著性检验。

4.3.2 Manner-Kendall(M-K) 非参数检验法

Manner-Kendall(M-K) 非参数检验法常用于分析降水、径流、气温等要素时间序列的变化趋势。

设有 n 个样本量的时间序列 x_1, x_2, \dots, x_n , 对于所有 $k, j \leq n$, 且 x_k 和 x_j 的分布是懂得, 计算检验统计量 s , 公式如下:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{Sgn}(x_j - x_k) \quad (6)$$

其中

$$\text{Sgn}(x_j - x_k) = \begin{cases} +1(x_j - x_k) > 0 \\ 0(x_j - x_k) = 0 \\ -1(x_j - x_k) < 0 \end{cases} \quad (7)$$

s 为正态分布, 均值为 0, 方差 $\text{var}(s) = n(n-1)(2n+5)/18$ 。当 $n > 10$ 时, 标准正态统计变量通过下式计算:

$$Z = \begin{cases} \frac{S-1}{\sqrt{\text{var}(S)}} > 0 \\ \frac{S+1}{\sqrt{\text{var}(S)}} < 0 \end{cases} \quad (8)$$

对于统计值 Z 来说, 大于 0 时, 是增加趋势; 小于 0 时则是减少趋势。 z 的绝对值在大于 1.28、1.64、2.32 时, 分别表示通过了置信度 90%、95% 和 99% 的显著性检验。

4.4 模型求解

表 2 M-K 趋势检验结果

指标	Z 统计量	趋势
Rx5day	1.7902	增加趋势显著
SDII	2.338	增加趋势显著
重现期	-0.1663	无明显趋势
湿日总降水	0.69454	无明显趋势
Rx1day	1.8097	增加趋势显著
CDD	-2.0641	下降趋势显著
CWD	-1.2032	无明显趋势

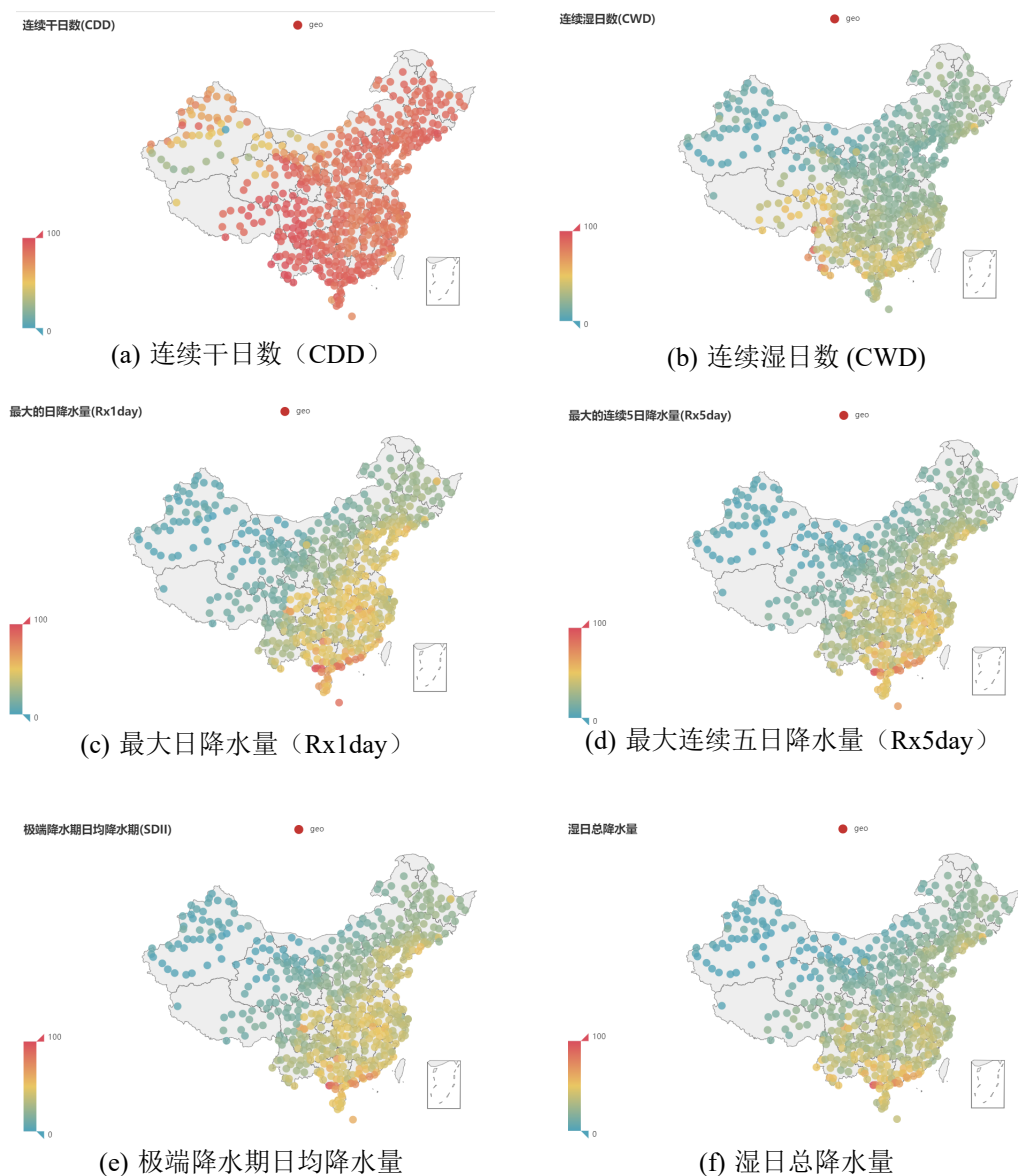


图4 指标地理分布示意

4.5 结果分析

根据上文中正交经验函数分解反映在地图上的结果, 根据图 3-(a) 显示, 我国各地在连续干日数上的差异并不明显。而对于除了连续干日数外的所有指标, 我国的极端降水特征呈现出东南地区与西北地区差异较大, 但内部趋势相近的情况。通过考查我国季风区与非季风区的分界线 (大兴安岭 - 阴山 - 贺兰山 - 巴颜喀拉山 - 冈底斯山), 发现上述结果与分界线几乎相同, 可信度较高。

4.6 模型检验

4.6.1 North 检验

为了确定各个模块之间是否相互独立, North 检验是最简单也是必须要做的检验。

North 检验，是计算特征值误差范围来进行显著性检验。特征值 λ 的误差范围

$$e_j = \lambda_j \left(\frac{2}{n} \right)^{\frac{1}{2}} \quad (9)$$

其中 n 为样本量，当相邻特征值 $\lambda_j + 1$ 满足

$$\lambda_{j+1} - \lambda_j \geq e_j$$

时，认为这两个特征值对应的经验正交函数是有价值的信号。

表 3 North 检验结果

North 检验	结果
湿日总降水量的第一空间模态	1.25911e+09>7.98117e+07
极端降水期日均降水量的第一空间模态	1.19139e+07>729173
连续干日数 (CDD) 的第一空间模态	1.2778e+06>81166.5
连续湿日 (CWD)	123176>7803.49
最大日降水量 (Rx1day)	2.82596e+07>1.777815e+06
最大连续五日降水量 (Rx5day)	8.68538e+07>5.51763e+06

通过 North 检验，可以发现每一个模块都能通过检验，因此可以说明每一个模块的数据都有价值，此模型可以较为成功的反应实际问题。

五、问题二模型的建立与求解

5.1 问题分析

根据相关文献，全球气温变暖加大了全球极端天气发生的频率，南京大学吴润琦分析了气温变化对我国极端天气的影响。同时也发现太平洋厄尔尼诺系数（NIO）^[2]、全球海陆温度指数^[3]、太平洋年代际振荡（PDO）、大气中二氧化碳含量、太阳黑子活动指数^[4]等均对极端天气的发生有影响。确定指标后，从中国气象网、NOAA 等网站上获取了相应的资料。相应的，本文选用极端降水强度（SDII）作为观测的响应指标，并针对极端降水指标拟合的 Gamma 概率密度函数，选取了 α 和 β 作为概率分布的响应指标。

另外，本文拟采用 Pearson 相关系数对影响因素进行相关性分析，对降水的影响因素与响应指标之间的相关关系进行分布。

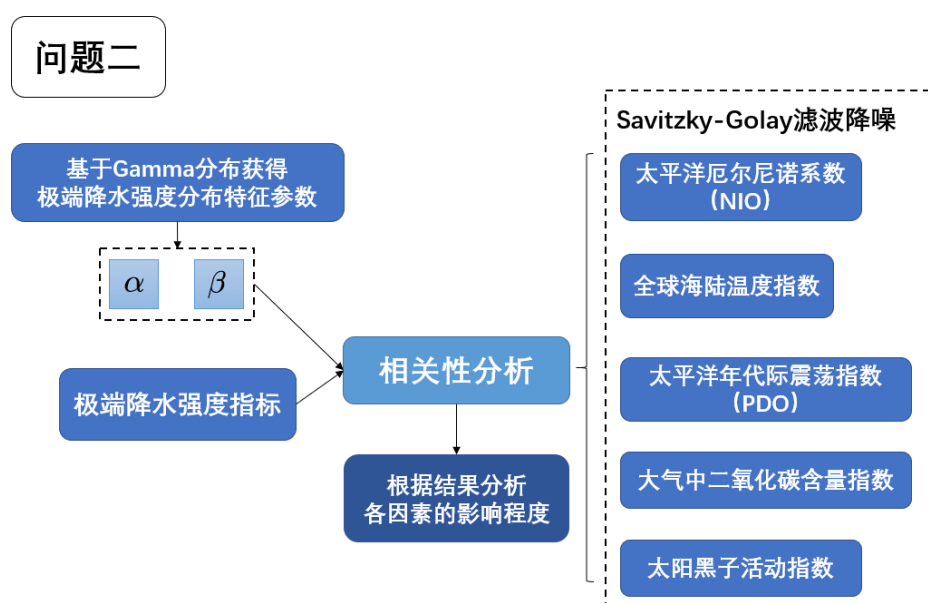


图 5 问题二流程图

5.2 预备工作

将收集到的数据画出三维分布图，据图 6 观察数据在时间上存在准周期震荡趋势。由于收集到的数据可能出现个别误差或呈现短周期震荡的趋势，于是选择对收集到的数据进行平滑化处理。本文采用 S-G 滤波拟合方法对数据进行处理。

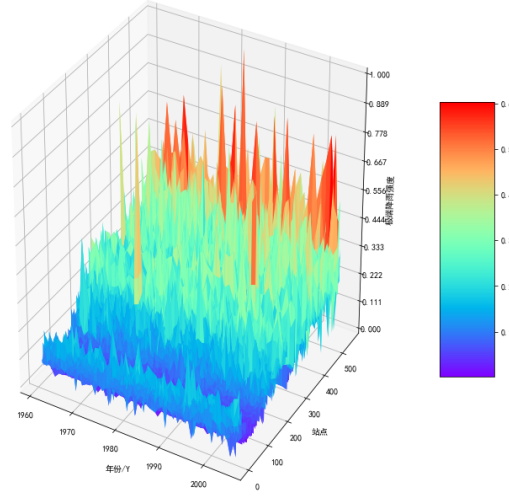


图 6 极端降水强度分布图

S-G 滤波拟合方法是由 Savitzky 等在 1964 年提出的一种基于平滑时间序列数据和最小二乘原理的卷积算法，它是一种移动窗口的加权平均算法，但其加权系数不是简单的常数窗口，而是通过在滑动窗口内对给定高阶多项式的最小二乘拟合得出，其表达式为：

$$Y_j^* = \frac{\sum_{i=-m}^m C_i \times Y_{j+i}}{N} \quad (10)$$

其中， Y_j^* 为拟合值， Y_{j+i} 为像原始值， C_i 为第 i 个值滤波时的系数， m 为半个滤波的宽度， N 为滤波的长度，等于滑动数组的宽度 $2m + 1$ 。

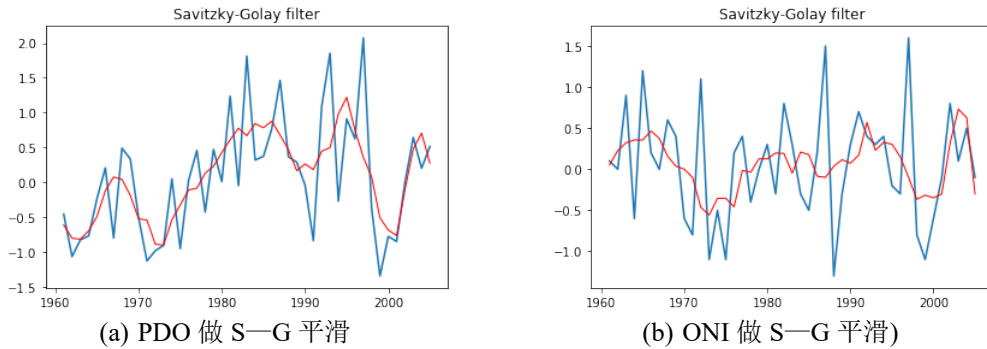


图 7 S-G 滤波拟合结果

并选取 Gamma 分布的两个参数和 SDII 作为判断极端降水情况的指标。

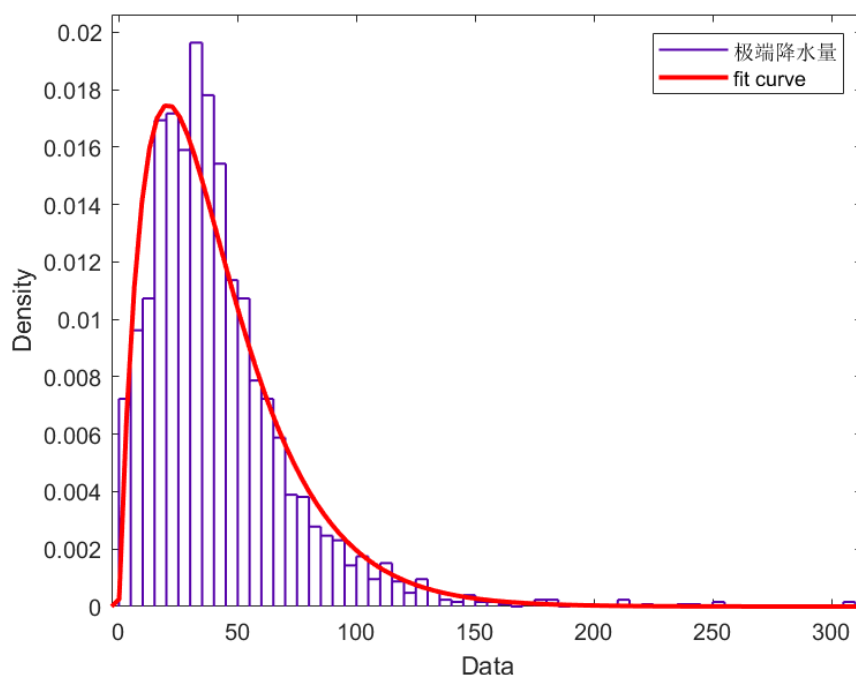


图 8 第一年极端降水量分布与 Gamma 曲线拟合图

5.3 模型建立

5.3.1 Pearson 相关性分析

Pearson 相关性分析是反映两个变量的类型与程度，通过求出相关系数能够精确的反应变量之间的相关程度。

Pearson 相关性分析的计算

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(Y - \bar{X})^2 \sum(Y - \bar{Y})^2}} = \frac{l_{XY}}{\sqrt{l_{XX}l_{YY}}} \quad (11)$$

其中 X 距离均差的平方和： $l_{XX} = \sum(X - \bar{X})^2$

X 距离均差的平方和： $l_{YY} = \sum(Y - \bar{Y})^2$

X 与 Y 间的离均差积和： $l_{XY} = \sum(X - \bar{X})(Y - \bar{Y})$

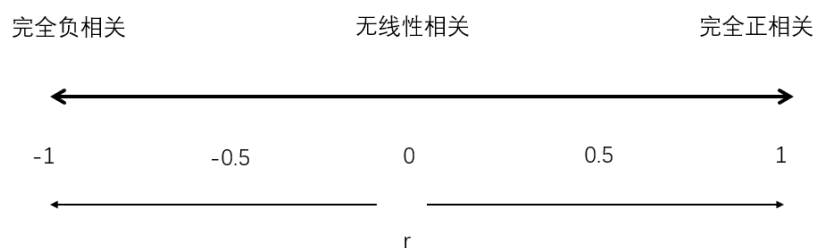


图 9 相关性示意图

5.4 模型求解

表 4 相关性系数

		CO2	PDO	ONI	海陆温度指数	太阳黑子相对活动指数
SDII	相关系数	0.242	0.23	0.125	-0.262	-0.013
	显著性	0.019	0.026	0.225	0.013	0.899
α	相关系数	-0.259	-0.061	-0.016	0.238	-0.233
	显著性	0.012	0.557	0.876	0.024	0.024
β	相关系数	0.222	0.008	-0.032	-0.194	0.128
	显著性	0.031	0.938	0.754	0.065	0.214

5.5 结果分析

根据表 3 的结果显示，二氧化碳、PDO、以及海陆温度指数都对极端降水强度具有显著的影响，其中可以观察到二氧化碳作为温室气体的指标，与极端降水强度（SDII）具有正相关的影响，同时太平洋年际震荡指数和太平洋厄尔尼诺系数都对极端降水有着促进作用。结合历史上大暴雨与厄尔尼诺现象的相关性，可以认为结论基本符合实际情况。

α 和 β 两个参数对 Gamma 分布的形状有着明显的影响，并且根据观察可以发现，随着 α 和 β 的增加，极端降水强度的分布趋于平缓，这意味着极端降水强度的区间变广，在一定程度上说明强极端降水时间的增加。根据相关性分析结果可以看出，海陆温度指数、太阳黑子相对活动指数都对其形状有着显著的影响，虽然 β 参数受到海陆温度指数负相关，但总体上趋势符合实际情况。

六、问题三模型建立与求解

6.1 问题分析

查阅资料后发现，气候动力学模式在区域气候的模拟上存在着去多系统偏差，例如，绝大部分气候模式对我国降水模拟存在着湿偏差或者干偏差，其中东部夏季极端降水普遍存在低估，西部则存在高估；此外，模式还普遍存在着太多低强度降水日以及不正确的季节变化。

而统计偏差订正方法是进一步减少模式偏差的一个重要手段。近年来出现了许多偏差订正的方法例如简单的对数据进行放缩或者线性变换、方差变换等，其次是基于经验分布函数的分位点投影变换，以及考虑空间一致性的基于 EOF、SVM 等的订正方法。

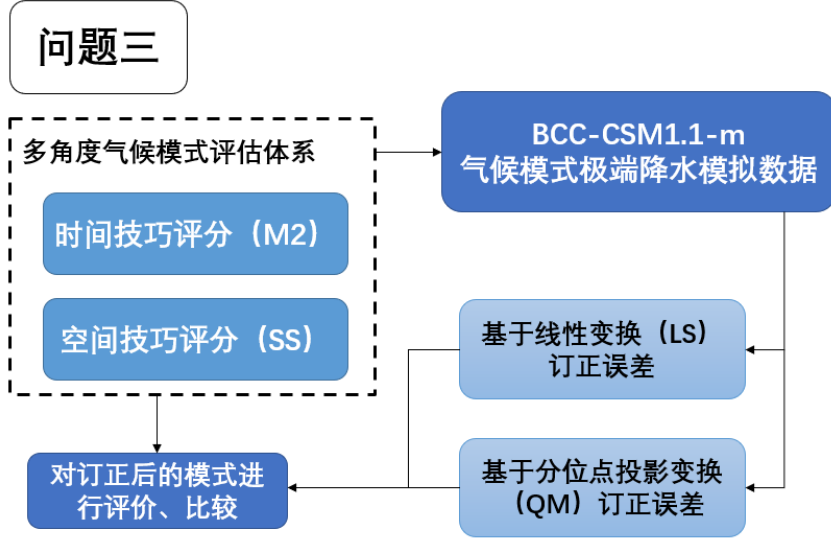


图 10 问题三流程图

6.2 模型的建立

6.2.1 线性变化模型

线性变换 (LS) 主要基于观察降水和模式模拟之间的差异对模式资料进行订正, 即将差异作为订正因子乘到模式数据上, 从而使得订正后的模式资料能够更好的与观测资料均值温和。同时假定所获得的订正因子即时在未来气候条件下也是不变的。具体做法如下 (Lenderink, 2007):

$$P_{mod,ctr}^*(d) = P_{mod,ctr}(d) \cdot \Delta \quad (12)$$

$$P_{mod,cor}^*(d) = P_{mod,cor}(d) \cdot \Delta \quad (13)$$

$$\Delta = \frac{\mu(P_{obs,ctr}(d))}{\mu(P_{mod,ctr}(d))} \quad (14)$$

其中 $P_{obs,ctr}(d)$ 、 $P_{mod,ctr}(d)$ 为建模时期观测和模式逐日降水数据, $P_{mod,cor}(d)$ 则对应独立验证时期的数据; Δ 为订正因子, 是观测与模式在建模时期夏季降水均值的比值, 反应两者气候均值特征的差异。 $P_{mod,ctr}^*(d)$ 、 $P_{mod,cor}^*(d)$ 则分别为乘以订正因子 Δ 后建模时期和验证时期模式订正逐日降水。

6.2.2 分位点投影变换模型

基于累积经验分布函数的分位点投影变换 (QM) 方法是由 Michelangei 等 (2009) 为考虑江水的分布特征提出的模型, 其认为观测 P_0 和模式模拟降水 P_m 两者分布函数存在以下对应关系:

$$F_0(P_0) = F_m(P_m) \quad (15)$$

依据公式 (15), 通过反解上述方程得到观测降水的一个转换函数:

$$P_0 = F_0^{-1}(F_m(P_m)) \quad (16)$$

QM 认为对于模式降水的未来模拟 (14) 关系依然成立, 应用其可得模式偏差订正降水值。

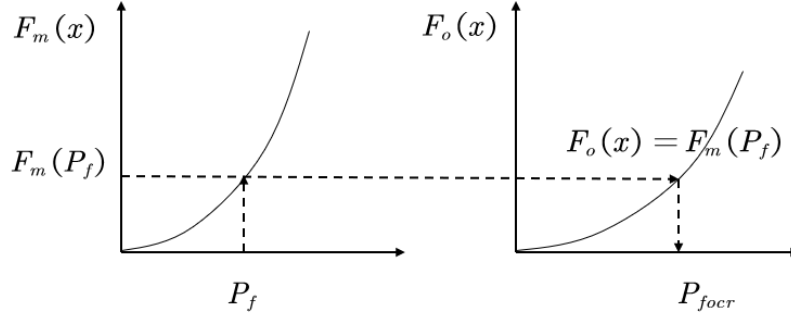


图 11 分位点投影变换 QM 流程

($F_0(x)$ 和 $F_m(x)$ 分别为观测或模拟降水低于或等于 x 的概率, P_0, P_f 和 P_{fcor} 分别为观测、模式模拟和模式偏差订正后降水)

图 10 给出了 QM 偏差订正流程示意图, 对于某个站点, 为了获取模式模拟降水 P_f 的偏差订正降水 P_{fcor} , 首先计算得到建模时期观测和模式模拟的分布 $F_0(x)$ 和 $F_m(x)$; 其次, 带入 P_f 和 $F_m(x)$ 可得对应的分布函数值 $F_m(P_f)$; 最后依据公式 (15), 通过反向求解, 可得符合观测分布 $F_0(x)$ 的偏差订正降水 P_{fcor} 。应用 QM 偏差订正流程, 不难得到模式建模期和验证期偏差订正降水值为:

$$P_{mod.ctr}^*(d) = F_0^{-1}(F_m(P_{mod.ctr}(d))) \quad (17)$$

$$P_{mod.cor}^*(d) = F_0^{-1}(F_m(P_{mod.cor}(d))) \quad (18)$$

6.3 模型的求解

通过对预测数据的线性变化以及分位点投影变化, 得到以下结果

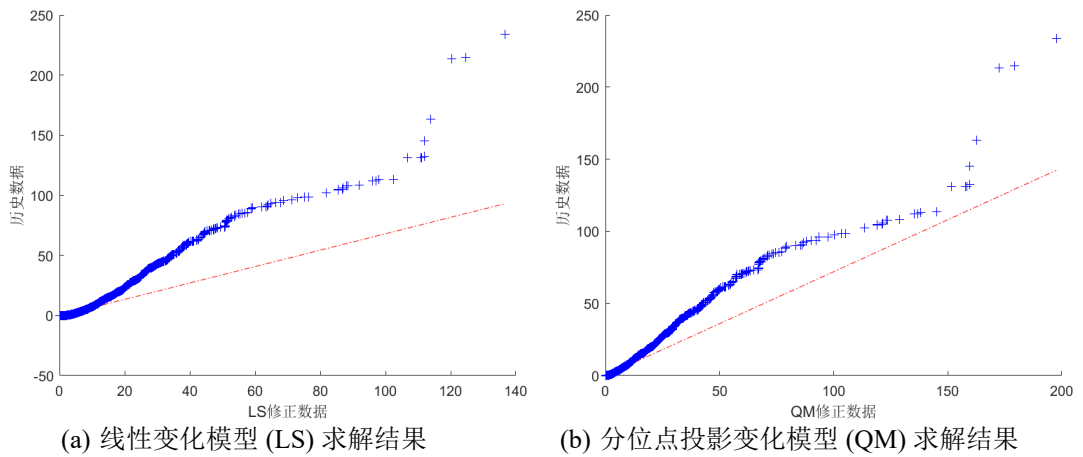


图 12 结果呈现

图 12 的 Q-Q 图显示经过 QM 订正之后的数据比经过 LS 订正的数据更接近历史真实数据的分布，表明 QM 订正效果的显著，而 LS 订正效果微弱。

6.4 结果分析

6.4.1 时空检验指标

1) 空间技巧评分 (SS)

引入 Picerce 等 (2007) 提出的综合考虑模拟场与观测长相关系数以及平均偏差的指数：空间技巧评分 (spatial skill score, 简称 SS)

$$SS = 1 - \frac{MSE(m, o)}{MSE(\bar{o}, o)} = r_{m,o}^2 - [r_{m,o} - \frac{s_m}{s_o}]^2 - [\frac{(\bar{m} - \bar{o})}{s_o}]^2 \quad (19)$$

其中

$$MSE(m, o) = \frac{1}{N} \sum_{k=1}^N (M_k - O_k)^2 \quad (20)$$

考虑上式的无量纲化，得到无量纲空间技巧评分 SS。式中 M_k 和 O_k 分别为各模式极端指数的模拟场和观测场， N 为空格点数。式中 $MSE(\bar{o}, o)$ 表示观测场方差。进一步展开 SS 的定义，其中首项 $r_{m,o}^2$ 为两个场的空间相关系数。第二项、第三项中的 s_m 和 s_o 分别为模拟与观测的空间场均方差，它们主要反映模拟场的偏差。可见，SS 值综合考虑了空间场相关系数和模拟场偏差，模拟与观测场的相关系数越大，模拟场偏差越小，SS 值越接近 1. 即模式对该极端指数场的模拟能力越好。

2) 时间技巧评分 (M2)

M2 指数用于评价气候场上每一个点上的时间序列相对于该点上观测时间序列的年际变率的插值。主要以标准差来衡量，公式如下

$$M2 = (\frac{STD_m}{STD_o} - \frac{STD_o}{STD_m})^2 \quad (21)$$

其中 STD 表示的是模式和观测变量的标准差。当模拟年际变化率与观测变率差相等时，M2 为 0，表明模拟效果最好。而 M2 越大说明模拟变率与观测变率差的越多，模拟效果越差。

表 5 误差订正效果

	模式评估指标	计算结果	提升百分率
BCC-CSM1.1-m	SS	-3.0456	
	M2	1.2722	
分位投影变换订正 (QM)	SS	-2.607	14.40%
	M2	0.2816	77.82%
线性变换订正 (LS)	SS	-2.607	14.40%
	M2	0.4575	64.04%

根据订正效果来看分为投影订正的效果更好，对原数据的提升更高。

七、问题四的求解与分析

7.1 问题分析

利用问题三订正后的数据求出历史时期（1961-2005）、21 世纪中期（2036-2065 年）、末期（2071-2100）在概率和数量上的分布特征。并通过数据对比得出夏季极端降水的变化特征。根据变化趋势给出相应的气候变、化应对、防灾减灾措施制定方面的政策建议

7.2 问题四求解

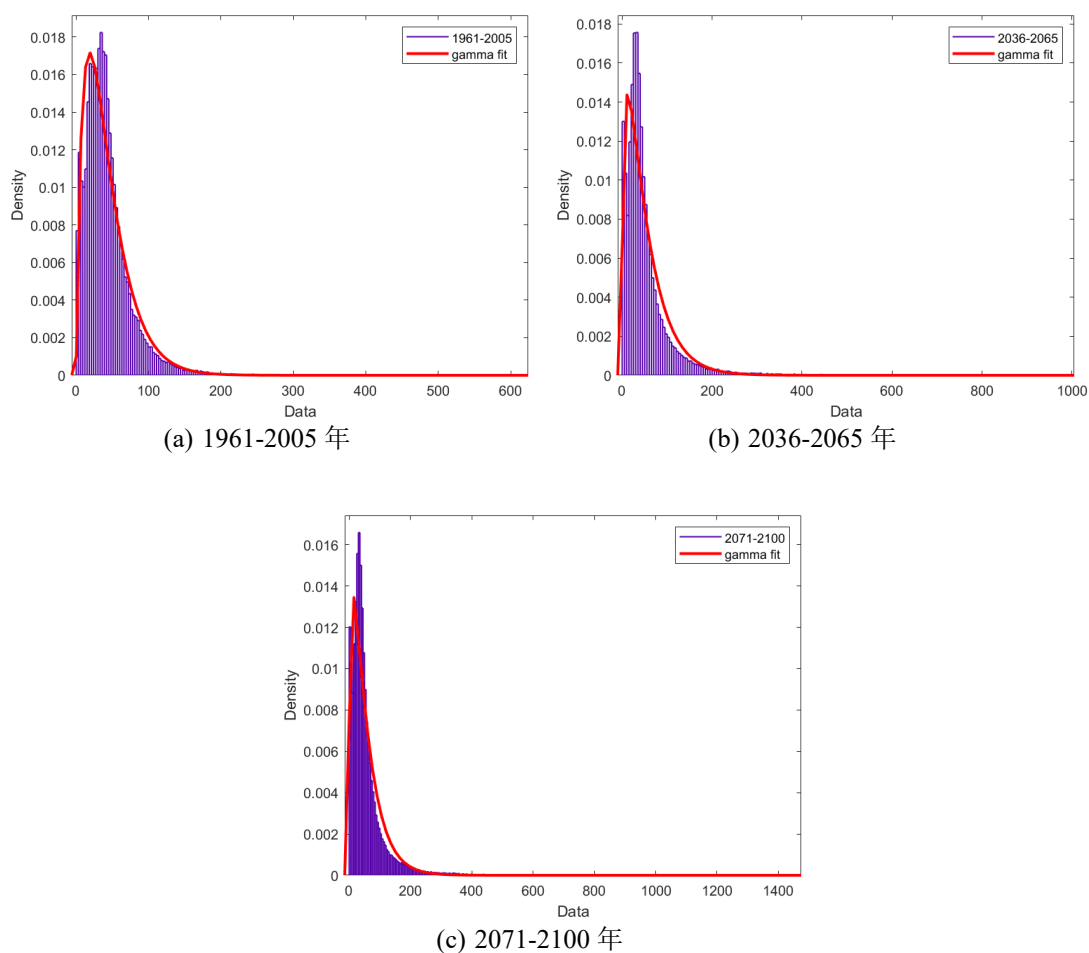


图 13 利用 Gamma 函数拟合降水分布

表 6 预估降水数据概率分布对比

	1961-2005 年	2036-2065 年	2071-2100 年
α	1.7837	1.2809	1.2483
β	23.6923	40.7751	44.5318
降雨量平均值	42.2603	52.2303	55.5883
降雨量 95 分位点	103.9799	143.5315	154.1209
降雨量 75 分位点	57.3155	72.0171	76.7159
降雨量 50 分位点	34.6786	39.4409	41.648
降雨量 25 分位点	19.061	18.7849	19.5806

表 7 预估降水数据数量分布对比

	1961-2005 年	2036-2065 年	2071-2100 年
极端降水总天数	115340	76867	76788
553 站点年均值	2563	2562	2560

7.3 结果分析

图表中的数据显示，从 1961-2005 年到中世纪到后世纪，平均值与分位点数据逐年上升，但夏季极端降水的天数在数量上波动不大，每年极端降水发生的数量保持平稳。而对极端降水情况概率分布的分析可知，高概率密度对应的降雨强度值逐渐向高强度方向变动。极端降水的最大强度不断建立新的最高点。

综上表明我国夏季极端降水情况，有着向强度更大，范围更广的情况发展。

7.3.1 给相关部门的政策与建议

通过对三个问题的解答，我们发现夏季极端降雨的情况有很明显的空间和时间分布特征，其中东南地区受到夏季极端降雨天气的影响要远远大于西北地区；极端降水量在分布法与 Gamma 曲线类似。同时夏季极端降雨量受到多方面环境的影响，其中全球气温变暖，和海陆温度指数对极端降水的影响较大。在对未来极端降雨趋势预估之后，可以明显的看出，极端降水情况的频率和强度预期都会增加。

查阅中国天气网，极端降水天气事件的发生呈现出明显增加的趋势，同时极端降水事件发生的区域范围也有增大的趋势。我国每年因极端天气影响损失的财产较大，因此即时的预测极端天气的时间，以及即时的做有针对性的应对措施十分有必要。首先要做好对建筑结构，河道排水装置全面清查，确保建筑具有较好的稳固性，和排水装置的有效性。其次即时发布天气预警，确保市民在极端天气的人身安全。最后在极端天气时停止户外施工作业，密切监视建筑和路面情况，对山体滑坡和建筑坍塌等重大险情即时报告。在极端天气引起的洪灾下我们希望有关部门能够加强相关防洪的基础设施建设以及完善紧急救援流程。如，在近期在湖北省发生的特大洪水灾害，若能提前预估今年的极端降水情况针对地区实际地形做好备案措施，包括对水道的疏通，重灾地区居民的提前转移。

八、模型总结与评价

8.1 模型优点

1. 使用了经验正交函数分解法，基于实际地理位置分析了极端降水的时空特征，结果全面且与实际情况相符。
2. Manner-Kendall 趋势检验不依赖原始数据的分布，能够适应极端降水特征偏态的性质。
3. 误差订正效果评估采取了多种指标，分别考虑了时空特点进行分析。

8.2 模型缺点

1. 在相关性分析时未能全面的考虑数据的周期性特点，虽然 S-G 滤波保证了分析结果的可靠性，但是仍有一部分数据信息被忽略。

8.3 模型改进

本文的极端降水阈值采用了传统的第 95 百分位数进行确定，而实际上这种方法确定的阈值存在离散程度较大等问题^[5]，若时间允许，可以尝试使用趋势波动分析法（Detrend Fluctuation Analysis, DFA）与替代数据法相结合的随机重排去趋势波动分析确定阈值，使得后续分析结果可信度进一步提升。

参考文献

- [1] 程诗悦, 秦伟, 郭乾坤, 徐立荣. 近 50 年我国极端降水时空变化特征综述 [J]. 中国水土保持科学, 2019, 17(03): 155-161.
- [2] 管习权, 林娜, 王雨. 基于 10 年 GPCP 降水资料的全球极端降水分布分析 [J]. 气象研究与应用, 2013, 34(01): 13-16+109.
- [3] 孟清, 高翔, 白红英, 张扬, 王辉源. 1960-2015 年秦岭地区极端降水的时空变化特征 [J]. 水土保持研究, 2019, 26(06): 171-178+183.
- [4] 袁瑞强, 王亚楠, 王鹏, 王仕琴, 陈宇宏. 降水集中度的变化特征及影响因素分析——以山西为例 [J]. 气候变化研究进展, 2018, 14(01): 11-20.
- [5] 侯威, 钱忠华, 杨萍, 封国林. 基于随机重排去趋势波动分析的全国极端日降水事件综合指标 [J]. 干旱气象, 2012, 30(02): 149-157.
- [6] Changes of daily climate extremes in southwestern China during 1961–2008 [J]. Li Zongxing, Yuanqing He, Puyu Wang, Wilfred H. Theakstone, Wenling An, Xufeng Wang, Aigang Lu, Wei Zhang, Weihong Cao. Global and Planetary Change . 2011
- [7] 朱连华. 中国地区极端降水的统计建模及其未来概率预估 [D]. 南京信息工程大学, 2017.

附录 A 代码

A.1 (EOF) matlab 源程序

```
function [V,PC,D,G]=eof_test(X)

%V eigen vector 空间函数(), 必须是的, 为变量个数VpXpp

%PC 时间函数 Y=V'X

%X 原始变量, 列对应原始变量类型, 行对应样本原始变量的值都需要标准化处理, 即减去均值, 除以 std

%size(X)
[V,D]=eig(X*X'); % 特征向量V 对角阵, 即主成分D
[diagonal,I] = sort(diag(D),'descend');
V=V(:,I);
PC=(V'*X)';
G = diagonal/sum(diagonal);

end
```

A.2 problem1.m matlab 源程序

```
% 频率就是极端降雨天数, 强度可以采用之前的强度, 持续性就是现在求的这个
clear;
[data,txt] = xlsread("../2_obs_sta_jja_pre_553_1961_2005.csv");
dating = data(:,1:3);
data = data(:,4:end);
%持续性%
%% total_rain 总降水量(>1才计数mm)
for i = 1:45
    for j = 1:553
        temp_data = data(1+92*(i-1):92*i,j);
        total_raining(i,j) = sum(temp_data(find(temp_data>1)));
        %total_rain(i,j)=sum(data(find(data(1+92*(i-1):92*i,:)>1)));
    end
end

%% CDD
for i = 1:45
    for j = 1:553
        temp_data = data(1+92*(i-1):92*i,j);
        maxcnt = 0;
        cnt = 0;
        flag = 0;
        for k = 1:92
            if temp_data(k) < 1
                if flag == 0
                    flag = 1;
```

```

        cnt = cnt + 1;
    else
        cnt = cnt + 1;
    end
    if cnt > maxcnt
        maxcnt = cnt;
    end
else
    if flag == 1
        flag = 0;
        cnt = 0;
    end
end
end
cdd(i,j)=maxcnt;
end
end

%% CWD
for i = 1:45
    for j = 1:553
        temp_data = data(1+92*(i-1):92*i,j);
        maxcnt = 0;
        cnt = 0;
        flag = 0;
        for k = 1:92
            if temp_data(k) >= 1
                if flag == 0
                    flag = 1;
                    cnt = cnt + 1;
                else
                    cnt = cnt + 1;
                end
                if cnt > maxcnt
                    maxcnt = cnt;
                end
            else
                if flag == 1
                    flag = 0;
                    cnt = 0;
                end
            end
        end
        cwd(i,j)=maxcnt;
    end
end

%强度%
%%Rx1
for i = 1:45
    for j = 1:553

```



```

    temp_data = data(1+92*(i-1):92*i,j);
    rx1(i,j) = max(temp_data);
    %total_rain(i,j)=sum(data(find(data(1+92*(i-1):92*i,:)>1)));
end
end

%%Rx5
for i = 1:45
    for j = 1:553
        temp_data = data(1+92*(i-1):92*i,j);
        for k = 1:92-4
            tempp_data(k) = sum(temp_data(k:k+4));
        end
        rx5(i,j) = max(tempp_data);
        %total_raining(i,j) = sum(temp_data(find(temp_data>1)));
        %total_rain(i,j)=sum(data(find(data(1+92*(i-1):92*i,:)>1)));
    end
end

%频率%
%%
NumOfDay = 92;
NumOfYear = 45;
NumOfSite = 553;
jiduan_days = zeros(NumOfYear, NumOfSite);
for i = 1:NumOfSite
    [val,valord] = sort(data(:,i));
    threshold(i) = val(round(NumOfYear*NumOfDay*0.95));
    for j = 1:NumOfYear
        tempdata = data(1+NumOfDay*(j-1):NumOfDay*j,i);
        jiduan_days(j,i) = length(tempdata(find(tempdata >= threshold(i))));
    end
end
jiduan_days = fillmissing(jiduan_days,'constant',0);

```

A.3 (problem2.m 源程序

```

clear;
[data, txt] = xlsread("2_obs_sta_jja_pre_553_1961_2005.csv");
NumOfDay = 92;
NumOfYear = 45;
NumOfSite = 553;
data = data(:,4:end);
for i = 1:NumOfSite
    [val,valord] = sort(data(:,i));
    threshold(i) = val(round(NumOfYear*NumOfDay*0.95));
    for j = 1:NumOfYear
        tempdata = data(1+NumOfDay*(j-1):NumOfDay*j,i);
        overval{j,i} = tempdata(find(tempdata >= threshold(i)));
    end
end

```

```

end
for i = 1:45
    for j = 2:553
        overval{i, 1} = [overval{i,1};overval{i,j}];
    end
end
for i = 1:45
    overvals{i,1} = overval{i, 1};
end
for i = 1:44
    overvals{1,1}= [overvals{1,1};overvals{i+1,1}];
end
phat = gamfit(overvals{1,1});
%overval = fillmissing(overval,'constant',0);
%save intens

```

A.4 North 检验 –matlab 源程序

```

clc
D=sort(diag(D1),'descend');
e1=D(1)*sqrt(2/length(D));
d1=D(1)-D(2);
fprintf('湿日总降水量的第一空间模态("检验north,%g>%g\n",d1,e1);

D=sort(diag(D2),'descend');
e2=D(1)*sqrt(2/length(D));
d2=D(1)-D(2);
fprintf('极端降水期日均降水期的第一空间模态("检验north,%g>%g\n",d2,e2);

D=sort(diag(D3),'descend');
e3=D(1)*sqrt(2/length(D));
d3=D(1)-D(2);
fprintf('的第一空间模态检验cddnorth,%g>%g\n",d3,e3);

D=sort(diag(D4),'descend');
e4=D(1)*sqrt(2/length(D));
d4=D(1)-D(2);
fprintf('的第一空间模态检验cwndnorth,%g>%g\n",d4,e4);

D=sort(diag(D5),'descend');
e5=D(1)*sqrt(2/length(D));
d5=D(1)-D(2);
fprintf('的第一空间模态检验rx1north,%g>%g\n",d5,e5);

D=sort(diag(D6),'descend');
e6=D(1)*sqrt(2/length(D));
d6=D(1)-D(2);
fprintf('的第一空间模态检验rx5north,%g>%g\n",d6,e6);

```

A.5 QM 订正 –matlab 源程序

```
%% 订正qm
%load data_fore;
%data_fore = data_fore - 0.5;
for i = 1:553
    newfit_fore(:,i) =
        gaminv(gamcdf(data_fore(:,i),fit_gamcoef(i,1),fit_gamcoef(i,2)),raw_gamcoef(i,1),ra
end
newfit_fore(find(newfit_fore<0))=0;
```

A.6 SS,M2–matlab 源程序

```
%load newfit_data
load data_raw
% data_fit_norm = (newfit_data -
    min([min(newfit_data);min(data_raw)])) ./ (repmat(max([max(newfit_data);max(data_raw)])-mi
% data_raw_norm = (data_raw -
    min([min(newfit_data);min(data_raw)])) ./ (repmat(max([max(newfit_data);max(data_raw)])-mi
% mse_m_o1 = sum((data_raw_norm-data_fit_norm).^2,2)/553;
% mse_o_o1 = var(data_fit_norm, 0, 2);
% SS1 = 1-mse_m_o1./mse_o_o1;
% plot(SS1);
%set(gca,'YLim',[0 1])

%% M2
% M21=(std(newfit_data)./std(data_raw)-std(data_raw)./std(newfit_data)).^2;
% plot(M21);
%set(gca,'YLim',[0 1])

%% LS method evaluation
load ls_data
data_fit_norm = (ls_data -
    min([min(ls_data);min(data_raw)])) ./ (repmat(max([max(ls_data);max(data_raw)])-min([min(1
data_raw_norm = (data_raw -
    min([min(ls_data);min(data_raw)])) ./ (repmat(max([max(ls_data);max(data_raw)])-min([min(1
mse_m_o2 = sum((data_raw_norm-data_fit_norm).^2,2)/553;
mse_o_o2 = var(data_fit_norm, 0, 2);
SS2 = 1-mse_m_o1./mse_o_o1;

%% M2
M22=(std(ls_data)./std(data_raw)-std(data_raw)./std(ls_data)).^2;
%plot(M22);
```