



用外部强迫因子对近百年陆地降水变化的统计建模试验

李庆祥^{1,2}

1. 中山大学大气科学学院, 广东省气候变化与自然灾害重点实验室, 广州 510275;

2. 南方海洋科学与工程广东省实验室(珠海), 珠海 519082

E-mail: liqingx5@mail.sysu.edu.cn

2020-03-21 收稿, 2020-05-06 修回, 2020-05-08 接受, 2020-05-12 网络版发表

国家重点研发计划(2018YFC1507705, 2017YFC1502301)和国家自然科学基金(41975105)资助

摘要 利用多元线性回归模型拟合历史气候变化经常会遇到残差自相关以及影响因子之间存在多重相关性的问题, 从而导致模型出现病态. 本研究利用多个自然强迫、人为活动因子的辐射强迫(有效辐射强迫)作为自变量, 采用较为简单的组合统计模型将近百年陆地降水变化分解为自然强迫、人为活动的影响和内部变率(噪声), 取得了较为合理的结论. 首先采用多元线性回归(MLR)+差分滑动平均自回归(ARIMA)组合模型对全球及各纬度带陆地降水距平变化序列中人类活动和自然强迫因子的重要性进行了分析, 发现全球和北半球中、高纬度陆地降水的模型解释方差较高(超过40%), 且方程中人类活动因子的拟合系数具有高显著性, 说明人类活动对这些区域降水变化具有显著贡献. 进一步, 利用偏最小二乘回归(PLSR)模型, 量化了人为强迫多个分量对上述降水变化的各自影响(贡献), 结果表明: 火山爆发和多数人为因子对全球陆地降水的距平变化非常重要; 只有人为气溶胶、尾迹卷云对于北半球中纬度、高纬度陆地降水距平变化一致性地表现为正贡献, 而其他人为因子对不同纬度带的贡献是相反的, 体现出一定的不确定性.

关键词 陆地降水, 数据集, 统计建模, 人为强迫, 自然强迫

气候变化是外强迫(信号)和内部气候变率(噪声)共同作用的结果. 气候系统的外强迫变化包括自然强迫(太阳辐射、火山等)和人类活动外强迫(温室气体、气溶胶排放、土地利用等)的变化. 因此, 器测资料及气候系统模式模拟所揭示的气候变化也应该包括上述3个组成部分, 但在气候变化序列中分离上述三部分气候变化却一直是现代气候学研究的核心议题与难点之一.

20世纪后期以来, 基于模式与观测对比的气候变化归因研究为气候变化成因研究提供了重要支撑. 其代表性的方法是利用基于最优指纹的统计推理法, 即先利用最优指纹法估计气候变化“信号”对外部强迫的响应系数, 然后对响应系数做推理检验, 在一定置信度

水平上确认气候变化的原因^[1,2]. “最优指纹法(optimal finger print, OFP)”^[3,4]是一种量化鉴别气候变化信号并作归因分析的方法, 它可以用广义多元回归来实现, 即把观测的气候变化看作是外部强迫引起的气候变化信号的线性叠加, 再加上气候系统内部变率. 利用模式模拟出的气候变化, 对多初值成员做集合平均来消除内部变率的影响来估计气候变化信号及其响应系数; 在一定的置信度水平上确认气候变化的原因, 需要对最优指纹法估计得到的气候“信号”响应系数进行推理检验. 一般多采用经典的概率论方法(标准法)进行假设检验^[5,6]. 随着模式的发展, 其复杂程度也不断加大, 目前这种归因也已延伸到区域尺度, 对气候变化的影响

引用格式: 李庆祥. 用外部强迫因子对近百年陆地降水变化的统计建模试验. 科学通报, 2020, 65: 2266–2278

Li Q X. Statistical modeling experiment of land precipitation variations since the start of the 20th century with external forcing factors (in Chinese). Chin Sci Bull, 2020, 65: 2266–2278, doi: 10.1360/TB-2020-0305

和极端事件的归因研究也取得了较大的进展^[7,8]。

显然,气候模式的模拟能力就成为了气候变化归因效果的重要基础。一般认为,气候系统模式在其物理基础、对过去和当前气候的模拟效果方面均取得了长足的进步,这也提高了指纹归因的信心和可靠性。但模式在区域性问题上至今还存在较大误差;即使在大尺度范围有时也存在重大的问题,比如对一些重要的气候现象(模态),如Atlantic Multi-decadal Oscillation (AMO)、Quasi-Biennial Oscillation(QBO)等还存在无法准确模拟的问题;特别是由于云反馈导致的不确定性问题还存在较大的不确定性。也有研究指出,这种归因方法由于没有使用气候模式提供的物理信息来约束对强迫的估计响应,因而无法处理建模的不确定性^[9]。

随着气候变化研究的不断深入,IPCC(Intergovernmental Panel on Climate Change)系统地对影响气候变化的所有外部因子的辐射强迫(有效辐射强迫)进行了评估,使得对气候变化影响因子的贡献进行定量化研究提供了非常便利的条件^[10]。近年来也出现了许多将气候变化直接进行分解的研究,其主要思路就是将观测或模式模拟气候变化序列进行外部强迫和内部变率部分的统计分解^[11,12]。从统计上讲,如果某种影响因子(非气候反馈)的辐射强迫(有效辐射强迫)都是准确定量的,它总能在气候变化观测中得到一定程度的体现,那么如何分离这些因子的具体贡献,就成了关键的科学问题。

我们发现,在已知过去气候变化特征和量化的各种辐射强迫(有效辐射强迫)下,利用正确设置的统计模型,也可以在某种程度上对各种强迫因子的各自和总体贡献进行分离,并得出其在不同统计显著性水平上的最优估计。本研究选取目前检测归因研究认为较为困难的降水要素^[13~17],开展了系统的统计建模和归因分析,得出了一些较为令人鼓舞的结果。

1 数据及分析方法

1.1 数据

本研究使用的降水资料是我们研发的全球逐月降水数据集(global precipitation dataset, GPD),该数据集经过了系统的质量控制,对部分不连续的站点观测降水序列进行了均一化。经对比研究,该数据集在全球和区域性降水气候变化趋势方面和美国国家环境信息中心、德国全球降水中心的降水数据集较为一致,具有

较好的准确性^[18]。该数据集时间覆盖范围为1900年1月~2013年12月,2013年后,该数据集尚没有进行系统性升级更新,因此本文只研究了上述时间段范围。

对自然和人为的辐射强迫进行评估是IPCC第五次评估报告(AR5)的一项重要工作^[19]。此次报告中在瞬时辐射强迫和平流层调整的辐射强迫外,新引入了有效辐射强迫的概念,使得对于此前的报告中辐射强迫的应用更加全面。同时,AR5还公开了各种自然因子、人为因子的辐射强迫(有效辐射强迫)数据(图1),为气候变化检测归因等研究提供了基础数据。本研究采用AR5公布的全球平均的逐年各类自然因子(主要包括太阳辐射和火山等)和人为因子(包括二氧化碳(CO₂)、其他温室气体、人为气溶胶、臭氧(平流层、对流层)、平流层水汽、地表反照率变化(土地利用、冰雪覆盖)和尾迹卷云等)的辐射强迫(有效辐射强迫)数据,为与降水资料年份一致,这些辐射强迫因子数据的时间覆盖亦截取1900~2013年逐年数据。

1.2 研究方法

1.2.1 多元线性回归及差分自相关移动组合模型

多元线性回归(multiple linear regression, MLR)是指一个因变量(预报对象),多个自变量(预报因子)的回归模型,只要假定自变量和因变量之间的关系是线性的,则可以用多元线性回归拟合。一般将气候变化归因为自然强迫、人为强迫和气候系统内部变率。IPCC历次评估报告认为,人类活动(人为强迫,用全球CO₂含量导致的辐射强迫代表(其他温室气体折算成CO₂辐射强迫当量))越来越肯定地被证明是当前气候变暖的主要贡献者。另外,外部自然强迫(如火山、太阳常数变化,以相应的辐射强迫代表)等也对气候变化产生一定的贡献^[19,20],而这种贡献一般也可以用线性关系进行表达。但我们知道,人为强迫中各种分量彼此之间存在多重相关性,也就是说它们不完全是相互独立的,如果直接进行多元回归分析,必然使得方程产生病态^[21,22]。因此,为避免这种多重相关性带来的影响,本研究首先以全球及不同纬度带陆地平均降水序列作为因变量 y ,用外部自然影响 x_1 、人类活动的辐射强迫作为自变量 x_2 ,采用多元线性拟合进行回归拟合:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon, \quad (1)$$

$\beta_0, \beta_1, \beta_2$ 为回归系数,用最小二乘法估计得到,但需要对模型参数进行统计检验, ε 为拟合后的残差;另外,由

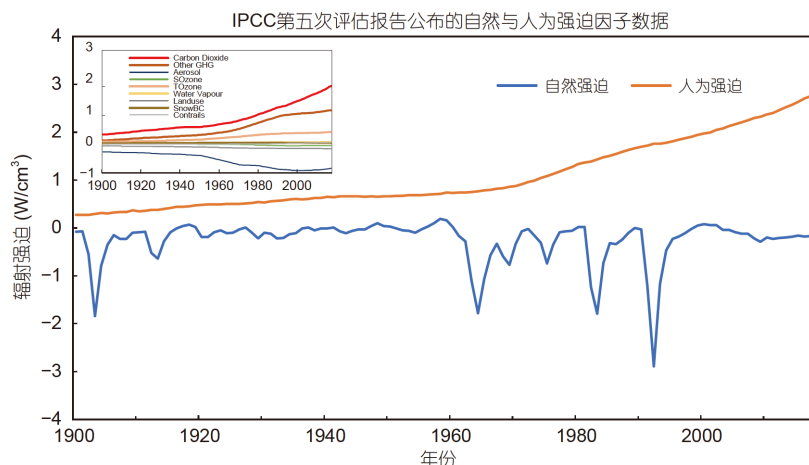


图1 自然与人为辐射强迫数据. 插图显示9种人为强迫

Figure 1 The natural and anthropogenic radiative forcings data. The inset shows the 9 types of anthropogenic radiative forcings

于上述两种自变量都是线性项, 因此自然变率部分必然包括在残差之中, 同样, 残差 ε 对于评估多元线性回归模型的可靠性也非常重要. 因此, 进一步对残差序列进行分析也非常必要.

残差分析表明, 自然强迫和人为强迫, 二者无共线性现象(其方差膨胀系数(variance inflation factor, VIF)接近1). 因此, 本研究只考察了另一个残差是否存在自相关现象的指标, 即德宾-沃森(Durbin-Watson, DW)数. DW数过大或过小时, 表明残差存在明显的负或正的自相关, 这就导致线性模型不再是最优无偏估计(best linear unbiased estimator, BLUE), 变量的显著性统计失去意义, 即其模型的模拟(预测)是失效. 这时候需要对残差进行进一步差分移动自回归(autoregressive integrated moving average model, ARIMA)建模^[23], 即建立混合模型, 以矫正各个自变量的边际影响水平, 其原理与步骤如下.

将 ∇ 记为差分算子, 有

$$\nabla^2 y_t = \nabla(y_t - y_{t-1}) = y_t - 2y_{t-1} + y_{t-2}. \quad (2)$$

对于延迟算子 B , 有

$$y_{t-p} = B^p y_t, \forall p \geq 1, \quad (3)$$

因此可以得出

$$\nabla^k = (1 - B)^k. \quad (4)$$

对于非平稳时间序列, 在消去其局部水平或者趋势之后, 其显示出一定的同质性, 也就是说, 此时序列的某些部分与其他部分很相似. 这种非平稳时间序列

经过差分处理后可以转换为平稳时间序列, 那么称这样的时间序列为齐次非平稳时间序列, 其中差分的次数 d 就是齐次的阶.

设有 d 阶齐次非平稳时间序列 y_t , 那么有 $\nabla^d y_t$ 是平稳时间序列, 则可以设其为ARMA(p, q)模型, 即

$$\lambda(B)(\nabla^d y_t) = \theta(B)\varepsilon_t. \quad (5)$$

其中,

$$\lambda(B) = 1 - \lambda_1 B - \lambda_2 B^2 - \dots - \lambda_p B^p,$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q,$$

分别为自回归系数多项式和滑动平均系数多项式. ε_t 为零均值白噪声序列. 可以称所设模型为自回归求和滑动平均模型, 记为ARIMA(p, d, q).

当差分阶数 d 为0时, ARIMA模型就等同于ARMA模型, 即这两种模型的差别就是差分阶数 d 是否等于零, 也就是序列是否平稳, ARIMA模型对应着非平稳时间序列, ARMA模型对应着平稳时间序列.

1.2.2 偏最小二乘回归模型

上述模型可以分离出自然和人为强迫分别对于气候变化的影响(贡献), 但我们仍然需要分离各种辐射强迫本身的单独贡献, 这时候可以采用一种新的偏最小二乘回归(partial least squares regression, PLSR)模型. PLSR是一种新型的多元统计数据分析方法, 它于1983年由Wold等人^[24]首次提出, 用以解决化学样本分析中存在的变量多重相关, 以及解释变量多于样本点等实际问题. 由于PLSR能解决以往普通多元回归方法

所无法解决的一些问题,因而得到了研究人员的重视.近30年来,偏最小二乘回归在理论和方法上都得到迅速的发展,而其应用领域也从最初的化学领域快速扩展到更多的自然科学和社会科学领域^[25].

偏最小二乘回归方法提供了一种多因变量对多自变量的回归建模方法,当变量之间存在高度相关性时,所建立的模型,其分析结论更可靠,整体性更强.适合在样本容量小于变量个数的情况下进行回归建模.可以实现多种多元统计分析方法综合应用,把建模类型的预测分析与非模型式数据内涵分析有机结合.它还可以同时实现回归建模(回归分析)、数据结构简化(主成分分析)以及两组变量间相关分析(典型相关分析).它适用于当变量之间存在高度相关性时,建立回归模型,其模型改进传统多元回归分析的不足,使得分析结论更可靠,整体性更强.它还适合在样本容量小于变量个数的情况下进行回归建模.可以实现多种多元统计分析方法(包括主分量分析方法等)的综合应用.与传统回归分析不同,它对多个自变量和因变量进行标准化处理,形成自变量和因变量标准化矩阵 E_0 和 F_0 .对自变量作主分量分析,提取其与因变量有最密切关系的最大特征值对应的主分量,和对应的荷载向量,然后以此主分量与因变量作回归,分别求各自残差矩阵,再对残差矩阵作类似选取,如此逐次,求自变量的主分量和对因变量的回归,最后得到

$$\begin{aligned} E_0 &= t_1 p'_1 + t_2 p'_2 + \cdots + t_s p'_s, \\ F_0 &= t_1 r'_1 + t_2 r'_2 + \cdots + t_s r'_s + F_s, \end{aligned} \quad (6)$$

式中, t_i 表示逐次($i=1, 2, \cdots, s$)提取的自变量矩阵主分量向量, p_i 为自变量荷载向量, r_i 为因变量在主分量轴上的投影向量,符号($'$)表示转置.通过残差计算出预报因变量的解释方差,作为偏回归估计的方程稳定性的度量.

2 近百年全球及区域降水观测变化

图2为全球陆地及几个纬度带(北半球高纬度($60^\circ \sim 90^\circ \text{N}$)、北半球中纬度($30^\circ \sim 60^\circ \text{N}$)、低纬度($30^\circ \text{S} \sim 30^\circ \text{N}$)和南半球中纬度($60^\circ \sim 30^\circ \text{S}$))114年间的年降水距平变化序列.类似于IPCC(2013),利用受限极大似然回归(REML)方法^[26]计算了全球及各个纬度带($60^\circ \sim 90^\circ \text{N}$, $30^\circ \sim 60^\circ \text{N}$, $30^\circ \text{S} \sim 30^\circ \text{N}$, $60^\circ \sim 30^\circ \text{S}$)陆地降水的长期趋势及其95%信度下的显著性(表1).REML方法是IPCC TAR以来计算气候变化趋势的基本方法.它由于考虑到了降水序列的自相关性,故比普通线性趋势对极端

值更不敏感,因此,较为适合作为气候变化趋势的计算方法,特别是对于气温、降水等存在自相关性的气候要素.显然,从全球尺度上看,1900年以来降水略有增加,平均每10年增加3.54 mm,北半球高纬度降水增加更为迅速,每10年为10.73 mm,中纬度次之,低纬度略低,而南半球中纬度则为略下降趋势(但后二者均未通过95%信度上的显著性水平).1951年以来也遵循类似的规律,即由北向南,降水的增加趋势有递减的规律,低纬度趋势不明显,而南半球中纬度为显著下降;上述的分析结果与IPCC(2013)^[19]引用的几个数据集CRUTS 3.10.01、GHCN v2、GPCC v6和Smith等人的结果^[27~30]相比,本研究估计的降水趋势普遍增大,可能是本研究估计的时段(1900~2013年)及采用的数据集存在一定的差别所致.20世纪90年代后低纬度年降水的趋势显著增多,这可能和低纬度的温度上升^[31],以及Sahel地区降水显著上升有关^[19].

3 全球及区域降水变化影响因子建模及归因分析

3.1 人类活动和自然强迫总体影响及其显著性

3.1.1 多元线性回归拟合的效果

分别利用全球和各纬度带陆地降水距平序列为因变量,以自然强迫总量(Natural)和人为活动总量(Tan)为自变量,进行多元线性回归建模,结果见表2.从模型本身来看,5个MLR模型均有显著性的意义;但显然,全球陆地降水距平的MLR模型标准误差为最小,其次是北半球中纬度,再次为北半球高纬度,低纬度和南半球中纬度的误差较高;但是从德宾-沃森数(DW)来看,只有南半球中纬度降水通过了无自相关的假设检验.也就是说,前4个MLR拟合模型的残差存在显著的自相关(正自相关),说明模型仍然存在可以预测的信息.因此,还需要进一步对残差序列进行分析,提取可预测信息.一般有两种方法:一是用广义线性回归(generalize linear model, GLM)模型;二是用差ARIMA模型,本研究采用后者进行组合.

此外,从表征拟合效果的调整后 R^2 来看,北半球高纬度最高,达到0.582,中纬度次之,为0.455;全球再之,为0.308,低纬度和南半球中纬度最低,分别只有0.183和0.078.这说明,对于低纬度及南半球中纬度陆地降水变化来说,即使自然强迫和人为强迫联合起来都只能解释整个区域降水变化的一小部分(<20%),说明对其

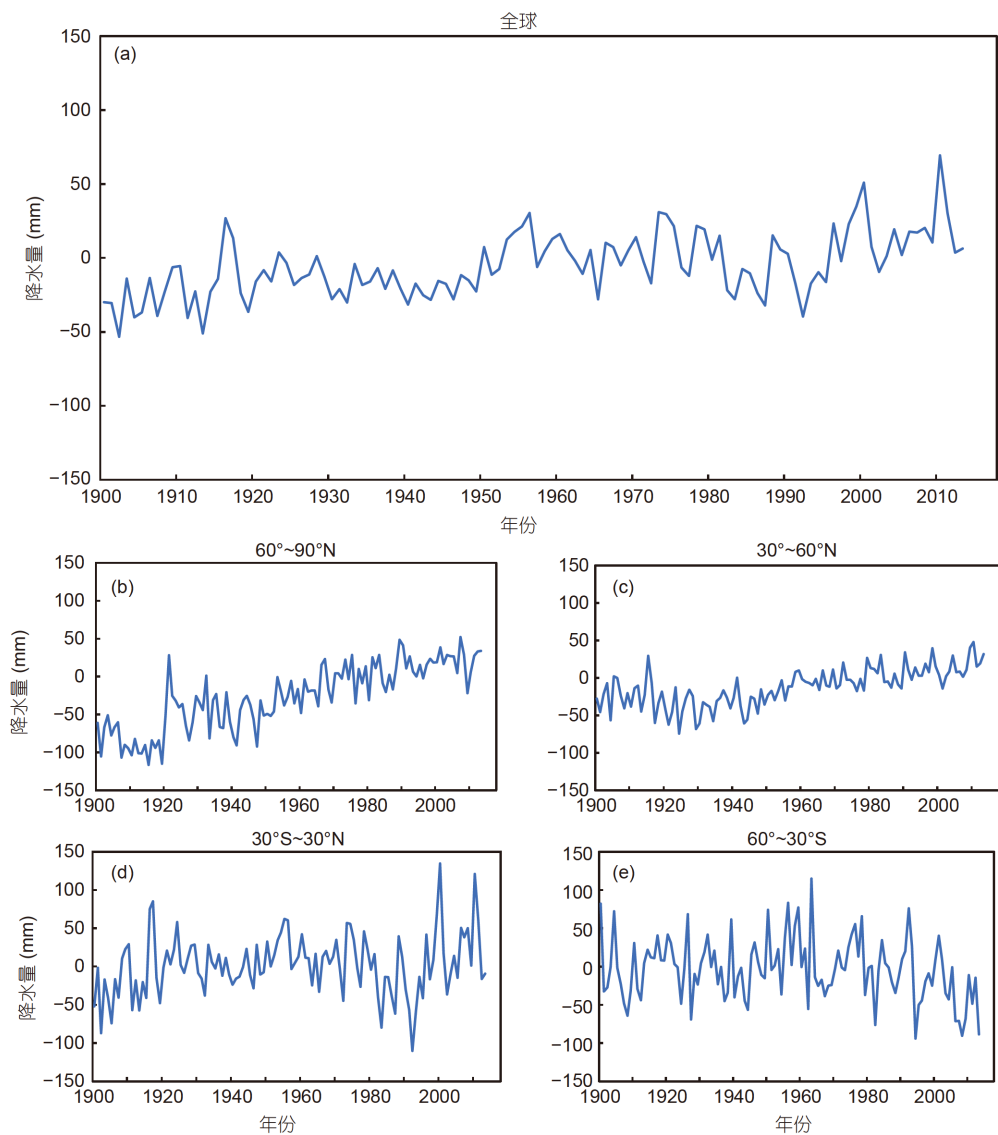


图 2 (网络版彩色)全球(a)及各纬度带(b~e)平均降水距平序列. (b) 60°~90°N; (c) 30°~60°N; (d) 30°S~30°N; (e) 60°~30°S
Figure 2 (Color online) The global (a) and regional (b~e) annual average precipitation series. (b) 60°~90°N; (c) 30°~60°N; (d) 30°S~30°N; (e) 60°~30°S

表 1 全球及区域性降水变化趋势(mm/10 a)及其5%显著性^{a)}

Table 1 Global and regional land precipitation change trends and their uncertainties at 5% level

	1900~2013年	1951~2013年
全球	3.54±1.46**	1.68±3.87
60°~90°N	10.73±1.87**	9.32±3.70**
30°~60°N	5.22±1.47**	5.51±2.78**
30°S~30°N	2.17±3.21	-0.48±8.69
60°~30°S	-2.42±3.04	-10.63±7.15**

a) **表示该趋势超过95%信度的显著性水平

最重要的影响因子并没有找到. 因此, 下文的分析只侧重了全球和北半球中高纬度降水的变化.

3.1.2 MLR残差的ARIMA模型拟合

下面以全球陆地降水为例, 来说明ARIMA模型的建立. 从3.1.1节中全球降水的MLR模型来看, 显然模型中自变量之间不存在共线性(VIF=1.007), 但DW=1.32, 明显偏低, 表明残差中可能存在自相关. 因此对残差进行深入分析, 计算残差的滞后自相关(auto-correlation function, ACF)和偏自相关(partial autocorrelation function, PACF)(图3). 显然, 无论是ACF还是PACF, 无论是滞后多少步, 二者都超过了置信度的范

表 2 多元线性回归(MLR)模型概述

Table 2 Overview of the multiple linear regression (MLR) model

区域	R^2	调整后 R^2	标准误差	显著性	德宾-沃森
全球	0.320	0.308	17.683	0.000	1.325
60°~90°N	0.589	0.582	27.658	0.000	1.145
30°~60°N	0.465	0.455	18.687	0.000	1.492
30°S~30°N	0.197	0.183	35.597	0.000	1.590
60°~30°S	0.094	0.078	39.000	0.000	1.687

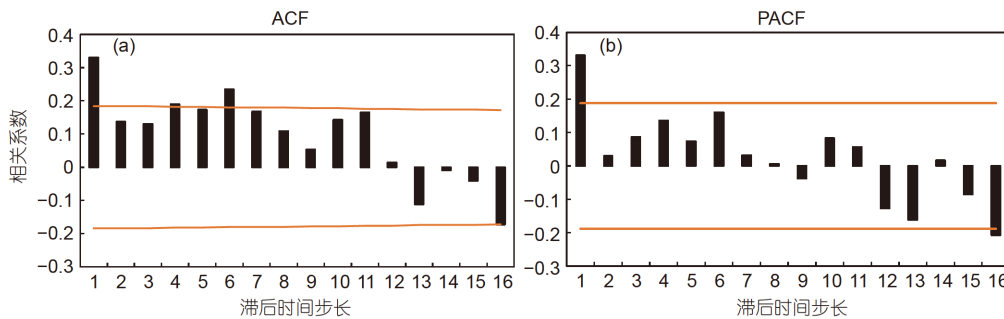


图 3 (网络版彩色)全球陆地降水多元线性回归残差的滞后自相关(a)和偏自相关(b). 横线为95%置信度范围的最高、最低值
Figure 3 (Color online) Lag autocorrelation (a) and partial autocorrelation (b) plots of the MLR residuals of global land precipitation. The orange line in the figure is the highest and lowest values in the 95% confidence range

围, 即均是拖尾的. 因此, 需要采用ARIMA模型对残差进一步拟合. 经参数选择对比, ARIMA(1,1,1)(表3)的标准均方误差为最小.

3.1.3 MLR, ARIMA 组合模型

根据上述讨论重新建模, 在上述MLR方程中加上ARIMA(1,1,1)模型, 新的组合模型拟合方程为

$$Y = -18.443 + 13.730 \times \text{Natural} + 16.626 \times \text{Tan} + 0.684 \times \text{ARIMA}(1, 1, 1) + \varepsilon. \quad (7)$$

重新计算VIF=1.007(仍然无共线性), DW=1.861(残差已经不存在自相关), 说明方程(7)为最优无偏估计, 其中, ε 为随机误差. 前3项为确定性部分, 即在该预测模型中, 该部分是由关于预测自变量的函数组成, 其中包含了回归模型中所有可解释、可预测的信息. 即在全球陆地降水距平的变化这个因变量中, 自然强迫、人为强迫因子和一个残差的ARIMA(1,1,1)模型可以分别解释全球陆地降水距平变化中可预测部分的方差贡献(约38.3%, 高于表2中的30.8%), 其均方误差为16.67 mm, 比表2中略低. 对回归系数的显著性检验表明, 常数和以上3个确定性项的系数均通过了99.9%的显著性水平. 显然, 模型是合理的, 各个自变量或模型的系数均具有非常高的显著性. 也就是说, 自然强迫、

表 3 ARIMA 模型参数

Table 3 The parameters of the ARIMA model

参数	估算	标准误差	t	显著性
常量	12.705	17.159	0.740	0.461
AR 延迟 1	0.196	0.114	1.713	0.090
差分			1	
MA 延迟 1	0.881	0.057	15.477	0.000

人为强迫均是影响全球陆地降水变化的重要因子, 加上一个ARIMA模型, 一共解释了约40%的全球陆地降水变化的方差贡献.

根据同样的思路, 对60°~90°N, 30°~60°N纬度带降水距平分别建模为

$$Y = -75.561 - 3.732 \times \text{Natural} + 48.120 \times \text{Tan} + 0.990 \times \text{ARIMA}(1, 0, 9) + \varepsilon, \quad (8)$$

$$Y = -38.582 - 3.515 \times \text{Natural} + 25.408 \times \text{Tan} + 1.036 \times \text{ARIMA}(3, 0, 0) + \varepsilon. \quad (9)$$

上述(8)、(9)式两个模型分别解释了60°~90°N、30°~60°N纬度带降水距平变化的68.5%、52.3%的总方差, 但从几个自变量的系数来看, 自然强迫的系数均没有达到90%以上的显著性水平. 这说明北半球中高纬度

降水不能用自然强迫(太阳、火山等)来预测. 但正如前所述, 人类活动及ARIMA模型对两个纬度带降水距平变化的预测有着确定性的影响. 这和利用模式模拟进行“指纹法”的分析结论是较为一致的^[18,19].

3.2 不同因子各自的影响分离

从上述分析中可以看出, 人类活动对于近百年全球及北半球中高纬度陆地的降水变化有一定的确定性贡献, 但如何量化每一种人类活动的具体贡献, 前人做了大量的研究. 但由于线性回归方法无法量化评估存在多重相关性的多种人类活动所造成的贡献, 而偏最小二乘回归(PLSR)在这方面具有明显的优势, 因此本研究用它来继续分离各种人类活动因子对全球和北半球中高纬度陆地降水变化各自的影响及其显著性水平.

对前一节的前3个陆地降水变化序列(全球、北半球高纬度、北半球中纬度)进行PLSR分析, 以研究各种人类活动的分量对其的贡献水平.

3.2.1 全球陆地降水变化中各种自然与人为强迫的影响与贡献

以全球陆地降水距平变化序列作为因变量, 以所有外部强迫作为自变量(包括自然强迫的太阳活动和火山, 人为强迫的CO₂、其他温室气体(OGHG)、气溶胶、土地利用、雪盖、对流层臭氧、平流层臭氧、平流层水汽和尾迹卷云等共11个因子)进行PLSR建模, 结

果见表4. 图4为PLSR中对自变量主分量分析提取的前10个潜在因子(特征值)的解释方差变化情况. 显然, 前6个潜在因子之后引入的新因子对模型的解释方差已经没有太多的增加, 故取前6个作为建模的因子, 共解释方差为46.3%. 为了体现各个自变量的相对重要性, 计算了所有变量标准化后的模型系数(图5), 从图中的系数可以看出各个变量的相对重要性对比. 也就是说, 对于全球陆地降水距平变化, 自然强迫的贡献是正贡献, 使得全球陆地降水距平有增加的趋势, 但这种贡献较小; 考虑到辐射强迫本身的变化趋势, 火山CO₂和其他温室气体、平流层水汽和对流层臭氧对全球陆地降水变化为负贡献, 气溶胶、平流层臭氧和地球表面反照率(包括土地利用和雪盖)、尾迹卷云等均为正贡献, 共同作用导致了全球陆地降水变化的正趋势. 显然从各个自变量的重要性指标来看, 取前6个潜在因子的累积重要性, 除了太阳辐射略低, 其他都接近或大于1, 表明除太阳辐射外, 几乎所有的这些要素对全球陆地降水变化的影响都是非常重要的, 其中最为重要的因子依次为: 火山、尾迹卷云、雪盖和平流层水汽.

3.2.2 北半球中、高纬度陆地降水变化中各种人为强迫的影响与贡献

类似上面的分析, 分别对北半球高纬度和中纬度陆地降水距平变化进行PLSR建模. 考虑到上一节的分析中表明, 外部强迫对此两个纬度带降水变化的影响

表4 全球陆地降水距平变化PLSR建模结果

Table 4 The PLSR modeling results of global land precipitation anomalies

自变量	因变量	前6个潜在因子累积	权重					
			1	2	3	4	5	6
常量	-83.87	Y*	0.201	0.337	0.187	0.210	0.288	0.948
太阳辐射	31.95	0.73	0.162	0.237	-0.563	-0.148	0.424	-0.014
火山	12.36	1.46	0.145	0.884	0.205	-0.262	-0.130	0.026
二氧化碳	69.66	0.93	0.341	0.013	0.148	0.073	0.113	-0.387
其他温室气体	-168.29	1.00	0.324	-0.095	-0.168	-0.401	-0.372	-0.464
气溶胶	-135.16	0.96	-0.341	0.022	-0.030	-0.083	-0.083	-0.496
平流层臭氧	-830.14	0.95	-0.307	0.121	0.170	0.364	0.051	-0.545
对流层臭氧	-182.93	0.91	0.336	-0.053	-0.111	-0.214	-0.268	-0.131
平流层臭氧	-368.15	1.10	0.335	-0.052	-0.272	-0.504	-0.624	-0.479
土地利用	-104.28	0.94	-0.355	-0.063	-0.009	-0.101	0.055	-0.196
雪盖	1353.05	1.12	0.250	0.404	0.268	0.905	0.066	0.062
尾迹卷云	786.96	1.39	0.327	0.112	0.696	0.911	1.127	0.247

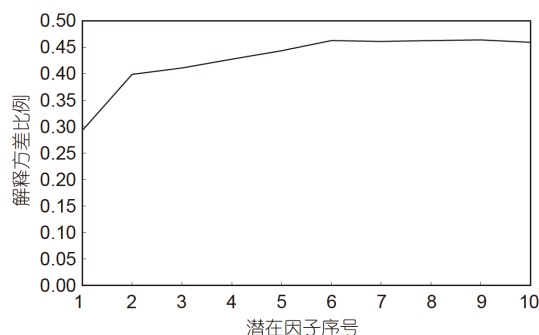


图4 全球陆地降水PLSR潜在因子解释方差

Figure 4 The explained variances of potential factors in PLSR model of the global land precipitation anomalies

不显著, 在此为了简化模型, 我们只分析了人类活动对上述两个纬度带降水变化的影响. 类似于3.2.1, 分别对北半球高纬度和中纬度陆地降水变化, 及自变量主分量分析取前4个潜在因子进行建模(分别解释了高纬度和中纬度陆地降水70.3%和52.2%的方差贡献). 表5、6分别为北半球高纬度和中纬度陆地降水变化建模的结果. 图6显示, 从自变量的重要性来看, 差别不大, 各个自变量对于北半球中高纬度的降水变化均很重要, 不同要素略有差别; 但值得注意的是, 其中只有像气溶胶、尾迹卷云这两个要素对高纬度和中纬度陆地降水的贡献是同号的, 均有利于降水的增加. 其余自变量对高纬度和中纬度降水的影响正好相反, 这也说明模型还存在一定的不确定性. 当然, 这些辐射强迫因子也是对全球平均而言的, 对于各个区域的情况, 这些因子对

降水的影响本身也具有一定的不确定性, 还需要根据本区域的具体情况进一步针对性研究. 总之, 利用纯统计手段, 对降水变化进行建模和分析研究, 还需要结合其物理规律, 做进一步的优化和完善. 但总的来说, 关于北半球中、高纬度的降水受到人为因子的确定性影响的结论, 与已有的成果是高度一致的^[31-35].

4 建模的数学物理本质

辐射强迫定义为气候系统辐射能量收支平衡中外部强加的扰动. 在气候系统模式中, 通过气候灵敏度参数, 将全球地表温度响应和全球平均辐射强迫进行了线性关系表达. 这种同一模式中的气候灵敏度的相对不变性使得辐射强迫成为估计全球年平均的地表温度(当然也包括了和温度相关的其他气候要素)响应的方便工具. 本研究利用统计模型直接对气候变化进行分解的方法仍主要基于气候变化和各自辐射强迫因子之间的线性关系.

将自然强迫和人为强迫分别作为整体来进行建模分析, 目的是从总体上将二者对全球或各纬度带陆地降水变化的影响的显著性进行初步判断, 也为下一步简化模型, 更为准确地聚焦研究对象奠定基础. 大多数气候变化要素(如全球陆地降水、表面温度等)序列等都存在自相关, 从物理上讲, 也可以理解为各种强迫因子对气候变化的影响(作用)有一定的滞后性(延续性), 这也表现在平时仅采用普通多元线性回归拟合时经常出现残差的自相关现象. 差分滑动平均自回归(ARI-

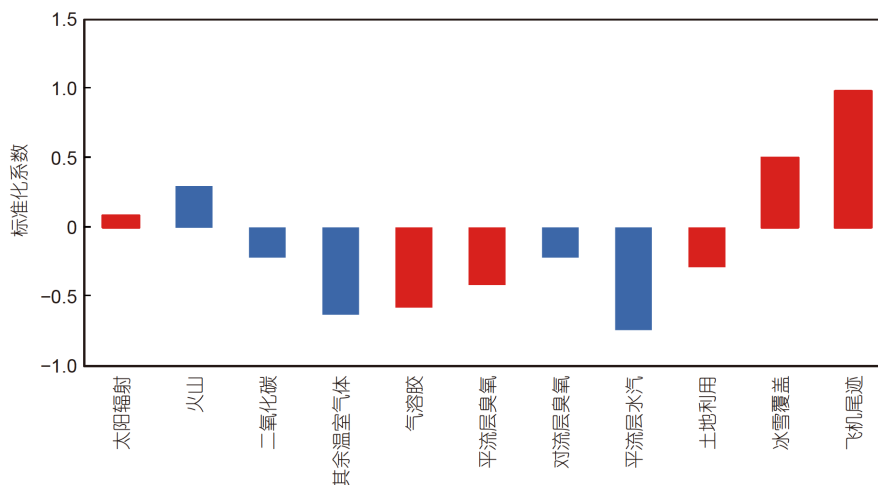


图5 (网络版彩色)全球陆地降水距平变化PLSR模型中各个辐射强迫的标准化系数. 暖色直方图表示对降水增多为正贡献, 反之为负贡献

Figure 5 (Color online) The normalized coefficients of various radiative forcings in the PLSR model of global land precipitation anomalies. The warm color histogram indicates a positive contribution to the increase in precipitation, and a negative contribution otherwise

表 5 北半球高纬度陆地降水距平变化PLSR建模结果

Table 5 The PLSR modeling results of the land precipitation anomalies at high latitudes in the northern hemisphere

变量	因变量	前4个潜在因子累积	权重			
			1	2	3	4
常量	-148.78		0.297	0.153	0.211	0.677
二氧化碳	16.25	1.00	0.339	-0.134	0.281	0.039
其他温室气体	10.65	1.02	0.348	-0.107	-0.076	0.021
气溶胶	-101.28	1.07	-0.361	-0.071	-0.101	-0.655
平流层臭氧	316.27	0.97	-0.327	0.227	0.501	0.171
对流层臭氧	-15.96	1.05	0.359	0.010	-0.186	-0.165
平流层臭氧	835.37	1.06	0.361	0.074	0.043	0.391
土地利用	594.00	1.08	-0.361	-0.155	0.375	0.751
雪盖	1912.94	0.84	0.226	0.897	0.300	0.051
尾迹卷云	388.39	0.89	0.294	-0.292	0.633	-0.067

表 6 北半球中纬度陆地降水距平变化PLSR建模结果

Table 6 The PLSR modeling results of the land precipitation anomalies at middle latitudes in the northern hemisphere

自变量	因变量	前4个潜在因子累积	权重			
			1	2	3	4
常量	-46.86		0.257	0.352	0.213	0.241
二氧化碳	-0.37	1.19	0.347	-0.304	-1.107	-0.515
其他温室气体	-130.05	1.17	0.353	-0.306	-1.018	-0.267
气溶胶	-34.43	1.39	-0.364	-0.510	-1.606	-0.418
平流层臭氧	-468.13	1.08	-0.343	0.189	0.671	-0.716
对流层臭氧	333.97	1.09	0.361	0.155	0.513	0.209
平流层臭氧	-423.89	1.20	0.349	-0.399	-1.135	-0.462
土地利用	-173.73	1.19	-0.353	-0.291	-1.074	-0.060
雪盖	-200.11	0.45	0.147	-0.294	0.002	-0.071
尾迹卷云	764.54	1.06	0.326	0.404	0.806	-0.128

MA)模型包括了自回归(AR)模型、移动平均(MA)模型、自回归移动平均(ARMA)模型和差分自回归滑动平均(ARIMA)模型. 其中, 自回归(AR)模型主要解决用历史数据本身进行预报的问题, 但要求数据本身有平稳性, 同样也只适用于预测与自身前期相关的现象, 这恰好是符合气候变量的自相关性特征与预报的原理; 滑动平均(MA)模型主要关注上述AR模型误差项的累加, 滑动平均法消除了预测中的随机波动; 而差分方法主要解决非平稳序列的平稳化. 三者的结合, 将MLR中可以预报的自变量的重要性进行了修正, 同时对无法预报的部分进行了补充.

分离各种自然和人为强迫对气候变化的各自影响, 即多个因子(自变量)之间的相互“混淆”问题, 历来是气候变化归因问题的一个难题. 比如城市化和温室气体对气温变化的影响非常相似(平均温度升温, 日较差变小), 用统计语言来说, 这两种因子可能具有非常明确的相关性(实际上这种情况在多种因子的情况下也普遍存在, 即因子间存在多重相关性或共线性); 而偏最小二乘回归(PLSR)模型则通过集成了主成分分析、典型相关分析、线性回归分析的优点, 分别在自变量和因变量中提取出主成分(潜在因子), 为了回归分析的需要, 主成分有下列两个要求: (1) 因变量和自

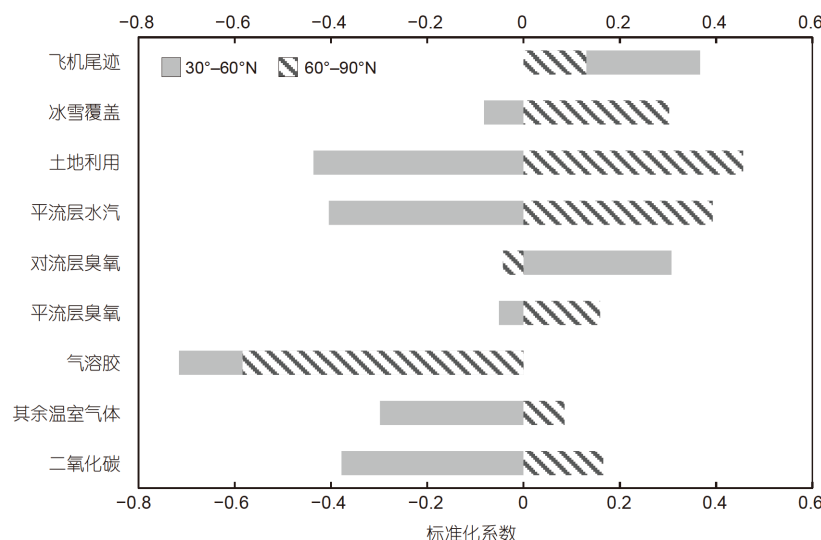


图6 北半球高纬度和中纬度陆地降水距平变化PLSR模型中各个辐射强迫的标准化系数

Figure 6 The normalized coefficients of various radiative forcings in the PLSR model of land precipitation anomaly changes at high and mid-latitudes in the northern hemisphere

变量的主成分各自应尽可能大地携带本身的变异信息; (2) 两个主成分的相关程度能够达到最大。在此基础上实施自变量和因变量对各自主成分的回归, 如果回归方程已经达到满意的精度, 则算法终止; 否则, 将利用残余信息进行第二轮的主成分提取; 如此往复, 直到能达到一个较满意的精度为止。其系数即为各个自变量对因变量变化的贡献的相对大小, 并给出其对因变量变化的重要性(一般认为, 重要性因子大于1, 表示该自变量对因变量的变化非常重要)。在这个步骤中, 可以通过PLSR模型的自变量、研究要素和研究区域的选取和相应的建模结果的对比, 又可以较为量化地解释其物理意义, 以达到对某种(或多种)辐射强迫因子本身对气候要素变化的各自贡献的检测和成因分析的目的。

5 总结与展望

本研究以1900~2013年全球和各个纬度带陆地降水距平变化为因变量, 以整体的自然强迫、人为强迫为自变量, 结合差分滑动平均自回归(ARIMA)模型进行了组合建模, 初步分离了人为强迫对于北半球中高纬度和全球降水的影响信号的显著性水平; 进而利用偏最小二乘回归分别对人类活动的各个自变量(辐射强迫/有效辐射强迫)对北半球降水变化的各自具体贡献水平进行了分析, 得出以下一些初步结论。

(1) 多元线性回归和差分滑动平均自回归组合模

型, 较为明确地分离出了全球和北半球中、高纬度陆地降水距平变化中的人为强迫的影响信号, 显著性水平较高, 解释方差较大。

(2) 偏最小二乘回归模型在一定程度上克服了普通线性回归方程的病态, 较好地分离了各个自然强迫、人为强迫因子对全球和北半球中、高纬度陆地降水的具体影响(贡献), 并给出了各个因子贡献的相对强弱判别, 对于自变量之间有多重相关性的建模和影响研究具有很好的应用价值。

(3) 上述多元线性回归和差分滑动平均自回归组合模型对全球、北半球高纬度、中纬度地区陆地降水距平变化的解释方差接近40%, 其中北半球高纬度地区的解释方差接近60%, 拟合效果较好; 从模型拟合结果来看, 人类活动和自然强迫对于全球陆地降水来说, 都是确定性因子; 但对于北半球中高纬度来说, 自然强迫的作用不显著, 人类活动解释了很大一部分方差贡献, 并具有高显著性; 另外, 全球和区域降水距平变化的自相关性, 也是降水的重要因子之一。

(4) 无论是对全球陆地, 还是北半球中、高纬度陆地降水距平变化, 气溶胶、尾迹卷云都是明确的显著正贡献; 其余各因子的贡献则出现明显的正、负分化, 说明这些因子对于全球和区域性降水变化的贡献(影响)存在一定的不确定性, 需要根据更为精确的各区域的实际因子数据, 对因子和模型进行进一步针对性优化完善。

(5) 由于上述整个模型过程都是对历史观测数据进行分解建模. 一方面, 它对基准数据的“纯洁性”提出了一定的要求; 另一方面, 它可以不受区域的限制, 甚至可以对局地气候变化中存在的人为影响进行建模分析; 此外, 它可以直接对历史气候序列中的气候信号进

行检测和分析, 得出人类活动或自然强迫对气候变化的相对贡献及其显著性, 应该说是一个较为便捷的方法, 是对气候模式归因的有力补充. 但同时, 如何结合气候系统模式的模拟结果, 进一步对统计结果进行深入的机理分析, 还需要做大量的深入研究.

参考文献

- 1 Zhou T J, Li L J, Li H M, et al. Progress in climate change attribution and projection studies (in Chinese). *Chin J Atmos Sci*, 2008, 32: 906–922 [周天军, 李立娟, 李红梅, 等. 气候变化的归因和预估模拟研究. *大气科学*, 2008, 32: 906–922]
- 2 Wang S W, Luo Y, Zhao Z C, et al. Attribution of climate warming to the causes (in Chinese). *Adv Clim Chang Res*, 2012, 8: 308–312 [王绍武, 罗勇, 赵宗慈, 等. 气候变暖的归因研究. *气候变化研究进展*, 2012, 8: 308–312]
- 3 Hasselmann K. On the signal-to-noise problem in atmospheric response studies. In: Shaw D B, ed. *Meteorology Over the Tropical Oceans*. Bracknell: Royal Meteorological Society, 1979. 251–259
- 4 Hasselmann K. Multi-pattern fingerprint method for detection and attribution of climate change. *Clim Dyn*, 1997, 13: 601–611
- 5 Allen M R, Tett S F B. Checking for model consistency in optimal fingerprinting. *Clim Dyn*, 1999, 15: 419–434
- 6 Hegerl G C, Hasselmann K, Cubasch U, et al. Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. *Clim Dyn*, 1997, 13: 613–634
- 7 Stott P A, Kettleborough J A. Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature*, 2002, 416: 723–726
- 8 Abreu R C, Tett S F B, Schurer A, et al. Attribution of detected temperature trends in Southeast Brazil. *Geophys Res Lett*, 2019, 46: 8407–8414
- 9 Ribes A, Zwiers F W, Azaïs J M, et al. A new statistical approach to climate change detection and attribution. *Clim Dyn*, 2017, 48: 367–386
- 10 Folland C K, Boucher O, Colman A, et al. Causes of irregularities in trends of global mean surface temperature since the late 19th century. *Sci Adv*, 2018, 4: eaao5297
- 11 Wu T, Hu A, Gao F, et al. New insights into natural variability and anthropogenic forcing of global/regional climate evolution. *npj Clim Atmos Sci*, 2019, 2: 1–13
- 12 Zhang H, Huang J P. Interpretation of the IPCC Fifth Assessment Report on anthropogenic and natural radiative forcing (in Chinese). *Adv Clim Chang Res*, 2014, 10: 40–44 [张华, 黄建平. 对第五次评估报告关于人为和自然辐射强迫的解读. *气候变化研究进展*, 2014, 10: 40–44]
- 13 Zhang X, Zwiers F W, Hegerl G C, et al. Detection of human influence on twentieth-century precipitation trends. *Nature*, 2007, 448: 461–465
- 14 Sun Y, Solomon S, Dai A, et al. How often will it rain? *J Clim*, 2007, 20: 4801–4818
- 15 Hegerl G C, Black E, Allan R P, et al. Challenges in quantifying changes in the global water cycle. *Bull Amer Meteorol Soc*, 2015, 96: 1097–1115
- 16 Sarojini B B, Stott P A, Black E. Detection and attribution of human influence on regional precipitation. *Nat Clim Chang*, 2016, 6: 669–675
- 17 Ma S, Zhou T, Stone D A, et al. Detectable anthropogenic shift toward heavy precipitation over eastern China. *J Clim*, 2017, 30: 1381–1396
- 18 Yang S, Xu W, Xu Y, et al. Development of a global historic monthly mean precipitation dataset. *J Meteorol Res*, 2016, 30: 217–231
- 19 Intergovernmental Panel on Climate Change (IPCC). The Physical Science Basis: Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press, 2013
- 20 Intergovernmental Panel on Climate Change (IPCC). The Physical Science Basis: Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press, 2007
- 21 Huang J Y, Li Q X. A new diagnosis method for non-homogeneity on a series (in Chinese). *Plateau Meteorol*, 2007, 26: 62–66 [黄嘉佑, 李庆祥. 一种诊断序列非均一性的新方法. *高原气象*, 2007, 26: 62–66]
- 22 Li Q X, Huang J Y, Ju X H. Experimental study on reconstruction of maximum temperature data in shanghai (in Chinese). *J Tropic Meteorol*, 2008, 24: 349–353 [李庆祥, 黄嘉佑, 鞠晓慧. 上海地区最高气温资料的恢复试验. *热带气象学报*, 2008, 24: 349–353]
- 23 Rui X J, Zhong B L, Yan J P, et al. Study on the processing method of periodic and tendentious signal in time series model-building (in Chinese). *J Appl Sci*, 1995, 13: 195–201 [芮小健, 钟秉林, 颜景平, 等. 时序建模中周期信号和趋势项的处理问题研究. *应用科学学报*, 1995, 13: 195–201]
- 24 Wold S, Albano C, Dunn M. Pattern Regression Finding and Using Regularities in Multivariate Data. London: Analysis Applied Science Publication, 1983
- 25 Wang H W. Partial Least Squares Regression Method and Its Application (in Chinese). Beijing: National Defense Industry Press, 1999 [王惠文. 偏最小二乘回归方法及其应用. 北京: 国防工业出版社, 1999]
- 26 Diggle P J, Liang K Y, Zeger S L. Analysis of Longitudinal Data. Oxford: Oxford University Press, 1994

- 27 Mitchell T D, Jones P D. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *Int J Climatol*, 2005, 25: 693–712
- 28 Vose R S, Oak Ridge National Laboratory. Environmental Sciences Division, U.S. Global Change Research Program, United States. Dept of Energy Office of Health and Environmental Research, Carbon Dioxide Information Analysis Center (US), and Martin Marietta Energy Systems Inc. The Global Historical Climatology Network: Long-Term Monthly Temperature, Precipitation, Sea Level Pressure, and Station Pressure Data. Carbon Dioxide Information Analysis Center. Available to the public from N.T.I.S., 1v. (various pagings), 1992
- 29 Becker A, Finger P, Meyer-Christoffer A, et al. A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth Syst Sci Data*, 2013, 5: 71–99
- 30 Smith T M, Arkin P A, Ren L, et al. Improved reconstruction of global precipitation since 1900. *J Atmos Ocean Technol*, 2012, 29: 1505–1517
- 31 Li Q, Sun W, Huang B, et al. Consistence of global warming trends strengthened since 1880s. *Sci Bull*, 2020, doi: <https://doi.org/10.1016/j.scib.2020.06.009>
- 32 Min S K, Zhang X B, Zwiers F. Human-induced arctic moistening. *Science*, 2008, 320: 518–520
- 33 Fyfe J C, Gillett N P, Marshal G J. Human influence on extratropical Southern Hemisphere summer precipitation. *Geophys Res Lett*, 2012, 39: L23711
- 34 Noake K, Polson D, Hegerl G, et al. Changes in seasonal land precipitation during the latter 20th century. *Geophys Res Lett*, 2012, 39: L03706
- 35 Polson D, Hegerl G C, Zhang X, et al. Causes of robust seasonal land precipitation changes. *J Clim*, 2013, 26: 6679–6697

Summary for “用外部强迫因子对近百年陆地降水变化的统计建模试验”

Statistical modeling experiment of land precipitation variations since the start of the 20th century with external forcing factors

Qingxiang Li^{1,2}

¹ Guangdong Province Key Laboratory for Climate Change and Natural Disasters, School of Atmospheric Sciences, Sun Yat-sen University, Guangzhou 510275, China;

² Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519082, China
E-mail: liqingx5@mail.sysu.edu.cn

Fitting of the historic climate change with multiple linear regression (MLR) models often encounters the problem of residual autocorrelation and multiple correlation among the dependent factors, which leads the model not to be the Best Linear Unbiased Estimator (BLUE). To solve these problems, a set of sophisticated statistical modeling procedures are developed in this paper to detect the contributions of the external forcing to the global and regional land precipitation changes during the recent century. The radiative forcing (effective radiative forcing) data of multiple natural and anthropogenic factors are taken as independent variables and a combined model of MLR and Autoregressive Integrated Moving Average (ARIMA) is first used to separate the impacts of natural forcing and human activities (signals) and internal variability (noise) due to land precipitation changes since the start of the 20th century. Based on the results, a partial least squares regression (PLSR) model is further adopted to quantify the respective contributions of the each components of anthropogenic forcing. Some preliminary conclusions were drawn:

(1) The combined model of MLR and ARIMA has clearly separated the influence of anthropogenic forcing signals in the global and regional (middle and high latitude in Northern Hemisphere) land precipitation anomaly at 5% significance level, and the explained variance is relatively large.

(2) The PLSR model overcomes the morbidity of ordinary linear regression equations to a certain extent, and better separated the respective impact (contribution) of various natural and anthropogenic forcing factors on global and regional (middle and high latitude in Northern Hemisphere) land precipitation anomaly; and the relative strength of each factor's contribution are determined. It has good application value for modeling and impact studies with multiple correlations among different independent variables.

(3) The above combined model explains about 40% of the total variance of land precipitation anomalies change over global and middle latitude regions in Northern Hemisphere, and nearly 60% of the total variance in high latitude regions. Based on the fitting results, both the human activities and natural forcing are the deterministic impact factors of global land precipitation variations. But the natural forcing's contribution is not significant, while human activities still explain a large part of the variance and thus have high significance for the regional land precipitation variations. In addition, the autocorrelation should also be considered as one of the important impact factors for global and regional land precipitation anomaly changes.

(4) The anthropogenic aerosol and contrails have clear and significant positive contributions to the global and regional land precipitation anomalies, while the contributions of the remaining factors show obvious positive and negative differentiation, indicating that there is certain degree of uncertainty. Therefore, it is necessary to further optimize and improve the factors and models based on more accurate factor datasets for each region.

(5) The modeling approach described above is primarily based on the idea to decompose and model the historical climate series. On one hand, it may put forward certain requirements for the “integrity” of the observational data; while on the other hand, it is not restricted by the study domains, climate variables and the independent variables. Moreover, it may even model/separate the anthropogenic/natural effects of climate changes in local scales. In addition, it can independently detect and analyze the relative contribution and significance of human activities or natural forcing to climate change in the historical observational series without using the climate system model. Therefore, it is proved to be a convenient approach for detecting the causes of climate change and a powerful supplement to model attribution.

land precipitation, dataset, statistical modeling, anthropogenic forcing, natural forcing

doi: [10.1360/TB-2020-0305](https://doi.org/10.1360/TB-2020-0305)