

我国东部季风区主要流域气候变化情景下日平均 温度和日降水量随机模拟数据集

杨 赤

北京师范大学全球变化与地球系统科学研究院, 北京 100875

摘 要: 基于历史观测日值资料和耦合模式比较计划第五阶段的 15 个全球气候模式对历史排放情景和三种未来典型浓度路径的模拟月值结果, 综合应用广义加性模型、广义线性模型、贝叶斯模型平均等多种统计方法, 分别建立了流域尺度日平均温度和日降水量的多模式统计降尺度模型。在此基础上, 针对我国东部季风区海河、淮河、辽河、太湖、渭河和珠江六个流域, 运用随机模拟方法, 得到各流域在历史 (1970–2000 年) 和未来 (2020–2050 年) 排放情景下, 空间分辨率为 $0.5^\circ \times 0.5^\circ$ 的日平均温度和日降水量的样本集合。数据集充分反映了局地气候态的分布, 特别是极端温度、极端降水等位于气候态分布尾端的小概率事件对气候变化的响应, 可用于水文、生态和环境等领域的工程设计、灾害风险评估等。本数据集的数据文件可由 GrADS 软件直接读取。原始文件数 56 个, 数据量 25.4 GB (压缩为 20 个文件, 11.3 GB)。

关键词: 东部季风区; CMIP5; 广义加性模型; 贝叶斯模型平均

DOI: 10.3974/geodp.2018.01.05

1 前言

高分辨率的气候数据是水文、生态和环境等领域数值模型的基本驱动条件。特别是用于水文工程设计、灾害风险评估等目的时, 气候数据应能够真实反映局地的极端天气和气候条件, 如极端降水、极端温度等。从概率分布的角度看, 极端天气和气候事件是位于气候态分布尾端的小概率事件, 无法藉由气候平均态, 以确定性的方式表现出来, 因此, 需要通过随机模拟的方法, 以样本集合的方式表现气候态分布及其不确定性。在气候变化背景下, 区域的气候态分布不仅在空间上非均一, 在时间上也非平稳。因此, 首先要建立时空统计模型来描述气候要素的时空变化, 再根据模型运用随机模拟方法得到其样本集合。

为研究人类活动导致的气候变化, 耦合模式比较计划第五阶段 (CMIP5) 设计了一系列有关历史和未来温室气体排放的情景实验, 并以“典型浓度路径” (RCP) 来表示^[1]。参与 CMIP5 的诸多全球气候模式 (GCM) 依据实验设计对历史和未来气候进行了模拟, 其

收稿日期: 2018-01-30; 修订日期: 2018-03-20; 出版日期: 2018-03-25

基金项目: 中华人民共和国科学技术部 (2010CB428400)

作者信息: 杨赤 E-4997-2017, 北京师范大学全球变化与地球系统科学研究院, chi@bnu.edu.cn

数据引用方式: [1] 杨赤. 我国东部季风区主要流域气候变化情景下日平均温度和日降水量随机模拟数据集[J]. 全球变化数据学报, 2018, 2(1): 23–28. DOI: 10.3974/geodp.2018.01.05.

[2] 杨赤. 我国东部季风区主要流域气候变化情景下日平均温度和日降水量随机模拟数据集[DB/OL]. 全球变化科学研究数据出版系统, 2018. DOI: 10.3974/geodb.2018.01.02.V1.

输出结果成为评估未来气候变化影响的重要依据。GCM 能够较好地模拟大尺度的大气环流特征。如果能将有关输出变量通过统计模型与局地气候要素联系起来,使其对局地气候变化起到指示作用,就能预估局地气候对全球气候变化的响应,即所谓“降尺度”^[2]。再依据降尺度模型进行随机模拟,得到局地气候要素随时空变化的样本集合。

本工作针对我国东部季风区海河、淮河、辽河、太湖、渭河和珠江六个流域的日平均温度和日降水量这两个气候要素,基于历史观测日值资料和 15 个 CMIP5 模式对历史排放情景和三种未来排放情景(RCP2.6、RCP4.5 和 RCP8.5)的模拟月值结果,首先分别建立各个 GCM 的降尺度模型,然后再运用贝叶斯模型平均(BMA)^[3]方法集成 15 个降尺度模型,最后通过随机模拟得到各流域在历史排放情景和未来三种排放情景下气候要素的样本集合。降尺度模型充分考虑了气候要素的空间相关结构。样本集合的分布由各单一 GCM 降尺度模型中气候要素的分布,以及 15 个 GCM 之间的差异所造成的不确定性共同决定。

2 数据集元数据简介

我国东部季风区主要流域气候变化情景下日平均温度和日降水量随机模拟数据集(EAMA-CMIP5-Tas&Pr)^[4]的名称、作者、地理区域、数据年代、时间分辨率、空间分辨率、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表 1。

3 数据研发方法

3.1 数据来源

建立降尺度模型所需的训练样本,包括输入变量和响应变量两部分。输入变量样本为 CMIP5 模式对历史排放情景的模拟月值输出,下载自 CMIP5 的分布式数据节点网络(<https://esgf-node.llnl.gov/search/cmip5/>)^[5]。响应变量为日平均温度和日降水量的站点观测数据,由中国气象局气象数据中心(<http://data.cma.cn/>)提供。

3.2 算法原理

作为随机变量,日平均温度和日降水量有着截然不同的分布特征。日平均温度近似服从正态分布;而日降水量是非负的但包含了大量 0 值,可以用 Tweedie 分布来描述^[7]。经反复试验,两者的降尺度模型采用不同的方法建立:用单一的正态分布的广义加性模型(GAM)^[8]建立整个流域内多站点日平均温度的降尺度模型,而用 Tweedie 分布的广义线性模型(GLM)^[9]逐一建立流域内单站点日降水量的降尺度模型。根据降尺度模型的随机化分位数残差(RQR)^[10]构造流域内日平均温度或日降水量的空间相关模型,用于流域内网格点上多变量联合分布的随机模拟。BMA 方法用于集成多个 GCM 的降尺度模型,确定各自的权重,以此为概率每日抽取某一 GCM 的降尺度随机模拟作为当日的结果。有关方法的详细描述请参阅文献[11]。

表 1 我国东部季风区主要流域气候变化情景下日平均温度和日降水量随机模拟数据集元数据简表

条目	描述
数据集名称	我国东部季风区主要流域气候变化情景下日平均温度和日降水量随机模拟数据集
数据集短名	EAMA-CMIP5-Tas&Pr
作者信息	杨赤 E-4997-2017, 北京师范大学全球变化与地球系统科学研究院, chi@bnu.edu.cn
地理区域	海河流域: 34.5 °N-42.5 °N, 112.5 °E-119.5 °E; 淮河流域: 30.5 °N-36.5 °N, 111.5 °E-121.5 °E; 辽河流域: 38.5 °N-45 °N, 116.5 °E-125.5 °E; 太湖流域: 29.5 °N-32.5 °N, 118.5 °E-121.5 °E; 渭河流域: 33.5 °N-37.5 °N, 104 °E-110 °E; 珠江流域东部: 20.5 °N-27.5 °N, 109.5 °E-116.5 °E; 珠江流域西部: 20.5 °N-27.5 °N, 102.5 °E-109 °E
数据年代	历史排放情景 (Historical): 1970-2000; 未来排放情景 (RCPx.x): 2020-2050
时间分辨率	1 d
空间分辨率	0.5°×0.5°
数据格式	.nc (NetCDF)
数据量	11.3 GB (压缩后)
数据集组成	数据集文件以 “[变量]_[流域]_[情景]_[起始年]_[终止年].nc” 的格式命名, 其中[变量]为日平均温度 (tas) 或日降水量 (pr), 以四维数组的形式存储。四个维度分别为集成员 (ens)、时间 (time)、纬度 (lat) 和经度 (lon)。每个变量集合均包含 50 个成员。集合的不确定性由集合的第 5、10、50、90 和 95 百分位数来表征 (pr 只用后三个), 变量名为 q5、q10、q50、q90 和 q95, 都是维度为 time、lat 和 lon 的三维数组。文件可由 GrADS 软件 (http://cola.gmu.edu/grads/) 直接读取
基金项目	中华人民共和国科学技术部 (2010CB428400)
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101, 中国科学院地理科学与资源研究所
数据共享政策	全球变化科学研究数据出版系统的“数据”包括元数据 (中英文)、实体数据 (中英文) 和通过《全球变化数据学报》(中英文) 发表的数据论文。其共享政策如下: (1) “数据”以最便利的方式通过互联网系统免费向全社会开放, 用户免费浏览、免费下载; (2) 最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源; (3) 增值服务用户或以任何形式散发和传播 (包括通过计算机服务器) “数据”的用户需要与《全球变化数据学报》(中英文) 编辑部签署书面协议, 获得许可; (4) 摘取“数据”中的部分记录创作新数据的作者需要遵循 10% 引用原则, 即从本数据集中摘取的数据记录少于新数据集总记录量的 10%, 同时需要对摘取的数据记录标注数据来源 ^[6]

3.3 技术路线

生成降尺度随机模拟数据的技术路线如图 1 所示。首先用 M 个 GCM 输出的数据集和流域内 N 个站点的观测数据集, 分别拟合 M 个日平均温度降尺度 GAM, 或 $M \times N$ 个日降水量降尺度 GLM (即每个站点拟合一个 GLM), 然后根据模型拟合得到的 RQR 构造 M 个空间相关模型。与此同时, 用 BMA 方法得到 M 个模型加权平均的权重, 以此为概率抽取每日用于降尺度的 GCM, 最后用相应的空间相关模型在流域内网格点上模拟当日的平均温度或降水量。

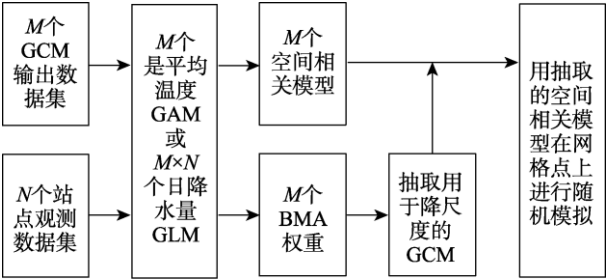


图 1 日平均温度和日降水量降尺度随机模拟技术路线图

4 数据结果与验证

4.1 数据集组成

随机模拟结果为在历史（Historical）和未来（RCP2.6、RCP4.5、RCP8.5）气候变化情景下，各流域内 0.5°×0.5° 的日平均温度和日降水量包括 50 个成员的集合，以期能够反映气候要素分布的变化，捕捉气候极值变化的信息。同时还提供了集合的第 5、10、50、90 和 95 百分位数（日降水量只用后三个）。数据集包含的数据文件见表 2。

图 2 为根据历史情景模拟结果得到的各流域在历史时期内多年平均的日平均温度和年降水量，反映了各流域的基准气候态。

表 2 数据集所含数据文件

流域	日平均温度	日降水量
海河	tas_haihe_historical_1970-2000.nc	pr_haihe_historical_1970-2000.nc
	tas_haihe_rcp26_2020-2050.nc	pr_haihe_rcp26_2020-2050.nc
	tas_haihe_rcp45_2020-2050.nc	pr_haihe_rcp45_2020-2050.nc
	tas_haihe_rcp85_2020-2050.nc	pr_haihe_rcp85_2020-2050.nc
淮河	tas_huaihe_historical_1970-2000.nc	pr_huaihe_historical_1970-2000.nc
	tas_huaihe_rcp26_2020-2050.nc	pr_huaihe_rcp26_2020-2050.nc
	tas_huaihe_rcp45_2020-2050.nc	pr_huaihe_rcp45_2020-2050.nc
	tas_huaihe_rcp85_2020-2050.nc	pr_huaihe_rcp85_2020-2050.nc
辽河	tas_liaohe_historical_1970-2000.nc	pr_liaohe_historical_1970-2000.nc
	tas_liaohe_rcp26_2020-2050.nc	pr_liaohe_rcp26_2020-2050.nc
	tas_liaohe_rcp45_2020-2050.nc	pr_liaohe_rcp45_2020-2050.nc
	tas_liaohe_rcp85_2020-2050.nc	pr_liaohe_rcp85_2020-2050.nc
太湖	tas_taihu_historical_1970-2000.nc	pr_taihu_historical_1970-2000.nc
	tas_taihu_rcp26_2020-2050.nc	pr_taihu_rcp26_2020-2050.nc
	tas_taihu_rcp45_2020-2050.nc	pr_taihu_rcp45_2020-2050.nc
	tas_taihu_rcp85_2020-2050.nc	pr_taihu_rcp85_2020-2050.nc
渭河	tas_weihe_historical_1970-2000.nc	pr_weihe_historical_1970-2000.nc
	tas_weihe_rcp26_2020-2050.nc	pr_weihe_rcp26_2020-2050.nc
	tas_weihe_rcp45_2020-2050.nc	pr_weihe_rcp45_2020-2050.nc
	tas_weihe_rcp85_2020-2050.nc	pr_weihe_rcp85_2020-2050.nc
珠江（东）	tas_zhujiang-east_historical_1970-2000.nc	pr_zhujiang-east_historical_1970-2000.nc
	tas_zhujiang-east_rcp26_2020-2050.nc	pr_zhujiang-east_rcp26_2020-2050.nc
	tas_zhujiang-east_rcp45_2020-2050.nc	pr_zhujiang-east_rcp45_2020-2050.nc
	tas_zhujiang-east_rcp85_2020-2050.nc	pr_zhujiang-east_rcp85_2020-2050.nc
珠江（西）	tas_zhujiang-west_historical_1970-2000.nc	pr_zhujiang-west_historical_1970-2000.nc
	tas_zhujiang-west_rcp26_2020-2050.nc	pr_zhujiang-west_rcp26_2020-2050.nc
	tas_zhujiang-west_rcp45_2020-2050.nc	pr_zhujiang-west_rcp45_2020-2050.nc
	tas_zhujiang-west_rcp85_2020-2050.nc	pr_zhujiang-west_rcp85_2020-2050.nc

4.2 数据结果

随机模拟结果用集合的均值、标准差、极值等统计量，从空间分布、季节变化和年际变化等角度进行了分析验证，结果详见文献[11]。图 3 给出了以 1981–2000 年为基准期，在

三种 RCP 情景下,2031–2050 年各流域平均的日平均温度和年降水量相对于基准期的变化。由图可见，各流域日平均温度均有上升，RCP8.5 情景下上升幅度最大，而珠江流域东部

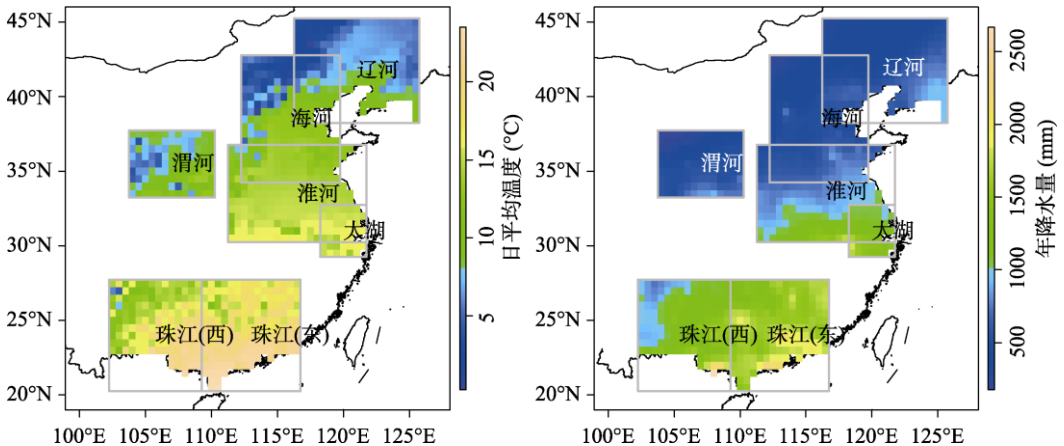


图 2 1970–2000 年历史情景模拟结果的各流域多年平均的日平均温度和年降水量

居首，达 0.94 °C。太湖流域的上升幅度最小，RCP2.6 情景下仅 0.20 °C。除珠江流域东部外，各流域年降水量亦均有增加，同样在 RCP8.5 情景下增加幅度最大，而珠江流域西部居首，达 291.3 mm。珠江流域东部的年降水量在各种情景下均有小幅减少，在 RCP2.6 情景下减少幅度最大，约 28.3 mm。

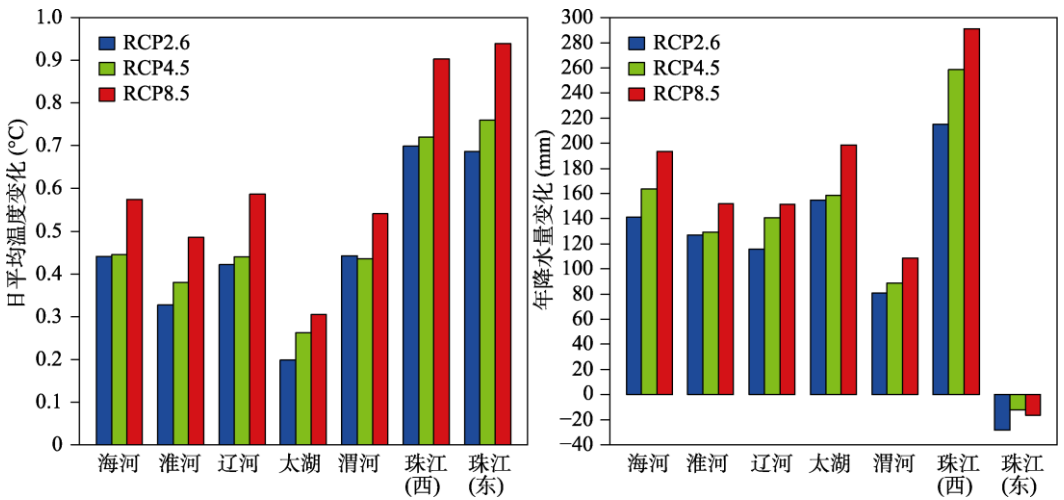


图 3 2031–2050 年各流域平均的日平均温度和年降水量相对于基准期（1981–2000 年）的变化

5 讨论和总结

本数据集应用了诸多较为复杂的统计模型，如 Tweedie 分布、GAM、GLM、BMA 和 Kriging 等，并经过了较大规模的统计计算方得以生成，旨在能够反映日或日以下尺度的气候要素在气候变化情景下其分布的变化以及多模式预测的不确定性，特别是气候极值变化

的信息。数据集可作为水文、生态和环境等领域的基础数据,用于水文工程设计、灾害风险评估等目的。其不足之处是在不同区域的接壤或重叠地区数据间可能存在不一致性。在模型和算法方面还有诸多可改进之处,但会受制于计算能力。对数据集的初步分析表明,在三种 RCP 情景下,2031–2050 年各流域平均的日平均温度和除珠江流域东部以外的年降水量相对于基准期均有增加,在 RCP8.5 情景下增幅最大;珠江流域东部的年降水量有所减少,在 RCP2.6 情景下减幅最大。

参考文献

- [1] Taylor, K. E., Stouffer, R. J., Meehl, G. A. An overview of CMIP5 and the experiment design [J]. *Bulletin of the American Meteorological Society*, 2012, 93: 485–498.
- [2] Schoof, J. T. Statistical downscaling in climatology [J]. *Geography Compass*, 2013, 7(4): 249–265.
- [3] Raftery, A. E., Gneiting, T., Balabdaoui, F., *et al.* Using Bayesian model averaging to calibrate forecast ensembles [J]. *Monthly Weather Review*, 2005, 133: 1155–1174.
- [4] 杨赤. 我国东部季风区主要流域气候变化情景下日平均温度和日降水量随机模拟数据集[DB/OL]. 全球变化科学研究数据出版系统, 2018. DOI: 10.3974/geodb.2018.01.02.V1.
- [5] <https://esgf-node.llnl.gov/search/cmip5/>.
- [6] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策 [OL]. DOI: 10.3974/dp.policy.2014.05 (2017 年更新).
- [7] Dunn, P. K. Occurrence and quantity of precipitation can be modelled simultaneously [J]. *International Journal of Climatology*, 2004, 24: 1231–1239.
- [8] Hastie, T. J., Tibshirani, R. J. Generalized Additive Models [M]. London: Chapman and Hall, 1990.
- [9] McCullagh, P., Nelder, J. A. Generalized Linear Models (2nd Edition) [M]. London: Chapman and Hall, 1989.
- [10] Dunn, P. K., Smyth, G. K. Randomized quantile residuals [J]. *Journal of Computational and Graphical Statistics*, 1996, 5(3): 236–244.
- [11] 杨赤. 气候变化情景下气象要素的随机模拟[M]. 段青云, 徐宗学等著. 未来水文气候情景预估及不确定性分析与量化. 北京: 科学出版社, 2017: 73–95.