

基于梁氏信息流的因果人工智能对7月降水预报

容逸能¹, 马继望², 赵宇慧^{2,3}, 张殷宸⁴, 张蓝心⁴, 付冠琦³, 何蔚邦³, 郭子彧⁴, 徐芬⁵, 杨洋⁶, 李渊⁷, 马伟翔⁸

¹复旦大学大气与海洋科学系, 上海, 200438; ²南方海洋科学与工程广东省实验室(珠海), 珠海, 519000; ³南京信息工程大学海洋科学学院, 南京, 210044; ⁴南京信息工程大学大气科学学院, 南京, 210044; ⁵国防科技大学气象海洋学院, 长沙, 410073; ⁶厦门大学海洋与地球学院, 厦门, 361005, ⁷台州市气象局, 台州, 318000; ⁸华为软件研发中心, 深圳, 518219

关键字: 信息流, 因果分析, 降水预报, 人工智能

受季风气候影响, 我国降水主要集中在夏季汛期。特别是近年来, 极端降水事件越发严重^[1], 并多发生在7月^[2-5]。因此, 对针对7月的次月降水预报是十分重要的。根据1990-2021年的历史数据, 我们利用梁氏信息流因果理论和人工智能算法对我国7月降水展开预报。

预报采用中/美国家气候中心的130项/40项每月气候检测监测指数、26项NSSTC的海温指数等, 基于NCEP再分析数据计算的各类土壤指数(约200余项)和NOAA的海温数据计算的16个海区的海温指数, 结合国家气候中心提供的2288观测站月降水数据。其中各类指数共计411项, 这些指数之间往往高度共线性并且有很高的相关性。因此需要采用一种合适的方法来筛选真正能够用于预报的因子。

不同于其他半经验性的统计学因果推断方法, 基于梁氏信息流的因果分析是由第一性原理出发严格推导的一套近乎无任何假设的因果推断工具^{[6]-[8]}, 这套源自动力气象的理论因其严谨性已大量应用于不同的学科领域, 尤其是最近的量子信息领域^[9]。对于一个包含 d 个节点 $(X_1, X_2, \dots, X_d) \equiv \mathbf{X}$ 的网络, 其演化可由

d 维系统 $d\mathbf{X}/dt = \mathbf{F}(\mathbf{X}, t) + \mathbf{B}(\mathbf{X}, t) \dot{\mathbf{W}}$ 描述，其中 $\dot{\mathbf{W}}$ 为噪声， $\mathbf{B} = (b)_{ij}$ 为随机扰动振幅矩阵。Liang^[10]证明了由 X_j 至 X_i 的信息流为：

$$T_{j \rightarrow i} = -E \left(\frac{1}{\rho_i} \int_{\mathbb{R}^{d-2}} \frac{\partial F_i \rho_{\mathbf{X}}}{\partial x_i} d\mathbf{x}_{\mathbf{X}} \right) + \frac{1}{2} E \left(\frac{1}{\rho_i} \int_{\mathbb{R}^{d-2}} \frac{\partial^2 (g_{ii} \rho_{\mathbf{X}})}{\partial x_i^2} d\mathbf{x}_{\mathbf{X}} \right). \quad (1)$$

其中， F_i 为 $\mathbf{F}$ 的分量， $g_{ii} = \sum_k b_{ik} b_{ik}$ ， ρ 为概率密度函数， E 为数学期望算子，其余的符号含义参见[10]。Liang^[10]还证明了若控制 X_i 演变的 F_i 和 g_{ii} 不依赖于 X_j ，即 X_j 不是 X_i 的因，那么 X_j 对 X_i 的信息流严格为零。在（1）式中，这一特性是在不需要任何约束条件的情况下，作为定理被证明出来，这也是量化因果中最重要的**零因果准则**。在实际应用中，为方便起见，通常对（1）采用其线性化的估计^[11]：

$$\hat{T}_{j \rightarrow i} = \frac{C_{ij}}{C_{ii} \det \mathbf{C}} \sum_{k=1}^d \Delta_{jk} C_{k, di}. \quad (2)$$

其中 C_{ij} 表示 x_i 和 x_j 之间的协方差， $C_{i, dj}$ 表示 x_i 和 $\dot{x}_j$ 之间的协方差， $\dot{x}_j$ 是 dx_j/dt 的欧拉前差近似， Δ_{ij} 是矩阵 $\mathbf{C} = (C_{i, j})$ 的代数余子式。这样，利用各气候因子的时间序列即可以快速得到它们对降水的信息流，从而判断各个因子对降水的定量因果。

考虑到我国观测站点分布不均，在沿海地区远大于西部地区，为了尽可能捕捉到数据的空间特征，本文采用Rong 和 Liang^[12]提出的因果人工智能方法，利用主成分分析，得到预报的主成分时序，再利用因果推断得到主成分时序的重要因子，最后代入机器学习中训练模型。类似地，对于降水站点数据，取前90%方差贡献率的主成分的时序用于模型训练。这样，次月降水预报问题由对站点数据的预报转化成对各个主成分的降水预报。由于不同主成分的影响因子不同，我们对每个主成分的序列单独建立预报模型。

表 1: 2021-2022 年 7 月降水预报的距平符号一致率（AS），距平相关系数（ACC）和平均方差技巧（MSSS）。第 1 行是我们的因果人工智能方法（IF-AI）；第 2 到 4 行分别是欧洲中心（EC）、美国国家环境预报中心（NCEP）和中国气象局（CMA）的次季节-季节预测模型结果。

	2021			2022		
	AS	ACC	MSSS	AS	ACC	MSSS
IF-AI	0.64	0.48	-0.64	0.62	0.40	0.17
EC	0.58	0.08	-1.65	0.53	0.07	-0.72
NCEP	0.60	0.19	-1.63	0.48	0.02	-1.70
CMA	0.59	0.16	-2.21	0.60	0.27	-1.06

以对7月降水第一主成分的预报为例，利用梁氏信息流诊断各个气候因子对第一主成分的因果。注意到，我们降水数据是月尺度的，以全年完整的时序来计算得到的因果将是在全部时段上，各个因子对降水的影响。而这很可能掩盖掉气候因子单单对7月降水的因果。因此，本文采用面板数据形式下的估计方式^[13]来讨论对7月的降水影响。那么以探讨6月Niño 3.4区海温指数对7月降水第一主成分的影响为例，将每年6-7月的海温数据和降水第一主成分的时序分别依序排列，得到两条序列，并记录海温序列每个序号对应的自然年月份。这样形式的数据带入至[13]提供的算法，就可以得到对特定月份降水主成分的定量因果。

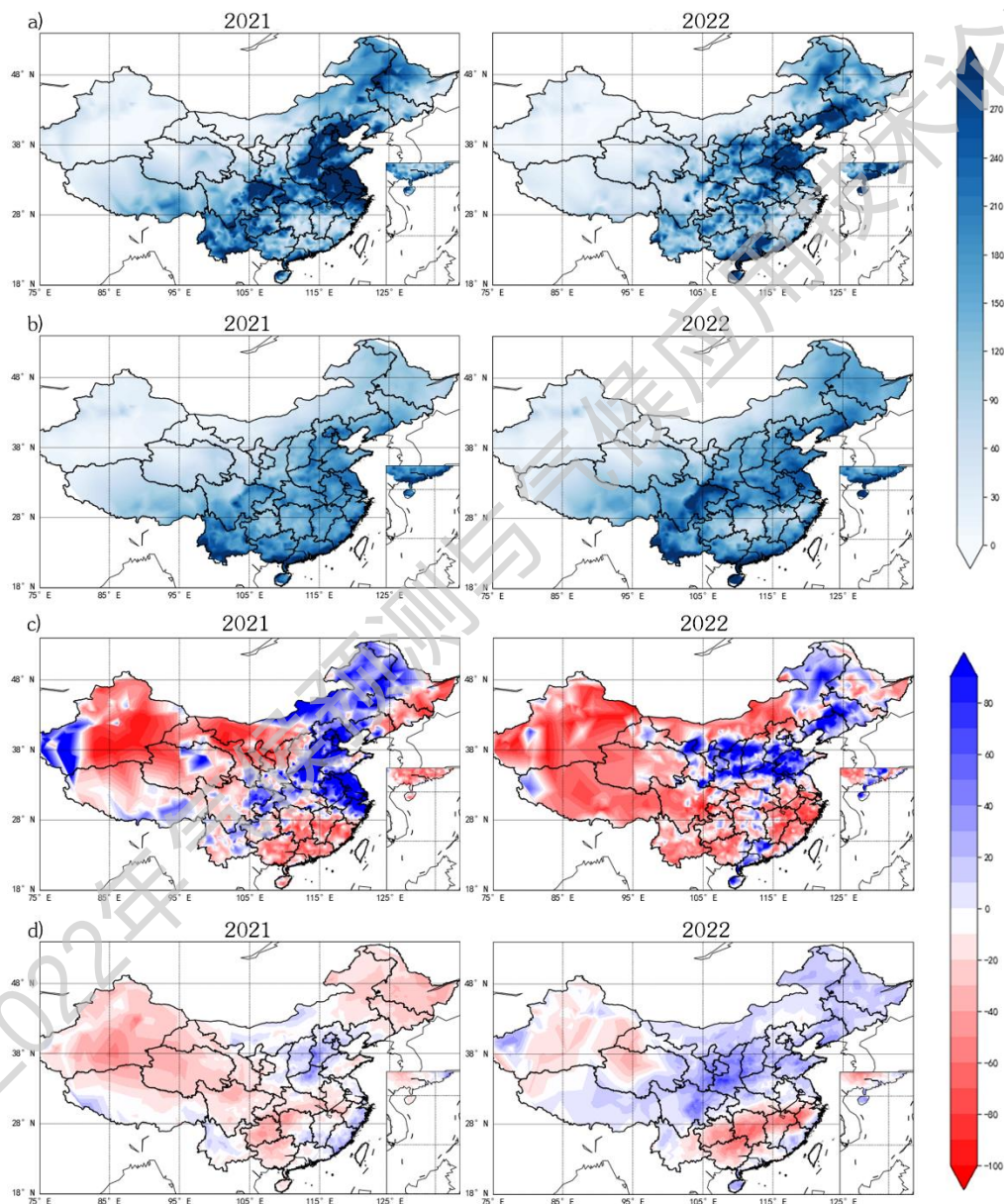


图1: 2021-2022年7月降水量的观测 (a) 和基于因果人工智能方法的预测 (b) ;c, d 同a和b但为降水距平百分比。

根据定量因果的结果，我们使用随机森林算法训练1990-2019年的降水数据，

并使用2020年为验证集，最后得到预报模型。图1给出了2021–2022年7月的因果人工智能模型预报和站点观测。从降水异常百分比中（图1c和1d）可以看出，我们的模型能够大体正确成功预报各个月份的降水的正负异常（AS评分大于0.5，表1），并且降水的空间分布与实际过程相似（ACC评分较高，表1），不确定性最小（MSSS评分最高，表1）。包括21年7月河南暴雨事件^[15–17]在内的极端降水都能被模型正确捕捉。特别对于今年长江中下游的罕见干旱，我们的模型也能够稳定地表征该地区持续性的降水负异常。相较于中国气象局（CMA）、欧洲中心（EC）和美国国家环境预报中心（NCEP）的模式预报，我们模型的效果均有明显提升。这显示了我们因果人工智能模型在次月降水预报中的巨大潜力。考虑到总体降水强度仍然显著低于实际观测，特别是对极端降水事件的预报强度和范围仍不及实际观测，未来将使用更精细的时空数据对该问题继续深入研究。

参考文献

- [1] G. Myhre *et al.*, “Frequency of extreme precipitation increases extensively with event rareness under global warming,” *Sci. Rep.*, vol. 9, no. 1, Art. no. 1, Nov. 2019, doi: 10.1038/s41598-019-52277-4.
- [2] 翟盘茂, 王萃萃, and 李威, “极端降水事件变化的观测研究,” 气候变化研究进展, no. 03, pp. 144–148, 2007.
- [3] D. Zhang, F. Guo-Lin, and H. Jing-Guo, “Trend of extreme precipitation events over China in last 40 years,” *Chin. Phys. B*, vol. 17, p. 736, Feb. 2008, doi: 10.1088/1674-1056/17/2/062.
- [4] D. Zhang, W. Wang, S. Liang, and S. Wang, “Spatiotemporal Variations of Extreme Precipitation Events in the Jinsha River Basin, Southwestern China,” *Adv. Meteorol.*, vol. 2020, p. e3268923, Oct. 2020, doi: 10.1155/2020/3268923.
- [5] X. Gu *et al.*, “Extreme Precipitation in China: A Review on Statistical Methods and Applications,” *Adv. Water Resour.*, vol. 163, p. 104144, May 2022, doi: 10.1016/j.advwatres.2022.104144.
- [6] X. S. Liang and R. Kleeman, “Information transfer between dynamical system components,” *Phys. Rev. Lett.*, vol. 95, no. 24, p. 244101, Dec. 2005, doi: 10.1103/PhysRevLett.95.244101.
- [7] X. S. Liang, “Information flow within stochastic dynamical systems,” *Phys. Rev. E*, vol. 78, no. 3, p. 031113, Sep. 2008, doi: 10.1103/PhysRevE.78.031113.
- [8] X. S. Liang, “The Liang-Kleeman information flow: theory and applications,” *Entropy*, vol. 15, no. 1, pp. 327–360, Jan. 2013, doi: 10.3390/e15010327.
- [9] B. Yi and S. Bose, “Quantum Liang Information Flow as Causation Quantifier,” *Phys. Rev. Lett.*, vol. 129, no. 2, p. 020501, Jul. 2022, doi: 10.1103/PhysRevLett.129.020501.
- [10] X. S. Liang, “Information flow and causality as rigorous notions ab initio,” *Phys. Rev. E*, vol. 94, no. 5, p. 052201, Nov. 2016, doi: 10.1103/PhysRevE.94.052201.
- [11] X. S. Liang, “Normalized multivariate time series causality analysis and causal graph reconstruction,” *Entropy*, vol. 23, p. 679, May 2021, doi: 10.3390/e23060679.
- [12] Y. Rong and X. S. Liang, “An information flow-based sea surface height reconstruction through machine learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, p. 4205909, 2022, doi: 10.1109/TGRS.2022.3140398.
- [13] Y. Rong and X. S. Liang, “Panel Data Causal Inference Using a Rigorous Information Flow

Analysis for Homogeneous, Independent and Identically Distributed Datasets,” *IEEE Access*, vol. 9, pp. 47266–47274, 2021, doi: 10.1109/ACCESS.2021.3068273.

[14] 齐道日娜, 何立富, 王秀明, and 陈双, ““7·20”河南极端暴雨精细观测及热动力成因,” *应用气象学报*, vol. 33, no. 01, pp. 1–15, 2022.

[15] J. Rao, J. Xie, Y. Cao, S. Zhu, and Q. Lu, “Record Flood-Producing Rainstorms of July 2021 and August 1975 in Henan, China: Comparative Synoptic Analysis Using ERA5,” *J. Meteorol. Res.*, 2022, doi: 10.1007/s13351-022-2066-6.

[16] H. Qin *et al.*, “Climate change attribution of the 2021 Henan extreme precipitation: Impacts of convective organization,” *Sci. China Earth Sci.*, vol. 65, no. 10, pp. 1837–1846, Oct. 2022, doi: 10.1007/s11430-022-9953-0.

2022年气候预测与气候应用技术论坛