

基于因果人工智能的2022年极端干旱预报

马继望^{1,2,3}, 容逸能^{2,1}, 赵宇慧^{4,1}, 付冠琦⁴, 张殷宸⁵,
郭子彧⁵, 何蔚邦⁵, 杨洋⁶, 徐芬⁷, 李渊⁸, 马伟翔⁹

1. 南方海洋实验室前沿研究中心, 珠海, 519000
2. 复旦大学大气与海洋科学系, 上海, 200438
3. 山东省气象台, 济南, 250031
4. 南京信息工程大学海洋科学学院, 南京, 210044
5. 南京信息工程大学大气科学学院, 南京, 210044
6. 厦门大学海洋与地球学院, 厦门, 361005
7. 国防科技大学气象海洋学院, 长沙, 410073
8. 台州市气象局, 台州, 318000
9. 华为软件研发中心, 深圳, 518219

关键字: 信息流, 因果分析, 极端干旱, 人工智能

2022年夏季, 长江流域遭受了极端干旱事件, 甚至出现了水位在汛期不断退落的反常现象, 鄱阳湖和洞庭湖更是在汛期就进入了枯水期^[1]。此次干旱给长江沿线的农业、畜牧业以及人们的生产生活都造成了显著的不利影响^[1]。众所周知, 提前预报并部署防范, 是减轻灾害性天气不利影响的有效途径。然而, 不同于一周以内的天气预报, 跨月的季节内预报是现在预报业务中的瓶颈, 通常难以准确预报。其中, 最根本的原因是大气混沌特性, 该特性使得利用数值模式预报天气的时效难以突破两周的上限^[2]。近年来, 机器学习尤其是深度学习技术迅猛发展, 能否利用该技术在跨月的天气预报中取得突破成为一个有待回答的热点问题, 本文将简要介绍作者针对该问题所做的一些工作及其应用前景。

对各类不同类型动力系统的预报是机器学习研究的一个前沿问题, 其中, 混沌系统是非常特殊且具有挑战的一类, 常见的机器学习方法难以对混沌系统进行预报。可喜的是, 最新的研究对其背后的原因有了新的认识——混沌系统无

法预报的根本原因是训练中缺乏因果关系的约束^[3]。大气系统是典型的混沌系统，这意味着利用机器学习进行天气预报必须考虑因果的约束。为此，我们开发了因果机器学习算法，在机器学习中加入了因果的约束。其中，所用的因果分析方法是从小概率原理出发严格推导出的一套近乎无任何假设的因果推断工具^[4, 7]，具有清晰明确的物理意义，这与其他半经验性的统计学因果推断方法有本质不同。因其严谨性和普适性，这套源自动力气象领域的理论已大量应用于不同的学科领域，尤其是最近广受关注的量子信息领域^[5]，下面对该方法进行简要介绍。

对于一个包含 n 个节点 $(X_1, X_2, \dots, X_n) \equiv \mathbf{X}$ 的网络，其演化可由 n 维系统 $\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}, t) + \mathbf{B}(\mathbf{X}, t) \dot{\mathbf{W}}$ 描述，其中 $\dot{\mathbf{W}}$ 表示噪声， $\mathbf{B} = (b)_{ij}$ 为随机扰动振幅矩阵。

Liang (2016) 证明由 X_j 至 X_i 的信息流为^{[6], [7]}：

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

其中， F_i 为 \mathbf{F} 的分量， $g_{ii} = \sum_k b_{ik} b_{ik}$ ， ρ 表示概率密度函数， E 为数学期望，其他符号含义参见^[7]。Liang (2016) 还证明了如果控制 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，则 X_j 对 X_i 的信息流严格等于零。在 (1) 式中，这一特性在不需要任何约束的情况下，作为定理被证明出来，这即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) 采用其线性化的最大似然估计^[8]：

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

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 为 dx_j/dt 的欧拉前差， Δ_{ij} 为矩阵 $\mathbf{C} = (C_{i, j})$ 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，从而判断各因子对降水的定量因果。

利用基于上述因果分析工具建立的因果人工智能算法，本研究完全按照业务实践的标准，对2022年9月的月降水量进行了预报，发现该方法能够提前一个月以上准确地预报出长江流域的极端干旱事件的持续，如图1所示（图中红色部分表示干旱）。预报结果显示，长江中下游沿江一带有显著的干旱，包括湖北、湖南、江西、安徽、浙江和江苏等省份，这与后来的实际干旱情况相符。值得一提的是，该预报还准确报出了江西北部鄱阳湖区域的干旱中心，这说明基于因果人工智能算法的预报不仅报出了总体的干旱分布，对于局地的干旱中心也具有指示意义。

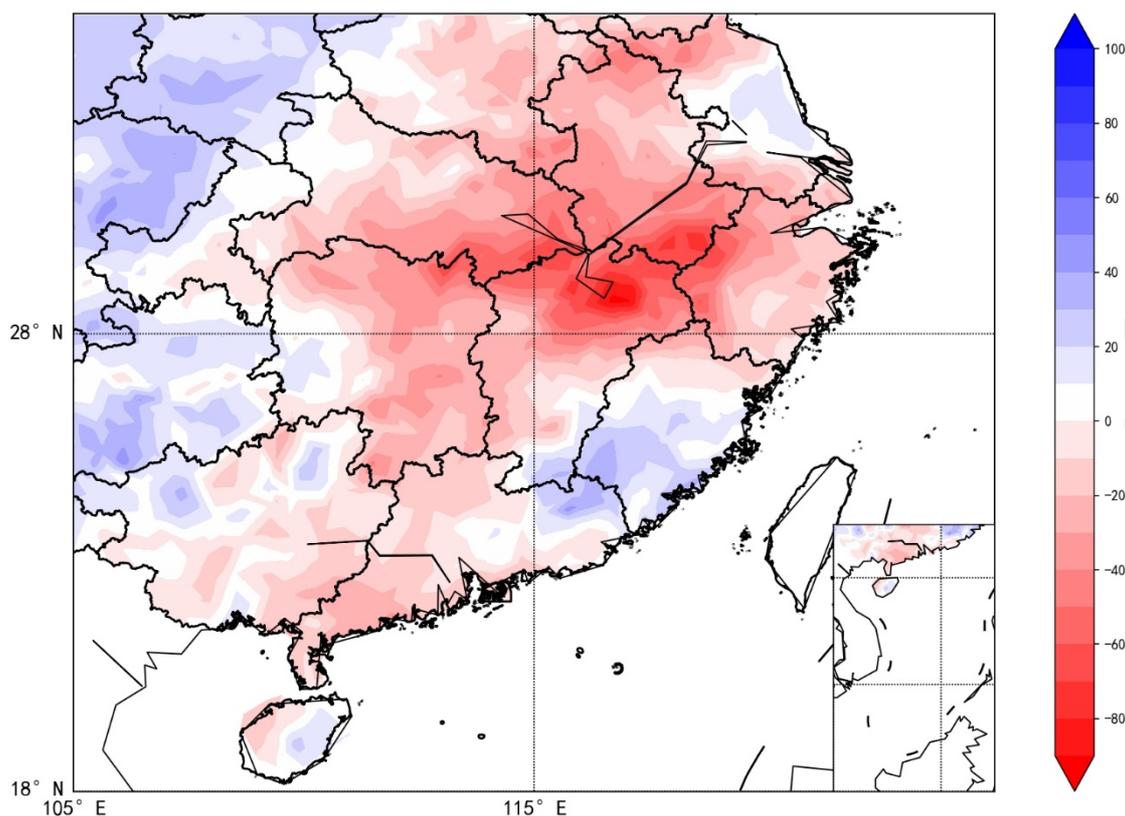


图1: 基于因果人工智能方法预报的2022年9月降水距平百分比。

综上所述, 本研究仅从数据角度出发, 基于国家气候中心的气候监测指数、海温指数等资料和因果人工智能算法, 建立了次月降水预报模型, 其结果具有较高的准确性, 准确预报出了今年的极端干旱事件, 有望为极端事件预报、决策服务和防灾减灾提供信息支撑。此外, 该方法不限于降水的预报, 将来还可以应用到强对流、极端高低温、灾害性大风等多要素的高影响天气预报中。

参考文献

- [1] 夏军, 陈进, 余敦先, “2022 年长江流域极端干旱事件及其影响与对策[J],” 水利学报, 2022, 53, 1-11.
- [2] E. N. Lorenz, “Deterministic Nonperiodic Flow,” *J. Atmospheric Sci.*, vol. 20, no. 2, pp. 130–141, 1963.
- [3] S. Wang, S. Sankaran, and P. Perdikaris, “Respecting causality is all you need for training physics-informed neural networks,” *ArXiv220307404 Nlin Physicsphysics Stat*, 2022.
- [4] X. S. Liang and R. Kleeman, “Information transfer between dynamical system components,” *Phys. Rev. Lett.*, vol. 95, no. 24, p. 244101, 2005.
- [5] B. Yi and S. Bose, “Quantum Liang Information Flow as Causation Quantifier,” *Phys. Rev. Lett.*, vol. 129, no. 2, p. 020501, 2022.
- [6] X. S. Liang, “The Liang-Kleeman information flow: theory and applications,” *Entropy*, vol. 15, no. 1, pp. 327–360, 2013.

- [7] X. S. Liang, "Information flow and causality as rigorous notions ab initio," *Phys. Rev. E*, vol. 94, no. 5, p. 052201, 2016.
- [8] X. S. Liang, "Normalized multivariate time series causality analysis and causal graph reconstruction," *Entropy*, vol. 23, p. 679, 2021.

2022年气候预测与气候应用技术论坛