

古气候数据同化研究进展及展望

宁亮¹, 刘健^{1,2*}, Zhengyu LIU³, 邢芳淼¹, 吴芬¹, 严蜜¹, Zilu MENG⁴, 陈可凡¹, 覃燕敏¹, 孙炜毅¹, 温琴¹

1. 南京师范大学气候系统预测与变化应对全国重点实验室, 南京师范大学虚拟地理环境教育部重点实验室, 江苏省地理环境演化与区域应对国家重点实验室培育建设点, 江苏省地理信息资源开发与利用协同创新中心, 南京师范大学地理科学学院, 南京 210023

2. 江苏省大规模复杂系统数值模拟重点实验室, 南京师范大学数学科学学院, 南京 210023

3. Department of Geography, The Ohio State University, Columbus 43210, USA

4. Department of Atmospheric and Climate Science, Seattle 98195, USA

* 通讯作者, E-mail: jliu@njnu.edu.cn

2025-08-21收稿, 2025-12-08修回, 2025-12-22接受, 2026-02-09网络版发表

国家重点研发计划项目(2023YFF0804704)和国家自然科学基金项目(42130604、42575050、42475051、42575051)资助

摘要 重建古气候特征并认识古气候演变规律是地球系统科学与全球变化领域的热点问题之一, 有助于认清现代暖期的历史地位, 理解暖期背景下气候变化的特征及机制, 从而改进未来气候预估精度。代用资料和数值模拟是当前古气候研究的两种主要手段, 而古气候数据同化作为一种新兴的研究方法, 有机结合古气候代用资料和数值模拟, 以充分融合两者优势, 进而提高古气候重建的准确性。本文系统总结了近年来国内外古气候数据同化领域的最新研究进展。首先回顾了古气候数据同化方法的历史沿革, 介绍了主要同化方法的优缺点和适用性, 以及最近几年同化方法的改进, 例如机器学习方法的应用及在线同化的发展; 然后按照古气候特征时段介绍了古气候数据同化的应用, 特别是不同代用资料中的相应难点, 并总结了当前开源的数据集和算法平台; 然后用一个具体案例给出了耦合氧同位素的模拟结果在古气候同化的应用过程; 最后, 本文还探讨了古气候数据同化研究中未解决的问题及挑战, 展望了未来的研究方向。

关键词 古气候数据同化, 代用资料, 模拟结果, 代用资料系统模型, 不确定性

1 引言

古气候研究有助于理解气候变化机制、确定影响环境的关键气候因子、评估当代气候数值模式, 从而改进未来气候和环境变化预测, 服务可持续发展(王会军, 2022)。古气候代用资料和模式模拟是古气候研究中的两种主要研究方法。两者各有自己的优势, 古气候代用资料是古气候史实的反应, 而模拟结果中包括了动力机制。

古气候代用资料是用来重建古气候的各种生物、物理、化学指标, 过去两千年时段代用资料数据集包括PAGES2k数据集(Past Global Changes 2k)(PAGES2k Consortium, 2017)、Mann09数据集(Mann等, 2009), 全新世的代用资料数据集包括Temperature 12ka(Kaufman等,

2020)、极地全新世代用资料气候数据集(Arctic Holocene Proxy Climate Database)(Sundqvist等, 2014)、LegacyClimate 1.0数据集(Herzschuh等, 2023)等。此外还有各种代用资料专题的集成数据集, 例如珊瑚记录CoralHydro2k数据集(Walter等, 2023)、石笋记录SASIL数据集(Comas-Bru等, 2020)等。这些数据集为过去不同特征时段的古气候重建做出了贡献。

在古气候模拟方面, 国际古气候模拟比较计划(Paleoclimate Modelling Intercomparison Project, PMIP)一直致力于通过模拟来理解过去气候的状态和对外强迫的响应, 目前, 最新的PMIP4模拟结果已发布, 包含了过去千年的瞬变模拟(past1000), 以及中全新世(Mid-Holocene, MH)、末次盛冰期(Last Glacial Maximum,

中文引用格式: 宁亮, 刘健, Liu Z, 邢芳淼, 吴芬, 严蜜, Meng Z, 陈可凡, 覃燕敏, 孙炜毅, 温琴. 2026. 古气候数据同化研究进展及展望. 中国科学: 地球科学, doi: 10.1360/SSTe-2025-0255

英文引用格式: Ning L, Liu J, Liu Z, Xing F, Wu F, Yan M, Meng Z, Chen K, Qin Y, Sun W, Wen Q. 2026. Progress and prospects of paleoclimate data assimilation. Science China Earth Sciences, <https://doi.org/10.1007/s11430-025-1810-2>

LGM)、末次间冰期(Last Interglacial, LIG)、中上新世(mid-Pliocene)、始新世早期,以及古新世晚期EEOC(Early Eocene)、PETM(Paleocene-Eocene Thermal Maximum)和pre-PETM(pre-Paleocene-Eocene Thermal Maximum)三个典型暖期等平衡态模拟(Kageyama等, 2018)。而连续的瞬变模拟包括了过去千年多成员集合模拟试验(Community Earth System Model-Last Millennium Ensemble, CESM-LME)(Otto-Bliesner等, 2016)、南京师范大学过去两千年模拟试验(NNU-2ka)(王志远等, 2016)等。全新世时段的连续瞬变模拟包括南京师范大学全新世模拟试验(NNU-Holocene)(万凌峰等, 2020)、中国科学院大气物理研究所的HT-11.5ka试验(Tian等, 2020)、德国马克斯-普朗克气象研究所基于MPI-ESM模式的试验(Bader等, 2020)、英国哈德莱中心基于HadCM3模式的试验(Hopcroft和Valdes, 2021, 2022)、瑞典斯德哥尔摩大学基于EC-Earth模式的试验(Zhang等, 2021)等; LGM以来的连续瞬变模拟试验包括TraCE-21ka试验(Simulation of the Transient Climate of the Last 21000 Years)(Liu等, 2009),以及最新的耦合氧同位素的iTraCE-21ka试验(Isotope-enabled Simulation of the Transient Climate of the Last 21000 Years)(He等, 2021); 还有涵盖过去30万年的瞬变加速模拟试验(Xie等, 2019; Yan等, 2023)。这些模拟结果为理解不同特征时段的古气候变化机制提供了帮助。

但两者又有着各自的劣势,例如古气候代用资料多是单点资料,时空上不连续且分布不均匀;存在定年误差,且部分代用资料的气候指示意义存在多解性和争议(陈发虎等, 2023);主要反映温度或降水这些地面变量。除此之外,古气候代用资料还包含两类误差,即采样和分析过程中由于人为因素或者测量分析仪器本身所存在的不确定性而导致的器测误差,和代用资料所能代表的空间范围与气候模式网格所代表的空间范围不一致而导致的空间代表性误差(李新, 2013)。而模拟结果多为模式本身的内部变率,无法准确反映古气候中的真实内部变率的位相,并且对外强迫的响应强度存在不确定性。而且,尽管古气候模拟致力于在外强迫和反馈方面更加真实,但当前主流模式研发主要用于未来气候预估和预测,并非重新为古气候而开发和校准的(Kageyama等, 2018),因此在模拟结果的准确性方面可能存在偏差。所以,在古气候研究中需要将两者的优势结合并弥补各自的不足,来获得更准确的古气候特征和规律,这就是古气候数据同化的初衷(von Storch等, 2000)。古气候数据同化提供了一个数学框架,从代用资料和模拟结果中分别提取有用信息,其中代用资料提供真实的气候变化史

实,模拟结果提供基于物理方程的动力约束框架,通过对代用资料和模拟结果误差的定量估计,对古气候模式运行进行约束(或对模拟结果进行直接修正),从而获得更准确的时空连续的古气候重建(Hakim等, 2013)。由此可见,古气候数据同化跟现代气候的数据同化具有共同的动机,即利用空间上不连续的观测来生成具有空间规则格点的再分析数据,特别是没有直接观测的变量。但古气候数据同化又具有其独特的挑战,即同化的对象为代用资料,其代表的物理意义不清楚,由此带来的转换算子和不确定性计算等同化步骤方面的挑战。例如,石笋 $\delta^{18}\text{O}$ 是指示季风强度的代用资料,但它的具体气候表征意义(环流强度、水汽源地变化等)存在争议,而我们希望用它来重建季风降水变化,因此需要借助耦合氧同位素的模拟结果构建非线性的代用资料系统模型(proxy system model, PSM),这也带来了与现代同化有显著差别的特殊挑战。

由于其特有的优势,很多记录和模拟间的争议,例如全新世温度悖论问题(Liu等, 2014),也可以通过同化来进行回答。同化可以通过给出连续的空间场,给出更精准的全球和区域气候变化空间分布,从而有助于理解全球和区域气候对不同强迫的响应。除此之外,同化还可以针对古气候关注的科学问题(例如气候敏感性),通过给出更准确的结果,为未来预估提供更可靠的参考。从而,可以更好地认识现代暖期的历史地位及影响,促进古今气候环境研究的融合(王会军, 2022)。

之前学者已针对古气候同化的原理、方法、应用等方面做了详细的总结(方苗和李新, 2016; 张昊勋等, 2025; Tierney等, 2025b),近年来,古气候数据同化的算法及“在线”同化等方面又有了长足的进步(Sun等, 2022; Meng和Hakim, 2024),特别是氧同位素模拟的引入可以改进非线性PSM及在线同化等,引起了学界的广泛关注。因此,本文将简单回顾古气候数据同化的原理、方法和应用等方面的发展历史,重点介绍最近几年的古气候数据同化新方法和新技术(例如,在线同化策略),然后对当前古气候数据同化中所面临的理论、技术、数据等方面的瓶颈问题进行探讨,在此基础上对古气候数据同化在古气候关键科学问题上的应用及未来应该重点攻关的研究方向进行展望。

2 古气候数据同化方法及沿革

简单来说,古气候数据同化的基本理念是,由代用资料来约束模式运行,结合前一时刻的模拟结果,对当前的气候状态做出最优的估计。其核心基于传统的贝叶斯理论:

$$P(\mathbf{x} | \mathbf{y}) \propto P(\mathbf{y} | \mathbf{x}) \cdot P(\mathbf{x}) \quad (1)$$

其中, \mathbf{x} 表示重建的气候变量, \mathbf{y} 表示代用资料, $P(\mathbf{x})$ 表示模式结果提供的先验概率, $P(\mathbf{y} | \mathbf{x})$ 表示似然函数, 即给定气候状态下的代用资料概率, $P(\mathbf{x} | \mathbf{y})$ 表示同化之后得到的气候变量的后验概率, 即同化结果。由公式可以看出, 最终同化的结果与先验分布和似然有关。古气候数据同化的努力方向也集中于对这两者的估计和求解后验的分布。因此, 古气候数据同化主要包括四个组成部分, 古气候模拟、代用资料、代用资料系统模型、同化算法。具体流程为(图1), 针对某一时刻, 首先, 针对所要同化的变量, 通过古气候模拟生成所需要的数据, 作为同化的先验估计; 然后, 利用代用资料系统模型将先验估计转换到代用资料空间, 并计算该时刻真实的代用资料值与通过代用资料系统模型估算的代用资料值之间的差异, 这个差异在数据同化中叫作“增益”(即innovation); 最后, 基于模式协方差和代用资料的误差来计算差异的权重, 并将其用于对先验估计进行更新以此得到

后验估计(Posterior estimate)(Talagrand, 1997)。然后对下一时刻, 重复上述步骤, 最终生成同时具有气候史实和物理机制的同化结果。

在实际的古气候数据同化研究当中, 主要的区别在先验的估计和优化。按照其出现的时间先后, 当前主流的古气候数据同化算法包括牛顿松弛逼近、粒子滤波、离线集合卡尔曼滤波, 及最近的在线同化方法(图2)。下面简单介绍以下几种方法的原理及应用。

2.1 牛顿松弛逼近

牛顿松弛逼近法是一种在预测模式中添加一个强迫项, 从而使得模式状态逐步逼近观测的同化方法(Hoke和Anthes, 1976)。其具体公式为

$$\psi^n = f(\psi^{n-1}) + \alpha H^T (d^{n-1} - H(\psi^{n-1})) + \zeta^n \quad (2)$$

其中, ψ^n 是 t_n 时刻的模式状态, 是 t_{n-1} 时刻的状态 ψ^{n-1} 的函数 f , α 是松弛逼近参数, H 是将模式状态转换到观测

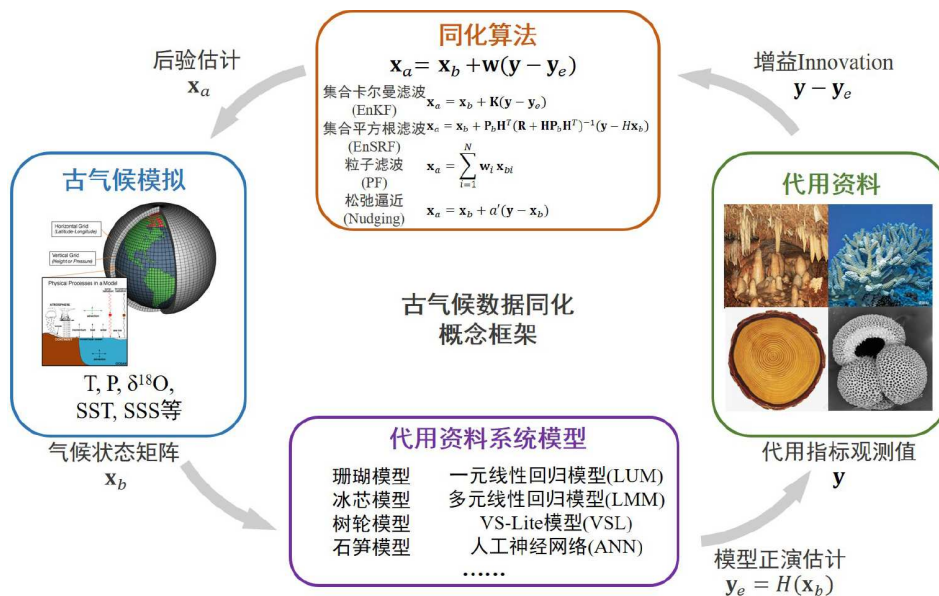


图1 古气候数据同化概念框架

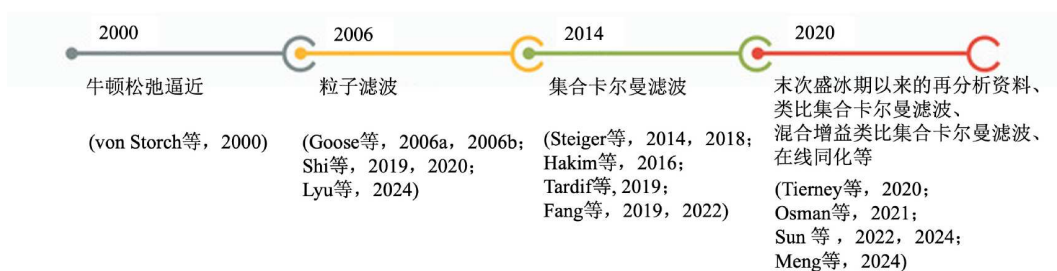


图2 古气候数据同化方法的沿革

空间的算子, d^n 是 t_n 时刻的观测数据, ξ^n 是随机噪声。

松弛逼近的主要优势是方法简单直接, 容易实现, 且约束效果好。而其缺点也同样明显, 松弛逼近只能同化模式输出的变量, 因此在同化时需要将观测数据转换成模式输出的变量。而且, 由公式可以看出, α 决定了松弛逼近的强度, 过强的松弛逼近可能会由于过快的收敛而产生错的动力, 而过弱的松弛逼近会导致结果不能受到观测的约束 (Dubinkina 和 Goosse, 2013)。除此之外, α 的选取一般基于经验来进行选择, 缺少物理基础。

在应用方面, 松弛逼近是最早应用的古气候数据同化方法。von Storch 等 (2000) 利用松弛逼近对历史时期的 NAO 指数进行了重建, 结果显示同化结果相对模拟结果可以更好地复现历史气候的真实变化。Widmann 等 (2010) 在对比包括松弛逼近在内的三种同化方法在过去千年北欧气候同化中的表现时, 认为其可以有效地重建过去千年的气候变化, 但却很难重建与模式内部变率不同的目标模态。Dubinkina 和 Goosse (2013) 对比了松弛逼近、粒子滤波和融合松弛逼近的粒子滤波三种方法在过去 150 年南半球高纬度气候重建中的表现, 发现单纯的松弛逼近在无直接观测的变量 (如海表盐度) 同化方面表现不如其他两种方法, 因为其无法反映海洋动力过程。之后, 松弛逼近在古气候数据同化研究中就较少应用了。

2.2 粒子滤波

粒子滤波的基本思想是, 基于贝叶斯的似然估计对一组随机模式样本粒子进行加权, 去逼近状态的后验概率分布。其后验概率分布的计算公式 (Dubinkina 和 Goosse, 2013) 为

$$p(\psi^n | d^n) = \sum_{i=1}^M \omega_i^n \delta(\psi^n - \psi_i^n) \quad (3)$$

其中, δ 是核密度, ω_i^n 是每个粒子的权重, 其计算公式为

$$\omega_i^n = K^{-1} p(d^n | \psi_i^n) \quad (4)$$

其中, K 是均一化系数, $p(d^n | \psi_i^n)$ 是模式状态下观测的似然。

粒子滤波的优势是, 它不需要假设先验是高斯分布 (Dubinkina 和 Goosse, 2013), 也不需要假设观测和先验估计存在线性关系。其缺点为, 对观测数据的数量和质量要求高, 并且会出现权重向少量粒子聚集的现象。

在粒子滤波的早期应用中, 多采用简化的粒子滤波, 即仅选取最接近观测的一个模拟作为最优粒子, 作为下

一步同化的初始条件 (Goosse 等, 2006; Widmann 等, 2010)。Goosse 等 (2006) 用了简化的粒子滤波方法对过去千年北半球的气候进行了模拟, 仅基于少量粒子 (30 个) 和简单的权重计算就可以生成与记录一致的气候状态。在后来的一些研究中, 还将粒子滤波跟松弛逼近进行比较, 例如前面提到的两个工作 (Widmann 等, 2010; Dubinkina 和 Goosse, 2013)。单就粒子滤波的表现而言, Widmann 等 (2010) 将简化的粒子滤波应用到北欧的温度同化中, 复现了温度的多年代际变率, 虽然他仅仅用了 11 个粒子 (模拟结果)。这也说明, 并未出现粒子滤波常见的粒子退化问题。Dubinkina 和 Goosse (2013) 也发现粒子滤波可以很好地重建无直接观测的变量 (如海表盐度), 特别是与松弛逼近相结合的粒子滤波。在粒子滤波的改进方面, Dubinkina 等 (2011) 和 Annan 和 Hargreaves (2012) 在粒子滤波中加入了残差重采样方法, 发现采用残差重采样的标准粒子滤波, 同化结果比简化的粒子滤波有显著提高。

在算法方面改进的同时, 除了针对温度和环流场的重建, 近年来粒子滤波还被应用到了东非、东亚和南美等更广区域的降水重建中 (Klein 和 Goosse, 2018; Shi 等, 2019; Lyu 等, 2024), 也取得了较好的效果。特别是最近的研究 (Lyu 等, 2024) 在将粒子滤波应用到南美季风降水及环流场的重建时, 选取了多于 600 个粒子, 发现粒子滤波可以很好地抓住 $\delta^{18}\text{O}$ 与降水的非线性动力关系, 从而得到更好的同化结果。

2.3 离线集合卡尔曼滤波

离线集合卡尔曼滤波 (ensemble Kalman filter, EnKF) 是近几年古气候数据同化领域应用比较广泛的一种方法 (Hakim 等, 2016; Tardif 等, 2019; Tierney 等, 2020; Li 等, 2024; Wu 等, 2025)。其思想为用古气候代用指标的值得来更新每个时刻的期望值, 并通过对比代用指标的不确定性与观测误差协方差来分配权重, 其集合思想体现在基于统计特征进行背景误差协方差矩阵的估计方面。其具体公式为

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}[\mathbf{y} - H(\mathbf{x}^b)] \quad (5)$$

其中, \mathbf{x}^a 是同化结果, \mathbf{x}^b 是先验估计, 通常是从一个静态来源中采样或条件性采样得到的, 例如已有的气候模式模拟结果, \mathbf{y} 是代用指标, H 是将先验估计转换到代用指标空间的代用资料系统模型, $\mathbf{y} - H(\mathbf{x}^b)$ 表征了观测数据与先验估计的差异, \mathbf{K} 为卡尔曼增益矩阵, 用来表征 $\mathbf{y} - H(\mathbf{x}^b)$ 的权重并将其转换到状态 (\mathbf{x}^b) 空间, 其计算公式为

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1} \quad (6)$$

其中, \mathbf{B} 为先验估计的协方差矩阵, \mathbf{R} 是观测的误差协方差矩阵, \mathbf{H} 是线性的代用资料系统模型。

离线集合卡尔曼滤波方法的主要优点为, 在满足假设情况下结果精度较高, 而且其求解比较容易, 易于并行计算, 系统方便搭建, 因此在近几年古气候数据同化领域应用比较广泛。其主要缺点为, 假设先验估计误差和观测误差为高斯分布, 而且假设观测和模式结果为线性关系。

在应用方面, 早期的研究针对EnKF在古气候数据同化中的应用进行了一系列的理想化试验。例如, Huntley和Hakim(2010)测试了EnKF对观测站点分布的敏感性, 发现当站点数量较少时, 少量的分布合理的站点的同化效果跟数量很多但随机分布的站点的同化效果接近。Pendergrass等(2012)发现, 在模式预报技巧长于代用指标分辨率, 且气候协方差与平均态显著相关的两种情况下, 同化技巧相对基于统计的重建显著提高。Steiger等(2014)将EnKF应用在过去千年的温度同化当中, 并与传统的主成分分析(PCA)方法进行了对比, 发现EnKF结果在空间特征方面更加可信, 特别是在代用资料更为稀疏的区域。之后, Hakim等(2016)和Tardif等(2019)进一步应用EnKF方法制作了过去千年再分析资料(LMR)。近年来, 新出现的过去两千年的再分析资料(Hu等, 2024; Wu等, 2025)、全新世温度重建(Erb等, 2022)、LGM以来的再分析资料(Tierney等, 2020; Osman等, 2021)均基于EnKF方法。

2.4 在线同化

由于离线同化方法先验分布的构建仅依赖一个静态来源, 如已有的气候模拟结果, 缺少对之前气候状态的记忆, 因此很多学者开始尝试用在线同化方法来克服这个不足(Perkins和Hakim, 2017, 2020)。例如Perkins和Hakim(2017, 2020)在年尺度的温度和环流场的同化中, 利用线性倒数模型(linear inverse model, LIM)对同化的后验结果进行在线预测, 生成下一同化时刻的先验分布。他们发现, 在线同化方法要优于离线同化方法, 其改进主要源自海气耦合系统的动力约束。相对于早期的在线同化方法多基于LIM, Meng和Hakim(2024)基于深度学习模型构建了一个在线EnKF同化系统, 进一步重建了月尺度热带太平洋海表温度、经向和纬向风应力及7层海洋上层温度。他们发现, 因为深度学习模型可以捕捉更多的当前和未来气候状态间的非线性关系并保留更多的预测能力, 所以相对于传统的LIM, 深度学习模型可以生成更准确的预测结果。而且, 这些改进与区域和变量有

关, 主要集中在赤道外的纬向风应力和海表温度、赤道附近的海洋温度及中太平洋的温跃层。在降水同化方面, 由于降水的记忆更低, 在线同化的降水同化技巧仍低于温度和环流场(Perkins和Hakim, 2020), 与离线同化类似。

由于海洋的记忆更长, 因此在线同化可以将海洋的记忆传递给大气, 从而改进大气的同化, 特别是当过去千年时段代用资料多为陆地指标的情况(Perkins和Hakim, 2020; Meng和Hakim, 2024; Meng等, 2025)。而针对更长时间段, 海洋的记忆可以在什么尺度上帮助在线同化提高预测技巧, 它对年代及更长尺度变化预测的可靠性如何, 有孔虫等深海代用指标对提高长尺度变化的预测技巧有什么帮助, 以及预测技巧是否存在对模式的依赖性, 这些方面都有待进一步研究。此外, 目前的在线同化多集中在过去两千年时段, 而更长时间段的在线同化所需计算量太大, 因此简单气候模式的可行性更高。当然, 未来随着算力的提高, 地球系统模式也有可能应用在更长时间段的在线同化。

3 古气候数据同化的应用

近年来, 古气候数据同化应用于不同特征时段的古气候重建领域, 取得很多成果。前人已对近年来的应用进行了详细的总结(张昊勋等, 2025), 这里从技术细节方面给出一个简单回顾。

3.1 过去两千年

过去两千年是古气候数据同化最为成熟的时段。从最早的古气候数据同化的引入及方法的改进, 都是围绕着过去两千年时段的气候重建进行的(例如, von Storch等, 2000; Goosse等, 2010)。近年来, 随着代用资料和模式的进步, 过去两千年的数据同化也取得了显著进展(Zhu等, 2023)。其中, 常用的代用资料包括PAGES2k和Mann09两组数据集, 以及最新的CoralHydrol 2k等数据集。常用的模拟结果包括PMIP的LM和CESM-LME数据。在同化方法方面, 从早期的松弛逼近、粒子滤波、EnKF等方法均有所应用, 近年来应用较多的是粒子滤波和EnKF方法。此外, 过去两千年的同化方法也有了很大的进步, 例如Sun等(2022, 2024)研发了新的类比集合卡尔曼滤波(analogue offline ensemble Kalman filter, AOEnKF)和混合增益类比集合卡尔曼滤波(hybrid gain analogue offline ensemble Kalman filter, HGAOEnKF)方法, 通过改进先验分布的选取, 提高了古气候数据同化技巧, 以及近年来的在线同化方法(Meng和Hakim, 2024; Meng等, 2025; Sun等, 2025)。

当前,主要的同化数据集包括过去千年再分析资料(Last Millennium Reanalysis, LMR)(Hakim等, 2016; Tardif等, 2019)、古水文动力数据同化产品(Paleo Hydrodynamics Data Assimilation Product, PHYDA)(Steiger等, 2018)、南京师范大学过去两千年再分析资料(NNU-2ka Reanalysis))(Hu等, 2024; Wu等, 2025)。这些数据集中除了包含温度、降水、环流场等常规变量,还包括了帕默尔干旱指数(Palmer Drought Severity Index, PDSI)、热带辐合带(Intertropical Convergence Zone, ITCZ)、厄尔尼诺/南方涛动(El Niño-Southern Oscillation, ENSO)、太平洋年代际涛动(Pacific Decadal Oscillation, PDO)、大西洋年代际涛动(Atlantic Multidecadal Oscillation, AMO)等指数。

由于过去两千年的同化技术比较成熟,因此除了重建气候特征外,古气候数据同化还被应用于多尺度气候变化的机制分析(Zhu等, 2022)。Erb等(2020)重建了过去千年美国地区的干旱及环流场,发现内部变率而非外强迫主导了多年干旱。Lyu等(2024)重建了过去千年的南美季风强度,发现在中世纪气候异常期(Medieval climate anomaly, MCA)到小冰期(Little Ice Age, LIA)的转换时段,南美季风存在百年尺度的增强,这与大西洋ITCZ的南移及太平洋沃克环流增强有关。Fang等(2022)基于北极地区的树轮、冰芯、湖泊沉积、历史文献代用指标对过去千年的北极放大指数进行了同化,发现AMO主导了其年代际的变化,而人为温室气体主导了其在工业革命以来的百年尺度减弱。

整体而言,过去两千年的温度及环流同化已相对完备,但因为降水相较于温度而言具有更高的空间异质性和局地性特征(Hancock等, 2023),且降水与代用资料的关系、主导机制更加复杂,所以降水的同化仍存在挑战(Wu等, 2025)。此外,由于过去两千年的代用资料丰富,而且PSM更成熟,因此过去两千年时段还承担着未来同化方法改进的任务。

3.2 全新世

相对过去两千年,全新世及更早时段中的古气候数据同化应用相对少一些,主要的同化算法为EnKF方法。全新世时段, Ertel等(2022)基于EnKF方法,重建了时空连续的全新世温度变化。所使用的代用资料为涵盖湖泊沉积、海洋沉积、泥炭、冰芯、石笋等指标的Temperature 12k数据集(Kaufman等, 2020),模拟结果为基于HadCM3模式的末次盛冰期瞬变试验(Snell等, 2022)和基于CCSM3模式的TraCE-21ka试验(Liu等, 2014)。结果显示,中全新世的温度为工业革命前最高,比过去千年

高0.09 °C,这个结果比之前的全新世重建结果(Marcott等, 2013; Kaufman等, 2020)要低,但比其他同化结果(Osman等, 2021)要高。此外, Ertel等(2022)还测试了季节性对全新世温度趋势的影响,发现即使在全部记录都考虑夏季误差的情况下,仍无法解释记录 and 模拟之间的偏差。

3.3 LGM以来

Tierney等(2020)和Osman等(2021)基于EnKF方法对LGM以来的温度变化进行了同化。所使用的代用资料为表征海表温度的海洋地化指标数据集($\delta^{18}\text{O}$ 、Mg/Ca、 $U_{37}^{K'}$ 、TEX₈₆),模拟数据为基于iCESM模式开展的4个切片试验的结果。同化结果显示, LGM时段全球平均温度下降-6.1 °C(95%置信区间为-6.5—-5.7 °C),由此推算的气候敏感性为3.4 °C(95%置信区间为2.4—4.5 °C)(Tierney等, 2020)。LGM以来温度变化的主要驱动因子为冰盖和温室气体导致的辐射外强迫,其次为AMOC和季节太阳辐射的变化(Osman等, 2021)。Annan等(2022)基于EnKF方法,利用了PMIP的多模式结果和3套集成的格点海表温度和表面气温数据,重建了LGM的海表温度和表面气温。结果显示, LGM时段全球平均气温相对工业革命前的距平为-4.5±0.9 °C,这个结果与Tierney等(2020)结果之间的差别主要来自于先验的选择。因此,他们推荐使用多模式集合来作为可靠的先验估计,但前提条件是模拟结果的范围要尽可能真实全面地表征不确定性的主要来源。

3.4 深时数据同化

近年来,古气候数据同化也被应用于深时气候的重建。例如,针对古新世-始新世极热事件(PETM, 56 Ma), Tierney等(2022)利用EnKF方法对PETM的气候状态进行了重建。所用的代用指标除了末次盛冰期再分析资料(Last Glacial Maximum Reanalysis, LGMR)同化所用的4种海洋地化指标外,还添加了陆地的代用指标MBT_{Me}^{5'},所使用的模拟资料为一组基于iCESM的早始新世模拟试验结果。同化结果显示, PETM时段的全球平均温度变化为5.6 °C(95%置信区间为5.4—5.9 °C),由此推算的气候敏感性为6.5 °C(95%置信区间为5.7—7.4 °C)(Tierney等, 2022)。

Li等(2024)也基于EnKF方法对PETM期间的碳循环扰动进行了重建,所使用的代用指标为海底沉积CaCO₃和海表温度指标($\delta^{18}\text{O}$ 、Mg/Ca、TEX₈₆),所使用的模拟资料是cGENIE模式的100个成员集合模拟结果。同化结

果显示, 大气 CO_2 从890 ppm增加到1980 ppm (1 ppm=1 $\mu\text{L/L}$), 而海水pH值下降0.46, 海水钙饱和度从10.2下降到3.8。

上新世(5.33–2.58 Ma)是距今最近的一次 CO_2 浓度接近400 ppm的地质时段, Tierney等(2025a)利用EnKF方法对上新世的气候状态进行了重建, 所使用的代用指标为海表温度指标($\delta^{18}\text{O}$ 、 Mg/Ca 、 TEX_{86}), 所使用的模拟资料为14个PlioMIP2试验、2个基于CESM2的上新世敏感性试验和21个基于CESM1的类-上新世试验结果。同化结果显示, 中上新世增温约为4.1 $^{\circ}\text{C}$ (95%置信区间为3.0–5.3 $^{\circ}\text{C}$), 由此推算的气候敏感性为4.8 $^{\circ}\text{C}$ (95%置信区间为2.6–9.9 $^{\circ}\text{C}$), 而且赤道太平洋海温梯度呈现出类El Nino的模式, 此外北太平洋盐度偏低而北大西洋盐度偏高。

Judd等(2024)进一步利用EnKF方法对显生宙(485 Ma)的全球平均温度进行了重建(PhanDA), 所使用的代用资料为表征海表温度的海洋地化指标数据集($\delta^{18}\text{O}$ 、 Mg/Ca 、 $U_{37}^{K'}$ 、 TEX_{86}), 所使用的模拟资料是基于iCESM的80个成员的切片模拟试验结果。同化结果显示, 全球平均温度变化范围为11–36 $^{\circ}\text{C}$, 由此推算的气候敏感性为8 $^{\circ}\text{C}$ 左右, 而且显生宙全球平均温度的主要驱动因子是 CO_2 浓度。

整体而言, 过去两千年的同化方法比较成熟且多样, 包括粒子滤波、EnKF及其改进型, 以及各种最新的在线同化方法, 而全新世及其之前更久远的时段, 同化方法以EnKF为主。在同化的初始场方面, 过去两千年和全新世时段的同化都是以瞬变积分模拟结果为主, 而LGM及其之前更久远的时段, 同化以切片模拟结果为主。在同化的代用指标方面, 过去两千年的同化以树轮、珊瑚 $\delta^{18}\text{O}$ 、 Sr/Ca 等指标为主, 全新世同化以石笋 $\delta^{18}\text{O}$ 、湖泊和海洋沉积等为主, 而LGM及更久时段的同化以海洋地化指标为主。

3.5 现有同化数据集与平台

过去两千年时段的同化数据集方面, Hakim等(2016)结合PAGES 2ka的重建数据集和过去千年的模拟结果, 制作了包括温度、降水、环流场数据的过去两千年的再分析资料LMR。Steiger等(2018)进一步重建了过去两千年的干湿指数及环流场PHYDA。Hu等(2024)制作了NNU-2ka Reanalysis。Erb等(2022)重建了全新世温度数据、Tierney团队重建了LGMR、PETM、上新世、显生宙的温度变化资料, 这些资料均已开放共享。

这些数据集中, 过去两千年的同化结果除了温度、

降水等常规变量之外, 还包括不同高度的环流场, 而全新世及更久时段的同化结果主要以温度重建为主, 因此这些时段的环流场同化也是未来的发展方向之一, 这就必须借助于对这些时段气候变化机理的更深入理解及模式模拟的改进。

在同化平台方面, 现在很多同化算法都是开源的, 这些算法以EnKF或其改进型为主, 主要的语言为Python或Matlab。主要的平台包括与LMR相对应的LMR(Hakim等, 2016), 改进版的LMR Turbo(LMRt)(Zhu等, 2021)和最新的Climate field reconstruction cfr)(Zhu等, 2024), 这些算法都是基于Python语言的。美国亚利桑那大学的Jessica Tierney团队也研发了与其LGMR相对应的同化软件包(DASH)(King等, 2023), 是基于Matlab软件的。南京大学雷荔傑团队研发的EnKF同化算法的改进型AOEnKF和HGAOEnKF软件包(Sun等, 2022, 2024), 也是基于Matlab软件的。北京大学李明松团队研发的deep-DA软件包(Li等, 2024), 是基于Python软件的。这些软件包的核心算法均是EnKF或其改进型。

3.6 案例分析

氧同位素 $\delta^{18}\text{O}$ 作为一种存在于多种载体的代用指标, 广泛应用于各个时段的古气候重建。之前基于氧同位素的重建, 多以线性回归模型为基础, 而忽视了氧同位素与气候变量关系背后的物理机制(Liu等, 2023)。现在, 有了耦合氧同位素的模拟试验, 可以更好地厘清氧同位素变化的物理机制, 使得基于非线性PSM针对各种载体中的氧同位素同化成为可能(Tierney等, 2020; Lyu等, 2024)。Lyu等(2024)认为, 同化过程中加入 $\delta^{18}\text{O}$ 后, 相比LMR和PHYDA, 可以更好地重建南美季风的变化。那么, 非线性PSM与传统的线性PSM的同化技巧有什么差别呢? 下面我们具体对比一下针对珊瑚 $\delta^{18}\text{O}$ 基于线性PSM和非线性PSM进行海表温度(SST)同化的差异。

图3A和3B分别为基于线性PSM同化结果的Nino3.4指数和非线性PSM同化结果的Nino3.4指数与观测的对比, 结果显示两种方法同化的结果均与观测有较高的相关性($r=0.82, 0.81, p<0.01$), 没有显著的差别。图3C和3D分别是以Palmyra岛记录为例的两种同化方法所用到的协方差空间场, 其中Palmyra岛的SST和 $\delta^{18}\text{O}$ 与全球的SST相关空间场类似, 呈现典型的ENSO模式, 说明两个变量均对Nino3.4区域SST有很好的表征, 可以用来进行Nino3.4指数的同化。同化结果显示, 两者相差不大, 主要体现在非线性PSM的同化结果分布更加大, 其CE值为0.35, 小于线性PSM同化结果的CE值0.55。说明两种方法在复现Nino3.4指数变率方面的技巧相差不大, 但非线性PSM同

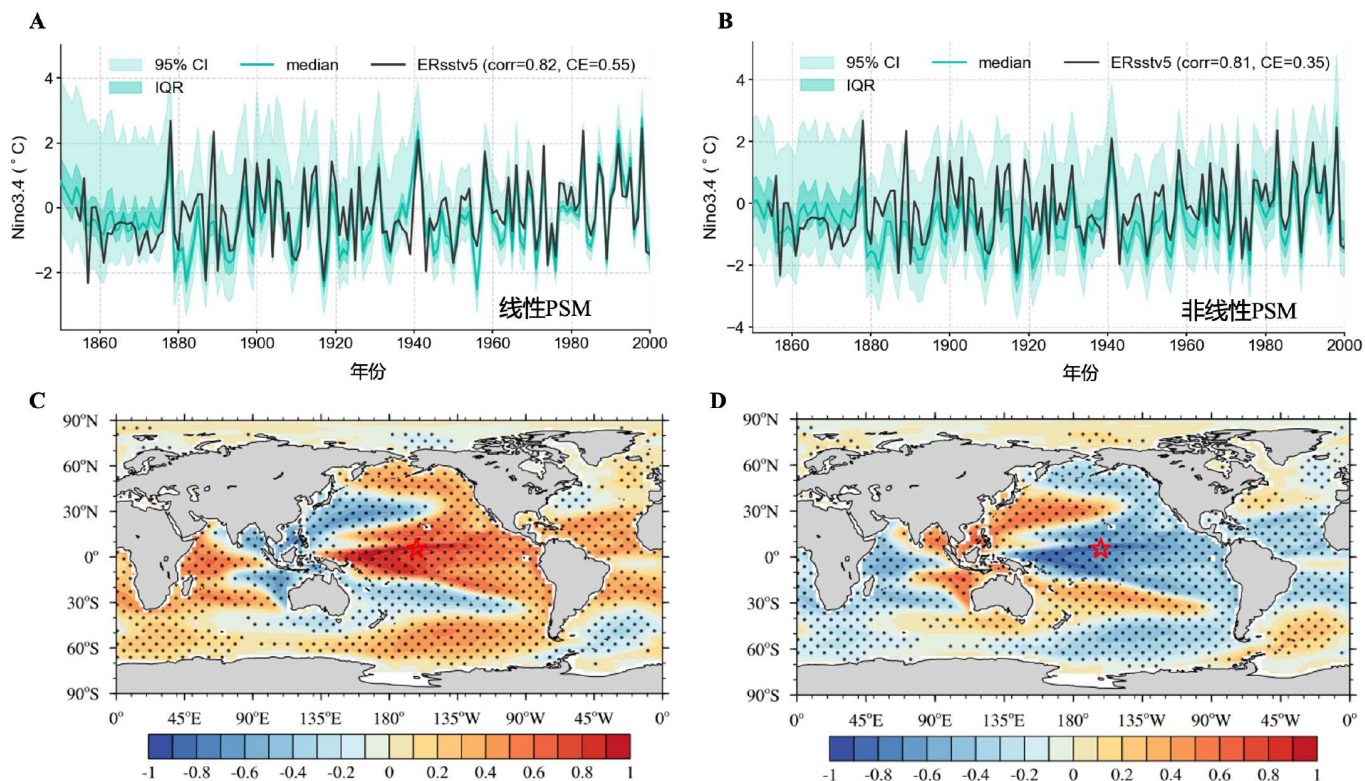


图3 基于线性PSM(A)和非线性PSM(B)的观测时段Nino3.4指数同化结果对比, 以及表征同化所用协方差矩阵的点面相关空间场(C, D), 图D中, 在计算相关系数之前, 海水的 $\delta^{18}\text{O}_{\text{sw}}$ 已转化为珊瑚 $\delta^{18}\text{O}_{\text{c}}$ 。

化结果的分布范围更大。

这说明, 基于耦合氧同位素模拟结果的非线性PSM同化效果与传统线性PSM类似, 那未来将基于模拟结果构建的非线性PSM应用与石笋氧同位素等代用指标的同化中, 也具有较好的前景。也可用于有孔虫同位素和树轮同位素等代用指标的线性PSM的矫正及同化结果验证。

4 展望

自2000年古气候数据同化概念提出以来, 古气候数据同化在古气候研究中取得了重要的进展。本文系统回顾了古气候数据同化的发展沿革、在各个时段的应用、解决的关键科学问题。可以说, 古气候数据同化在很大程度上提升了我们对历史时期不同时段气候时空特征和演化过程的认知, 在过去两千年这些同化应用比较成熟的时段, 同化方法还可以用来完善气候变化的成因机制。当前, 几个国际科学计划中的古气候同化相关的研究计划也在紧锣密鼓地筹备中, 例如CMIP7的古气候数据同化比较计划(Paleo-DA MIP)、PAGEs2ka的古气候再分析集成项目(PRISM)。

诚然, 古气候数据同化研究还非常初步, 即使当前常用的离线EnKF方法实际上是非常简化但有效的方法,

仍存在一些亟待解决或继续完善的方面, 下面就从古气候数据同化可解决的关键科学问题、优缺点及改进、未来发展方向等方面进行展望。

4.1 古气候数据同化可以解决的关键科学问题

如何通过结合代用资料和模拟结果从而更好的重建古气候的特征及规律, 是古气候数据同化关注的首要科学问题。除了基础的重建不同尺度的气候变化之外, 古气候数据同化还被应用于解决一些代用资料和模拟结果、代用资料之间的争议。其中一个典型的例子就是全新世温度悖论问题。Erb等(2022)认为中全新世温度比过去千年高, 但Osman等(2021)和Bova等(2021)的同化结果并未反映出明显的中全新世大暖期。Erb等(2022)还做了敏感性试验, 发现虽然代用指标的季节偏差可能指示的是夏季温度的趋势, 但是季节偏差的潜在影响不足以使重建的全球平均温度与瞬变模拟中看到的变暖趋势相一致。而Osman等(2021)认为, 南半球稀疏的代用资料的权重过高可能会导致中全新世暖期, 但这个只能解释部分暖异常, 如何缓解代用资料和模拟结果间的差别还需要进一步的研究, 特别是同化是否可以在这个方面有所贡献。最近, Hao等(2025)基于全球海洋沉积物代用指标对全新

世温度进行了重建,发现中全新世温度异常有显著的空间差异,其中欧洲和欧亚大陆高纬度的冬季和全年平均温度均为偏高,而其他区域温度偏低,之前的差异源自模拟结果对植被和海冰的反馈偏差所造成的高纬度冷偏差,而代用资料因为集中在欧洲地区所导致的暖偏差。近期的欧亚大陆中纬度的烯酮重建(Jiang等, 2024)也显示全新世温度趋势具有空间异质性,即中国东北地区呈现降温趋势,而西伯利亚西南部呈现增温趋势,这为不同区域代用资料的同化协方差矩阵局地化的选取提供了重要信息。最近的研究(Liu等, 2025)发现,代用资料中低估了温度的季节性,导致重建的全新世温度趋势被夏季温度主导,而同化是否可以通过耦合模拟结果来减小这个不确定性从而给出更可靠的全新世温度趋势重建,也是一个值得进一步深入探讨的问题。

除了用来重建气候特征之外,同化还开始被应用于机制研究,特别是对同化技术比较成熟的过去两千年时段。目前,过去两千年的同化已针对环流场进行了重建,而其他特征时段的同化结果还侧重温度等变量时空特征的重建。例如,过去千年针对不同尺度的季风降水或干旱变化及其驱动机制(Erb等, 2020; Lyu等, 2024)。因为过去两千年的代用指标同环流变化的PSM相对清楚,关键在于过去千年的代用指标在观测时段有数据,可以同观测数据构建PSM,另外过去千年的代用指标的气候表征意义更成熟。算法方面,粒子滤波同化的环流场跟温度和降水的变化更为一致,而其他方法由于环流场和温度降水是单独同化的,所以并不一致。但粒子滤波对环流场重建的可靠性,更依赖模式结果对机制的模拟。除了温度、降水、环流场之外,古气候数据同化还可以对更多的气候环境要素进行重建,例如Li等(2024)针对PETM时期碳循环和碳酸盐饱和状态的重建,为理解PETM时期的大洋酸化和海水碳酸盐饱和状态提供了帮助。这也为更古老的地质历史时期的气候环境重建提供了全新的思路。

此外,定量区分内部变率 and 外强迫的贡献,作为古气候研究领域关注的另外一个关键科学问题,也是古气候数据同化对机制研究的一个方面。之前的研究多基于单因子敏感性试验集合的方法来进行区分,而这种方法存在着模式依赖性的问题,而古气候数据同化通过结合多个模式结果给出更可靠的对外强迫的响应,可以为回答这个问题起到一定作用。例如,在给定外强迫的情况下,早期的同化方法所得到的结果没有超出模式结果内部变率的范畴(Widmann等, 2010)。在同化过程中,内部变率的来源包括多模式结果中随机抽取的初始场及代用数据中比信号更高频的变率(或噪声),对外强迫响应

的来源包括了多模式模拟的响应和代用数据中的信号。通过定量评估两个来源的可靠性,给出更可靠的对外强迫响应和内部变率,从而更好地定量区分两者对气候变化的相对贡献。

在服务未来气候预估方面,气候敏感性也是古气候数据同化经常用来研究的关键科学问题之一。气候敏感性的定义为,二氧化碳浓度升高2倍时地球系统的温度变化,其计算方式(Tierney等, 2020)为

$$ECS = \frac{\Delta GMST}{\Delta R} \times F_{2 \times CO_2} \quad (7)$$

其中, $\Delta GMST$ 是同化所得到的该特征时段全球平均温度的变化, ΔR 是该特征时段的辐射强迫, $F_{2 \times CO_2}$ 二氧化碳浓度升高2倍所带来的辐射强迫。在计算过程中, ΔR 和 $F_{2 \times CO_2}$ 由模式结果估算,同化结果提供 $\Delta GMST$ 的值。Tierney等(2020)基于LGMR估算的气候敏感性为 3.4°C (95%的置信区间为 $2.4\text{--}4.5^\circ\text{C}$)。Tierney等(2025a)基于上新世温度同化所估算的气候敏感性为 4.8°C (90%的置信区间为 $2.6\text{--}9.9^\circ\text{C}$)。Judd等(2024)基于显生宙温度同化所估算的气候敏感性为 $\sim 8^\circ\text{C}$ 。由此可见,气候敏感性的值随着特征时段温度的增加而升高,且置信区间的变化说明同化结果的分布随时段前推而逐渐增加。此外,特征时段的辐射强迫估算的不确定性,对气候敏感性的计算也有显著影响,说明了进一步提高特征时段外强迫重建及模式反馈精度。

4.2 古气候数据同化的不足及改进

近年来,科研人员针对古气候数据同化的代用资料系统模型、同化算法等组成部分的不足做了很多改进,这些改进体现在对同化初始场的选取、观测与模拟变量关系的构建、在线同化等方面的提高。

之前讲到,不确定性的计算是古气候数据同化的主要挑战之一,代用资料和模式结果的不确定性计算都存在着很多难题。在计算代用资料的不确定性时,应该涵盖记录自身的统计不确定性、定年误差、器测误差、空间代表性误差以及代用资料系统模型的表示误差,如何定量计算这些不确定性需要深入的探讨,并与重建领域的专家紧密合作。在计算模拟结果的不确定性时,需要大量的长期气候平均值的独立样本,即在观测误差范围内的由不同外强迫、边界条件和模式参数的长期模拟,当前的计算资源很难满足这个需求。而且,当前同化多采用同一模式的结果,未来的同化应采用多模式模拟结果,来更好的量化模拟结果的误差。

在代用资料的质量控制和误差估计方面,虽然更多

的代用资料可以带来更多的古气候信息,有利于得到更准确的古气候重建,但在将新的代用资料引入同化过程中时,不能简单地进行叠加,而是需要进行一定的质量控制,然后量化其不确定性,然后客观判断其在同化过程中所占的权重。

在同化过程中,针对不同时段、不同尺度、不同变量的同化,不同区域代用资料的贡献(Wu等, 2025)需要首先进行量化。然后,定年误差方面,在过去两千年时段,借助树轮定年准确的优势,可以借助树轮定年对年或更高分辨率的代用资料(如珊瑚等)来进行部分矫正(Hu等, 2024)。而在更长尺度上,当没有精准定年的代用资料进行矫正的情况下,如何将定年误差量化至同化过程中,需要与重建领域专家的经验充分结合起来。此外,除了已经应用比较成熟的代用资料,还有其他一些不常用记录特别是定性类的资料(如历史文献等)如何应用到同化中,也需要思考。从而在不同种类的代用资料误差的计算方法中,考虑每种资料的分辨率、定年精度等特征,实行多种资料的融合。

在代用资料系统模型的构建与完善方面,当前树轮、珊瑚氧同位素和海洋有孔虫、 $U_{37}^{K'}$ 等的代用资料系统模型已比较成熟,但石笋氧同位素、孢粉等的代用资料系统模型需要进一步构建及完善(Ning等, 2025a),特别是跟降水之间的PSM目前还处于初步阶段。在此过程中,借助氧同位素模拟结果,石笋氧同位素等同位素相关的代用资料系统模型可以构建出来(Ning等, 2025b)。但由此构建出来的非线性算子跟传统线性算子的差别,需要系统地进行对比其优缺点。除此之外,机器学习等方法也已被应用在非线性PSM的构建,未来可以进一步完善(Fang和Li, 2019; Wei等, 2024)。Fang和Li(2019)基于人工神经网络的方法构建了树轮宽度的非线性PSM,发现其同化效果比线性回归和VS-Lite模型更好,也证实了机器学习方法在未来其他变量同化当中应用的可行性。

在同化结果的验证方面,如果同化的时段涵盖了器测时段,那可以用观测数据进行验证,具体可使用相关系数、均方根误差、功效系数等指标来进行验证(张昊勋等, 2025)。但对更长时段的同化结果,由于观测数据的缺失,通常采用随机保留25%的代用资料来进行独立验证(Hakim等, 2016; Tierney等, 2020; Osman等, 2021; Wu等, 2025)。但这种验证方法将代用资料的不确定性包括在了其中,会对同化结果验证的客观性有一定影响。最近的研究也尝试使用独立的观测数据如钻孔温度进行验证(Meng等, 2025)。因此,未来需要探索更加客观的同化结果验证方法。

4.3 古气候数据同化的未来发展方向

在上述改进的基础上,古气候数据同化还可以在古气候在线同化、古气候动力学约束、深度学习和大数据的应用等几个方面进一步发展。

4.3.1 古气候在线数据同化

在同化算法的未来发展方面,除了针对传统的离线同化算法的改进(Sun等, 2022, 2024)之外,当前也开始了在线同化(Perkins和Hakim, 2020; Meng和Hakim, 2024; Sun等, 2025)。当然,古气候在线数据同化与现代气候中的在线积分同化有所不同,主要区别在于其当前时刻的初始场是由前面12个月的状态向量通过机器学习方法所生成的(Meng和Hakim, 2024)。相对于普通的离线同化和基于LIM模型的在线同化方法,基于机器学习的在线同化方法在过去两千年时段表现出了更高的准确性,特别是在代用资料更稀疏的情况下(Sun等, 2025)。目前的在线同化多集中在月尺度分辨率的过去两千年数据同化,是否继续提高分辨率及应用于更长时段的同化,如何提高在线同化的计算效率并充分发挥在线同化在初始场构建方面的优势,是一个关键方面。并且,基于深度学习的在线同化算法仍存在信号衰减的缺点,需要应用膨胀方法增加初始集合扰动以增大初始集合离散度(Meng和Hakim, 2024; Sun等, 2025)。

4.3.2 古气候动力学在古气候数据同化中的应用

作为同化过程中的主要动力约束,古气候动力机制的完善和应用在同化方法的发展中也起到重要的作用。首先是在进一步明晰代用指标的气候指示意义方面。当前,部分代用指标的气候指示意义多由相关系数来判断,但其背后的物理过程并不清楚。例如,通常认为南美安第斯山的冰芯 $\delta^{18}O$ 可以表征ENSO,但这种相关实际反映了轨道尺度赤道东太平洋海温异常对对流层中层水汽 $\delta^{18}O$ 的影响(Liu等, 2023)。此外,东亚地区的石笋 $\delta^{18}O$ 变化对局地降水量、水源地转换、上游衰减效应的表征,也可以通过耦合氧同位素的模拟结果来定量区分(Ning等, 2025b),从而得到更精准的代用资料系统模型。在借助模式来判断代用指标的气候指示意义之后,可以帮助定量划分同一代用资料对不同气候变量的表征,可以拓展同化在不同局地变量方面的应用,也可以通过遥相关关系对大尺度的海温或环流场进行同化。此外,在将不同类型代用资料在同化中的融合,也可以考虑不同代用指标对气候变量的定量表征。

然而,需要注意的是,在应用动力约束的时候,不

应该简单地套用模拟结果的协方差矩阵。因为模式本身存在着误差,例如(Sanchez等, 2021)等发现模式常见的双ITCZ误差,对SPCZ区域的珊瑚同化ENSO的结果有显著影响,需要在同化前进行误差订正。类似的,模式在遥相关方面的模拟误差也会影响同化结果。此外,大尺度环流场的遥相关影响在不同特征时段的强弱有所不同(Ning等, 2025a),因此在同化的时候,动力约束也需要随时间变化。当前,除了过去两千年(Hakim等, 2016; Wu等, 2025)和全新世(Erb等, 2022)时段采用瞬变模拟结果,更长时段同化(Tierney等, 2020; Li等, 2024)多采用切片模拟,未来研究中更长时段的瞬变模拟结果(如iTraCE结果等)可以使得同化初始场的选取更加符合外强迫对气候系统的影响。因此,瞬变模拟结果的应用有可能会提高长时段同化结果,但需要进一步对比验证。

此外,古气候数据同化结果中的动力约束,可以为重建的优化选点提供帮助。Huntley和Hakim(2010)通过一组敏感性试验发现,当观测比较稀疏的时候,站点的位置比站点的数量对同化结果的精度更加重要,即代用资料采集之前的选点可参考同化结果来进行优化选择。Wu等(2025)也发现,在代用资料数量接近时,记录的位置对同化结果的影响更大。近期的同化算法的改进,也侧重强调在稀疏代用资料情况下同化效果的提升(Sun等, 2025)。因此,可以通过设计敏感性试验的方法,定量评估代用资料的站点信息对气候重建的贡献。可以说,古气候数据同化相比直接的模拟结果,因为其涵盖了代用资料的气候指示意义等信息,可以更好地依据代用资料的类型和属性为其采集选点提供理论指导(方苗和李新, 2016)。不过,这些结论大多是在过去两千年时段的同化结果中得到的,更长时段的类似敏感性分析,目前还有待开展。

4.3.3 机器学习和大数据在古气候数据同化中的应用

机器学习和大数据在地球科学领域已有了广泛的应用,为很多复杂的地学问题提供了新的研究思路。在古气候数据同化研究中,深度学习首先应用在代用资料系统模型的构建方面,其优势在于不需要对代用指标和气候变量间的物理机制有清晰的认知(Fang和Li, 2019)。然而,其构建需要大量的数据来训练,因此目前主要还是集中在树轮(Fang和Li, 2019)和珊瑚(Wei等, 2024)这些在观测时段有足够数据的代用指标方面,而针对其他在观测时段只有少量时间节点的代用指标,基于深度学习构建代用资料系统模型还比较困难。当然,这种对物理机制的模糊认知也可能会对同化结果带来潜在的缺陷,而且,深度学习构建的代用资料系统模型同线

性代用资料系统模型类似,也存在过拟合的潜在风险(Fang等, 2022)。此外,代用资料质量对模型构建的影响,也需要进行评估。

除了深度学习之外,其他机器学习方法也在古气候数据同化中具有重要的应用潜力。例如,因果推断方法(苏建宾等, 2023)可以用于厘清代用指标与气候变量之间真实的因果关系,突破目前常用的线性相关关系,从而提高重建过程的可靠性和可解释性。迁移学习方法在构建代用资料系统模型方面具有很大的优势,可以突破训练数据稀缺、计算成本高等制约,但仍存在过拟合的风险。在将具有气候指示意义存在争议的代用资料进行同化之前,可以先使用因果发现算法来判断是某一种气候变量直接导致了代用指标的变化,还是多种指标存在混杂因素,这将有助于筛选出最可靠的代用指标进行同化。而物理信息神经网络方法可以利用已知的物理规律来指导神经网络的训练,使其在代用资料稀缺的区域也能构建出符合物理规律的代用资料系统模型。

数据驱动的气候模型的构建也是机器学习在同化领域的主要用途之一。例如,Meng和Hakim(2024)将深度学习引入古气候数据同化中,构建了一个在线同化系统,发现其在热带上层海温重建方面,比传统的LIM模型更加准确。Sun等(2025)进一步基于深度学习网络和集成混合集合卡尔曼滤波构建了过去两千年的在线同化,结果显示这种方法通过膨胀观测误差等手段,比常规的基于LIM模型的在线同化和离线同化的结果精度更高,特别是在早期代用资料比较稀疏的情况下。

此外,机器学习还被应用在模式误差估计方面,彭子怡等(2024)构建了基于卷积神经网络的模式误差估计方法,用来计算模式参数不准确所导致的模式误差及初始条件误差。在应用在简单模式的同化过程中,发现该方法可以很好地订正模式误差,因此可以考虑未来应用在古气候数据同化中。

参考文献

- 陈发虎, 段炎武, 郝硕, 等. 2023. 全新世温度大暖期模式与持续升温模式: 记录-模型对比问题及其研究展望. 中国科学: 地球科学, 53: 1699–1717
- 方苗, 李新. 2016. 古气候数据同化: 缘起、进展与展望. 中国科学: 地球科学, 46: 1076–1086
- 李新. 2013. 陆地表层系统模拟和观测的不确定性及其控制. 中国科学: 地球科学, 43: 1735–1742
- 彭子怡, 雷嘉俾, 谈哲敏. 2024. 一种估计模式误差的混合深度学习-资料同化方法. 中国科学: 地球科学, 54: 3691–3707
- 苏建宾, 陈都鑫, 郑东海, 等. 2023. 追索为什么? 地球系统科学中的因果推理. 中国科学: 地球科学, 53: 2199–2216
- 万凌峰, 刘健, 高超超, 等. 2020. 全新世火山喷发对温度变化趋势影响的模拟研究. 第四纪研究, 40: 1597–1610

- 王会军. 2022. 古今气候环境研究怎样融合? *地球科学*, 47: 3811–3812
- 王志远, 刘健, 王晓青, 等. 2016. 地球系统模式CESM1.0对太阳辐射和温室气体敏感性的差异研究. *第四纪研究*, 36: 758–767
- 张昊勋, 李明松, 胡永云. 2025. 古气候数据同化: 原理和展望. *中国科学: 地球科学*, 55: 416–434
- Annan J D, Hargreaves J C. 2012. Identification of climatic state with limited proxy data. *Clim Past Discuss*, 8: 481–503
- Annan J D, Hargreaves J C, Mauritsen T. 2022. A new global surface temperature reconstruction for the Last Glacial Maximum. *Clim Past*, 18: 1883–1896
- Bader J, Jungclauss J, Krivova N, et al. 2020. Global temperature modes shed light on the Holocene temperature conundrum. *Nat Commun*, 11: 4726
- Bova S, Rosenthal Y, Liu Z, et al. 2021. Seasonal origin of the thermal maxima at the Holocene and the last interglacial. *Nature*, 589: 548–553
- Comas-Bru L, Rehfeld K, Roesch C, et al. 2020. SISALv2: A comprehensive speleothem isotope database with multiple age-depth models. *Earth Syst Sci Data*, 12: 2579–2606
- Dubinkina S, Goosse H. 2013. An assessment of particle filtering methods and nudging for climate state reconstructions. *Clim Past*, 9: 1141–1152
- Dubinkina S, Goosse H, Sallaz-Damaz Y, et al. 2011. Testing a particle filter to reconstruct climate changes over the past centuries. *Int J Bifurcat Chaos*, 21: 3611–3618
- Erb M P, Emile-Geay J, Hakim G J, et al. 2020. Atmospheric dynamics drive most interannual U.S. droughts over the last millennium. *Sci Adv*, 6: eaay7268
- Erb M P, McKay N P, Steiger N, et al. 2022. Reconstructing Holocene temperatures in time and space using paleoclimate data assimilation. *Clim Past*, 18: 2599–2629
- Fang M, Li X. 2019. An artificial neural network-based tree ring width proxy system model for paleoclimate data assimilation. *J Adv Model Earth Syst*, 11: 892–904
- Fang M, Li X, Chen H W, et al. 2022. Arctic amplification modulated by Atlantic Multidecadal Oscillation and greenhouse forcing on multidecadal to century scales. *Nat Commun*, 13: 1865
- Goosse H, Renssen H, Timmermann A, et al. 2006. Using paleoclimate proxy-data to select optimal realisations in an ensemble of simulations of the climate of the past millennium. *Clim Dyn*, 27: 165–184
- Goosse H, Crespin E, de Montety A, et al. 2010. Reconstructing surface temperature changes over the past 600 years using climate model simulations with data assimilation. *J Geophys Res*, 115: D09108
- Hakim G J, Annan J, Brönnimann S, et al. 2013. Overview of data assimilation methods. *PAGES news*, 21: 72–73
- Hakim G J, Emile-Geay J, Steig E J, et al. 2016. The last millennium climate reanalysis project: Framework and first results. *J Geophys Res-Atmos*, 121: 6745–6764
- Hancock C L, McKay N P, Erb M P, et al. 2023. Global synthesis of regional Holocene hydroclimate variability using proxy and model data. *Paleoceanog Paleoclimatol*, 38: e2022PA004597
- Hao S, Zhang X, Duan Y, et al. 2025. Model seasonal and proxy spatial biases revealed by assimilated mid-Holocene seasonal temperatures. *Sci Bull*, 70: 2014–2022
- He C, Liu Z, Otto-Bliesner B L, et al. 2021. Hydroclimate footprint of pan-Asian monsoon water isotope during the last deglaciation. *Sci Adv*, 7: eaabe2611
- Herzschuh U, Böhmer T, Li C, et al. 2023. LegacyClimate 1.0: A dataset of pollen-based climate reconstructions from 2594 Northern Hemisphere sites covering the last 30 kyr and beyond. *Earth Syst Sci Data*, 15: 2235–2258
- Hoke J E, Anthes R A. 1976. The initialization of numerical models by a dynamic relaxation technique. *Mon Weather Rev*, 104: 1551–1556
- Hopcroft P O, Valdes P J. 2021. Paleoclimate-conditioning reveals a North Africa land-atmosphere tipping point. *Proc Natl Acad Sci USA*, 118: e2108783118
- Hopcroft P O, Valdes P J. 2022. Green Sahara tipping points in transient climate model simulations of the Holocene. *Environ Res Lett*, 17: 085001
- Hu W, Ning L, Liu Z, et al. 2024. Reconstructing tropical monthly sea surface temperature variability by assimilating coral proxy datasets. *NPJ Clim Atmos Sci*, 7: 261
- Huntley H S, Hakim G J. 2010. Assimilation of time-averaged observations in a quasi-geostrophic atmospheric jet model. *Clim Dyn*, 35: 995–1009
- Jiang J, Meng B, Wang H, et al. 2024. Spatial patterns of Holocene temperature changes over mid-latitude Eurasia. *Nat Commun*, 15: 1507
- Judd E J, Tierney J E, Lunt D J, et al. 2024. A 485-million-year history of Earth's surface temperature. *Science*, 385: 1316
- Kageyama M, Braconnot P, Harrison S P, et al. 2018. The PMIP4 contribution to CMIP6—Part 1: Overview and over-arching analysis plan. *Geosci Model Dev*, 11: 1033–1057
- Kaufman D, McKay N, Routson C, et al. 2020. A global database of Holocene paleotemperature records. *Sci Data*, 7: 1–34
- King J, Tierney J, Osman M, et al. 2023. DASH: A MATLAB toolbox for paleoclimate data assimilation. *Geosci Model Dev*, 16: 5653–5683
- Klein F, Goosse H. 2018. Reconstructing East African rainfall and Indian Ocean sea surface temperatures over the last centuries using data assimilation. *Clim Dyn*, 50: 3909–3929
- Li M. 2024. mingsongli/deepDA: V1.0.0-Nature Geoscience Release (v1.0.0). Zenodo, doi: 10.5281/zenodo.13777776
- Li M, Kump L R, Ridgwell A, et al. 2024. Coupled decline in ocean pH and carbonate saturation during the Palaeocene-Eocene Thermal Maximum. *Nat Geosci*, 17: 1299–1305
- Liu Z, Otto-Bliesner B L, He F, et al. 2009. Transient simulation of last deglaciation with a new mechanism for Bolling-Allerod warming. *Science*, 325: 310–314
- Liu Z, Lu Z, Wen X, et al. 2014. Evolution and forcing mechanisms of El Niño over the past 21,000 years. *Nature*, 515: 550–553
- Liu Z, Bao Y, Thompson L G, et al. 2023. Tropical mountain ice core $\delta^{18}\text{O}$: A Goldilocks indicator for global temperature change. *Sci Adv*, 9: eadi6725
- Liu Z, Cheng J, Zheng Y, et al. 2025. The seasonal temperature conundrum for the Holocene. *Sci Adv*, 11: eadt8950
- Lyu Z, Vuille M, Goosse H, et al. 2024. South American monsoon intensification during the last millennium driven by joint Pacific and Atlantic forcing. *Sci Adv*, 10: eado9543
- Mann M E, Zhang Z, Rutherford S, et al. 2009. Global signatures and dynamical origins of the Little Ice Age and Medieval Climate Anomaly. *Science*, 326: 1256–1260
- Marcott S A, Shakun J D, Clark P U, et al. 2013. A reconstruction of regional and global temperature for the past 11,300 years. *Science*, 339: 1198–1201
- Meng Z, Hakim G J. 2024. Reconstructing the tropical Pacific upper ocean using online data assimilation with a deep learning model. *J Adv Model Earth Syst*, 16: e2024MS004422
- Meng Z, Hakim G J, Steig E J. 2025. Coupled seasonal data assimilation of sea ice, ocean, and atmospheric dynamics over the last millennium. arXiv preprint, arXiv: 2501.14130
- Ning L, Liu Z, Mann M E, et al. 2025a. Decadal climate variability during the pre-industrial Common Era: Characteristics and mechanisms. *Sci Bull*, 70: 2190–2203
- Ning L, Xing F, Liu Z, et al. 2025b. Tripolar precipitation change accompanying water isotopes in the Holocene reanalysis of Asian monsoon hydroclimate. *Geophys Res Lett*, 52: e2025GL116451
- Osman M B, Tierney J E, Zhu J, et al. 2021. Globally resolved surface temperatures since the Last Glacial Maximum. *Nature*, 599: 239–244
- Otto-Bliesner B L, Brady E C, Fasullo J, et al. 2016. Climate variability and

- change since 850 CE: An ensemble approach with the Community Earth System Model (CESM). *Bull Amer Meteorol Soc*, 97: 735–754
- PAGES2k Consortium. 2017. A global multiproxy database for temperature reconstructions of the Common Era. *Sci Data*, 4: 170088
- Pendergrass A G, Hakim G J, Battisti D S, et al. 2012. Coupled air-mixed layer temperature predictability for climate reconstruction. *J Clim*, 25: 459–472
- Perkins W A, Hakim G J. 2017. Reconstructing paleoclimate fields using online data assimilation with a linear inverse model. *Clim Past*, 13: 421–436
- Perkins W A, Hakim G J. 2020. Linear inverse modeling for coupled atmosphere-ocean ensemble climate prediction. *J Adv Model Earth Syst*, 12: e2019MS001778
- Sanchez S C, Hakim G J, Saenger C P. 2021. Climate model teleconnection patterns govern the Niño-3.4 response to early Nineteenth-Century volcanism in coral-based data assimilation reconstructions. *J Clim*, 34: 1863–1880
- Shi F, Goosse H, Klein F, et al. 2019. Monopole mode of precipitation in East Asia modulated by the South China Sea over the last four centuries. *Geophys Res Lett*, 46: 14713–14722
- Snoll B, Ivanovic R F, Valdes P J, et al. 2022. Effect of orographic gravity wave drag on Northern Hemisphere climate in transient simulations of the last deglaciation. *Clim Dyn*, 59: 2067–2079
- Steiger N J, Hakim G J, Steig E J, et al. 2014. Assimilation of time-averaged pseudoproxies for climate reconstruction. *J Clim*, 27: 426–441
- Steiger N J, Smerdon J E, Cook E R, et al. 2018. A reconstruction of global hydroclimate and dynamical variables over the Common Era. *Sci Data*, 5: 180086
- Sun H, Lei L, Liu Z, et al. 2022. An analog offline EnKF for paleoclimate data assimilation. *J Adv Model Earth Syst*, 14: e2021MS002674
- Sun H, Lei L, Liu Z, et al. 2024. A hybrid gain analog offline EnKF for paleoclimate data assimilation. *J Adv Model Earth Syst*, 16: e2022MS003414
- Sun H, Lei L, Liu Z, et al. 2025. An online paleoclimate data assimilation with a deep learning-based network. *J Adv Model Earth Syst*, 17: e2024MS004675
- Sundqvist H S, Kaufman D S, McKay N P, et al. 2014. Arctic Holocene proxy climate database—New approaches to assessing geochronological accuracy and encoding climate variables. *Clim Past*, 10: 1605–1631
- Talagrand O. 1997. Assimilation of observations, an introduction. *J Meteorol Soc Jpn*, 75: 191–209
- Tardif R, Hakim G J, Perkins W A, et al. 2019. Last millennium Reanalysis with an expanded proxy database and seasonal proxy modeling. *Clim Past*, 15: 1251–1273
- Tian Z, Jiang D, Zhang R, et al. 2020. Transient climate simulations of the Holocene (version 1)—Experimental design and boundary conditions. *Geosci Model Dev*, 15: 4469–4487
- Tierney J E, Zhu J, King J, et al. 2020. Glacial cooling and climate sensitivity revisited. *Nature*, 584: 569–573
- Tierney J E, Zhu J, Li M, et al. 2022. Spatial patterns of climate change across the Paleocene-Eocene Thermal Maximum. *Proc Natl Acad Sci USA*, 119: e2205326119
- Tierney J E, King J, Osman M B, et al. 2025a. Pliocene warmth and patterns of climate change inferred from paleoclimate data assimilation. *AGU Adv*, 6: e2024AV001356
- Tierney J E, Judd E J, Osman M B, et al. 2025b. Advances in paleoclimate data assimilation. *Annu Rev Earth Planet Sci*, 53: 625–650
- von Storch H, Cubasch U, Gonzalez-Rouco J F, et al. 2000. Combining paleoclimatic evidence and GCMs by means of data assimilation through upscaling and nudging (DATUN). In: *Proceeding of the 11th Symposium on Global Change Studies*. 1: 28–31
- Walter R M, Sayani H R, Felis T, et al. 2023. The CoralHydro2k database: A global, actively curated compilation of coral $\delta^{18}\text{O}$ and Sr/Ca proxy records of tropical ocean hydrology and temperature for the Common Era. *Earth Syst Sci Data*, 15: 2081–2116
- Wei Y, Deng W, Chen X, et al. 2024. A comprehensive evaluation of machine learning on coral trace element paleothermometers for sea surface temperature reconstruction. *Paleoceanog Paleoclimatol*, 39: e2024PA004885
- Widmann M, Goosse H, van der Schrier G, et al. 2010. Using data assimilation to study extratropical Northern Hemisphere climate over the last millennium. *Clim Past*, 6: 627–644
- Wu F, Ning L, Liu Z, et al. 2025. A new last two millennium reanalysis based on hybrid gain analog offline EnKF and an expanded proxy database. *NPJ Clim Atmos Sci*, 8: 62
- Xie X, Liu X, Chen G, et al. 2019. A transient modeling study of the latitude dependence of East Asian winter monsoon variations on orbital timescales. *Geophys Res Lett*, 46: 7565–7573
- Yan M, Liu Z, Han J, et al. 2023. Relationship between the East Asian summer and winter monsoons at obliquity time scales. *J Clim*, 36: 3993–4003
- Zhang Q, Berntell E, Axelsson J, et al. 2021. Simulating the mid-Holocene, last interglacial and mid-Pliocene climate with EC-Earth3-LR. *Geosci Model Dev*, 14: 1147–1169
- Zhu F, Emile-Geay J, Hakim G J, et al. 2021. LMR Turbo (LMRt): A lightweight implementation of the LMR framework (0.8.0) [Software]. Zenodo. <https://doi.org/10.5281/zenodo.5205223>
- Zhu F, Emile-Geay J, Anchukaitis K J, et al. 2022. A re-appraisal of the ENSO response to volcanism with paleoclimate data assimilation. *Nat Commun*, 13: 747
- Zhu F, Emile-Geay J, Anchukaitis K J, et al. 2023. A pseudoproxy emulation of the PAGES 2k database using a hierarchy of proxy system models. *Sci Data*, 10: 624
- Zhu F, Emile-Geay J, Hakim G J, et al. 2024. cfr (v2024.1.26): A Python package for climate field reconstruction. *Geosci Model Dev*, 17: 3409–3431

(责任编辑：张仲石)