不同平流层气溶胶注入策略的可探测性分类与信号显现时间

Sai Charan Petchetti[1] 和 Daniele Visioni[2,*]

[1]加利福尼亚大学洛杉矶分校物理与天文学系，美国加利福尼亚州洛杉矶[2]康奈尔大学地球与大气科学系，美国纽约州伊萨卡
*联系作者。

电子邮箱： dv224@cornell.edu

K关键词： 平流层气溶胶注入、太阳辐射调节、信号显现时间、机器学习

arXiv:2611.28455v1 [physics.ao-ph] 21 Nov 2026

摘要

平流层气溶胶注入（SAI）可以采用不同的注入策略，在实现相同全球平均降温目标的同时，产生各异的区域气候响应。能否从观测到的气候中识别所采用的策略，取决于区域差异相对于内部变率有多大。本研究探讨：GAUSS集合（CESM2(WACCM6)，SSP2-4.5）中的五种反馈控制型平流层气溶胶注入策略，是否能根据近地表气温、降水和极地海冰比例的年度全球及区域平均值加以区分；这些差异又会在何时显现。这五种策略的目标均为使温度比工业化前水平高 1.0°C。我们采用多项逻辑回归分类器，并进行留一成员交叉验证；通过受限置换零假设计算家族错误率阈值，再结合经校准的持续性要求来定义出现年份。结果显示，这些策略在 2040 年代中后期（部署后约 8–14 年）变得可区分；同时使用全球和区域特征的模型，其峰值准确率为 0.87。此外，我们发现，基于单个区域—变量对训练的模型，其表现均不及使用全球和区域平均值的模型。这表明判别信号较弱，且分散在不同区域和变量中。赤道注入策略和高纬度注入策略最容易区分。对于较低的温度目标，分类器更容易识别多纬度策略。我们的结果表明，部署约十年或更久之后，年度气候特征便可帮助我们有把握地识别出所采用的协调注入策略。

1 引言

全球减少温室气体排放的努力单凭自身力量，不太可能避免气候变化带来的一些影响 [1], ，尤其是在二氧化碳移除（CDR）方法无法如期扩大规模的情况下。这促使人们研究气候干预方法，例如平流层气溶胶注入（SAI），将其作为减排和其他缓解策略的可能补充 [2] ，以避免未来几十年出现温度超调。平流层气溶胶注入部署情景将涉及多项选择，例如目标降温幅度、开始部署的日期以及注入策略本身。

文献中早已确立，不同的注入策略可能产生相同的全球平均降温，却带来不同的区域结果。赤道注入往往会导致热带地区残余过度冷却，而两极冷却不足 [3]。在其他纬度注入，既可以采用单一纬度对来维持半球对称性 [4, 5], ，也可以遵循更复杂的多纬度反馈控制设计，例如 Kravitz等 [6], 提出的方案，该方案同时平衡全球平均温度、半球间温度梯度以及赤道至两极的温度梯度。此外，还可以采用目标和权衡各异的其他策略，例如以降水为目标的策略 [7]。Zhang等 [4] 和 Bednarz等 [5] 在共同情景下比较了这些对称策略和多目标策略，发现即使所有策略的目标都是相同的全球平均温度，策略选择仍会影响气溶胶分布、注入效率以及广泛的地表响应。由于部署未必由单一行动方协调 [8] ，且不同策略会带来不同的区域权衡 [4], ，因此，能否根据观测到的气候响应判断所部署的策略，与监测、归因和治理息息相关。解决这一问题的一个有效框架是信号显现时间。这是气候研究中一种成熟的方法，可用于估计强迫信号何时能够从内部变率中区分出来 [9], ；近年来，研究者也开始利用机器学习模型来研究这一问题 [10]。

一项新兴研究利用机器学习分类器量化平流层气溶胶注入气候信号的可探测性，但迄今探讨的都是较容易的问题，其类别可分性高于策略归因。举例来说，Mamalakis 等人 [11], 训练了一个神经网络，用以量化部署后平流层气溶胶注入气候与部署前气候之间的可区分程度。同样，Labe 等人 [12] 采用与本研究相同的模型、背景情景和部署年份，并训练了一个逻辑回归分类器，逐年检测是否部署了单一平流层气溶胶注入策略，发现该信号在约十年内即可被检测到。与此同时，Dong 等人 [13] 对七种 单方面、未经协调的注入纬度进行分类，以判定哪一种导致了给定的气候响应，并发现它们在约两年内即可区分。这些研究均未探讨：在若干种 协调一致的策略中，哪一种策略——这些策略各自都以相同的全球平均温度为目标——导致了所观测到的气候状态；它们也没有提供显著性框架，以控制同时检验多个候选预测变量和多个候选年份所带来的影响。

在本研究中，我们针对这两项空白展开研究。我们使用不确定性、情景与策略地球工程评估（GAUSS）数据集中的近地表气温、降水和极地海冰比例的年度区域和全球平均值，训练一个多项逻辑回归分类器，以区分五种反馈控制型平流层气溶胶注入策略；这些策略均在共同的背景情景下，将相同的 1.0°C 高于工业化前水平的温度作为目标。我们通过留一成员交叉验证评估分类器性能，并将结果与一种保留每次模拟时间结构的受限置换零假设进行比较。该零假设产生一个族错误率显著性阈值和一个经校准的连续运行长度判据，二者共同定义出现年份。该框架使我们能够量化协调策略是否具有可区分性，以及这种信号何时出现、由哪些区域和变量承载。此外，我们还将训练好的分类器应用于两个其他温度目标下的多纬度模拟，以检验其能否推广到训练条件之外。

2 数据集

本研究使用 GAUSS 数据集 [2]。GAUSS 包含一系列在 CESM2(WACCM6) 的中层大气配置下开展的平流层气溶胶注入模拟实验 [14, 15] ，旨在以共享社会经济路径（SSP）2-4.5 作为背景排放情景，使全球变暖达到不同水平时的温度维持在相对于工业化前水平的特定值。我们采用先前在 [4] 中介绍的模拟，其目标是将全球平均温度维持在高于工业化前水平 1.0°C。 在这些模拟中，温度目标从部署开始时（2035 年）的约 1.5°C 逐步下降，并在前十年内降至 1.0°C [4]。

我们比较了五种不同的注入策略：赤道单点注入（0N）、在 15N+15S、30N+30S 和 60N+60S 进行对称注入，以及一种多纬度反馈控制策略。每种策略的模拟均包含三个集合成员。本研究使用 2035–2069年近地表气温、降水和极地海冰比例的月度输出，并将其平均为年值（海冰比例仅使用 9 月和 3 月的数值）。为检验检测信号的泛化能力，我们还纳入了 GAUSS 多纬度模拟，其中温度高于工业化前水平 0.5°C 和 1.5°C（每种温度目标各有三个集合成员）。

3 方法

对于每个集合成员，我们计算了近地表气温和降水的年度区域和全球平均值，以及 9 月（北极）和 3 月（南极）的海冰比例极地平均值（图 1）。我们对月度输出进行了重新居中。对于 GAUSS 多纬度策略中三个温度目标下的集合成员，由于未记录时间边界，我们将数据向前回移了 15 天。重新居中可确保每个日历年的月度序列覆盖 1 月至 12 月，并保留完整的 2035 年至 2069 年时段。

我们计算并使用了全球和区域平均值。区域定义如下（图 2）：

- 热带平均值： 20°S–20°N，所有经度 [4]
- 亚马孙盆地： 15°S–5°N， 78°W–50°W [4]
- 刚果盆地： 10°S–8°N， 12°E–31°E [4]
- Ni˜3.4 区域： 5°S–5°N， 170°W–120°W [4]
- 大西洋主要发展区（MDR）： 5°N–20°N， 85°W–15°W [4]

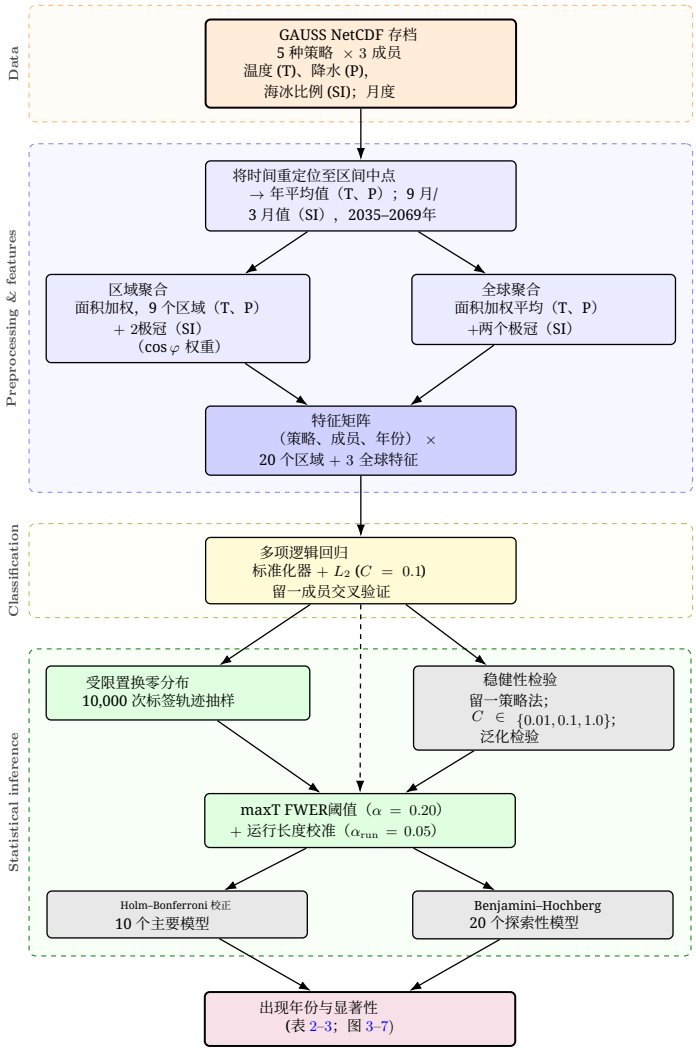

图 1. 分类与涌现检测流程的系统架构。虚线箭头表示输入显著性检验的经验准确率曲线，该检验同时使用零分布。

- 北大西洋中纬度地区： 30°N–60°N， 80°W–0° [5]
- 极冠： 70°N–90°N，所有经度 [5]
- 阿留申低压区： 30°N–70°N， 140°E–120°W [5]
- 东南亚季风区： 0°–30°N， 90°E–110°E [16]
- 北极： 65°N–90°N，所有经度（仅限九月海冰比例）
- 南极： 65°S–90°S，所有经度（仅限3月海冰比例）

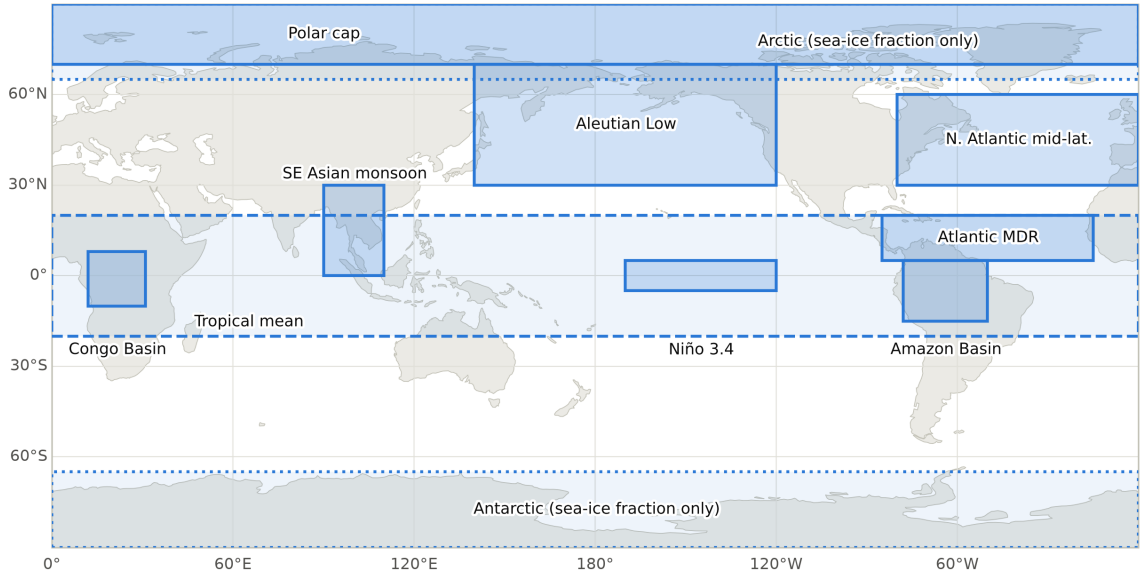

图2. 用于表征区域特征的区域。温度和降水取九个区域的平均值（实线边界；热带平均带以虚线表示）。海冰比例取北极（65°N–90°N，9月）和南极（65°S–90°S，3月）极冠的平均值（点线边界）。

区域定义参考了相关文献，这些研究指出，这些区域可能是受平流层气溶胶注入驱动变化影响最大的区域之一。Zhang等人 [4] 发现，注入位置越靠近赤道，每升高一度热带冷却所导致的热带平均降水减少幅度越大。他们发现，赤道策略导致的降幅最大，其次依序为15N+15S、30N+30S和60N+60S。我们采用Zhang等人 [4], 定义的同一热带平均区域，旨在捕捉不同策略之间类似的判别能力。同样，我们采用Zhang等人[4], 提出的亚马孙盆地、刚果盆地、Ni˜ño 3.4区和大西洋主要发展区；他们指出，由于热带环流变化及厄尔尼诺-南方涛动（ENSO）相关变率的影响，每个区域对注入策略的选择都尤为敏感。关于亚马孙盆地和刚果盆地，Simpson等 [17] 分析了它们对平流层气溶胶注入的水文气候敏感性。在刚果盆地，Abiodun等 [18] 和Obahoundje等 [19] 表明，平流层气溶胶注入的实施会改变干旱风险以及降水-温度复合极端事件。Ni˜ño 3.4区通常用于监测厄尔尼诺-南方涛动 [20]。虽然Gabriel和Robock [21]发现平流层气溶胶注入不会显著改变厄尔尼诺-南方涛动的频率或振幅，但他们记录的热带太平洋平均态响应仍取决于所施加强迫，这支持了我们在研究中使用Ni˜ño 3.4区。大西洋主要发展区最初被定义为大西洋主要飓风的主要起源区域 [22], ，后来研究发现该区域对平流层气溶胶注入的实施也很敏感 [23]。此外，我们采用Bednarz等 [5], 提出的北大西洋中纬度区、极冠区和阿留申低压区；他们利用这些区域构建冬季海平面气压指数，显示注入纬度与北半球地表气候之间存在关联响应。最后，我们采用Narenpitak等 [16], 提出的东南亚季风区；他们表明，这个受季风影响的区域对温室气体排放情景和不同太阳辐射调节（SRM）方法下的降水变化都尤为敏感。

九个区域的面积加权年平均值采用各网格单元中心纬度余弦值成比例的权重计算。仅计算北极地区的海冰比例

和南极地区的海冰比例，分别对应九月和三月，因此得到 20 个区域–变量特征。北极和南极海冰特征分别作为区域特征使用，二者合起来则构成全球海冰特征。

多项逻辑回归分类器（L_2 惩罚项，逆强度为 $C = 0.1$）使用在每个训练折上标准化为零均值和单位方差的输入特征进行训练，以根据年度特征平均值识别注入策略。模型采用留一集合成员交叉验证进行评估：三个折中的每一折均使用两个集合成员进行训练，并在留出成员上进行测试，同时分别计算每个日历年的分类准确率。例如，当留出集合成员 001 时，模型使用成员 002 和 003（全部五种策略）进行训练，并使用成员 001 中五种策略按年度整理的数据进行测试。由于每个折-年份组合中只有五个测试点，因此单个折-年份的准确率只能取六个可能值中的一个。为减小这种小样本噪声的影响，将三个折的准确率按日历年取平均，并进一步计算所得曲线的 5年居中滑动平均。出现年份旨在捕捉能够可靠区分这些策略的最早年份。根据 5年平滑平均准确率计算朴素出现年份，将平滑准确率首次超过 0.5 的年份定义为该年份。还根据未经平滑的平均准确率曲线计算基于运行的涌现年份，其显著性阈值由进行 10,000 次抽样的零置换检验得出。

每次 10,000 次零抽样都会将五个标签双射映射到完整轨迹上——全部 35 年的数据一起移动到新的策略标签——且对每个集合成员独立抽取。最大统计量（maxT）置换检验在 $\alpha = 0.20$ 水平上进行：将阈值设为零曲线在各年份的最大准确率分布的第80百分位数。随后，通过要求连续高于阈值的年份运行长度足够长，使得长度至少为该值的运行在至多 5% 的零曲线中出现（$\alpha_{\mathrm{run}} = 0.05$），来确定出现年份。在主要分析中，所有主要模型的这一校准运行长度均为两年。因此，若未经平滑的折叠平均准确率连续两年超过第80百分位数阈值，则将这两年中的第一年定为显现年份。随后计算并报告采用 $+1$ 校正的 Monte Carlo p 值。 $+1$ 校正会在分子和分母上各加 1，以避免 p 值为零。

采用 Holm-Bonferroni方法对预先指定的十个主要模型进行校正，以控制家族错误率（FWER）。这十个主要模型包括：区域温度和降水模型、全变量区域模型、仅全球模型、全球加区域模型、三个单一全球变量模型，以及北极和南极海冰模型。采用保守程度较低的 Benjamini-Hochberg 方法，对 20 个单一区域—变量探索性模型进行校正，以控制错误发现率（FDR）（2 个变量 $\times$ 9 区域 $+$ 2 极地海冰比例特征）。通过逐一排除各策略（10,000 次零假设置换），并改变 C 至$\{0.01,\ 0.1,\ 1.0\}$ 的范围（1,000 次零假设置换），评估稳健性。

在泛化检验中，分类器使用 1。0℃ GAUSS模拟进行训练，并在未用于训练的多纬度模拟上进行测试（表 1）。每个分类器均在其留出集成成员的多纬度模拟上进行测试，测试温度目标为 1。0° C（训练目标，作为基线）、0。5℃ 和 1。5℃。例如，使用成员 002 和 003 训练的分类器，会在每个温度目标下对成员 001 进行测试。

泛化检验仅针对多纬度策略进行评估。由于 GAUSS集合成员仅有三个，每年被正确分类的多纬度 GAUSS模拟所占比例只能取 0、1/3、2/3 或 1，这对于显著性检验而言过于粗略。为此，在 1.0° C GAUSS数据集上训练的五策略分类器会在泛化检验测试数据集上进行测试，并根据分类器分配给多纬度策略的概率对每年进行评分，该概率取三个折的平均值。此外，还会基于 10,000 次零置换抽样进行显著性检验。在每次抽样中，按照主要分析中的方法，在 1.0℃ GAUSS模拟上重新训练分类器，并在每个集合成员内部打乱策略标签，然后在泛化检验测试数据集上进行测试。随后，将使用未打乱标签训练的分类器的年度得分与 10,000 条零得分曲线进行比较。峰值 p值是指以下抽样的比例：其最高年度得分至少与未打乱标签分类器的最高年度得分一样高，并采用 $+1$校正。泛化检验结果未进行 Holm-Bonferroni 校正。

出现年份的计算方法与主要分析相同，但校准运行长度通常更长，且在不同测试之间有所不同（两年至五年）。由于测试数据集仅包含多纬度模拟，每次零抽样针对单一策略给出平滑变化的概率，而不是对全部五种策略计算准确率，因此，越过阈值的零曲线往往会在许多年里连续保持在阈值之上。因此，这些出现年份总体上比主要分析中的结果更保守，且不能直接进行比较。

表1. 主要分析和跨目标泛化测试的设计。

	主要分析	跨目标
问题	五种策略中哪一种 已部署?	多纬度策略是否 在不同的 温度目标?
训练数据	GAUSS，1.0°C 目标，全部五种策略中的两种 成员（留一成员交叉验证）	
测试数据	留出成员 ，全部五种 策略，1.0°C	留出成员的 多纬度模拟于 1.0°C（基线）、0.5°C，以及 1.5°C
年度得分	准确率（ 正确分类的策略占比）	分配给 多纬度策略
平均值	三个折	三个折
零假设：训练 标签	每个集成内随机打乱模拟的策略标签 集合成员（10,000 次抽取）	
零假设：测试 标签	与留出部分一同随机打乱 该成员自身的标签	每次模拟都会取其 该成员打乱后的 多纬度标签
显著性	峰值 p 值：零曲线中年度最高值所占的比例 得分至少与真实曲线的最高年度 得分（+1 校正）	
显现的运行长度	2 年	2–5 年（按每项测试校准）
显现 跨模型 校正	Holm–Bonferroni 校正（10 个主要检验）， Benjamini–Hochberg（20 个探索性检验）	None

4 结果

表2。 十个预先指定特征集（5种策略，随机基准 $= 0.20$）的标题分类结果。EY 表示显现年份。各行按峰值置换 p 值排序；最后一列报告在 $\alpha = 0.05$ 处对 FWER 进行 Holm–Bonferroni 控制后的显著性。Holm–Bonferroni 校正仅适用于峰值 p 值； † 标记未通过 Holm–Bonferroni 校正的模型的显现年份。

特征集	朴素出现年份	基于运行的出现年份	峰值准确率	峰值 p	Holm 显著性
全球 + 区域（所有变量）	2043	2043	0.87	0.0006	Yes
所有区域（所有变量）	2044	2045	0.80	0.0009	Yes
仅全局（所有变量）	2047	2049	0.73	0.0028	Yes
海冰比例 ，南极	–	2063	0.60	0.0068	Yes
降水，区域	2051	2055†	0.67	0.0127	No
温度，区域	2053	2050†	0.67	0.0132	No
降水，全球	–	2049†	0.53	0.0366	No
海冰比例，全球	–	2059†	0.53	0.0926	No
海冰比例 ，北极	–	–	0.40	0.4084	No
温度，全球	–	–	0.40	0.4675	No

表2 展示了十个预先指定的主要模型的结果。三种全变量主要模型（全球加区域、全区域和仅全球）以及南极海冰模型的峰值 p 值在经过 Holm–Bonferroni 校正后仍具有显著性，校正阈值为 $\alpha = 0.05$。区域降水与温度模型，以及全球降水与海冰模型在 2049 至 2059 年间显现，但未能通过 FWER 校正。北极海冰和全球温度模型均未显现，也未能通过 FWER 校正。

上述全球平均温度的结果反映了 GAUSS实验的设计，该实验包含一种反馈机制，每年改变注入量，以保持全球平均

地表温度对所有注入策略均保持不变。除全球平均地表温度外，多纬度注入策略还旨在保持半球间梯度和赤道至两极梯度不变。在全球平均属性中，地表温度正如预期，是最不具判别性的特征（$p = 0.47$；表 2）。其他全局变量——海冰比例和降水——比全球温度携带更多的判别信号（两者的峰值准确率均为 0.53，而全球温度为 0.40），但未通过家族错误率校正。

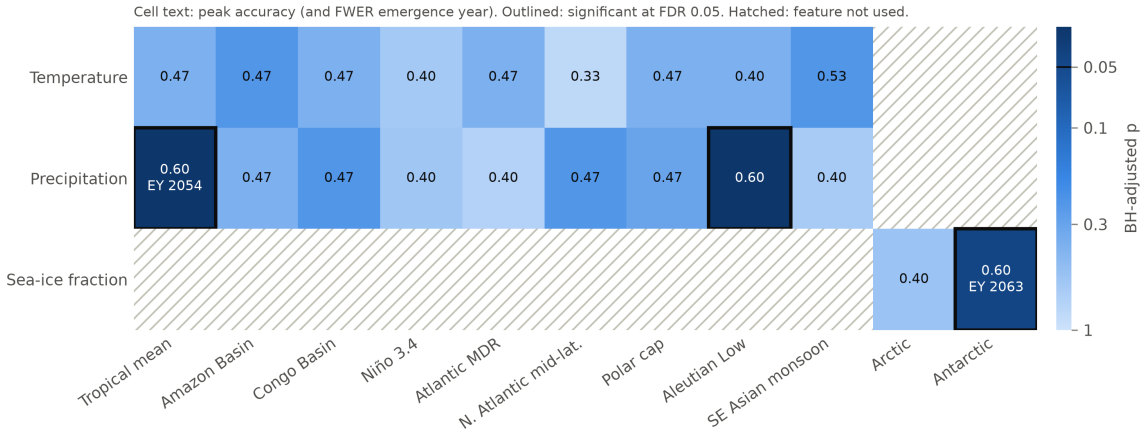

图3. 20 个探索性单区域 × 变量模型（5 种策略，随机基准 $= 0.20$）的分类结果。单元格颜色表示经 Benjamini–Hochberg 校正的峰值 p 值；单元格文本给出峰值准确率，以及（如有）基于运行的涌现年份。带边框的单元格表示在错误发现率 $= 0.05$ 下显著。带斜线的单元格表示未使用的区域–变量组合。对于任何单一特征，朴素运行平均准确率均未超过 0.5，因此未报告朴素出现年份。图标题中的“FWER涌现年份”指基于运行的涌现年份。

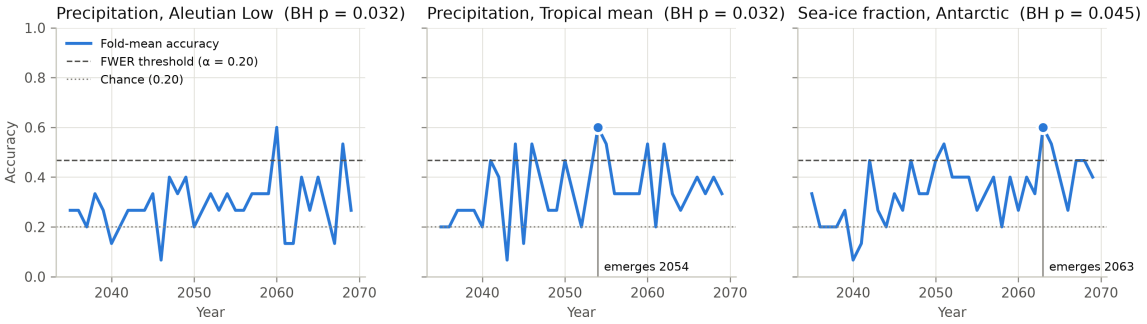

图4. 通过错误发现率控制的三个单区域 × 变量模型在三个折上的平均年度准确率。虚线表示家族错误率阈值（$\alpha = 0.20$），点线表示随机基准（0.20）。标记表示基于运行的涌现年份，即连续两年高于阈值的首年。阿留申低压降水虽超过阈值，但从未连续两年超过，因此未报告涌现年份。

图中 20 个探索性单一区域—变量模型3中，只有三个模型通过错误发现率控制（图4），其中两个是降水模型，一个是南极海冰比例模型；没有一个区域温度模型通过错误发现率控制。考虑到降水具有较高的年际变率，这一点值得注意，也可能反映了策略驱动的环流和热带辐合带（ITCZ）变化 [4, 5, 7]，不过，三个显著模型提供的证据有限。更值得注意的是，没有任何单一特征的表现能与组合的主要模型相当：单一特征达到的最高峰值准确率为 0.60，而表2中的全球加区域模型为 0.87；并且没有任何单一特征的平滑准确率超过 0.5。这表明策略间判别信号较弱，且分散在多个特征中，这正是策略设计旨在尽量减少策略间显著气候差异时预期会出现的结果。

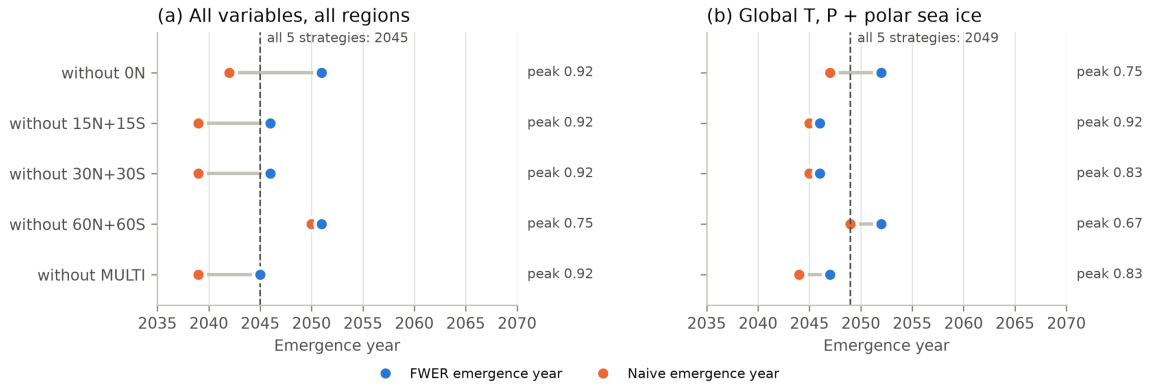

图5。 留一策略稳健性检验。依次移除每种策略，并使用剩余四种策略重新训练分类器（随机基准 $= 0.25$）。面板 (a) 使用所有变量和所有区域；面板 (b) 使用三个全局变量。标记显示每个留出策略的朴素出现年份和基于运行的涌现年份，右侧文本给出峰值准确率。虚线表示相应五策略模型的基于运行的涌现年份。十个模型中有九个相对于其置换零分布具有显著性（峰值 $p \leq 0.033$）。不包含 60N+60S 的仅全局模型是唯一不显著的模型（$p = 0.092$）。图例中的“FWER涌现年份”指基于运行的涌现年份。

根据图5所示的留一策略分析，我们发现，在全变量全区域模型（不含全球温度和降水）及仅全球模型中，即使剔除任意一个策略，判别信号仍会显现。剔除 0N 或 60N+60S 会将显现推迟至 2051–2052 年；剔除 60N+60S 还会降低峰值准确率，而剔除 15N+15S、30N+30S 或多纬度策略，全区域显现年份最多只变化一年，并会使仅全球显现提前。

这些结果表明，赤道（0N）和高纬度（60N+60S）策略是最容易区分的，而低纬度半球对称策略（15N+15S、30N+30S）和多纬度策略最难区分，这与 Zhang等人 [4], 的发现一致；鉴于这两类策略处于所测试注入纬度范围的相反两端，其产生的气候影响与其他策略差异最大，这一结果也是预期之中的。赤道（0N）策略在很大程度上将气溶胶限制在热带地区，导致单位全球降温对应的降水减少幅度最大，并使赤道至极地温度梯度得到不足补偿。另一方面，60N+60S 注入策略导致单位全球降温对应的降水减少幅度最小，使北极过度降温，而热带地区降温不足 [4]。

表3。 逻辑回归模型对逆正则化强度 C 的敏感性。四个特征集分别在 $C \in \{0.01, 0.1, 1.0\}$ 下重新运行（随机基准 $= 0.20$； $n_{\text{perm}} = 1000$； p-值未经校正）。

特征集	C	基于运行的出现年份	峰值准确率	峰值 p
仅全局	0.01	2052	0.67	0.0040
仅全局	0.10	2049	0.73	0.0020
仅全局	1.00	2049	0.73	0.0040
全球 + 区域	0.01	2049	0.80	0.0010
全球 + 区域	0.10	2043	0.87	0.0010
全球 + 区域	1.00	2039	0.93	0.0010
所有区域	0.01	2051	0.73	0.0060
所有区域	0.10	2045	0.80	0.0010
所有区域	1.00	2039	0.87	0.0010
降水，区域	0.01	2055	0.73	0.0030
降水，区域	0.10	2055	0.67	0.0200
降水，区域	1.00	2049	0.73	0.0060

正则化强度敏感性检验（表 3）证实，上述发现的可分性并非由所选的正则化强度驱动，尽管出现年份对其较为敏感。对于 $C \in \{0.01, 0.1, 1.0\}$，相较于 $C = 0.1$，出现年份最多可提前或推迟约六年，并且每种配置在 $p < 0.05$ 时仍具有显著性。较弱的正则化会使显现提前。

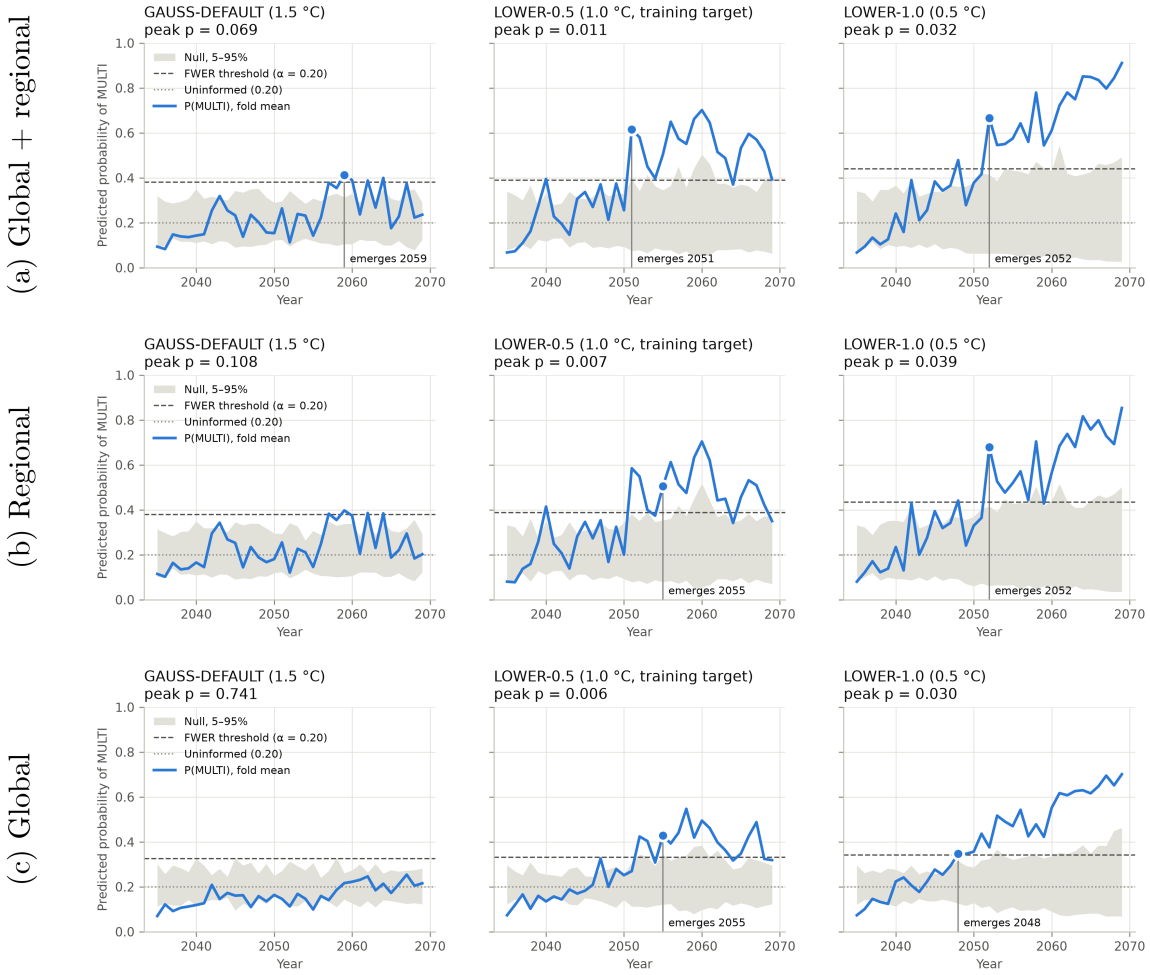

图 6. 在 1.5°C（GAUSS- DEFAULT）、1.0°C 训练目标（LOWER-0.5）和 0.5°C（LOWER-1.0）下，留出模拟中多纬度策略（MULTI）被赋予的概率，从左到右分别使用全球和区域（a）、区域（b）以及全球（c）特征。阴影表示 5–95% 置换零分布，虚线表示家族错误率阈值，点线表示随机基准（0.20）。面板标题包含未经校正的峰值 p 值。无论峰值 p 值是否显著，均显示涌现年份。

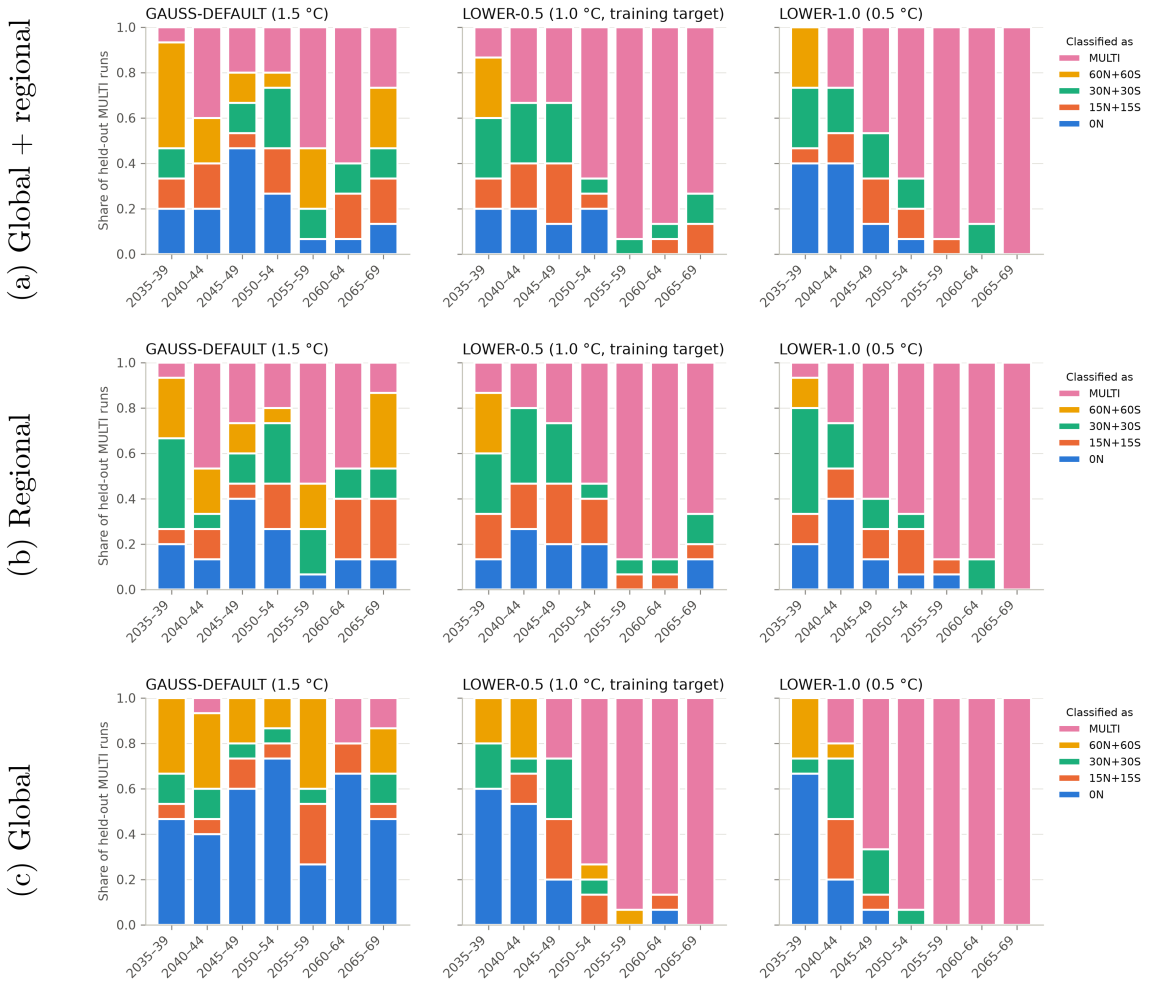

图 7. 从左到右，分别使用全球和区域（a）、区域（b）以及全球（c）特征，对留出组的多纬度（MULTI）模拟在 1.5℃（GAUSS-DEFAULT）、1.0℃ 训练目标（LOWER-0.5）和 0.5℃（LOWER-1.0）下进行分类。柱形图显示了 MULTI 模拟在各个五年时期中的分类结果。

图表 6 和 7 展示了分类器在未参与训练的模拟中识别多纬度策略的表现。随着温度目标降低，分类器似乎越来越擅长识别多纬度策略。较低的温度目标需要更多降温，因此需要注入更多气溶胶，信号也更强。正如 Visioni 等人所示[24]，，平流层气溶胶注入造成的各种地表影响与降温幅度近似呈线性关系，这表明在较低目标下，该策略的判别信号更强。对于全部三个特征集，分配给多纬度策略的峰值概率在 1.5°C 目标下最低（0.25–0.41），在 1.0°C 训练目标下居中（0.55–0.70），在 0.5°C 目标下最高（0.70–0.91）（图 6）。0.5°C 的结果对于全部三个特征集均具有显著性（未校正 $p \leq 0.039$），而 1.5°C 的结果（$p \geq 0.069$）均不显著。分类器预测的策略也呈现出相同模式（图 7）。从 2050 年代开始，大多数 0.5°C 模拟都被正确分类为多纬度策略，而 1.5°C 模拟则分散在几种策略中，被归类为多纬度策略的情况也较少。

5 讨论

我们的研究与 Labe 等人 [12], 的工作最为接近。他们使用相同的模型、背景情景和部署年份（CESM2(WACCM6)、SSP2-4.5、2035 年），采用集合成员留出法训练逻辑回归分类器，并按年进行预测。Labe 等人试图回答一个二元问题：使用 ARISE 1.5°C 多纬度策略 [25]时，预测平流层气溶胶注入是否已部署。他们发现，判别信号足够强，仅使用全球陆地温度训练的模型就能在最初约五年后完美区分平流层气溶胶注入与未注入情景（在整个 2035–2069年期间的准确率为 92.6%）。识别部署的是五种策略中的哪一种，属于信号较弱的问题，因为 GAUSS 反馈控制器会使每种策略都维持在相同的全球平均温度目标。在我们的研究中，只有到 2040 年代中后期才开始显现可区分性——也就是部署后约 8–14 年。

我们观察到显现较慢，这反映出所有策略的气候目标大致相同，也与更严格的族错误率置换检验有关，还可能与集合的规模有关。我们将文献中此前讨论的二元检测框架扩展到多分类问题，即确定部署了若干反馈控制策略中的哪一种。我们还提出了一种涌现判据，将族错误率置换阈值与经过校准的两年持续性要求相结合，以识别可认为可区分性开始显现的首个年份。

上述结果受到若干局限因素的影响。最重要的局限在于集合规模。每种策略有三个成员，因此置换参考集包含 $(5!)^3 \approx 1.7 \times 10^6$ 种重新标记方式，但移除对类别标签进行统一重新标记的置换后，统计上不同的方式只有 $\approx 1.4 \times 10^4$ 种，因此可达到的最小 p值为$\approx 10^{-4}$。这一下限从两个方面限制了我们的统计分析。首先，它限制了任何单个模型的显著性：可分性最强的两个特征集，其值已与 p值下限相差不到一个数量级（表2），这使我们的方法更难判断一个模型是否比另一个模型具有更强的判别能力。其次，该方法几乎没有余地进行多重比较家族错误率校正，因为在多项检验之间分配固定的显著性预算，会使所需阈值逼近下限，不过这可能不会实质性影响本研究考察的规模相对较小的模型族。

此外，集成规模较小也限制了统计功效。真实存在但较弱的判别信号可能无法连续两年达到显著性阈值。因此，受限于集成规模，涌现年份可能会推迟，而一些未通过校正的单变量模型仍可能包含有用的判别信号。然而，由于大多数模型在其耦合模式比较计划（CMIP）式模拟中提供的集合成员不超过三个，我们认为我们的结果与许多其他 CMIP 相关应用具有相关性。

我们的所有模型均使用单一地球系统模式（CESM2(WACCM6)）生成的数据进行训练和测试。由于地球系统模式在大气、海洋和冰系统的处理方式上存在显著差异， [26], 涌现年份以及我们对最具判别力的变量和区域的讨论可能取决于所用模型。研究表明，模型的内部变率会导致模型模拟气候与观测气候之间存在差异，即便时间跨度长达数十年 [27]。由于出现年份取决于可区分信号相对于内部变率的大小，因此应用于实际观测气候时，显现情况可能有所不同。将再分析资料中的观测气候变率纳入这一框架，是未来研究的一个有前景方向。另一个有前景的方向是，将本研究提出的框架应用于面向 CMIP7 [28], 提出的地球工程模型比较计划（GeoMIP）试验；这些试验将纳入

在共同的 1.5℃ 温度目标下，不同模型中的多种平流层气溶胶注入实施方案和海洋云增亮（MCB），以检验我们的结果是否适用于不同模型，以及能否准确区分不同类型的干预措施（例如平流层气溶胶注入与海洋云增亮）。

我们所有的模型都基于线性分类器（多项逻辑回归）进行训练，使用按区域平均的年度均值；这种方法可能会忽略卷积网络在全空间分辨率的地图上训练时能够学会利用的非线性特征。此外，这些区域是根据非穷尽性的文献综述事先选定的，而不是事后选择以最大化可分性。

主要分析检验了一个降温目标（低于 1.5 ℃ 0.5 ℃）和五种半球对称策略，旨在分析一组目标相似且采取“平衡”干预的策略。较低的温度目标和半球不对称策略可能会使显现更早发生。泛化检验仅针对多纬度策略进行，因为其他策略未在其他温度目标下进行模拟。因此，我们的泛化检验衡量的是多纬度策略是否仍然可检测，而不是在不同温度目标下五种策略是否仍可区分。泛化检验中的 p 值也未针对多重比较进行校正。

最后，基于连续运行的出现年份指标取决于显著性带（$\alpha = 0.20$）和运行长度校准（$\alpha_{\mathrm{run}} = 0.05$）的选择。这些选择并未经过全面变动。正则化强度则单独进行了调整（表 3），结果显示出现年份对这一选择较为敏感。

6 结论

本研究探讨了：在使用地表温度、降水和极地海冰比例的年度区域和全球平均值进行训练和测试时，五种反馈控制的 GAUSS 注入策略是否能够区分。我们还尝试确定判别信号在何时变得具有统计可检测性。我们使用多项逻辑回归分类器，并通过留一成员交叉验证，将其与受限置换零假设进行比较。各策略在 2040 年代中后期开始呈现可区分性，即部署约 8–14 年后、为达到 1.0℃ 目标而进行的十年逐步减排期即将结束时；交叉验证准确率峰值为 ~ 0.87。在本研究中，表现最佳的模型对三个变量均采用了全球和区域特征集。仅使用区域温度和降水的模型，其显现时间晚于同时使用全球和区域特征的模型，且未通过族错误率校正。由于反馈控制器的目标是实现共同的全球平均温度，全球温度特征通常不具信息量，因此基于该特征训练的模型是表现最差的主要模型。随着纳入更多全球和区域尺度的特征，我们的结果有所改善，这表明判别信号较弱，并分散在各个特征中。

本研究报告的显现时间晚于此前旨在检测是否部署了平流层气溶胶注入 [12], 或区分非受控单边注入 [13]的文献所报告的时间。鉴于所有策略都以相同的全球平均温度为目标，并采用共用的反馈控制器，再加上严格的显著性检验，这一结果在预期之中。尽管报告的涌现年份应视为取决于该模型、集合规模和部署时间表，我们的结果表明，若仅能获得区域尺度、年度分辨率的地表气候观测数据，那么大约需要十年或更久的数据，才能仅凭某一年的数据有把握地判断部署的是哪种协调策略；不过，整合多样化的监测气候特征有助于缩短检测时间。我们还提出了一种具体方法来量化显现，将基于受限置换零分布得出的族错误率阈值与经过校准的持续性要求（主要分析中为两年）相结合。在泛化检验中，随着温度目标降低，分类器识别多纬度策略的表现会更好。因此，策略分类可能取决于训练时采用的温度目标。

数据可用性

GAUSS模拟的输出 [4] 可向通讯作者索取；更多信息请见
h
t
tps://climate-engineering.mae.cornell.edu/gauss-geoengineering-assessment-across-uncertainty-scenarios-and-strategies。
用于开展分析和生成图表的代码可在以下网址获取：https://github.com/astrocharan/sai-classification/。

参考文献

[1]联合国环境规划署 2026年 限制超调：应对超过 1.5°C 的情况及回归路径（内罗毕：联合国环境规划署）https://doi.org/10.59117/20.500.11822/49857[2] MacMartin D G、Visioni D、Kravitz B、Richter J H、Felgenhauer T、Lee W R、Morrow D R、Parson E A 和 Sugiyama M 2022 太阳辐射调节建模情景 Proc. Natl Acad. Sci. USA 119 e2202230119 https://doi.org/10.1073/pnas.2202230119[3] Kravitz B 等 2019 比较地球工程对地表和对流层的影响：不同 SO_2 注入策略 J. Geophys. Res. Atmos. 124 7900–18 https://doi.org/10.1029/2019JD030329[4]Zhang Y、MacMartin D G、Visioni D、Bednarz E M 和 Kravitz B 2024 平流层气溶胶注入的半球对称策略 Earth Syst. Dyn. 15 191–213 https: // doi.org/10.5194/esd-15-191-2024[5] Bednarz E M、Butler A H、Visioni D、Zhang Y、Kravitz B 和 MacMartin D G 2023 注入策略——驱动大气环流及平流层气溶胶地球工程的臭氧响应 Atmos. Chem. Phys. 23 13665–84 https://doi.org/10.5194/acp-23-13665-2023[6] Kravitz B、MacMartin D G、Mills M J、Richter J H、Tilmes S、Lamarque J-F、Tribbia J J 和 Vitt F 2017 首次模拟平流层硫酸盐气溶胶地球工程设计，以实现多个同步气候目标 J. Geophys. Res. Atmos. 122 12616–34 https: //doi.org/10.1002/2017JD026874[7] Lee W、MacMartin D、Visioni D 和 Kravitz B 2020 扩展平流层气溶胶地球工程的设计空间，纳入基于降水的目标并探讨权衡 Earth Syst. Dyn. 11 1051–72 https://doi.org/10.5194/esd-11-1051-2020[8] M¨a¨att¨anen A、Ravetta F、Bureau J、Lurton T 和 Boucher O 2026 由两个非合作参与者实施平流层气溶胶注入部署情景的理想化建模 Environ. Sci.: Atmos. 6 324–37 https://doi.org/10.1039/D5EA00022J[9] Hawkins E 和 Sutton R 2012 气候信号的显现时间 Geophys. Res. Lett. 39 L01702 https://doi.org/10.1029/2011GL050087[10] Barnes E A、Toms B、Hurrell J W、Ebert-Uphoff I、Anderson C 和 D Anderson 2020 人工神经网络学习到的强迫变化指标模式 J. Adv. Model. Earth Syst. 12 e2020MS002195 https://doi.org/10.1029/2020MS002195[11] Mamalakis A、Barnes E A 和 Hurrell J W 2023 利用可解释人工智能量化平流层气溶胶注入后的“气候可区分性” Geophys. Res. Lett. 50 e2023GL106137 https://doi.org/10.1029/2023GL106137[12] Labe Z M、Barnes E A 和 Hurrell J W 2023 识别 ARISE-SAI-1.5模拟中的气候模式区域显现 Environ. Res. Lett. 18 044031 https://doi.org/ 10.1088/1748-9326/acc81a[13] Dong C、Connolly C 和 Barnes E A 2026 利用可解释人工智能区分不同平流层气溶胶注入纬度导致的短期气候响应 J. Geophys. Res. Atmos. 131 e2025JD045560 https://doi.org/10.1029/2025JD045560[14] Danabasoglu G 等 2020 社区地球系统模型第2版（CESM2） J. Adv. Model. Earth Syst. 12 e2019MS001916 https://doi.org/10.1029/2019MS001916[15]Gettelman A 等 2019 全大气社区气候模型第6版（WACCM6） J. Geophys. Res. Atmos. 124 12380–403 https://doi.org/10.1029/ 2019JD030943 [16] Narenpitak P、Kongkulsiri S、Tomkratoke S 和 Sirisup S 2024 太阳辐射调节对东南亚大陆地区及邻近海洋地表温度和降水的区域影响 Sci. Rep. 14 22713 https://doi.org/10.1038/s41598-024-73149-6

[17] Simpson I R, Tilmes S, Richter J H, Kravitz B, MacMartin D G, Mills M J, Fasullo J T 和 Pendergrass A G 2019年 平流层硫酸盐地球工程的区域水文气候响应及平流层加热的作用 J. Geophys. Res. Atmos. 124 12587–616 https://doi.org/10.1029/2019JD031093[18] Abiodun B J, Odoulami R C, Sawadogo W, Oloniyo O A, Abatan A A, New M, Lennard C, Pinto I, Egbebiyi T S 和 MacMartin D G 2021年 平流层气溶胶注入对非洲主要流域干旱风险管理的潜在影响 Clim. Change 169 31 https://doi.org/10.1007/s10584-021-03268-w[19] Obahoundje S, Nguessan-Bi V H, Diedhiou A, Kravitz B 和 Moore J C 2023年 平流层气溶胶地球工程对非洲复合降水和温度极端事件的影响 Sci. Total Environ. 863 160806 https://doi.org/10.1016/j.scitotenv.2022.160806[20] Trenberth K E 1997年 厄尔尼诺的定义 Bull. Am. Meteorol. Soc. 78 2771–7 https: //doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2[21] Gabriel C J 和 Robock A 2015年 平流层地球工程对厄尔尼诺/南方涛动的影响 Atmos. Chem. Phys. 15 11949–66 https://doi.org/10.5194/acp-15-11949-2015[22] Goldenberg S B, Landsea C W, Mestas-Nuñez A M 和 Gray W M 2001年 大西洋飓风活动近期增加：原因和影响 Science 293 474–9 https://doi.org/ 10.1126/science.1060040[23] Moore J C 等 2015年 地球工程对大西洋飓风风暴潮的影响 Proc. Natl Acad. Sci. USA 112 13794–9 https://doi.org/10.1073/pnas.1510530112[24] Visioni D, Bednarz E M, MacMartin D G, Kravitz B 和 Goddard P B 2023年 基线时段的选择会影响对平流层气溶胶注入结果的评估 《地球的未来》11 e2023EF003851 https://doi.org/10.1029/2023EF003851[25] Richter J H, Visioni D, MacMartin D G, Bailey D A, Rosenbloom N, Dobbins B, Lee W R, Tye M 和 Lamarque J-F 2022年 使用平流层气溶胶注入评估太阳辐射气候干预对地球系统的响应和影响（ARISE-SAI）：方案及首批模拟的初步结果 Geosci. Model Dev. 15 8221–43 https://doi.org/10.5194/gmd-15-8221-2022[26] Henry M, Haywood J, Jones A, Dalvi M, Wells A, Visioni D, Bednarz E M, MacMartin D G, Lee W 和 Tye M R 2023年 使用相同多目标平流层气溶胶注入策略的 UKESM1 和 CESM2 模拟比较 Atmos. Chem. Phys. 23 13369–85 https://doi.org/10.5194/acp-23-13369-2023[27] Jain S, Scaife A A, Shepherd T G, Deser C, Dunstone N, Schmidt G A, Trenberth K E 和 Turkington T 2023年 内部变率对气候模型评估的重要性 npj Clim. Atmos. Sci. 6 68 https://doi.org/10.1038/s41612-023-00389-0[28] Visioni D 等 2026年 地球工程模型比较计划（GeoMIP）对 CMIP7 的贡献——新实验方案说明及初步结果 Geosci. Model Dev. 19 8469–99 https://doi.org/10.5194/gmd-19-8469-2026