arXiv:2301.11921v1 [physics.data-an] 30 Nov 2022

使用不确定性感知机器学习模型研究气溶胶-云相互作用

Maëlys Solal牛津大学计算机科学系，英国
maelys.solal@ens.psl.eu

Andrew JessonOATML 牛津大学计算机科学系，英国

Yarin GalOATML 牛津大学计算机科学系，英国

Alyson DouglasAOPP 牛津大学物理学系，英国

摘要

气溶胶-云相互作用（ACI）包括气溶胶进入云并影响云属性所产生的各种效应。总体而言，气溶胶浓度增加会导致液滴尺寸变小，进而形成更大、更明亮、寿命更长的云，这些云会反射更多阳光并使地球降温。然而，这种效应的强度具有异质性，即取决于周围环境，使得气溶胶-云相互作用成为当前气候模型中最不确定的效应之一。在本研究中，我们利用因果机器学习，通过将问题重新表述为处理（气溶胶）和结果（液滴半径的变化），从卫星观测中估计气溶胶-云相互作用。我们预测气溶胶对云的因果效应，并给出取决于可能影响气溶胶作用的未知因素的不确定性界限。在所评估的三个气候模型中，我们发现只有一个能够合理地再现这一趋势，从而使其对气溶胶-云相互作用导致冷却的估计更具可信度。

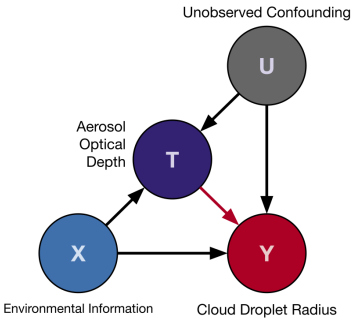
1 引言

以人类排放污染形式存在的气溶胶进入大气，并最终与云发生相互作用，从而产生气溶胶-云相互作用（ACI）。当气溶胶进入云中时，会催化一系列因果链事件。首先，气溶胶颗粒活化为云滴凝结核，增加云内液滴数量，重新分配水蒸气，使云滴的平均半径减小，并最终提高云的亮度（图1(a)） [1]。总体而言，大气中气溶胶的增加会导致形成更大、更明亮、寿命更长的云，从而反射更多入射阳光。因此，气溶胶-云相互作用是一个净冷却过程，能够抵消因CO_2水平上升所导致的一部分变暖。然而，该效应的强度取决于云周围的局地环境。由于当前模型在模拟具有如此环境异质性的气溶胶-云相互作用方面能力有限，气溶胶-云相互作用仍是当前气候模型中最不确定的效应之一[2, 3]。鉴于空间分辨率较低且参数化方案有限，气候模型只能对气溶胶-云相互作用进行近似模拟，且通常仅依赖少数环境参数，例如网格单元内的相对湿度。这些因素增加了未来预测的不确定性。目前，最先进的气候模型估计，气溶胶-云相互作用所导致的冷却范围可能抵消温室气体排放所致变暖的0%-50%。

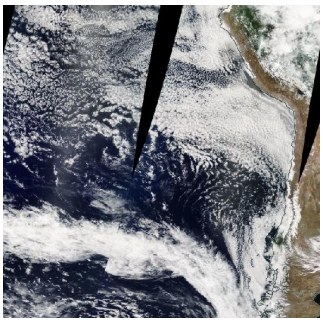
本研究采用因果机器学习，基于卫星观测估算气溶胶-云相互作用（ACI），并将问题重新表述为处理（气溶胶）与结果（云滴半径的变化）。我们预测气溶胶对云的因果效应，并给出不确定性界限，将其与

用机器学习应对气候变化：NeurIPS 2022 研讨会。

气候模型 ACI 的参数化方案进行比较。我们考虑两项假设遭到违背时产生的不确定性：正性（或重叠）和无混杂性（或不存在隐性混杂）。正性违背是由于数据未能充分涵盖所有处理水平，例如用气溶胶“处理”云。未测量混杂指未观测到、同时影响处理和结果的因素，例如湿度导致气溶胶膨胀并改变云属性。为了更好地理解这些不确定性的各个来源，我们采用 Overcast [4], , 将其作为一个典型案例，展示了像气溶胶-云相互作用（ACI）这样的社群需求如何推动机器学习方法的发展。与 [5], 等先前研究相比，我们将气溶胶光学厚度（AOD）——气溶胶浓度的代理指标——视为连续处理，而非离散处理，并开展考虑不确定性的敏感性分析，以研究正性和无混杂性可能遭到违背所带来的后果。



(a) 气溶胶-云相互作用的简单因果图


(b) 东南太平洋

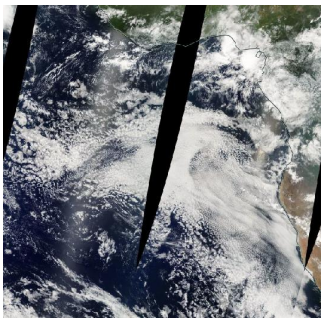

(c) 南大西洋

图 1：概述我们对因果图的认识，以及所分析两个区域的卫星影像。这两个区域因其独特的气溶胶-云相互作用以及可供借鉴的大量既有研究而被选中。

2 方法

参照 [4], , 我们采用潜在结果框架，估计连续处理 $\mathrm{T} \in \mathcal{T}$（气溶胶）对感兴趣结果 $\mathrm{Y} \in \mathcal{Y}$（云属性）的影响；该研究单位由协变量$\mathbf{X} \in \mathcal{X}$（环境信息）描述，如图 1(a) [6, 7, 8, 9]所示。我们将潜在结果记为 $\mathrm{Y_t}$，表示处理为 t 时可能观察到的结果。所考虑的协变量包括 900、850 和 700 毫巴处的相对湿度、海表温度、500 毫巴处的垂直运动、低层大气稳定度以及有效逆温强度。处理变量为气溶胶光学厚度（AOD），它是气溶胶浓度的代理指标。所考虑的结果是云滴大小（r_e）。为估计处理效应，我们研究条件平均潜在结果（CAPO）和平均潜在结果（APO）

$$\text{CAPO} = \mu(\mathbf{x}, \mathrm{t}) \coloneqq \mathbb{E}\left[\mathrm{Y_t} \mid \mathbf{X} = \mathbf{x}\right], \quad \text{and} \quad \text{APO} = \mu(\mathrm{t}) \coloneqq \mathbb{E}\left[\mu(\mathbf{X}, \mathrm{t})\right],$$

可利用观测分布 $P(\mathbf{X}, \mathrm{T}, \mathrm{Y})$ 识别，方法是

$$\tilde{\mu}(\mathbf{x}, \mathrm{t}) \coloneqq \mathbb{E}\left[\mathrm{Y} \mid \mathrm{T} = \mathrm{t}, \mathbf{X} = \mathbf{x}\right] \quad \text{and} \quad \tilde{\mu}(\mathrm{t}) \coloneqq \mathbb{E}\left[\tilde{\mu}(\mathbf{X}, \mathrm{t})\right],$$

并依赖进一步的假设（无混杂性、正性、无干扰性和一致性）。在此，我们研究处理效应估计对正性和无混杂性假设违背的稳健性（详情见附录 A）。我们计算与用户指定的假设放宽程度相对应的不确定性界限。例如，参数 Λ 由用户设定，用于描述假定的未测量混杂程度 [4, 10, 11]。有些混杂因素无法通过卫星直接测量，例如湿度会导致气溶胶膨胀并改变云属性；参数 Λ 可用于编码专家对这类混杂因素影响的判断。

我们使用卫星观测的日均值，1° x 1°，以统一东南太平洋和南大西洋的数据（图1(b)和1(c)）。在本文所示的所有实验中，我们将MODIS仪器测得的平均液滴半径（r_e）作为结果。我们采用MERRA-2的气溶胶光学厚度来近似气溶胶浓度。环境混杂因素包括900、850和700毫巴处的相对湿度、大气稳定度、海表温度以及500毫巴处的垂直运动，这些数据也均来自MERRA-2。有关数据和实现的更多细节，请参阅附录B和附录D。

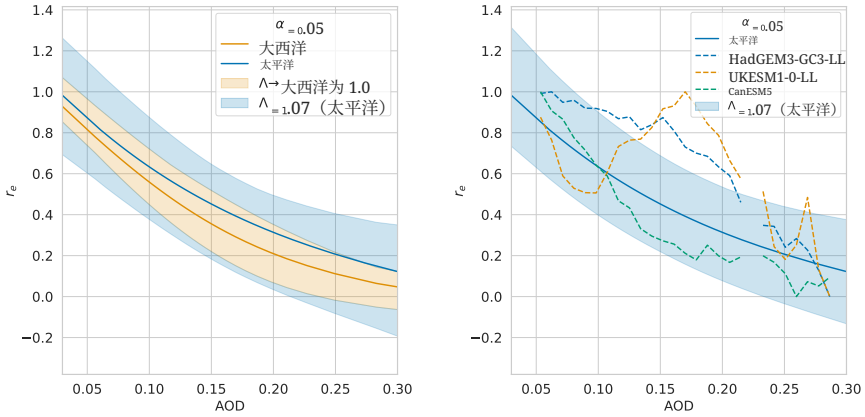
3 结果

3.1 运用领域知识推导合理的处理效应界限

与以往仅粗略估计可量化影响所导致的不确定性范围的研究不同，我们能够通过改变 Λ，推导出取决于混杂影响的置信区间。由于仅凭观测数据无法得知混杂效应的强度，我们提出一种通过对比两个地理区域来选择合理 Λ 的方法。我们对比东南太平洋和南大西洋，因为这两个区域的气溶胶-云相互作用（ACI）环境混杂因素不同，例如气溶胶类型、气溶胶吸湿性和气溶胶尺寸。这些都是重要的混杂因素，但遗憾的是，现有数据中并未包含它们。因此，我们为太平洋区域选择参数 Λ，使得在假设南大西洋区域不存在隐性混杂（$\Lambda \to 1$）的情况下，太平洋区域的处理效应界限能够涵盖南大西洋区域的效应界限，如图2(a)所示。将 Λ 设为1.07后，太平洋区域的界限便能合理反映未建模混杂因素可能造成的偏差。尽管由于其他混杂驱动因素，较大的 Λ 或许也合理，但领域知识告诉我们，这些是主要缺失的物理机制。

3.2 运用机器学习评估气候模型

现在，我们已经利用真实的观测结果得出了气溶胶-云相互作用的可能范围，因此可以通过查看气候模型的响应是否落在我们推导出的区间内，来判断它们对这一观测趋势的再现效果如何（图2(b)）。我们发现，加拿大模型 CanESM5 对气溶胶-云相互作用的模拟优于英国模型 HadGEM3-GC3-LL 和 UKESM1-0-LL。我们训练的机器学习模型不仅利用真实观测到的关系推导效应大小，还能考虑环境背景和混杂影响，从而得出真实、可量化的不确定性边界。因此，以 Overcast 发现的曲线作为真实响应，我们可以判断，那些落在通过对比不同区域得出的界限之外的模型，其模拟可能不符合物理规律，也极不可能在现实世界中出现。目前，CanESM5 估算气溶胶-云相互作用造成的总冷却效应，约可抵消温室气体导致的升温效应的一半；根据我们的结果，我们认为这一估算很可能最接近地球上观测到的真实值 [12]。



(a) 选择 Λ 使用两个地理区域　　(b) 与太平洋地区的地球系统模式进行比较

图2：气溶胶（气溶胶光学厚度）对平均液滴半径（r_e）影响的合理范围。

4 讨论与结论

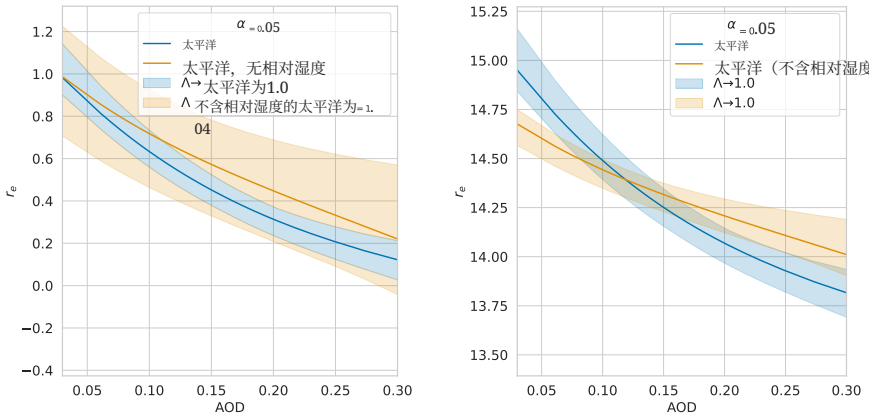
4.1 机器学习在气候模型验证中的作用

在这项研究中，我们展示了机器学习方法能够客观评估全球气候模型对气候过程的再现效果，例如气溶胶对平均液滴半径的影响。利用卫星观测的历史研究有一个缺点，那就是无法

量化周围环境可能如何影响气溶胶-云相互作用的强度。Overcast 会考虑此类情境混杂，并根据专家对隐藏混杂影响的判断，给出处理效应的范围。运用这种方法，我们得以了解气候模型是否再现了气溶胶光学厚度与 r_e 之间观测到的关系。目前，气候模型只能较好地再现那些可明确计算的大尺度过程；而气溶胶-云相互作用等发生在小于网格尺度上的过程，则仍未得到良好的参数化和近似。要改进气候模型，我们必须用更直观的方式了解其表现，例如将模型结果与观测结果进行比较。机器学习不仅提供了评估这些结果的方法，未来还可以对 Overcast 等模型学到的关系进行微调，以取代我们现有的参数化方案 [13]。

4.2 C跨领域协作与纯数据驱动方法的比较 驱动

尽管区域差异导致的不同混杂来源会改变结果，但也可以使用 Overcast 研究哪些环境因素是混杂的主要来源。我们进行了两项实验，分别纳入和不纳入 900、850 和 700毫巴处的相对湿度，以得到不同形态的结果，并将 Λ 拟合到两条剂量-反应曲线上。当 Λ 设为 1.04 时，两条曲线都落在不确定性边界内，因此我们可以观察到，由于气象不确定性而非区域不确定性，反应在这些边界内可能如何变化；而在后一种情况下，需要$\Lambda = 1.07$ （图3(a)）。缺乏真实值时，纯数据驱动技术无法判断应选择纳入还是不纳入相对湿度的模型；但结合领域知识后，我们知道，纳入湿度的曲线才是真实反应曲线（图3(b)）。对于气溶胶-云相互作用等气候过程，纯数据驱动方法可能并非最合适，因为选择正确的输入并验证结果离不开领域知识。将数据与理论相结合，汇聚机器学习和气候过程领域的专家，才能构建出最稳健的模型。



(a) 经缩放，采用适当的 Λ (b) 未缩放

图 3：从协变量中省略相对湿度时，效应的合理范围。

4.3 局限性与未来工作

本研究使用气溶胶光学厚度作为气溶胶浓度的代理指标，这可能会使处理效应估计产生偏差。例如，已知处理[14]中的测量误差可能导致偏差。此外，我们依赖低分辨率数据，无法完美捕捉微物理过程。未来的研究可考虑对基础因果模型采用不同假设，并尝试纳入气溶胶的其他属性，例如粒径、吸湿性和类型。

5 致谢

本项目获得欧盟地平线2020研究与创新计划资助，资助协议编号为821205（FORCeS）；同时获得欧洲研究理事会（ERC）项目“气溶胶对降水影响的约束”（RECAP）资助，该项目隶属于欧盟地平线2020研究与创新计划，资助协议编号为724602；此外还获得图灵研究所博士后拓展奖。

参考文献

[1] Sean A Twomey、M Piepgrass 和 TL Wolfe。《污染对全球云反照率影响的评估》。载于：*Tellus B*36.5（1984），第356–366页。[2] Olivier Boucher、D Randall、P Artaxo、C Bretherton、G Feingold、P Forster、V-M Kerminen、Y Kondo、H Liao、U Loh mann、P Rasch、S.K Satheesh、B Stevens 和 X.Y Zhang。《云与气溶胶》。载于：《*2013*年气候变化：自然科学基础》政府间气候变化专门委员会第五次评估报告第一工作组的贡献。T.F. Stocker、D. Qin、G.-K. Plattner、M. Tignor、S.K. Allen、J. Boschung、A. Nauels、Y. Xia、Bex V. 和 P.M. Midgley 编。英国剑桥和美国纽约：剑桥大学出版社，2013年，第571–658页。[3] 政府间气候变化专门委员会。《决策者摘要》。载于：《*2021*年气候变化：自然科学基础》政府间气候变化专门委员会第六次评估报告第一工作组的贡献。V. Masson-Delmotte、P. Zhai、A. Pirani、S.L. Connors、C. Péan、S. Berger、N. Caud、Y. Chen、L. Goldfarb、M.I. Gomis、M. Huang、K. Leitzell、E. Lonnoy、J.B.R. Matthews、T.K. Maycock、T. Waterfield、O. Yelekçi、R. Yu 和 B. Zhou 编。英国剑桥和美国纽约：剑桥大学出版社，2021年，第3–32页。[4] Andrew Jesson、Alyson Douglas、Peter Manshausen、Maëlys Solal、Nicolai Meinshausen、Philip Stier、Yarin Gal 和 Uri Shalit。连续值干预因果效应估计的可扩展敏感性与不确定性分析。2022年。[5] Andrew Jesson、Peter Manshausen、Alyson Douglas、Duncan Watson-Parris、Yarin Gal 和 Philip Stier。《运用非线性因果模型研究东南太平洋的气溶胶-云相互作用》。载于：*2021*年神经信息处理系统大会“用机器学习应对气候变化”研讨会。2021年。[6] Donald B. Rubin。《随机化与非随机化研究中处理因果效应的估计》。载于：《教育心理学杂志》66.5（1974），第688–701页。[7] Donald B. Rubin。《实验数据的随机化分析：关于费舍尔随机化检验的评论》。载于：《美国统计学会杂志》75.371（1980年9月），第591页。[8] Jasjeet Sekhon。因果推断与匹配方法估计的内曼-鲁宾模型。Janet M. Box-Steffensmeier、Henry E. Brady 和 David Collier 编。第1卷。牛津大学出版社，2009年9月。[9] Jerzy Splawa-Neyman、D. M. Dabrowska 和 T. P. Speed。《概率论在农业实验中的应用：原理探讨，第9节》。载于：《统计科学》5.4（1990年11月）。[10] Andrew Jesson、Sören Mindermann、Yarin Gal 和 Uri Shalit。《量化隐性混杂下个体层面因果效应估计中的无知》。载于：第*38*届国际机器学习会议论文集139（2021），第4829–4838页。[11] Andrew Jesson、Sören Mindermann、Uri Shalit 和 Yarin Gal。《利用不确定性感知模型识别因果效应推断失效》。载于：《神经信息处理系统进展》（2020）。[12] Christopher JSmith、RyanJ Kramer、Gunnar Myhre、KariAlterskjær、William Collins、Adriana Sima、Olivier Boucher、Jean-Louis Dufresne、Pierre Nabat、Martine Michou 等。《CMIP6模式中的有效辐射强迫及其调整》。载于：《大气化学与物理》20.16（2020），第9591–9618页。[13] Andrew Gettelman、David John Gagne、C-C Chen、MW Christensen、ZJ Lebo、Hugh Morrison 和 Gabrielle Gantos。《运用机器学习研究暖雨过程》。载于：《地球系统建模进展》13.2（2021），e2020MS002268。

[14] Yuchen Zhu、Limor Gultchin、Arthur Gretton、Matt Kusner 和 Ricardo Silva。处理测量误差下的因果推断：一种非参数工具变量方法。2022年。[15] Bryan A. Baum 和 Steven Platnick。《MODIS云产品简介》。载于：地球科学卫星遥感。John J. Qu、Wei Gao、Menas Kafatos、Robert E. Murphy 和 Vincent V. Salomonson 编。柏林、海德堡：Springer Berlin Heidelberg，2006年，第74–91页。[16] Ronald Gelaro 等。《现代时期研究与应用回顾性分析第二版（MERRA-2）》。载于：《气候学杂志》30.14（2017年7月），第5419–5454页。[17] Olivier P. Prat、Brian R. Nelson、Elsa Nickl 和 Ronald D. Leeper。《NOAA气候数据记录计划网格化卫星降水产品的全球评估》。载于：《水文气象学杂志》22（2021年9月），第2291–2310页。[18] James L. Cogan 和 James H. Willand。《利用NOAA 2卫星测量海表温度》。载于：《应用气象学杂志》15（1976年2月），第173–180页。[19] Michael Bosilovich。《MERRA-2：气候的初步评估》。载于：43（），第145页。[20] Robert Wood 和 Christopher S. Bretherton。《层状低云覆盖与对流层低层稳定度之间的关系》。载于：《气候学杂志》19.24（2006年12月），第6425–6432页。[21] Ilan Koren、Guy Dagan 和 Orit Altaratz。《从气溶胶受限到暖对流云增强》。载于：《科学》344.6188（2014年6月），第1143–1146页。[22] Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、GregoryChanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga、Alban Desmaison、Andreas Kopf、Edward Yang、Zachary DeVito、Martin Raison、Alykhan Tejani、Sasank Chilamkurthy、Benoit Steiner、Lu Fang、Junjie Bai 和 Soumith Chintala。《PyTorch：一种命令式风格的高性能深度学习库》。载于：《神经信息处理系统进展》第*32*卷。H. Wallach、H. Larochelle、A. Beygelzimer、F. d’Alché-Buc、E. Fox 和 R. Garnett 编。第32卷。Curran Associates, Inc.，2019年，第8024–8035页。[23] F. Pedregosa、G. Varoquaux、A. Gramfort、V. Michel、B. Thirion、O. Grisel、M. Blondel、P. Prettenhofer、R. Weiss、V. Dubourg、J. Vanderplas、A. Passos、D. Cournapeau、M. Brucher、M. Perrot 和 E. Duchesnay。《Scikit-learn：Python中的机器学习》。载于：《机器学习研究杂志》12（2011年），第2825–2830页。[24] Philipp Moritz、Robert Nishihara、Stephanie Wang、Alexey Tumanov、Richard Liaw、Eric Liang、Melih Elibol、Zongheng Yang、William Paul、Michael I. Jordan 和 Ion Stoica。*Ray*：面向新兴人工智能应用的分布式框架。2018年9月。[25] Richard Liaw、Eric Liang、Robert Nishihara、Philipp Moritz、Joseph E. Gonzalez 和 Ion Stoica。*Tune*：用于分布式模型选择与训练的研究平台。2018年7月。[26] Stefan Falkner、Aaron Klein 和 Frank Hutter。《结合Hyperband与贝叶斯优化》。载于：（），第6页。[27] Matthew W. Christensen 等。《用于约束气溶胶有效辐射强迫的机会性实验》。载于：《大气化学与物理》22.1（2022年1月），第641–674页。

A 理论背景：无混杂假设与正值性假设 s

混杂变量是同时影响处理 T 和结果 Y 的因素。无混杂假设认为，所有混杂变量都已观测到，并通过 $\mathbf{X}$ 加以控制，因此处理组之间具有可比性，即 $\mathrm{Y}_{\mathrm{T}} \perp\!\!\!\perp \mathrm{T} \mid \mathbf{X}$。

正值性假设认为，对于具有不同协变量的数据，各子组接受任意处理剂量的概率均大于零，即 $p(\mathrm{t} \mid \mathbf{x}) > 0$ 对任意 $\mathrm{t} \in \mathcal{T}$ 以及任意 $\mathbf{x} \in \mathcal{X}$ 满足 $p(\mathbf{x}) > 0$ 的情况均成立。

实践中，由于维数灾难，正值性与无混杂性之间存在权衡：当 $\mathbf{X}$ 维度较高且处理为连续变量时，我们不太可能观测到每个 $\mathbf{x} \in \mathcal{X}$ 对应的所有处理水平。

B 数据与预处理

我们使用从卫星观测再分析数据中获取的数据。搭载于 Terra 和 Aqua 卫星上的中分辨率成像光谱仪（MODIS）仪器以约 1 km $\times$ 1 km 的分辨率观测地球 [15]。这些观测数据被输入现代回顾性研究与应用分析第2版（MERRA-2）实时模型，以模拟大气及其组成部分，例如气溶胶 [16]。MERRA-2 计算全球温度、相对湿度和气压的垂直廓线，并同化高光谱和被动微波卫星观测，以提升其模拟地球大气的能力。本文研究的数据是 Aqua 和 Terra 卫星的 MODIS 观测数据，与环境的 MERRA 再分析数据进行了时空匹配。我们使用两个不同的数据集，分别是 $1^\circ \times 1^\circ$ 2004年到2019年南大西洋和东南太平洋观测数据的日均值。数据来源见表 1。

表 1：卫星观测数据来源

产品名称	描述
平均液滴半径（r_e）	MODIS（1.6、2.1、3.7 μm 通道） [15]
降水	NOAA CMORPH 气候数据记录 [17]
海表温度（SST）	NOAA WHOI 气候数据记录 [18]
低层大气稳定度（LTS）	MERRA-2 [16]
500毫巴处垂直运动（ω500）	MERRA-2 [19]
估算逆温强度（EIS）	MERRA-2 [16, 20]
x毫巴处相对湿度（RHx）	MERRA-2 [16]
气溶胶光学厚度（AOD）	MERRA-2 [16]

我们通过应用一些筛选条件 [21]，将观测范围限制在处于“气溶胶受限”状态的云。在“气溶胶受限”状态下，我们假设云的发展受云凝结核（因而也受气溶胶）供应量的限制。筛选条件的选择参考了领域知识。云液态水路径（CWP）仅保留低于 250μm 的值，r_e 仅保留低于 30μm 的值。气溶胶光学厚度（AOD）仅保留介于 0.03 和 0.3 之间的值。我们还筛除了降水云，以避免因果图中出现环路。最后，将所有特征归一化后再输入模型。

C 模型架构

这些模型是由两个组件构成的神经网络架构：特征提取器 $\phi(\mathrm{x}; \theta)$ 和密度估计器 $f(\phi, \mathrm{t}; \theta)$，其表示形式见附录 C。协变量 x 作为输入提供给特征提取器，其输出与处理 t 拼接后作为输入提供给密度估计器；密度估计器输出高斯混合密度 $p(\mathrm{y} \mid \mathrm{t}, \mathbf{x}\ \theta)$，我们可以从中进行采样。特征提取器利用注意力机制，根据观测值的地理坐标，对给定日期协变量之间的时空相关性进行建模。模型架构如图 4 所示。

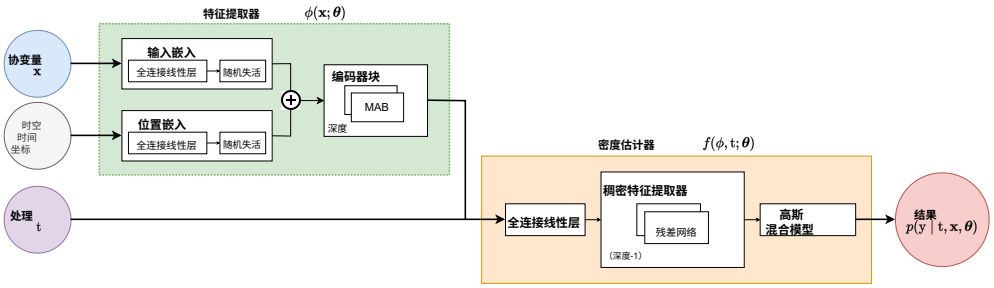


图 4： **Overcast 模型架构**。 输入以圆圈表示，蓝色 表示协变量，灰色 表示时空坐标，紫色 表示处理。在红色 圆圈中是模型的输出，即结果分布。模型包含一个绿色 特征提取器，以及一个密度估计器（以橙色表示）。

D 实现细节

我们遵循 [4] 中的实现。代码使用 Python 编写。所用的软件包包括 PyTorch [22], Scikit-learn [23], Ray [24],NumPy、SciPy 和 Matplotlib。

我们使用 Ray Tune [25]，通过 HyperBand贝叶斯优化 [26] 搜索算法优化网络超参数。调优过程中考虑的超参数见 [4]。各数据集最终采用的超参数见表2。超参数优化的目标是：对于随机种子为 1331 的单个数据集实现，计算验证数据中所有结果的批次级皮尔逊相关系数，并取其平均值。

我们按不同日期将数据划分为训练集、验证集和测试集。 [4] 按以下方式划分数据：周一至周五的数据点属于训练集，周六的数据点属于验证集，周日的数据点属于测试集。在我们的实现中，我们保持数据集之间相同的比例，但采用随机方式划分，使用随机种子 42，并将 $5/7$ 的数据用于训练集、 $1/7$ 用于验证集、 $1/7$ 用于测试集。采用随机划分的原因是气溶胶光学厚度 [27] 存在明显的每周周期。模型通过最大化 $p(\mathrm{y}\mid\mathrm{t},\ \mathbf{x}\ \boldsymbol{\theta})$ 的对数似然进行优化。

表2：每个数据集和模型的最终超参数

超参数	东南太平洋	南大西洋
隐藏单元	128	128
网络深度	3	4
GMM T分量	27	7
GMM Y分量	22	24
注意力头	8	8
负斜率	0.28	0.19
丢弃率	0.42	0.16
层归一化	否	True
批大小	128	160
学习率	0.0001	0.0001
训练轮数	500	500