arXiv:2301.03360v1 [stat.ML] 9 Jan 2023

风力涡轮机上的上行闪电：基于大尺度气象的风险评估

Isabell Stucke[1,2], Achim Zeileis[1], Georg J. Mayr[2],Thorsten Simon[1],Gerhard Diendorfer[3],Wolfgang Schulz[3], Hannes Pichler[3], 和 Deborah Morgenstern[1,2]
[1]因斯布鲁克大学统计系，奥地利[2]因斯布鲁克大学大气与冰冻圈科学系，奥地利[3]OVE Service GmbH，ALDIS（奥地利闪电探测与信息系统）部门，奥地利维也纳

通讯作者： Isabell Stucke (isabell.stucke@uibk.ac.at)

摘要。随着越来越多的风力涡轮机投入使用以生产可再生电力，上行闪电已成为风力涡轮机面临的日益重要的威胁。风力涡轮机越高，人造结构遭受上行闪电袭击的风险就越大。由于上行闪电的初始连续电流持续时间较长，会在闪电放电过程中产生大量电荷转移，因此其破坏性可能远大于下行闪电。目前，风力涡轮机闪电风险评估标准主要考虑夏季闪电活动，而这类活动是根据闪电定位系统（LLS）推断得出的。真实雷电流测量表明，闪电定位系统（LLS）可能只能探测到不足50%的上行闪电。这会导致严重低估风力涡轮机上无法被LLS探测到的上行闪电比例，而这种闪电类型在寒季占主导地位。本研究利用奥地利盖斯贝格塔和瑞士森蒂斯塔的直接上行闪电测量数据，旨在评估风力涡轮机上可被LLS探测到和无法被LLS探测到的上行闪电风险。我们将直接观测到的上行闪电与气象再分析数据关联，并采用强大的机器学习技术——随机森林进行分析。通过在塔架观测点训练的随机森林模型，确定了分别与可被LLS探测到和无法被LLS探测到的上行闪电发生与否相关的气象驱动因素；这些模型在独立数据上表现出很强的预测能力。第二步，我们将塔架训练模型的结果扩展到更大的研究区域（德国中部和北部）。我们使用该区域内风力涡轮机位置的数据，对用于诊断可被LLS探测到的闪电的塔架训练模型进行独立验证，发现该模型能够可靠地诊断此类上行闪电。基于案例研究事件绘制的风险图显示，研究区域内诊断概率较高的地点与实际发生上行闪电的地点相吻合。这支持将该模型推广至所有类型的上行闪电；推广后，风险和受影响区域均有所增加。

1 引言

可再生能源生产日益重要，近来促使风力涡轮机的数量显著增加（例如，Pineda 等人，2018）。由于这些结构通常高于 100 米，从高耸结构向云传播的上行闪电（UL）更容易形成（Berger，1967）。高耸结构更容易遭受

上行闪电（UL），因为与地面相比，它暴露于更强的电场中。低于 100 米的结构主要遭受下行闪电（DL），其先导从云向地表传播（例如，Rakov 和 Uman，2003）。

随着风力涡轮机不断增高，上行闪电（UL）已成为导致其遭受严重损坏的主要气象因素（例如，Rachidi 等，2008；Montanyà Puig 等，2016；Pineda 等人，2018；Matsui 等，2020；Zhang 和 Zhang，2020）。由于其初始连续电流（ICC）的持续时间约为下行闪电（DL）电流的十倍，因此破坏性可能大得多。对盖斯贝格山顶（奥地利萨尔茨堡）一座专门配备仪器的塔进行真实雷电流测量后发现，超过 50% 的上行闪电未被常规闪电定位系统（LLS）探测到。原因是，LLS 无法探测一种仅有初始连续电流（ICC）的上行闪电亚型（Diendorfer 等，2015；March 等，2016）。尽管瑞士的森蒂斯塔等塔能够提供可被闪电定位系统探测到的上行闪电（UL-LLS）的真实雷电流数据，但盖斯贝格塔是欧洲唯一一座能够全面提供 UL-LLS 和 LLS 无法探测到的上行闪电（UL-noLLS）发生情况信息的仪器化塔。

风力涡轮机的雷电防护标准（例如 IEC 61400-24, 2019）严重低估了风力涡轮机上 UL 的发生率，因为目前这些标准仅依据三个因素：风力涡轮机的高度、根据闪电定位系统得出的当地年闪电密度，以及涉及地形复杂度或海拔等因素的环境项（Rachidi 等人，2008；Pineda 等人，2018）。夏季的闪电活动显然主导当地年闪电密度，这是因为深对流会造成大量下行闪电。然而，预计 UL 将成为风力涡轮机上的主要闪电类型，而且在寒冷季节其重要性可能更为突出（Diendorfer，2020；Rachidi 等人，2008）。此外，风险评估标准无法计入 UL-noLLS，只能在存在高耸结构的情况下计入 UL-LLS。

本研究的主要目标是评估更大范围内风力涡轮机上 UL-LLS 和 UL-noLLS 的风险。只有极少数地点能够依据直接测量分析 UL 的实际发生情况。尽管已有雷电定位网络，或许能够用于分析高耸结构上的 UL-LLS，但闪电电流测量表明，仍有相当一部分未被探测到。鉴于传统 LLS 无法评估风力涡轮机上 UL 的全部风险，本研究采用了一种新方法。

该研究采用机器学习技术，将上行闪电的发生与大尺度气象环境联系起来。上行闪电的发生情况只能通过地面真实的闪电电流测量数据来确定。这些数据是建立和训练统计模型的基础，最终用于评估整个研究区域内上行闪电的风险。具体而言，本研究采用条件推断随机森林（Hothorn and Zeileis, 2015），以考察上行闪电在高耸结构上的发生率与大气之间高度非线性且复杂的相互作用。实现主要目标需要经过若干步骤。

基于奥地利（盖斯贝格塔）和瑞士（森蒂斯塔）两座仪器化塔的闪电电流测量数据，构建了两个模型：一个用于 UL-LLS，另一个用于 UL-LLS +UL-noLLS。首先，这些模型将判断大尺度气象变量与上行闪电的发生之间是否存在关联；其次，它们将展示大尺度气象在多大程度上可作为诊断工具，用于推断上行闪电的发生情况。

UL-LLS 数据的可用性有助于验证仪器化塔得出的结果是否具有可迁移性。我们的思路是提取研究区域内的风力涡轮机位置，并利用闪电定位系统数据识别寒冷季节（10月至次年4月）期间所有击中这些风力涡轮机的雷击。若能结合向上闪电地面真值电流测量，从大尺度气象中可靠地诊断 UL-LLS，则

最终评估无法利用闪电定位系统数据验证的 UL-noLLS 风险时，结果的可靠性会更高。

以下各节安排如下。第 2 节介绍为本研究提供必要地面实况数据的两座仪器化塔。第一座是盖斯贝格塔，提供 UL-LLS 和 UL-noLLS 数据；第二座是森蒂斯塔，仅提供 UL-LLS 数据。此外，本节还介绍研究区域内风力涡轮机上闪电的识别方法，以及所用的气象数据。

第 3 节概述两座仪器化塔的研究流程和主要发现。第 3.1 节介绍随机森林模型构建的基本原理。第 3.2 节展示模型在仪器化塔上的表现。此外，本节还介绍导致 UL 风险升高的最重要大尺度气象变量（第 3.3 节）。

随后，第 4 节将模型从仪器化塔推广至更大的研究区域，以找出 UL 风险较高的地区。第 4.1 节诊断风力涡轮机上的 UL-LLS，并介绍案例研究。随后，第 4.2 节利用整个考察时段，展示并讨论风力涡轮机上 UL-LLS 和 UL-LLS + UL-noLLS 的风险。

第5节对最重要的研究结果进行了总结。

2 数据

本研究结合了五种不同的数据源：在奥地利盖斯贝格塔（Diendorfer 等，2009）和瑞士森蒂斯塔（Romero 等，2012）直接测量的上行闪电数据；由欧洲闪电探测合作组织（EUCLID，Schulz 等，2016）远程测量的闪电定位系统数据；来自再分析数据库 ERA5（Hersbach 等，2020）的大尺度气象变量；以及使用 OpenStreetMap 数据库识别的风力涡轮机位置。

2.1 仪器化塔的向上闪电直接测量

图1展示了用于直接测量上行闪电电流的极少数仪器化塔中的两座。这两座塔分别是盖斯贝格塔（海拔1288米，$47°48′\mathrm{N}$，$13°60′\mathrm{E}$）和森蒂斯塔（海拔2502米，$47°14′\mathrm{N}$，$9°20′\mathrm{E}$）。仪器化塔处记录的闪电几乎全部为上行闪电。盖斯贝格塔在2000年至2015年期间共记录了819次上行闪电事件。森蒂斯塔在2010年至2017年期间记录了692次上行闪电事件。

盖斯贝格的高灵敏度分流器式传感器能够测量所有类型的向上闪电，不受电流波形影响，即 UL-LLS 和 UL-noLLS。然而，森蒂斯山使用的感应式传感器无法测量仅含初始连续电流（ICC）的向上闪电（约占 50%，Diendorfer 等，2015）。

直接测得的 UL 电流是构建随机森林模型的关键前提。这些模型在塔站数据上训练后，将扩展至更大的研究区域。结合两座塔站的数据，可构建两类模型，分别用于诊断 UL-LLS 和 UL-LLS + UL-noLLS。

2.2 风力涡轮机和研究区域中的 UL-LLS

LLS EUCLID 遥感探测到的闪电数据，以及从 OpenStreetMap 获取的风力涡轮机位置，可用于验证针对选定研究区域评估 UL-LLS 风险的统计模型。

截至2020年底，研究区域（北纬50°–54°、东经6°–16°）内已安装27 814台风力涡轮机（图1）。提取这些风力涡轮机的精确位置后，将 EUCLID 在0.003°圆形区域内（约为半径300米）探测到的雷击识别为上行闪电（UL）。EUCLID 对下行闪电（DL）的闪电探测效率超过90%（Schulz 等，2016）。如前所述，EUCLID 对 UL 的探测效率可能较低（< 50 % Diendorfer 等，2015）。

鉴于上行闪电具有破坏性，且在现行雷电防护标准中被严重低估，因此在研究区域内评估风力涡轮机遭受上行闪电的风险时，必须明确将其纳入考量。塔训练模型基于全年上行闪电数据构建。然而，与下行闪电相比，上行闪电在寒冷季节占主导地位，因此本研究仅在验证部分考虑2018年10月至2020年12月期间每年10月至次年4月的数据。此外，由于下行闪电在较暖季节占主导地位，若考虑全年数据，提取风力涡轮机遭受的雷击可能会导致下行闪电与上行闪电难以区分。

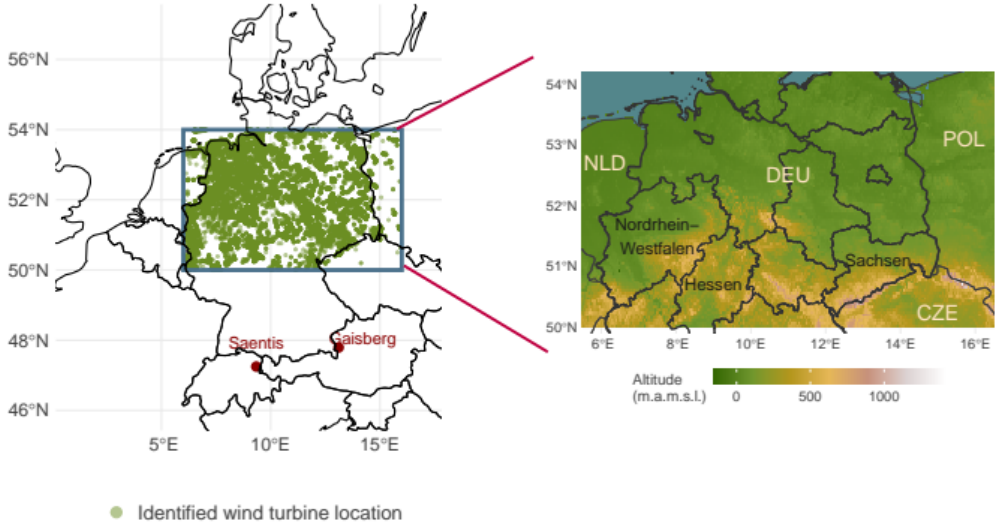


图1. 仪器化塔（盖斯贝格和森蒂斯山）位置及研究区域（方框）的地理概览。绿色圆点表示根据 © OpenStreetMap 2020 手动识别的风力涡轮机位置。右侧：研究区域地形图，显示平均海平面以上高度。数据取自航天飞机雷达地形测绘任务（Farr and Kobrick, 2000）。

2.3 气象数据

ERA5提供大气状态的逐小时再分析数据。其水平分辨率为31 km（网格单元大小为0.25 ° x 0.25 °），垂直方向有137个层次。本研究使用了35个可直接获取及推导的变量，涵盖地表、模式层以及垂直积分数据。这些变量反映了与云电化、闪电和雷暴相关的特征（Morgenstern等人，2022年）。变量完整列表见附录 A。数据经过空间和时间双线性插值，分别对应每次盖斯贝格和森蒂斯塔上上升闪电观测，以及本研究验证部分研究区域内的每个网格单元。

3 仪器化塔的方法与研究结果

本节介绍基本方法所需的背景信息，以及对仪器化盖斯贝格塔和森蒂斯塔进行分析所得的重要结果。以下将涵盖三个不同方面：首先介绍基本模型（即随机森林）的构建原理；其次介绍模型的性能；第三，说明哪些变量对于识别是否存在有利于上行闪电发生的条件最为重要。

3.1 塔训练随机森林的构建与验证

采用一种近年来已广泛应用于各个科学领域的机器学习技术，将大尺度气象与仪器化塔处上行闪电的发生联系起来。随机森林具有高度灵活性，能够处理非线性效应，并捕捉与所述建模问题相关的复杂交互作用（Strobl 等，2009）。

上行闪电的发生与不发生属于二元分类问题，采用35个大尺度气象变量（预测变量）进行处理。每个气象预测变量都通过随机森林与盖斯贝格塔或森蒂斯塔处有无上行闪电的情形建立联系。随机森林结合了多个决策树的预测结果，这些决策树基于从输入数据中随机选取的子样本进行训练。

具体而言，随机森林中的树通过置换检验（也称为条件推断，参见 Strasser 和 Weber（1999）），捕捉二元响应变量与各预测变量之间的关联来构建。在递归树构建的每一步中，选择与响应变量关联程度最高（最显著）的预测变量。随后，按照该预测变量对数据集进行划分，以尽可能区分不同的响应类别。对每个数据子集递归重复划分，直至满足某项停止准则（例如，关于显著性或子样本量的准则）。随机森林由 500 棵这样的树组成，每棵树都基于从完整数据集中随机抽取的三分之二数据进行训练，并且每次划分仅考虑随机选取的六个预测变量。最后，随机森林对树集成的预测结果取平均，从而使预测性能更加稳定并得到提升。有关算法和实现的更多详情，请参见 Hothorn 等（2006）以及 Hothorn 和 Zeileis（2015）。

为验证所得模型，将输入数据划分为训练数据样本和测试数据样本。模型在训练数据上进行训练，并在未见过的测试数据上评估其诊断能力。采用留一交叉验证来验证 UL-LLS 和 UL-LLS + UL-noLLS 的模型。第一个 UL-LLS 模型同时使用森蒂斯山和盖斯贝格的数据，以扩大训练数据规模。在训练过程中，将盖斯贝格数据中森蒂斯塔无法探测到的特定闪电类型排除，以确保一致性。第二个 UL-LLS +UL-noLLS 模型仅使用盖斯贝格数据，因为只有盖斯贝格塔提供了所有上行闪电亚型的完整信息。

2000年至2015年间，盖斯贝格塔记录到247个发生上行闪电的独立日期。2010年至2017年间，森蒂斯塔记录到186个独立日期。合并两座塔的上行闪电日期后，共有406个发生上行闪电的独立日期。每次训练时，都会留出247（406）个发生上行闪电的日期中的一天作为测试数据。重复此过程，直到247（406）个日期中的每一天都曾被留出一次，用于训练随机森林模型。最终得到247（406）个不同的模型，这些模型均在数量相同的有上行闪电和无上行闪电的情形上训练而成。

对输入的模型响应（即是否发生上行闪电）进行抽样，使两个类别保持平衡，也就是说，有上行闪电和无上行闪电的情形所占比例相同。在评估模型性能时，模型会针对训练过程中未纳入的数据（即各自被留出的那一天）诊断条件概率。由于模型响应的设置保持了平衡，我们将该概率称为条件概率。为了同样诊断无上行闪电日期的上行闪电条件概率，我们还会在2000年至2017年间随机抽取每个季节中没有上行闪电的日期。诊断能力较强，意味着在特定情形下盖斯贝格或森蒂斯山发生上行闪电时，模型给出的概率较高（即真阳性率较高）；而在未发生上行闪电时，模型给出的概率较低（即假阳性率较低）。

3.2 塔基训练的随机森林的性能

塔基训练的随机森林模型能够可靠地诊断 UL-LLS 和 UL-LLS + UL-noLLS，在使用来自塔站的未见留出数据进行验证时尤其如此。图2总结了随机森林对 UL-LLS + UL-noLLS（盖斯贝格）和 UL-LLS（盖斯贝格 +森蒂斯山）的交叉验证诊断能力。两个模型集成的诊断性能相近，且均表现良好。在相应情形（分钟）中观测到 UL 时，诊断所得的条件概率中位数约为 0.8。这表明真阳性率很高。同样，在没有闪电的情形下（右图），条件概率较低，表明假阳性率较低。

包含 UL-noLLS 的随机森林具有最高诊断能力，这表明传统 LLS 无法探测到的那部分闪电，确实可以仅凭大尺度气象可靠地诊断出来。这支持了这样一种观点：还应研究无法验证的 UL-noLLS 风险，而不应只关注 UL-LLS。

3.3 仪器化塔上 UL-LLS 的气象驱动因素

随机森林可用于评估单个变量对模型诊断性能的影响。具体做法是计算所谓的置换变量重要性。其原理是在评估模型诊断性能时，忽略某个预测变量所包含的信息，从而打破响应变量与该预测变量之间的关系。忽略某个预测变量的信息，是通过置换实现的，即随机打乱其取值，然后评估

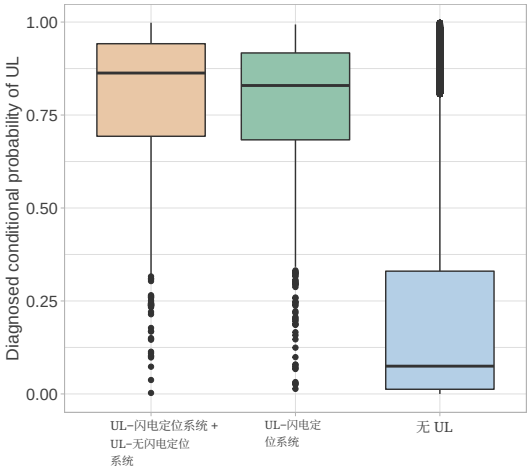


图2。 有无 UL事件情况下诊断条件概率的分布。左图：根据盖斯贝格数据（包括 UL的所有亚型），在特定分钟内观测到 UL时的条件 UL概率（真阳性）。中图：根据盖斯贝格和森蒂斯山数据的合并结果，在特定分钟内观测到 UL时的条件 UL概率。右图：在随机抽取的无 UL事件日期中，条件 UL概率（假阳性）。

诊断性能下降的程度。图3展示了根据100个不同随机森林模型计算得出的置换变量重要性的中位数，这些模型用于UL-LLS。每个模型都基于UL情形与随机选取的无UL情形之间的均衡比例。UL-LLS和UL-LLS + UL-noLLS模型的结果非常相似。

根据基于盖斯贝格和森蒂斯塔直接观测结果建立的随机森林， 对流降水对 UL发生的影响最大（图3）。忽略这一驱动变量的信息会使诊断性能下降最为明显。第二和第三重要的变量分别是最大上升气流速度和对流有效位能（CAPE）。对这三个最重要变量进行统计汇总表明，在森蒂斯塔和盖斯贝格塔，UL发生时的 CAPE都相对较低（中位数为 68 J kg^{-1}）。对流降水的中位数为 3.8 毫米，最大垂直上升气流速度的中位数为 -1.5 m s^{-1}。在所研究的时间范围内，与逐小时观测的“平均”情况相比，所有数值的绝对值都更大。然而，与深对流情况相比，其量级并不特别高；在深对流中，CAPE值通常尤其高于 500 J kg^{-1}。一个重要原因可能是，在仪器化塔处，UL全年大致均匀分布，而伴随深对流和高 CAPE值的强雷暴主要发生在夏季。因此，这也可能表明，UL的发生需要许多不同过程相互作用，以形成有利于 UL的条件；与形成有利于深对流的条件相比，这一过程可能更加复杂。

其他重要变量还包括最大降水率、雷暴云的垂直尺度、冰晶和固态水凝物的数量，以及2米露点温度。

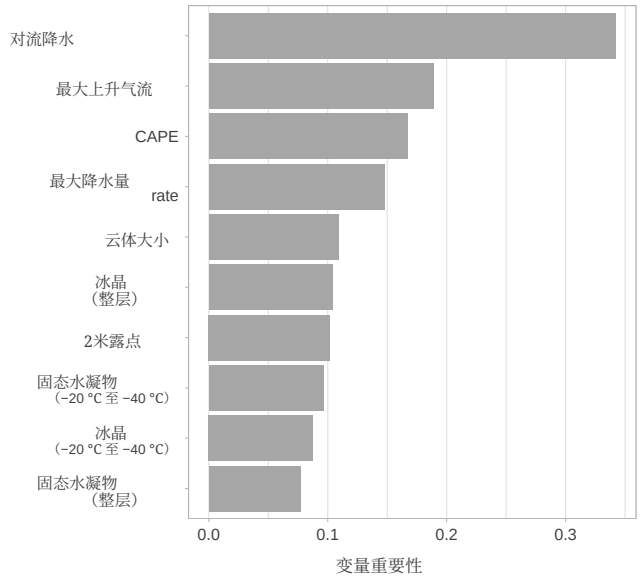


图3. 基于100个不同随机森林的置换变量重要性的中位数；这些随机森林依据有无上行闪电的情形比例均衡的数据构建，研究对象为盖斯贝格和森蒂斯塔。

4 风力涡轮机上的上行闪电

提取寒冷季节（10月至次年4月）的风力涡轮机位置，并识别距其300米以内的雷击，结果显示，研究区域内有些地区比其他地区更频繁地发生上行闪电（见图4）。有趣的是，较常受上行闪电影响的地区（面板（b），深粉色）与风力涡轮机较多的地区相吻合。不过，总体而言，风力涡轮机数量较多的地区（面板（a），深绿色）不一定与上行闪电数量较多的地区重合，例如研究区域的东北部地区。

下一节将介绍并讨论把仪器化塔的研究结果扩展至研究区域后的结果；在该区域中，提取了风力涡轮机的位置，并分析了这些位置的闪电活动。

4.1 根据大尺度气象条件诊断风力涡轮机处的 UL-LLS

基于两座仪器化塔数据建立的 UL-LLS 和 UL-LLS + UL-noLLS 随机森林模型，确定了在区分有上行闪电和无上行闪电的情形时最重要的大尺度气象变量。

现在，将塔基训练的随机森林模型应用于更大的研究区域，以评估风力涡轮机发生上行闪电的风险。随后将利用闪电定位系统数据中的闪电测量结果，对已确定的风力涡轮机位置进行验证。

以下结果基于与第 3.2 节所述类似的流程，但使用研究区域中的每个网格单元（31 km x 31 km）作为测试数据，而不是使用来自仪器化塔的交叉验证数据。

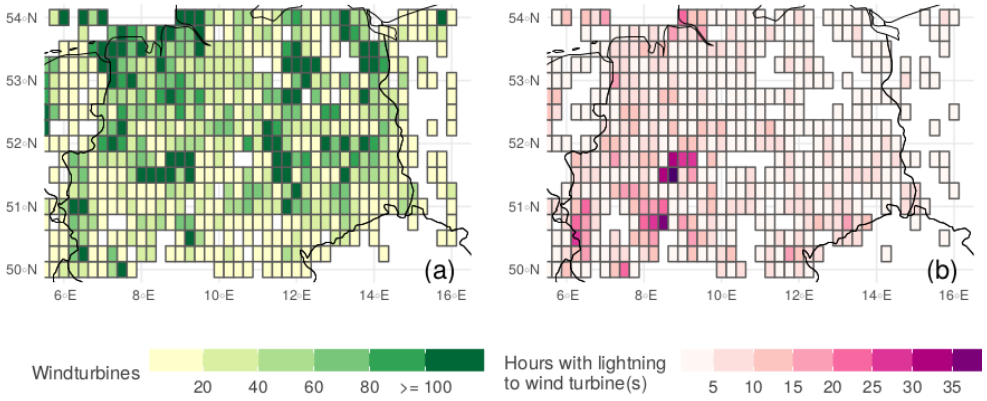


图4. 面板 (a)：根据 © OpenStreetMap 2020 年数据得出的每个网格单元内风力涡轮机的数量。面板 (b)：根据 EUCLID 数据得出的每个网格单元内风力涡轮机发生闪电的小时数。

在下文中，将塔基训练的随机森林模型应用于研究区域的每个网格单元。为提高结果的稳健性，再次使用基于盖斯贝格和森蒂斯塔观测数据建立的 100 个不同随机森林模型，诊断研究区域内所选案例研究中上行闪电的条件概率。本节中的结果展示了随机森林模型诊断出的条件概率中位数。

案例研究：风力涡轮机处的 UL-LLS

为展示塔架训练的随机森林在发生上行闪电事件的日期对 UL-LLS 的诊断能力，我们从 2018 至 2020 年研究区域的寒冷季节中选取了三个不同的案例日期。

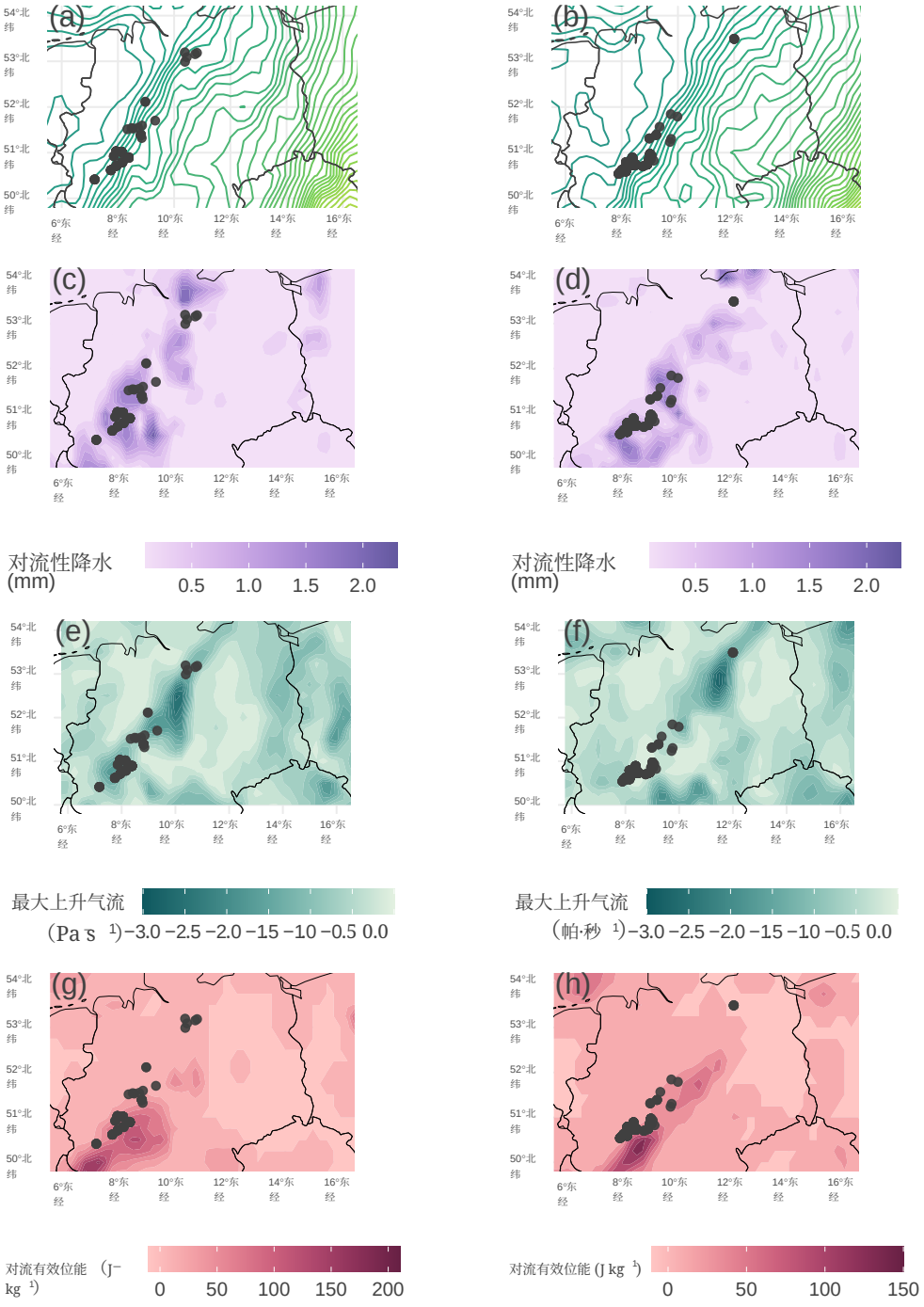


图 5. 2019年3月 *4th* 日研究区域的大尺度气象环境。左列展示13时（UTC）的环境，右列展示14时（UTC）的环境。从上到下依次为：850百帕气温等值线的空间分布（间隔为1 K）、对流降水、最大大尺度上升速度（负值表示向上运动）以及对流有效位能。颜色越深表示数值越大。所有图中的深灰色点表示根据闪电定位系统 EUCLID 数据确定的相应小时内、ERA5 网格单元中的闪电。

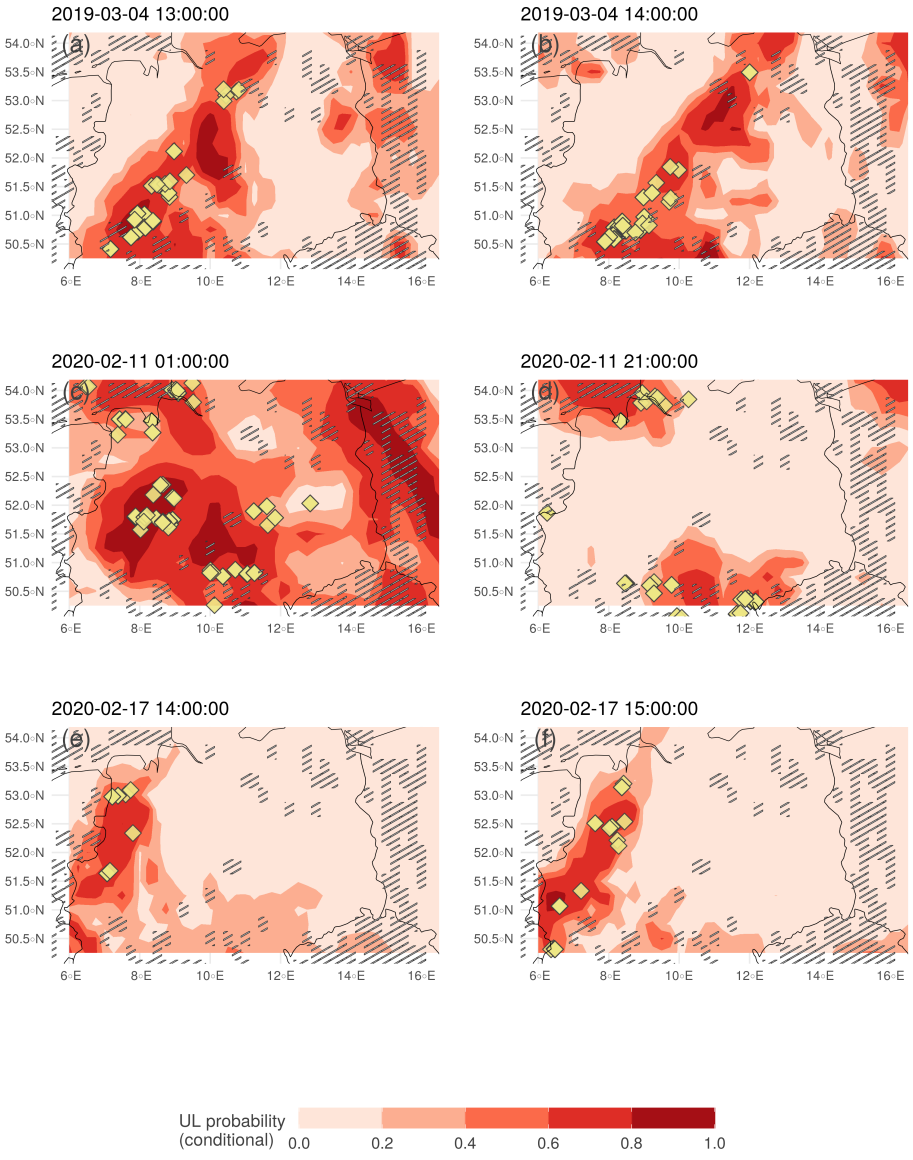


图6. 根据盖斯贝格和森蒂斯塔塔站数据训练的100个随机森林模型诊断得出的上行闪电条件概率中位数（红色区域）。黄色符号表示根据EUCLID数据得出的所考虑小时内的闪电。灰色阴影区域表示没有风力涡轮机的网格单元。

选取的案例研究日期具有中纬度寒冷季节的典型天气特征。过渡季节和冬季的大气往往变化剧烈，并受气旋与反气旋交替演变的影响，从而决定气象环境（Perry，1987年）。

尤其是中纬度气旋的形成与发展，为所谓的风场雷暴提供了有利条件（Morgenstern等人，2022年）。这类雷暴通常与强上升气流、大量降水以及较低但仍存在的对流有效位能有关。

针对仪器化塔观测到的驱动因素，对第一个案例进行了更详细的分析（图3）。图5展示了2019年 $4th$ March 13时（UTC）和14时（UTC）大尺度等温线的位置、对流降水的空间分布、最大上升气流速度以及对流有效位能。闪电定位系统在相应小时内检测到的、发生在所识别风力涡轮机附近的闪电事件，以深灰色圆点标示。

气象环境由中午前槽前方冷锋的过境决定。面板 (a) 和 (b) 中，850百帕等温线密集排列，自西向东横穿德国北部和中部，显示了冷锋的大致位置。冷锋意味着局地对流降水量增加，如面板 (c) 和 (d) 所示；强上升气流在面板 (e) 和 (f) 中表现为较大的负值；与夏季深对流相比，面板 (g) 和 (h) 中对流有效位能略有增加，但总体仍然较低。冷锋导致研究区域内三个变量的最大增幅分别出现在略有不同的区域。对流降水沿冷锋增多，而另外两个变量的最大值则局地集中在更小的区域（例如，德国北部和中部的最大上升气流速度）。

图6以红色显示了随机森林模型对三个案例研究日诊断得出的条件概率。面板 (a) 和 (b) 展示了图5所讨论的具体案例的结果。诊断出的分布模式是综合了三个驱动变量影响的结果。这表明，最终的概率图并非由单一变量决定，而是由不同影响因素相互作用形成，进而显示出上行闪电风险增加的区域。

黄色符号再次表示所考察小时内的雷击。黄色标记的已识别闪电事件要求附近300米范围内存在风力涡轮机，如第2节所述。本图未考虑所有其他可能发生上行闪电的高耸结构。因此，诊断出的概率不受风力涡轮机位置影响，这意味着即使没有安装风力涡轮机，也可能诊断出较高的概率。没有风力涡轮机的网格单元以灰色阴影标示。

图6中的三个案例研究日均显示，研究区域内上行闪电发生概率诊断值较高的区域，与相应小时内已识别的闪电事件分布吻合良好。在这三个案例研究中，发生上行闪电的诊断风险极低区域与风险极高区域之间界限分明。

在图6面板(c)和(d)所示的2020年 $11th$ 月，研究区域再次处于强西风气流中，并伴有局地对流降水、对流有效位能和强上升气流增加（此处未显示）。2020年 $17th$ 月，研究区域上空的高空（500百帕以上）有冷锋经过。尽管气象

形势不同，但这些条件再次与其他案例研究相似，三个驱动变量的数值均有所增加，而这三个变量对诊断出的条件概率有很大影响。

4.2 风力涡轮机上行闪电风险评估

识别因大尺度气象条件而导致上行闪电风险增加的区域，是评估风力涡轮机闪电风险的重要一步。案例研究清楚地表明，风力涡轮机处观测到的闪电与随机森林模型诊断出的高概率区域相吻合。以下分析考虑了所研究时段内所有识别出风力涡轮机处有闪电的事件。除了用于 UL-LLS 的模型将提供风险评估外，现在还将随机森林模型应用于研究区域和所研究时段，以评估 UL-LLS + UL-noLLS。

所研究时段涵盖2018年至2020年的过渡季节和冬季，共计185个事件日、1 027个单次闪电小时数和18 602次单次闪电事件。这些数字将作为衡量标准，用于验证随机森林模型诊断出的概率。请注意，这些数字是实际发生闪电次数的下限。考虑到人工识别风力涡轮机处闪电的不确定性，以及闪电定位系统探测上行闪电的不确定性，实际发生在风力涡轮机处的闪电事件数量可能显著更多。此外，这种验证方法仅考虑风力涡轮机处的闪电，而忽略了研究区域内所有其他可能受上行闪电影响的高耸结构，例如无线电塔。以下分析将所研究时段内的所有日期作为随机森林模型的新数据，用于按小时诊断条件概率。

研究旨在根据随机森林模型，找出相较于其他地区更频繁受到较高上行闪电风险影响的区域。为此，统计 ERA5 每个网格单元（0.25° × 0.25°）中条件概率超过0.5 的小时数。

风力涡轮机处 UL-LLS 风险评估

图7 (a) 显示，研究区域内有些地方比其他地方更常出现 UL-LLS，风险也更高。研究区域的西部和西南部发生 UL-LLS 的概率明显较高。这也与图4的面板 (b) 相符，该面板展示了各网格单元内实际观测到的风力涡轮机遭遇至少一次闪电事件的小时数。

有趣的是，图7中 UL-LLS 概率较高的区域，大致与研究区域南部三分之一区域内的高海拔地形相吻合（参见图1）。可能的解释包括风力涡轮机有效引雷高度增加（例如 Shindo，2018），以及地形抬升和热力驱动的风增加了雷暴形成的可能性（Kirshbaum 等，2018）。海风也可能解释了研究区域最西北部临海区域的概率较高这一现象。

将基于直接塔架测量的气象数据训练的 UL-LLS 模型成功推广至更大区域，并通过风力涡轮机上的独立验证，展现了我们的方法有望生成区域性

变化的风险图，进而可能促成因地区而异的（自愿或强制）风力涡轮机雷电防护标准。

风力涡轮机UL-LLS +UL-noLLS的风险评估

将基于塔架训练并经过验证的UL-LLS模型成功迁移至更大范围，为对所有上行闪电（UL-LLS和UL-noLLS）采用同样的基于塔架训练的模型提供了依据，尽管目前没有可用于独立验证的数据。

图7的面板（b）表明，若进一步考虑LLS不可探测的上行闪电类型，预计闪电次数会更多。尽管一些受影响更频繁的区域范围有所扩大，但上行闪电风险升高区域的分布模式相似。据此可以推测，大尺度气象导致UL-LLS或UL-noLLS闪电类型的形成机制相似。

该范围南部高海拔地形区域以及最西北部沿海区域的风险增幅最为显著。

5 结论

始于风力涡轮机等高耸结构的上行闪电（UL）比下行闪电（DL）更具破坏性。每次UL闪电都始于初始连续电流（ICC），其持续时间约为DL中初始连续电流的十倍，因此向高耸结构传递的电荷量大得多。此外，直接的上行闪电测量表明，大多数闪电定位系统（LLS）只能探测到不足50%的UL事件，因为它们无法探测仅包含ICC的UL。

在盖斯贝格仪器化塔直接测得的上行闪电季节变化很小。然而，当前的闪电防护标准以闪电定位系统数据推算的年闪电密度为依据，而该数据显然主要由暖季的下行闪电构成。UL-noLLS完全未被纳入考虑，寒季的上行闪电也被严重低估。关于上行闪电发生情况的基础知识仍不完整，阻碍了对风力涡轮机上行闪电风险的恰当评估。

因此，未考虑UL-noLLS以及寒季对上行闪电的重要性，将会大幅低估上行闪电对风力涡轮机造成的风险。本研究将罕见的上行闪电直接测量数据与大范围气象数据结合，利用机器学习模型估算风力涡轮机遭受所有上行闪电（UL-LLS和UL-noLLS）的风险。

第一步是基于两座专门建设的仪器化塔的长期观测，训练并验证两种不同的随机森林模型。一种模型仅考虑UL-LLS，另一种模型考虑UL-LLS + UL-noLLS。模型的输入数据为来自盖斯贝格塔（奥地利，2000—2015年）和森蒂斯塔（瑞士，2010—2017年）的上行闪电直接测量数据。盖斯贝格塔的传感器也能测量UL-noLLS，而森蒂斯塔的传感器则会漏测其中的大部分。

第二步，将随机森林模型扩展到更大的研究区域（北纬 50°—54°、东经 6°—16°），以识别寒冷季节（10月至次年4月）上行闪电风险较高的区域。作为验证，从闪电定位系统和 OpenStreetMap 数据中提取该区域内风力发电机上的所有闪电，并与随机森林诊断出的概率进行比较。

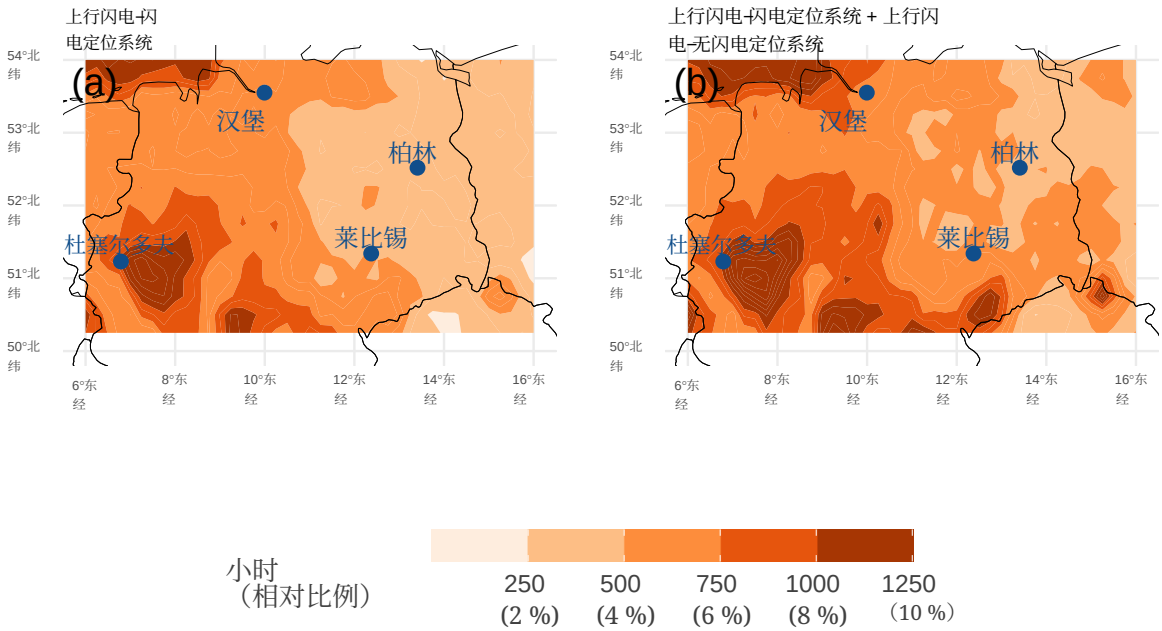


图 7. 面板 (a) 和 (b)：2018年至2020年寒冷季节（10月至次年4月）的上行闪电（UL）潜势图。橙色表示根据 100 个随机森林模型，条件概率超过 0.5 的网格单元中每小时数的中位数。面板 (a) 显示基于盖斯贝格和森蒂斯山数据合并训练的模型结果。面板 (b) 显示基于盖斯贝格数据（也包括 UL-noLLS）训练的模型结果。总计 12480 小时的相对比例作为参考。

结果表明，在盖斯贝格塔和森蒂斯塔训练的随机森林模型能够可靠地诊断上行闪电。较大尺度的气象驱动因素包括大量（对流性）降水、较强的垂直上升气流速度以及略微升高的对流有效位能。此外，云的垂直延伸范围，以及冰晶和固态水凝物的含量，也是重要变量。

将随机森林扩展至更大区域后发现，概率图与风力涡轮机处观测到的雷击相吻合。将基于盖斯贝格塔数据（包括 UL-noLLS 闪电）训练的模型扩展后显示，预计上行闪电（UL）风险增加的区域将更频繁地出现上行闪电。

德国西北部研究区域的西部和南部地形高海拔，西北端则为沿海地区，这些区域的风力涡轮机遭受上行闪电的风险最高。本研究表明，利用机器学习模型（随机森林），可以根据大尺度气象条件可靠地建模仪器化塔上的上行闪电直接测量数据。本研究还提出了一种新方法，通过使用风力涡轮机位置的 UL-LLS 数据，论证将该模型迁移至更大区域的合理性。因此，可以绘制区域细节丰富的风力涡轮机上行闪电风险图。

附录 A：补充材料

A1 数据可用性

ERA5数据可从 https://cds.climate.copernicus.eu 免费下载（Hersbach 等（2020））。EUCLID 数据以及盖斯贝格塔的直接观测数据仅可按需提供。详情请联系 Wolfgang Schulz。

A2 软件

所有计算以及最终数据集的构建、建模和预测均使用 R（R 核心团队，2021）完成，并使用了 netCDF4（Pierce，2019）、partykit（Hothorn 和 Zeileis，2015）和 ggplot2 软件包（Wickham，2016）。获取原始数据并从 ERA5 推导其他变量，则需要使用 Python3（VanRossum 和 Drake，2009）及 cdo（Schulzweida，2019）。

A3 使用更高概率阈值评估风力涡轮机上行闪电的风险

第 4.2 节介绍了 UL-LLS 和 UL-LLS + UL-noLLS 的风险评估模型结果，统计超过条件概率 0.5 的小时数。图 A1 展示了采用更高概率阈值（即 0.8）进行的风险评估。与概率阈值为 0.5 时相比，超过该阈值的小时数约减少一半。不过，区域分布格局仍然相似，最大值出现在研究区域以西和西南方向。

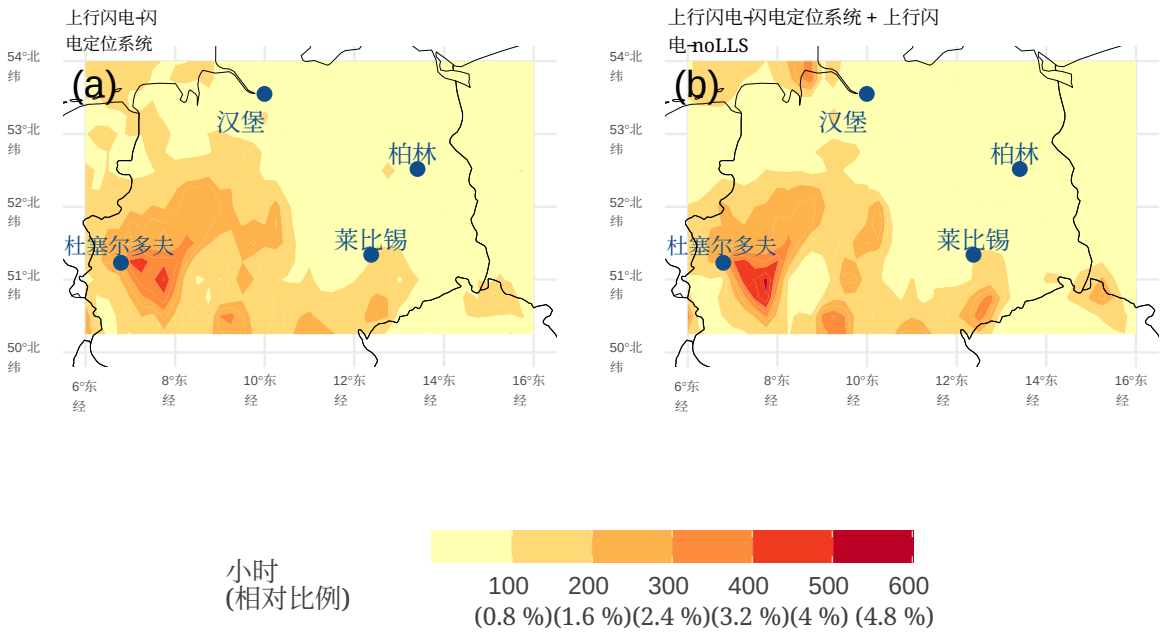


图 A1. 面板 (a) 和 (b)：2018年至2020年寒冷季节（10月至次年4月）上行闪电潜势的分布图。橙色表示每个网格单元中超过 0.8 条件概率的小时数中位数，该数值根据 100 个随机森林模型得出。面板 (a) 显示基于盖斯贝格和森蒂斯山数据合并训练的模型结果。面板 (b) 显示基于盖斯贝格数据（也包括 UL-noLLS）训练的模型结果。总计 12480 小时的相对比例作为参考。

表 A1. 取自 ERA5 的大尺度变量及由 ERA5 推导出的变量列表。建议认为这些推导变量（以斜体标示）可能对雷暴云的起电过程或对流发展具有重要作用。

大尺度变量	**Unit**
离地云底高度	米（离地）
对流降水（雨 + 雪）	m
大尺度降水	m
云体大小	m
最大降水率（雨 + 雪）	$\mathrm{kg\ m^{-2}\ s^{-1}}$
冰晶（整层，tciw）	$\mathrm{kg\ m^{-2}}$
固态水凝物（整层，tcsw）	$\mathrm{kg\ m^{-2}}$
过冷液态水（整层，tcslw）	$\mathrm{kg\ m^{-2}}$
水汽（整层）	$\mathrm{kg\ m^{-2}}$
散度垂直积分云中冻结水通量的	$\mathrm{kg\ m^{-2}\ s^{-1}}$
液体的垂直输送约 $-10℃$	$\mathrm{kg\ Pa\ s^{-1}}$
冰晶（$-10℃$ - $-20℃$）	$\mathrm{kg\ m^{-2}}$
冰晶（$-20℃$ - $-40℃$）	$\mathrm{kg\ m^{-2}}$
云水滴（$-10℃$ - $-20℃$）	$\mathrm{kg\ m^{-2}}$
固态水凝物（$-10℃$ - $-20℃$）	$\mathrm{kg\ m^{-2}}$
固态水凝物（$-20℃$ - $-40℃$）	$\mathrm{kg\ m^{-2}}$
固态物质（*cswc* + *ciwc*）约 $-10℃$	$\mathrm{kg\ m^{-2}}$
液态物质（云液态水含量 + 雨水含量）约 $-10℃$	$\mathrm{kg\ m^{-2}}$
2 米高度的露点温度	K

平均垂直积分 水汽辐合	$kg\ m^{-2}\ s^{-1}$
水汽 （−10℃ - −20℃）	$kg\ m^{-2}$
边界层高度	m
地表潜热通量	$J\ m^{-2}$
地表感热通量	$J\ m^{-2}$
向下地表太阳辐射	$J\ m^{-2}$
对流有效 位能	J 千克$^{-1}$
存在对流抑制	二元
平均海平面气压	Pa
−10 ℃ 等温线高度	米（距地面）
边界层耗散	$J\ m^{-2}$
最大垂直上升气流速度	$Pa\ s^{-1}$
总云切变	$m\ s^{-1}$
*10*米处风速	$m\ s^{-1}$
*10*米处风向	°
*10*米处与云底之间的切变	$m\ s^{-1}$

致谢。 我们感谢奥地利研究促进署（FFG）项目（编号：872656）和奥地利科学基金（FWF）项目（编号：P31836）为本研究提供资助。我们感谢瑞士联邦理工学院（EPFL）EMC集团提供森蒂斯塔雷击数据。最后，我们感谢 BLIDS 的运营方西门子提供 EUCLID 数据。

作者贡献。 Isabell Stucke 负责调查、软件编写、结果可视化及论文撰写。Deborah Morgernstern、Thorsten Simon 和 Isabell Stucke 负责数据整理、构建数据集，并基于 ERA5数据推导变量。Thorsten Simon 参与了编码概念的制定。Georg J. Mayr 为气象分析、数据组织和资金获取提供支持。Achim Zeileis 负责监督正式分析及统计方法的解释。Achim Zeileis、Georg J. Mayr 和 Thorsten Simon 担任项目管理员和指导者。所有作者均参与了本文的概念构思，讨论了方法论，评估了结果，并对论文提出意见。

利益冲突。 作者声明不存在利益冲突。

参考文献

Berger, K.：闪电放电的新观测：圣萨尔瓦托雷山研究结果，《富兰克林研究所期刊》，283，478–525，https://doi.org/10.1016/0016-0032(67)90598-4，1967年。Diendorfer, G.：利用对塔闪电验证闪电定位系统性能，《Impulse》，1，11，2010年。Diendorfer, G.：欧洲冬季月份风力涡轮机遭受雷击的概率，技术报告，Copernicus Meetings，https://doi.org/10.5194/egusphere-egu2020-3337，2020年。Diendorfer, G.、Pichler, H. 和 Mair, M.：盖斯贝格塔负极性上行先导闪电的一些参数（2000–2007），《IEEE电磁兼容汇刊》，51，443–452，https://doi.org/10.1109/TEMC.2009.2021616，2009年。Diendorfer, G.、Pichler, H. 和 Schulz, W.：闪电定位系统对上行先导闪电的探测，载于：第九届亚太国际雷电会议（APL）论文集，第1–5页，2015年。Farr, T. G. 和 Kobrick, M.：航天飞机雷达地形测绘任务产生了丰富的数据，《Eos：美国地球物理联合会汇刊》，81，583–585，2000年。Hersbach, H.、Bell, B.、Berrisford, P.、Hirahara, S. 和 Horányi, A.：ERA5全球再分析，《皇家气象学会季刊》，146，1999–2049，https://doi.org/10.1002/qj.3803，2020年。Hothorn, T. 和 Zeileis, A.：partykit：R语言递归分割模块化工具包，《机器学习研究杂志》，16，3905–3909，http://jmlr.org/papers/v16/hothorn15a.html，2015年。Hothorn, T.、Hornik, K. 和 Zeileis, A.：无偏递归分割：条件推断框架，《计算与图形统计学杂志》，15，651–674，https://doi.org/10.1198/106186006X133933，2006年。IEC 61400-24：风能发电系统——第24部分：雷电防护，标准，2019年。Kirshbaum, D.、Adler, B.、Kalthoff, N.、Barthlott, C. 和 Serafin, S.：湿润地形对流：物理机制及其与地表交换过程的联系，《大气》，9，80，https://doi.org/10.3390/atmos9030080，2018年。Mansoor, V.、Nagaveni, P.、Amudha, A.、Divyapriya, S.、Emayavaramban, G.、Ramkumar, M. S. 和 SivaramKrishnan, M.：揭示闪电之谜，《今日材料：会议录》，37，2572–2577，https://doi.org/10.1016/j.matpr.2020.08.500，2021年。March, V.、Montanyà, J.、Fabró, F.、van der Velde, O.、Romero, D.、Solà, G.、Freijó, M. 和 Pineda, N.：特定全球区域的冬季雷电活动及其对风力涡轮机和高耸结构的影响，载于：2016年第33届国际雷电防护会议（ICLP），第1–5页，电气电子工程师学会，https://doi.org/10.1109/ICLP.2016.7791447，2016年。Matsui, M.、Michishita, K. 和 Yokoyama, S.：日本云地闪密度及击中风力涡轮机的闪电数量，《电力系统研究》，181，106066，https://doi.org/10.1016/j.epsr.2019.106066，2020年。Montanyà Puig, J.、Fabró Tàpia, F.、Van der Velde, O. A.、March Nomen, V.、Williams, E. R.、Pineda Ruegg, N.、Romero Durán, D.、Solà de Las Fuentes, G. 和 Freijo Álvarez, M.：冬季雷电的全球分布：对风力涡轮机和飞机的威胁，《自然灾害与地球系统科学》，16，1465–1472，https://doi.org/10.5194/nhess-16-1465-2016，2016年。Morgenstern, D.、Stucke, I.、Simon, T.、Mayr, G. J. 和 Zeileis, A.：利用风场和质量场特征区分冬季和夏季闪电，《天气与气候动力学》，3，361–375，https://doi.org/10.5194/wcd-3-361-2022，2022年。Perry, A.：中纬度气候，第581–583页，Springer美国，波士顿，马萨诸塞州，https://doi.org/10.1007/0-387-30749-4_116，1987年。Pierce, D.：ncdf4：Unidata netCDF（第4版或更早版本）格式数据文件接口，https://CRAN.R-project.org/package=ncdf4，R软件包版本1.17，2019年。

Pineda, N., Montanyà, J., Salvador, A., van der Velde, O. A., 和 López, J. A.：有利于风力涡轮机遭受向下闪电和上行闪电的雷暴特征，《大气研究》，214，46–63，https://doi.org/10.1016/j.atmosres.2018.07.012，2018。R 核心团队：R：用于统计计算的语言与环境，R 统计计算基金会，维也纳，奥地利，https: //www.R-project.org/，2021。Rachidi, F., Rubinstein, M., Montanyà, J., Bermudez, J.-L., Sola, R. R., Sola, G., 和 Korovkin, N.：新一代风力涡轮机叶片雷电防护中的当前问题综述，《IEEE工业电子学汇刊》，55，2489–2496，https://doi.org/10.1109/TIE.2007.896443，2008。Rakov, V. A. 和 Uman, M. A.：《闪电：物理学与效应》，剑桥大学出版社，https://doi.org/10.1017/CBO9781107340886，2003。Romero, C., Paolone, M., Rubinstein, M., Rachidi, F., Rubinstein, A., Diendorfer, G., Schulz, W., Daout, B., Kälin, A., 和 Zweiacker, P.：森蒂斯塔闪电电流测量系统，《电力系统研究》，82，34–43，https://doi.org/10.1016/j.epsr.2011.08.011，2012。Schulz, W., Diendorfer, G., Pedeboy, S., 和 Poelman, D. R.：欧洲闪电定位系统 EUCLID——第1部分：性能分析与验证，《自然灾害与地球系统科学》，16，595–605，https://doi.org/10.5194/nhess-16-595-2016，2016。Schulzweida, U.：CDO 用户指南，https://doi.org/10.5281/zenodo.3539275，2019。Shindo, T.：高耸结构遭受闪电袭击的特征，《IEEJ电气与电子工程汇刊》，13，938–947，https://doi.org/10.1002/tee.22649，2018。Strasser, H. 和 Weber, C.：置换统计量的渐近理论，《统计学数学方法》，8，220–250，1999。Strobl, C., Malley, J., 和 Tutz, G.：递归分割导论：分类回归树、装袋法和随机森林的原理、应用与特征，《心理学方法》，14，323–348，https://doi.org/10.1037/a0016973，2009。Van Rossum, G. 和 Drake, F. L.：Python 3参考手册，Python文档手册第2部分，CreateSpace独立出版平台，斯科茨谷，加利福尼亚州，访问日期：2021-03-09，2009。Wickham, H.：ggplot2：数据分析的优雅图形，施普林格纽约出版社，https://ggplot2.tidyverse.org，2016。Zhang, Y. 和 Zhang, X.：风力涡轮机雷电暂态的统计分析，《可再生能源与可持续能源杂志》，12，063302，https://doi.org/10.1063/5.0031506，2020。