基于雷达的雨滴粒径分布预测：解析方法、神经网络与决策树方法的比较

R. J. Humphreys

摘要：可靠估算雨滴粒径分布（RSD）对于定量降水估计（QPE）、土壤侵蚀模拟和风力机叶片侵蚀等应用至关重要。雨滴谱仪等原位仪器虽然能够提供详细的雨滴谱测量值，但其空间覆盖范围有限，因此需要借助偏振雷达远程反演降雨微物理特性。本研究比较评估了利用解析方法和机器学习方法，从双偏振雷达观测量中反演雨滴谱参数的效果。在英国 Sheepdrove Farm，于2020年9月至2022年5月期间采集的 OTT Parsivel2雨滴谱仪分钟级测量值，经过与同址称重式雨量计和翻斗式雨量计的测量值对照后进行质量控制。研究将测得的雨滴谱拟合为归一化三参数伽马分布，并据此解析模拟了一系列双偏振雷达变量。随后，研究针对多种雷达特征集和模型架构，训练解析反演、神经网络和决策树模型，以估算伽马分布参数。为评估模型的稳健性和等效性，每种模型配置均采用随机的 70/30 训练—测试集划分训练 100 次，总计训练约 17,000 个模型。机器学习方法的整体表现通常优于解析方法；然而，没有任何一种模型类别或架构始终处于最优。模型性能在很大程度上取决于目标雨滴谱参数以及可用的雷达可观测量；在特征缩减的情况下，决策树表现尤为稳健。这些结果凸显了让反演模型结构适应实际运行中的数据约束的重要性，而非采用单一的通用方法。

关键词：雨滴粒径分布；定量降水估计；雷达；机器学习

1. 引言

从降水事件中可靠地估计雨滴粒径分布（RSD），对于土壤侵蚀、风力机叶片侵蚀和定量降水估计等各种实际问题至关重要。传统的雨滴粒径分布测量方法需要部署昂贵的地面实测设备（例如雨滴谱仪或光学设备）。然而，这些测量值只能提供特定点位的估计，无法捕捉雨滴粒径分布的空间变异性。因此，研究人员使用偏振雷达。偏振雷达可获取正交极化反射率测量值，用于远程估计雨的微物理特性，进而估计降雨率。多项研究已探索利用这项技术，以更低成本、更便捷的方式估计雨滴粒径分布，但仍需开展更多工作，以稳健地评估和比较这些模型。

本研究利用了2020年9月至2022年5月期间在斯温顿Sheepdrove Farm以一分钟间隔采集的 OTT Parsivel^{2}雨滴谱仪数据。除通过传感器的特定直径雨滴数量外，雨滴谱仪数据还包括降雨率和水凝物类型。水凝物类型用于剔除所有非液态降水（例如冰雹）。数据同一地点还可获得另外三台雨量计的数据：一台称重式雨量计和两台翻斗式雨量计（翻斗式雨量计）。因此，可以通过将其结果与其他雨量计提供的数据进行比较，来评估雨滴谱仪数据的有效性。

利用雨滴谱仪数据获得实测雨滴谱，并采用成熟的参数反演技术，将其拟合到雨滴谱的归一化三参数伽马模型。随后，利用该模型化雨滴谱通过解析方法模拟大量雷达参数，包括反射率和差分反射率。接着，采用文献中常见的多种方法训练模型，以估计归一化伽马模型的参数，这些方法包括解析方法、神经网络和决策树模型。对于神经网络和决策树方法，研究对模型结构进行了调整，以便比较不同架构的预测能力。此外，研究向模型提供了广泛的雷达产品，以适应现实应用中雷达产品可用性和可靠性的差异。由于这些模型以数据为中心，每个模型都使用随机抽取的 70% 模拟雷达参数训练 100 次，其余 30% 的数据则留作测试。研究共生成了 17000 个独特模型，从而能够广泛而严谨地洞察不同雨滴粒径谱反演模型的效能。

结果表明，在获取雨滴谱参数方面，机器学习方法比解析方法更准确。然而，没有任何一种机器学习模型在预测雨滴谱参数时始终表现出更强的预测能力。相反，模型结构应取决于可获取的雷达数据范围以及模型要估算的具体参数。例如，仅使用雷达反射率和差分反射率测量值时，具有 36 个叶节点的决策树被证明是最有能力准确获取中值体积直径的模型。

原创性声明：本文基于文献中提出的方法，训练了大量不同的雨滴粒径谱反演模型。输入数据经过了 100 次随机化，使得我们能够比以往研究更严格地分析模型的预测效用。此外，鉴于等效性这一概念，本文采用了更广泛的雷达参数来校准机器学习（ML）模型，从而考虑了真实世界数据中雷达产品可用性和不确定性的差异。我们还测试了范围广泛得多的机器学习结构，以评估模型架构对性能的影响。因此，本文全面深入地比较了不同模型类型、输入参数和机器学习结构对雨滴谱反演精度的影响。

2. 文献综述

准确估算雨滴粒径分布（RSD）对于解决各种现实问题至关重要。例如，了解雨滴粒径分布有助于更准确地认识降雨的动能，这对于预测风力涡轮机叶片前缘侵蚀程度具有重要价值。涡轮机叶片侵蚀是由雨滴撞击叶片并造成微小缺陷所致；这些缺陷会增加叶片表面的粗糙度，降低其空气动力性能，进而减少总体能量输出。侵蚀程度与撞击雨滴的动能有关（Tilg 等人，2022）。同样，Caracciolo 等人（2012）指出，土壤侵蚀也与降雨的动能有关，并且雨滴粒径分布“与测得的土壤流失量具有很强的相关性”。因此，包括 Dai 等人（2020）提出的模型在内的研究表明，可以利用高分辨率雨滴谱建立准确的土壤侵蚀模型。

了解雨滴谱对于定量降水估测（QPE）同样至关重要，因为它能揭示雨的微物理特性，从而有助于提高雷达降雨测量和预测的准确性（Tokay 和 Short，1996；Ulbrich，1983）。然而，雷达系统无法直接测量降雨率等特性，而是获取雷达参数的数值，随后可利用这些数值来估算降雨特性。例如，天气雷达系统可以通过测量反射率和差分反射率等雷达参数，估算降水的大小、形状和浓度。例如，反射率与雨滴谱的 6^{th} 阶矩有关（Seliga 和 Bringi，1976）。这一关系对定量降水估测（QPE）尤为关键，因为降雨率是雨滴谱的 3.67^{th} 阶矩（Zhang 等人，2001；Cluckie 和 Rico-Ramirez，2004）。因此，加深对雨滴谱的理解、对其进行建模并预测雨滴谱，能够显著改善雷达反演降雨率估算。

2.1 测量雨滴粒径分布

雨滴谱是对一次降水事件中雨滴粒径范围及其出现频率的统计表示。在二十世纪，人们开发了多种测量雨滴谱的方法。例如，染色法是一种常用的测量技术。实验人员将经过化学处理的纸张短时间暴露于雨中，随后记录纸上留下的斑迹（Kathiravelu 等人，2016）。Marshall 和 Palmer（1948）就是一个例子，他们使用两张染色滤纸记录雨滴粒径，并据此拟合出雨滴谱方程。另一种常见的雨滴谱测量技术是面粉颗粒法。该方法将黏结剂和面粉短暂暴露在雨中，使面粉受潮并形成坚硬的颗粒。这种方法常用于土壤侵蚀研究（Kathiravelu 等人，2016）。

近来，人们开发了多种类型的雨滴谱仪来测量雨滴谱。其中一种是 Joss-Waldvogel（JW）位移式雨滴谱仪。JW 雨滴谱仪已应用于多项研究：例如，Montopoli 等人（2008）利用 JW 雨滴谱仪数据分析并模拟归一化伽马雨滴粒径分布的时间演变。同样，Tokay 等人（2003）使用共址 JW 雨滴谱仪，研究了热带风暴中雷达校准和降雨验证的距离依赖性。

其他研究使用视频雨滴谱仪测量雨滴谱。例如，Adirosi 等人（2014）使用视频雨滴谱仪，分析三参数伽马雨滴粒径分布相对于自然雨滴粒径分布的准确性。此外，Bringi 等人（2002）使用视频雨滴谱仪数据，提出了一种根据雷达测量估计伽马雨滴尺寸分布模型参数的方法。

本研究将使用 OTT Parsivel^2 光学雨滴谱仪。Tokay 等人（2014）证明，与参考雨量计相比，Parsivel^2 的测量结果比 Parsivel^1 和上述 JW 雨滴谱仪都更加一致。Parsivel^2 已用于多项研究，包括 Tang 等人（2014）的研究，该研究将偏振雷达反演的雨滴谱与 Parsivel^{2}获得的测量值进行了比较。在该研究中，研究人员进一步证实，

不同气候区域之间的降雨微物理变化很大。

迄今为止，人们已经介绍了用于获取雨滴谱和降雨率的各种地基经验测量技术。然而，尽管这些技术具有实用价值，但它们无法捕捉大范围内降雨的空间变异性（Bringi 等人，2002；Bringi 和 Chandrasekar，2001）。为克服这一局限，雷达技术作为一种遥感技术得到了发展，能够远距离提供有关降水的详细信息。雷达系统主要分为两类：单偏振和双偏振。两类系统都会发射和接收无线电波脉冲，从而能够测量包括雷达反射率和多普勒速度在内的现象（Sokol 等人，2021）。

Seliga 和 Bringi（1976）的开创性研究证明了正交偏振雷达估算降雨率的潜力。这项研究促成了双偏振雷达系统的发展，相较于早期的单偏振雷达系统，双偏振雷达系统具有多项优势（Sokol 等人，2021）。例如，双偏振雷达提供了额外的雷达产品，包括：水平或垂直反射率（$Z_{h,v}$）、差分反射率（ Z_{dr}）、比差分相移（K_{dp}）、比衰减（A_H）、相关系数（ρ_{hv}）以及线性退偏振比（LDR_{vh}）。

这项技术已应用于地基和主动式卫星降雨传感器。主动式卫星降雨传感器是安装在卫星上的雷达仪器。它们能够提供广泛的全球覆盖，包括海洋和偏远地区等广阔且难以到达的区域，因此对于降雨率预测至关重要。例如，Bringi 等人（2002）利用热带降雨测量任务（TRMM）卫星，在巴西上空的一次飑线事件期间获取了雷达产品数据。

2.2 雨滴尺寸分布的建模与反演

雨的微观结构由其雨滴谱（RSD）定义。因此，人们投入了大量精力，试图开发一种能够准确描述雨滴谱形状的参数化模型。例如，Marshall 和 Palmer（1948）最早提出了这一尝试之一，他们认为雨滴谱遵循简单的双参数指数分布。Seliga 和 Bringi（1976）的后续研究表明，双偏振差分反射率参数（Z_{dr}）与液滴尺寸有关。因此，在他们的开创性论文中，研究人员证明了偏振雷达可用于估算指数模型所描述的雨滴谱中值体积直径（D_0）。然而，Ulbrich（1983）后来的研究表明，Marshall 和 Palmer（1948）的指数分布不足以充分描述雨滴谱；他转而提出，雨滴谱更真实地应采用三参数（N_w, D_0, μ）伽马模型进行建模。在此基础上，Testud 等人（2001）提出了（三参数）归一化伽马分布，以更好地表示全球各地自然形成的雨滴谱。最后，通过研究不同气候区（包括大陆、海洋和赤道区域）的雨滴谱，Bringi 等人（2003）得出结论：雷达反演得到的归一化伽马分布参数 N_w 和D_0, 总体上“与雨滴谱仪测量值一致”。不过，该论文承认，形状参数μ 难以准确预测。

自归一化伽马分布被引入以来，多项研究表明，可以根据极化雷达测量的Z_h、Z_{dr}和K_{dp}，反演伽马模型的参数（N_w， D_0， μ）。例如，Gorgucci 等人（2001；2002）针对 C 波段雷达提出了“β –方法”，随后 Bringi 等人（2002）将其扩展至 S 波段雷达。“β –方法”改进了早期基于先验假设推导的降雨算法，即平衡形状—尺寸关系。因此，“β –方法”表现出对形状—尺寸关系变化的免疫性。此外，“β –方法”表现良好，将雨滴中值直径（D_0）的估计精度控制在 10% 以内，并将（对数尺度的）等效截距参数（N_w）的估计精度控制在 6% 以内。然而，μ 的估计较为困难（Gorgucci 等人，2001；2002）。作为“β –方法”的替代方案，Zhang 等人（2001）提出了“约束伽马（CG）方法”，该方法假设N_w， D_0, 和 μ 并非相互独立。然而，Brandes 等人（2004）确定，与“CG 方法”相比，“β –方法”与实测雨滴谱分布观测的一致性更高，并且对特定差分相移（K_{dp}）误差不那么敏感。

Vulpiani 等人（2006年）较新的研究尝试使用神经网络（NN）反演伽马模型参数。通用逼近定理证明，具有单个隐藏层且采用连续、任意有界激活函数的神经网络，可以逼近任意连续函数（Cybenko，1989；Hornik 等人，1989）。因此，神经网络能够逼近高度非线性函数，包括描述雷达可观测量与雨滴谱参数之间关系的函数（Vulpiani 等人，2006年）。由此，Vulpiani 等人（2006年）得以利用神经网络，根据 Z_h、Z_{dr}和K_{dp} 估算归一化伽马模型的参数。这种神经网络方法改进了文献中描述的既有方法，在低比差分相移K_{dp}]值下也“表现相当不错”（Vulpiani 等人，2006年）。然而，这项研究既未探究能否使用其他雷达产品估算归一化伽马模型，也未尝试多种神经网络结构。

类似地，Pasarescu（2020）的另一项研究也开发了一种基于神经网络的模型，用于从极化天气雷达测量数据中反演雨滴谱参数。她的模型同样仅将Z_h、Z_{dr}和K_{dp}用作输入特征；尽管她尝试了多种神经网络架构，分析中也仅采用了一个隐藏层含22个神经元的隐藏层神经网络。

另一种 RSD 参数反演技术是使用决策树。例如，Conrick 等（2020）使用决策树集成（即随机森林），根据（S 波段）极化天气雷达测量数据（D_0, 、液态水含量（LWC）和降雨率）反演中值体积直径，以及降雨率。研究人员还根据 LWC 和D_0计算了N_w。该模型使用 Parsivel² 雨滴谱仪数据集进行了测试，结果发现其“比非线性回归反演更准确、无偏”，且“均方根误差（RMSE）的改善幅度有些超过 50%”。此外，该模型“更有利于获得物理上准确的雨滴谱”。然而，这项研究并未考察其他决策树结构和输入参数，也未将模型性能与一系列不同的解析模型和机器学习模型进行比较。因此，尽管 RSD 反演取得了显著进展，仍有若干亟待解决的关键研究空白：

1. Gorgucci（2002）的 RSD 反演方法与采用相同数据集的其他机器学习（ML）方法相比，表现如何？如果没有这项研究，就无法准确比较不同 RSD 反演方法的效果。2. 改变机器学习模型的结构会如何影响其表现？研究多种模型结构可能会得到更准确的机器学习模型。3. 最后，使用不同的输入集会如何影响 RSD 参数估计的准确性？

为回答这些研究问题，本文设定了以下目标：获取雨滴谱仪数据，筛除异常及无关结果，并使用雨量计数据进行验证。随后，确定实测雨滴谱的伽马模型，并计算雷达参数。最后，以雷达参数为输入，对各类模型进行迭代训练和测试，并确保尝试多种输入集和机器学习结构。

3. 数据与方法

本文所用数据收集于2020年9月22nd 日至2022年5月11th 日期间，来自位于英格兰南部Sheepdrove Farm的四台共址雨量计（见图1）。这些雨量计包括一台OTT Parsivel² 光学雨滴谱仪（见图2）、一台 OTT Pluvio² 称重式翻斗雨量计（见图3），以及两台翻斗式雨量计，其翻斗分辨率分别为0.1mm和0.2mm（见图4）。



图 1：Sheepdrove Farm 的位置

Sheepdrove Farm 英格兰地图



图2：OTT Parsivel² 光学雨滴谱仪雨量计



图3：OTT Pluvio² 称重式翻斗雨量计



图4：翻斗式翻斗（0.2mm）雨量计

翻斗式雨量计因其结构简单且成本相对较低，被广泛用于测量降雨（Strangeways，2000年）。它们由漏斗形收集器组成，将降水导入一对在水平轴上呈双稳态平衡的翻斗之一（美国气象学会，2012年）。当预先确定量的降水累积后，翻斗倾倒，排空其中的水，并触发开关，向记录设备发送信号（Nystuen等人，1996年）。翻斗式雨量计在低强度降雨事件期间往往会低估降雨率。部分原因是它们需要临界水量在翻斗中累积后才会倾倒（Ciach，2003年）。其测量值还会受到已落入雨量计的降水蒸发以及风扰动降水下落轨迹的影响（Nešpor和Sevruk，1999年）

最后，翻斗式雨量计在强降雨事件期间通常也会低估降雨率（Ciach，2003年），因为翻斗无法足够快地倾倒以捕获全部降水（Luyckx和Berlamont，2001年）。相比之下，称重式雨量计（如OTT Pluvio², ）较不易受到与高强度降雨事件相关的误差影响，但价格更高，且需要更多维护（Duchon和Biddle，2010年）。称重式翻斗雨量计的工作原理是持续称量累积在集水容器中的降水（Nemeth，2008年）。

光学雨滴谱仪（例如 Parsivel²）能够测量单个雨滴（Löffler-Mang and Joss, 2000），因此与

低强度和高强度降雨条件下的翻斗式雨量计相比，通常具有更高的准确性。然而，与翻斗式雨量计和Pluvio称重式雨量计不同，Parsivel² 雨滴谱仪不会直接提供降雨率数值。相反，它利用对准光学传感器的窄激光束测量单个雨滴的粒径和速度（Tokay 等人，2014）。当降水穿过仪器的采样面积时，会使光束发生扰动， resulting in light intensity attenuation causes the voltage output from the photodiode to reduce。随后，系统利用这一变化自动确定雨滴粒径和速度，从而估算降雨率（R_{disd}）和水凝物类型（Tokay 等人, 2014）。雨滴粒径以水凝物直径的形式提供，并划分为 0.062mm 至 24.5mm 之间的 32 个间距不均匀的粒径等级（n_i）。

Pluvio² 和翻斗式雨量计的降雨率（分别为 R_{pluvio} 和 R_{TBRG} ）将用于验证雨滴谱仪提供的数据。所有设备的累积时间间隔均为一分钟，这与包括 Tokay 等人（2003年）、Bringi 等人（2011年）、Adirosi 等人（2014年）和 Pasarescu（2020）在内的既有研究一致。

3.1 雨量计数据的筛选与验证 ata

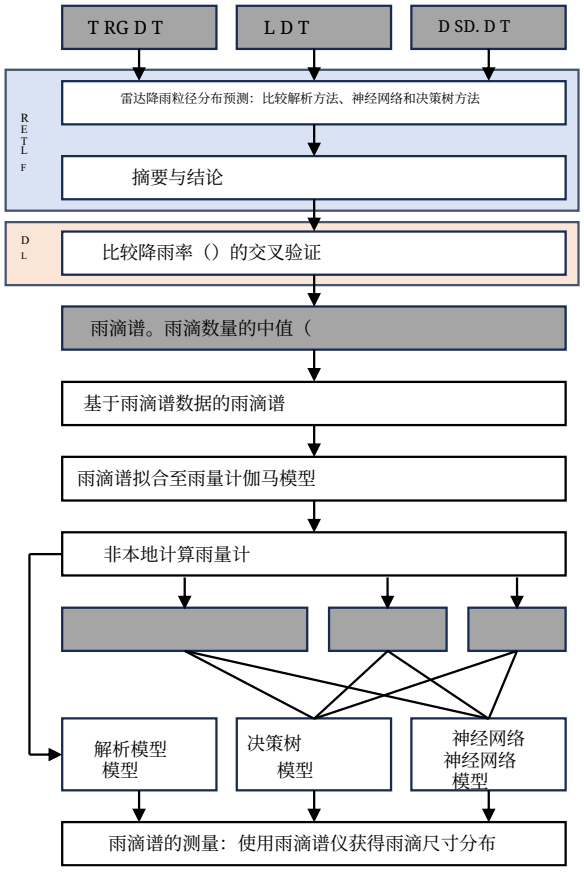


图5：从雨滴谱仪数据获取雨滴谱和雷达参数所遵循的流程

本研究使用来自地基雨量计的数据，生成一组通过解析方法推导出的

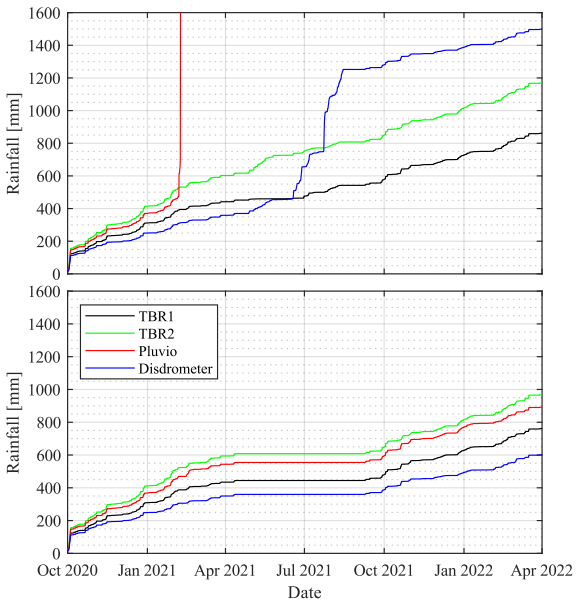


图6：降雨累积量之前（顶部）和过滤后（下方）。

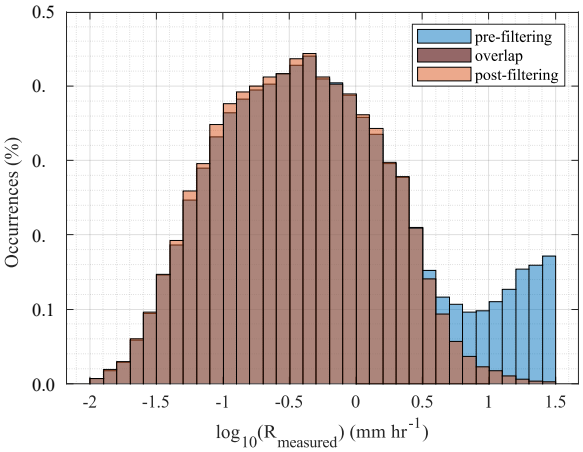


图7：雨滴谱仪降雨直方图：过滤前（蓝色/下方）与过滤后（橙色/上方）。

雷达参数。随后使用这些雷达参数拟合模型，以估算雨滴谱归一化伽马模型的三个参数。然而，为了提高模型的整体可靠性，训练数据必须具有尽可能高的质量。因此，我们制定了一个筛选与验证流程。图 5 概述了获取特征集（即通过解析方法

模拟雷达参数）。由于本研究仅限于分析雨，因此筛选过程的第一步是去除所有非液态水凝物。这是利用 OTT Parsivel² 雨滴谱仪提供的降水分类数据完成的。当降水类型为非液态（例如雪、冰雹等）时段内，所有雨量计的数据均予以剔除。

通过绘制数据采集时段（约597天）内的降雨累积量，可以识别出可能存在错误的数据

片段——见图6（上图）。例如，Pluvio雨量计记录到降雨量急剧增加（约发生于2021年2月6日），但其他雨量计并未证实这一情况。这一数据尖峰可能是由非降水实体落入雨量计造成的，例如一只小动物。为确保数据可靠，在降雨量增加期间，所有雨量计的全部数据均被剔除。TBR1数据（对应于$R_{TBRG\ 0.1毫米}$的累积量）在约2021年4月30日至2021年6月29日期间未能报告数据。翻斗机构可能因某种机械阻塞而无法翻转，因此剔除了这部分数据。最后，在约两个月的时间里（即2021年6月16日至2021年8月14日），雨滴谱仪记录到异常偏高的降水累积量，相当于该地区通常一年的降雨量。同期还记录到异常大的雨滴直径。经过一次维护检查后，雨滴谱仪开始返回更可信的结果。因此，这段错误数据最可能是由生长在激光传感器前方的植被造成的，导致仪器测量到下落速度几乎为零的大雨滴。因此，与发现的其他错误一样，所有传感器在这段错误时段内的全部数据均被剔除。最终的累积量数据绘于图6（下图）。经过这一处理，各雨量计记录的累积量明显更加吻合。图7显示了过滤前（蓝色）和过滤后（橙色）雨滴谱仪测得降雨量的直方图——请注意，异常偏高的降雨率有所减少。

为了评估雨滴谱仪数据的质量，有必要了解“真实”的降雨率。然而，所有经验测量技术都存在误差。例如，翻斗式雨量计存在系统误差，包括风以及强降雨和弱降雨所造成的误差（如蒸发、少测量和飞溅损失）（Habib等，2001年；Humphrey等，1997年；Nemec，1969年）。因此，需要利用现有数据构建对“真实”降雨率的近似值。

目前可用的数据来自共址雨量计，按一分钟累积间隔提供。然而，翻斗式雨量计的离散特性意味着，在一分钟间隔内，斗中的降水累积量可能尚未达到触发翻斗所需的临界阈值。这意味着，翻斗式雨量计数据无法反映该累积间隔内的正确降雨率。因此，将降雨率累积到更长的采样间隔（如30分钟）后，这些与采样相关的误差便不再重要（Ciach，2003年；Habib等，2001年）。此外，Ciach（2003年）证明，对多个雨量计的数据取平均值，是准确估算真实降雨率的有效方法。因此，通过计算两台翻斗式雨量计和Pluvio雨量计的降雨率算术平均值，构建了“真实”降雨率的估计值（R_T）。随后，将这一平均值按不同的时间间隔进行累积（T）。参见公式1。

$$R_T = R_{TBRG\ 0.1mm,T} + R_{TBRG\ 0.2mm,T} + R_{Pluvio,T} \quad (1)$$

确定了“真实”降雨率后，便可以深入了解雨滴谱仪数据的质量特征。方程 2 表示雨滴谱仪数据相对于“真实”降雨率的相对误差（e_T）。

$$e_T = \frac{R_{disdrometer,T} - R_T}{R_T} \quad (2)$$

图 8 展示了相对误差与“真实”降雨率的关系。定性来看，低降雨强度下相对误差的离散程度较高。然而，与 Ciach（2003年）的结果一致，随着降雨强度增加，这些误差会相对迅速地减小。

为了对数据得出更具定量性的结论，我们采用了非参数核回归估计器。该方法评估相对误差的标准差随“真实”降雨强度的变化情况。这一方法与 Ciach（2003年）采用的方法一致。方程 3 给出了 Nadaraya–Watson（N-W）核回归

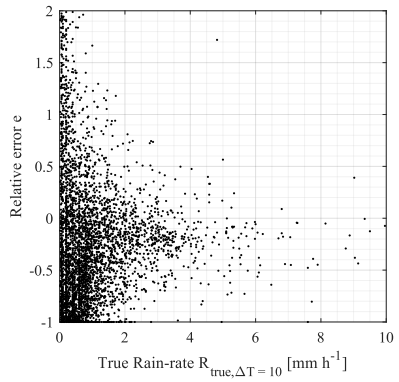


图 8：散点图，展示雨滴谱仪的相对误差相对于 a 处的“真实”降雨率 10分钟时间

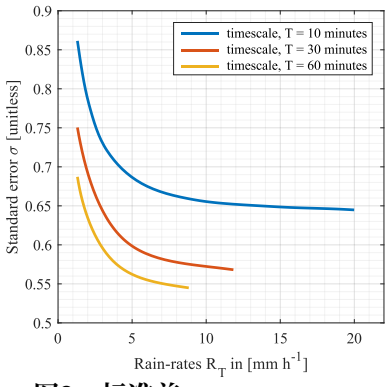


图9：标准差雨滴谱仪的相对误差相对于“真实”雨的率在10 ，30, 和60分钟。

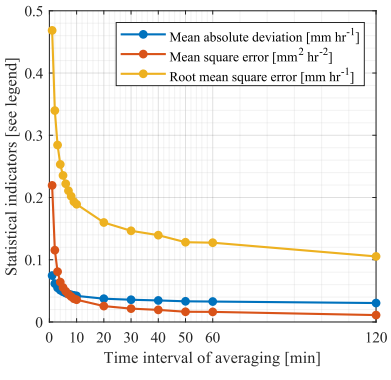


图10：不同累积时间间隔下 R_{disd} 的平均绝对误差、均方误差和均方根误差。

估计器，用于估算固定累积时间（T）下某一雨滴谱仪降雨值 R_{T_i} 对应的标准误差平方（σ_i^2），其表达式为：

$$\sigma_i^2(T,R_T)=\frac{\sum_i K\left(\frac{\boldsymbol{R}_T-R_{T_i}}{h}\right)e_{T_i}^2}{\sum_i K\left(\frac{\boldsymbol{R}_T-R_{T_i}}{h}\right)} \quad 3$$

其中，$\boldsymbol{R}_T$ 是完整的雨滴谱仪降雨数据集，$e_{T_i}^2$ 是特定“真实”降雨值的相对误差二阶统计矩，h是估计带宽，$K(\cdot)$是核函数。本分析采用了埃潘尼契尼科夫核，与 Ciach（2003）的方法一致——见方程 4。

$$K(x)=\begin{cases}0.75(1-x^2), & x\in[-h,h]\\ 0, & x\notin[-h,h]\end{cases} \quad 4$$

在 N-W回归中，较大的带宽会使拟合曲线更加平滑，但会牺牲细节。Ciach（2003）选择使用可变带宽，其值相当于特定雨滴谱仪降雨值的 20%；相比之下，本分析（见图 9）采用的带宽相当于特定雨滴谱仪降雨值的 100%。

与预期一致，图 9 所示的分析得出结论：相对误差随着降雨率的增加而降低。此外，还可以得出结论，采用较长的时间间隔进行累积会降低总体误差。然而，有必要扩展本分析，以确定累积间隔的变化如何影响误差。因此，为了量化雨滴谱仪测得的降雨率与真实降雨率之间的差异，将采用三个性能指标：平均绝对偏差（方程 5）、均方根误差（方程 6）和均方误差（方程 6）。

$$MAD=\frac{\sum_{i=1}^{N}(R_T-R_{disd,T})}{N} \quad 5$$

$$RMSE=\sqrt{MSE}=\sqrt{\frac{\sum_{i=1}^{N}(R_T-R_{disd,T})^2}{N}} \quad 6$$

其中，N 是记录的单个累积降雨率的数量（关于$R_{disd,T}$ 和R_T，请参见公式1和2）。

如前文 N-W回归所示，误差有随累积间隔增大而减小的趋势。图10进一步阐述了这一关系。可以看出，累积时间从一分钟开始增加时，误差显著减小；但随着时间继续增加，收益逐渐有限。因此，为了验证雨滴谱仪降雨数据的准确性，选择 60 分钟的间隔，因为它能显著降低误差，同时保留大量数据。

最后，图 11 展示了所有降雨率的散点图、均方根误差以及线性回归线。从这些结果可以明显看出，对于 60 分钟累积时间间隔，雨滴谱仪降雨率的均方根误差为 0.127 [毫米/小时$^{-1}$]，并且它倾向于低估降雨率。然而，就雨滴谱仪数据验证而言，显然，降雨率的“真实”值与雨滴谱仪的测量值之间存在密切关系（均方根误差 = 0.127 [mm hr^{-1}]）。因此，后续分析将假设雨滴谱仪数据足够准确，可用于训练模型。

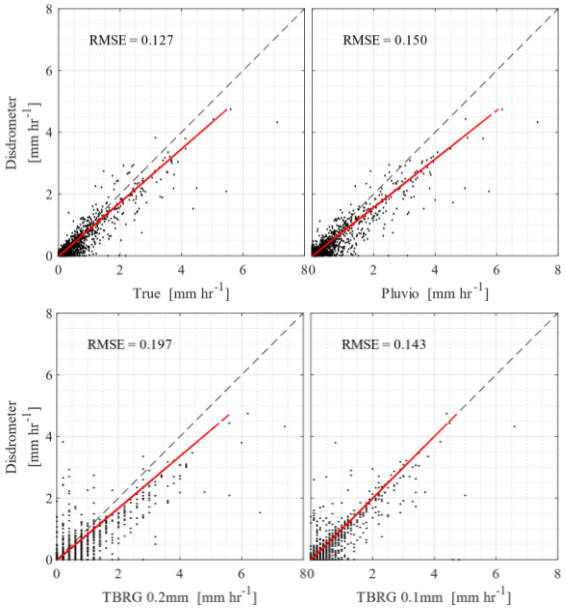


图 11：60 分钟累积时间间隔内，雨滴谱仪降雨率与其他降雨率的复合散点图。线性回归线以红色表示。

3.2 获取三参数归一化伽马模型的参数

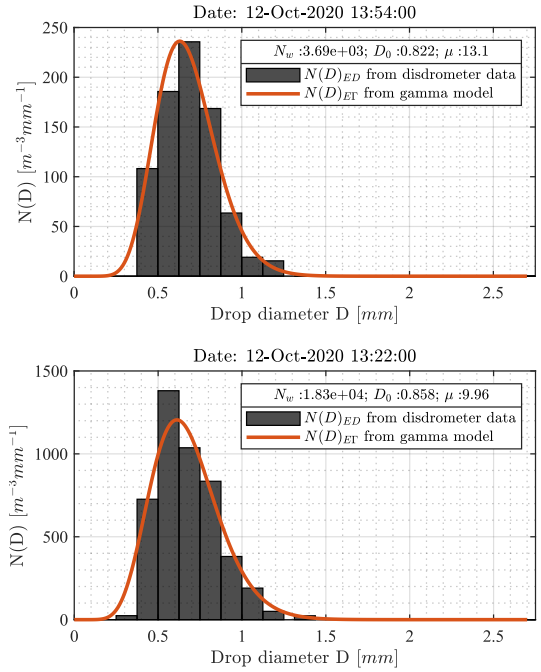


图 12：雨滴谱仪数据和三参数伽马模型的雨滴谱直方图

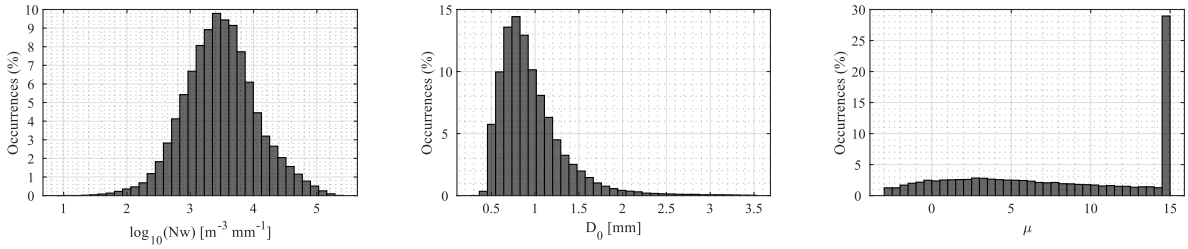


图 13：拟合得到的归一化伽马分布参数直方图（N_w， D_0， μ）

经过筛选和验证后，可以认为剩余的雨滴谱仪数据足够准确，可以继续进行分析。重要的是，这些数据包含按 32 个粒径等级（n_i）划分的水凝物直径，时间间隔为一分钟。为了将雨滴谱仪的降雨测量值与翻斗式雨量计进行比较，需要采用较长的时间间隔，以补偿翻斗式雨量计测量值中的误差。然而，获取雨滴谱时，使用原始的一分钟累积间隔更为合适。这是因为这一时间分辨率与瞬时雷达降雨观测更加一致，而且时间间隔足够短，能够捕捉不断变化的降雨变异性。

为获取三参数归一化伽马模型（下文简称3PNG模型）的参数，首先需要根据雨滴谱仪提供的数据，使用方程 7（Montopoli等人（2008））估算离散的一分钟时刻（t）的雨滴谱：

$$RSD, N_{ED}\,(D_i, t) = \frac{n_i(t)}{A \cdot \Delta t \cdot v_i \cdot \Delta D_i \cdot}\ [m^{-3} mm^{-1}] \qquad 7$$

其中， $N_{ED}\,(D_i, t)$表示尺寸类别（n_i）中的雨滴数量，下标ED 表示结果是根据雨滴谱仪数据估算得出的，D_i 表示该尺寸类别的直径宽度（单位：mm），Δt 表示时间间隔（一分钟）， A 表示雨滴谱仪测量面积， v_i 表示雨滴下落速度。然而，雨滴下落速度会随液滴大小而变化。对于本研究分析，某一雨滴的速度由方程 8（Atlas等人（1973））给出：

$$v_i = 9.65\ - e^{-0.6 D_i} \quad [m\ s^{-1}] \qquad 8$$

如第 1 节所述，文献中普遍认为应使用归一化伽马模型来表示雨滴谱。本研究将采用 Bringi等人（2003年）给出的形式——见方程 9：

$$RSD, N_{E\Gamma}\,(D) = N_w \cdot f(\mu) \cdot \left(\frac{D}{D_m}\right)^{\mu} \cdot e^{\frac{-(4+\mu)\cdot D}{D_m}} \quad [m^{-3} mm^{-1}] \qquad 9$$

这里，下标 $E\Gamma$ 表示雨滴谱是使用3PNG模型估算的，N_w 是归一化雨滴浓度参数（单位：mm^{-1} · m^{3}），D_m 是质量加权平均雨滴直径参数（单位：mm），μ 是伽马形状参数（无量纲），而 $f(\mu)$由公式10给出：

$$f(\mu) = \frac{6(4+\mu)^{\mu+4}}{4^4 \Gamma(\mu+4)} \qquad 10$$

这里， Γ 是伽马函数。最后，在本分析中，使用中值体积直径（D_0）代替参数 D_m ；Ulbrich（1983）给出了公式11所述的转换：

$$\frac{D_0}{D_m} = \frac{3.67+\mu}{4+\mu} \qquad 11$$

根据这些关系，可以通过迭代拟合，将雨滴谱的3PNG模型（$N_{E\Gamma}$）拟合到雨滴谱仪数据（N_{ED}）提供的雨滴谱。因此，得到了一组参数（N_w, D_0, μ），可用于表示整个记录时间段（2020年9月22日至2022年5月11日）内每个（未省略的）数据分钟对应的3PNG模型雨滴谱（$N_{E\Gamma}$）。图 12 显示了对任意选取的一组数据进行拟合的结果：3PNG模型雨滴谱（$N_{E\Gamma}$）（红线）与雨滴谱仪数据（N_{ED}）提供的雨滴谱直方图。图 13 中的直方图汇总了所有参数。

3.3 从归一化伽马分布雨滴谱中解析求取雷达参数

到目前为止，雨滴谱仪采集的数据已完成筛选和验证，并用于生成雨滴谱（N_{ED}）。随后，该雨滴谱被用于拟合雨滴谱的三参数归一化伽马模型（3PNG模型）（$N_{E\Gamma}$）。本节将使用解析方程，以3PNG模型的雨滴谱为输入，推导雷达参数的估计值。之后（在第3.4节中），这些雷达参数将用于拟合模型，以估计参数化的3PNG模型（$\widehat{N_w}$、 $\widehat{D_0}$、 $\hat{\mu}$）（方法结构概览见图5）。文献中提供了多种成熟的解析关系，可用于根据雨滴谱推导雷达参数。例如，Bringi 和 Chandrasekar（2001）给出了以下关系式（公式12、13和14），用于计算水平或垂直反射率（$Z_{h,\ v}$）、差分反射率（Z_{dr}）和比差分相移（K_{dp}）。

$$Z_{h,v} = \frac{4\lambda^4}{\pi^4 |K_w|^2} \int_{D_{min}}^{D_{max}} \left|S_{hh,vv}(D)\right|^2 N(D) dD \quad [mm^6 m^{-3}] \qquad 12$$

$$Z_{dr} = 10 \log_{10} \frac{Z_h}{Z_v} \quad [dB] \qquad 13$$

$$K_{dp} = 10^{-3} \frac{180}{\pi} \lambda \Re \left\{ \int_{D_{min}}^{D_{max}} [f_{hh}(D) - f_{vv}(D)] N(D) dD \right\} \qquad 14$$

其中，D 为雨滴等效体积直径，N（D）为雨滴粒径分布，D_{max} 和D_{min} 分别为粒径分布中的最大和最小直径，λ 为雷达波长，S（D）为雨滴的复散射矩阵，$S_{hh,\ vv}$（D）为S的后向散射共极化分量，R表示复数的实部，f_{hh} 和f_{vv} 分别为水平和垂直极化的前向散射振幅。

此外，Bringi 和 Chandrasekar (2001) 还给出了线性退偏振比（LDR_{vh}）以及 Z_h 与 Z_v 测量值之间相关系数（ ρ_{hv}）的方程。线性退偏振比用于衡量电磁信号在与目标降水粒子相互作用后，后向散射信号的极化状态发生了多大变化。线性退偏振比是一个介于零和一之间的无量纲量，其中零表示后向散射辐射完全未极化，一表示后向散射完全极化。通常，各向异性目标（例如冰晶）的线性退偏振比较高（Bringi 和 Chandrasekar, 2001）。

对于雨而言，Z_h 与Z_v 测量值之间的相关系数（ρ_{hv}）往往接近于一。然而，降雨越强，雨滴就越趋于扁椭球形。因此，Z_h 会大于 Z_v ，相关性也随之降低。对于其他非液态水凝物（例如雪、冰和融雪），相关性下降得更快（Bringi 和 Chandrasekar, 2001）。最后一个参数是比衰减（A_H），本文采用 Gu 等（2011）给出的形式。雷达信号衰减通常发生在强降雨期间，并会影响使用较高频率（例如 >5GHz）的雷达测量。

3.4 用于根据雷达参数估计伽马模型三个参数的模型

本节介绍利用第3.3节中建立的模拟雷达参数（Z_h, $Z_{dr}, K_{dp}, A_H, LDR_{vh}, \rho_{hv}$）进行训练、以估计3PNG模型参数（$\widehat{N}_w, \widehat{D}_0, \hat{\mu}$）的模型。这些模型包括解析方程（ANs）、神经网络（NNs）和决策树（DTs）。与Vulpiani 等人（2009年；2006年）一致，将使用模拟雷达数据中随机抽取的 70%来训练模型，其余30%用于测试模型。训练和测试过程将重复100次，以考虑模型等效性，并实现对模型结构的交叉验证（Beven，2006年）。

与解析方程不同，解析方程所用输入参数的选择取决于具体方程，而机器学习模型能够使用范围广泛的输入参数（或“特征”）。因此，提出了三种不同的“特征集”，以适应现实应用中雷达产品可用性和可靠性的差异。具体如下：

- Feature set 1: $\{Z_h, Z_{dr}, K_{dp}, A_H, LDR_{vh}, \rho_{hv}\}$
- Feature set 2: $\{Z_h, Z_{dr}, K_{dp}\}$
- Feature set 3: $\{Z_h, Z_{dr}\}$

水平极化雷达反射因子（Z_h）和差分反射率（Z_{dr}）常用于RSD参数反演（Gorgucci 等人，2002年）。此外，Vulpiani 等人（2006）指出，比差分相移（K_{dp}）是雨滴谱的一个合理预测因子，但其计算往往存在噪声，因此可能对结果产生不利影响。因此，与Vulpiani 等人（2006）一致，特征集2和3分别代表K_{dp} 可靠与不可靠时的选项。最后，特征集1扩展了分析范围，以评估更广泛的特征如何影响结果。

3.4.1 解析方程

给定某些天气雷达参数（Z_h、Z_{dr}、K_{dp}），Gorgucci 等人（2002）证明，可以估算 3PNG模型的参数（$\widehat{N}_w$、$\widehat{D}_0$、$\hat{\mu}$）。例如，中值体积直径（D_0）可由方程 15、16（Gorgucci 等人，2002）和 17（Bringi 等人，2002；Gorgucci 等人，2001）求得：

- $\widehat{D}_0 = AN_1(Z_h, Z_{dr}) = a_1 Z_h^{b_1}(\xi_{dr})^{c_1}$ [mm] 15
- $\widehat{D}_0 = AN_2(K_{dp}, Z_{dr}) = a_2 K_{dp}^{b_2}(\xi_{dr})^{c_2}$ [mm] 16
- $\widehat{D}_0 = AN_3(Z_{dr}) = a_3 Z_{dr}^{b_3}$ [mm] 17

此处，系数 a_i、b_i、c_i 通过将方程与测试数据进行曲线拟合来确定；$\xi_{dr} = 10^{0.1 Z_{dr}}$ 是线性尺度下的差分反射率，Z_{dr} 是以分贝表示的差分反射率，Z_h 是水平极化下的反射率因子（$mm^6 m^{-3}$）。3PNG模型的截距参数（N_w）由方程 18 和 19（Gorgucci 等人，2002）给出：

- $log_{10}\widehat{N}_w = AN_4(Z_h, Z_{dr}) = a_4 Z_h^{b_4}(\xi_{dr})^{c_4}$ 18
- $log_{10}\widehat{N}_w = AN_5(K_{dp}, Z_{dr}) = a_5 K_{dp}^{b_5}(\xi_{dr})^{c_5}$ 19

最后，方程 *20* 给出了形状参数（μ）（Gorgucci 等人，2002）：

- $\hat{\mu} = AN_6(\widehat{D}_0, Z_{dr}) = \frac{a_6(\widehat{D}_0)^{b_6}}{(\xi_{dr}-1)} - c_6(\xi_{dr})^{d_6}$ *20*

此处，$\widehat{D}_0$ 是通过方程 15、16 和 17 得到的中值体积直径估计值。因此，根据所使用的具体中值体积直径估计值，方程 20 有三个版本。

3.4.2 神经网络与决策树

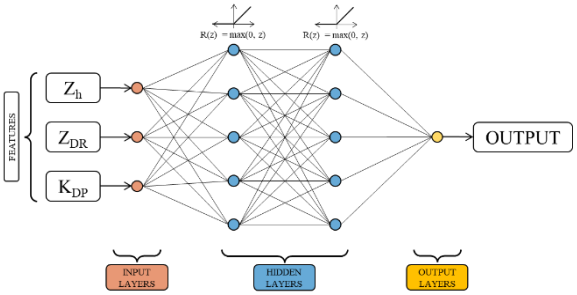


图14：具有三个特征（输入）的双隐藏层神经网络回归模型结构。图中展示了 ReLU 激活函数。

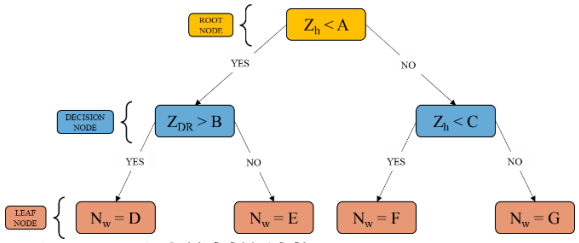


图15：四叶决策树的结构。根节点 ，决策节点 ，和叶节点均已标注，并用颜色区分。

图 13 展示了一个特定的回归神经网络（NN），该网络包含两个隐藏层、三个输入参数（称为特征）和一个输出节点。不过，本研究将采用范围更广的神经网络结构；具体而言，隐藏层数量有三种选择 {1,2,3}，每个隐藏层中的神经元数量有五种选择 {10,20,30,40,50}，并使用三组特征。这些选项组合对应 45 种独特的神经网络结构。因此，与采用 Sigmoid 激活函数的 Vulpiani 等人（2009；2006年）的研究不同，本研究需要使用计算效率更高（即速度更快）的修正线性单元（ReLU）激活函数（ Goodfellow 等人，2016）。每种神经网络结构都分别拟合到三组特征中的每一组伽马模型的参数重复 100 次：这相当于 13500 个神经网络。方程 21 描述了一个神经网络：

$$\hat{D}_0 = NN_{L,N}(Feature\ set) \quad 21$$

这里，L表示层数，N表示神经元数量，方程左侧表示输出，函数的自变量则是用于训练神经网络模型的特征集。

与神经网络模型一样，决策树（DT）也将具有多种结构。具体而言，其叶节点数量将有所不同 {4,12,36}。图15展示了一棵任意决策树。决策树模型将使用与神经网络模型相同的特征集。方程 22描述了一棵决策树：

$$\hat{D}_0 = DT_L(Feature\ set) \quad 22$$

此处，L表示叶节点的数量。

4. 结果

截至目前，已获得雨滴谱（N_{EF}）的3PNG模型，并利用该模型以解析方式计算出模拟雷达参数值。随后，利用这些雷达参数，从模拟雷达参数中随机抽取的 70%数据分别训练了大量解析模型、神经网络模型和决策树模型，每种模型重复100次。其余30%的数据留作测试这些模型之用。总计生成了17000个独特模型。

图16展示了根据雨滴谱仪数据（N_w、 D_0、 μ）获得的3PNG参数与使用1层、10个神经元，并将集合1作为输入特征——即$NN_{1,10}$（Z_h、 Z_{dr}、 K_{dp}、 A_H、 LDR_{vh}、 ρ_{hv}）——的神经网络模型获得的3PNG参数（$\hat{N}_w$、 $\hat{D}_0$、 $\hat{\mu}$）之间的散点图。需要特别指出的是，这些输出参数（$\hat{N}_w$、 $\hat{D}_0$、 $\hat{\mu}$）来自测试阶段，而非训练阶段。在测试阶段提供给模型的特征集，是未用于训练模型的30%数据。正如预期，如果将用于训练模型的70%特征重新输入模型，输出参数通常会更接近真实参数。

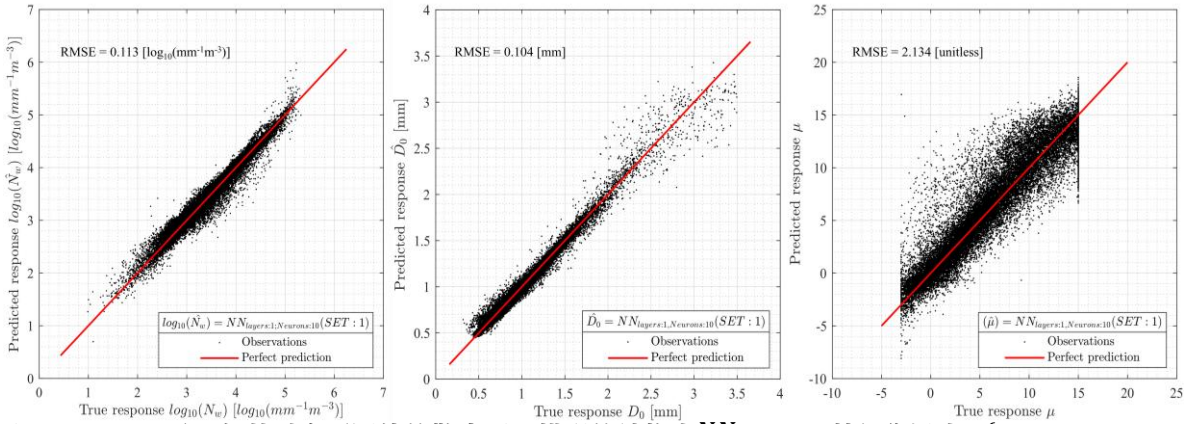


图16：N_w、 D_0和μ与其对应预测值的散点图。模型的结构为$NN_{L1,N10}$， 特征集设为1 $\{Z_h, Z_{dr}, K_{dp}, A_H, LDR_{vh}, \rho_{hv}\}$。

图16中的结果似乎表明，该模型在近似归一化雨滴浓度参数（N_w）和中值体积直径（D_0）的“真实”值方面相当有效，但在近似伽马形状参数（μ）方面效果较差。对于（以10为底的对数）N_w 散点图，该模型似乎具有很强的预测效用。然而，仔细观察会发现，模型难以预测较高的 N_w 值（即大于5的值）。散点图显示，数据呈现向上趋势，表明预测值（$\hat{N}_w$）高于“真实”的N_w 值。尽管如此，约33962个数据点中的绝大多数似乎都接近完美预测线。

同样，对于中值体积直径（D_0），模型似乎能够较好地预测较低的D_0 值，但在面对较高的D_0 值（>2毫米）时，预测能力较弱。在这种情况下，尽管数据点分布在完美预测线的两侧，但整体趋势似乎向下。这表明模型系统性地低估了较高的 D_0 值。与前文一样，值得注意的是，这只占33962个数据点中的少数。与 Bringi等人（2003年）和 Vulpiani 等人（2006年）的发现一致，使用该模型准确预测伽马形状参数（μ）更具挑战性。

μ 的“真实”值被限制在 -3 至 15 之间。这是因为，在自然雨滴谱中，超出该范围的值并不现实（Bringi 等人，2003年）；然而，预测值并未体现这一边界。

图16展示了所训练的17000个模型中的1个。为了适当地比较这大量的模型，需要进行更加定量且严格的分析。该分析如图17所示。与之前一样，这些结果仅针对测试阶段（而非训练阶段）。

如图17所示，解析模型（AN）的结果准确度稳定，这一点可从均方根误差值的范围较窄看出。结果表明，与 $AN_5(K_{dp}, Z_{dr})$ 相比， $AN_4(Z_h, Z_{dr})$ 对 N_w 的预测更准确，因此对于这一特定解析方程，预测 N_w 时，Z_{dr} 是比 K_{dp} 更有用的雷达参数。至于 D_0，结果显示 $AN_1(Z_h, Z_{dr})$ 和 $AN_2(K_{dp}, Z_{dr})$ 的结果相近，但两者的表现都不及 $AN_3(Z_{dr})$。最后，对于 μ，$AN_{6A}(\hat{D}_0, Z_{dr})$ 和 $AN_{6B}(\hat{D}_0, Z_{dr})$ 模型的表现大致相同。这或许并不令人意外，因为它们分别将性能相近的 AN_1 和 AN_2 的输出用作输入特征（分别）。有趣的是，与 AN_{6A} 和 AN_{6B} 相比，使用表现更出色的 AN_3 模型给出的 $\hat{D}_0$ 值的 $AN_{6C}(\hat{D}_0, Z_{dr})$ 表现较差。

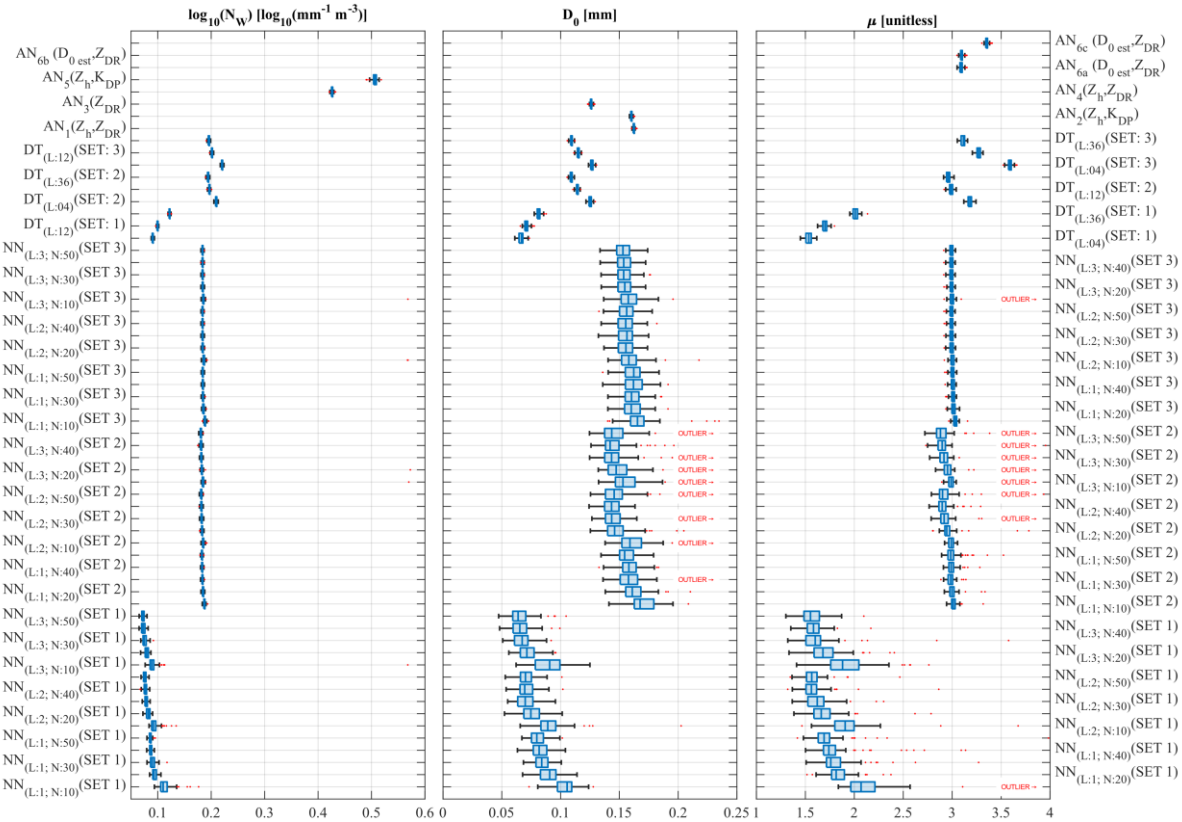


图17：各模型结构和特征集在测试过程中的均方根误差串行箱线图。每个箱线图由100个数据点构成，代表训练和测试过程的迭代。每一行对应一个模型，标签显示在左右两侧。

如图17所示，在预测 N_w、 D_0 和 μ 时，使用特征集1（SET1）的神经网络模型，其RMSE中位数始终低于使用SET2或SET3的神经网络。此外，对于使用SET1的神经网络，随着层数和神经元数量的增加，RMSE中位数往往会降低。这表明，在提供全部输入特征时，结构复杂度更高的模型表现更好。

对于使用 SET2 的神经网络，N_w 的均方根误差范围比 SET1 的结果更窄。此外，与 SET1 不同，随着模型结构复杂度的增加，预测性能提升微乎其微。例如，$NN_{L:1,\ N:10}$ 的均方根误差中位数为 0.184，而 $NN_{L:3,\ N:50}$ 的均方根误差中位数为 0.187。在预测 D_0 和 μ时，使用 SET2 的模型所得均方根误差比使用 SET1 的模型一致性更差。这可能是由与 K_{DP}相关的噪声造成的。此外，增加 SET2 神经网络的层数和节点数，降低了预测D_0 和 μ 时的均方根误差中位数。然而，结构更复杂的 SET2 神经网络产生了大量 D_0 和 μ 均方根误差异常值，这可能是过拟合所致。

在预测 N_w 时，SET3 神经网络的表现与 SET2 神经网络相似；在预测 D_0 和 μ时，其一致性高于 SET2。随着复杂度增加，SET3 神经网络仅有极其微小的改进。最后，尽管所有特征集和结构对 N_w 的预测都相当一致，但一些具有 2–3 层、每层 10–20 个神经元的神经网络模型，偶尔会出现远高于该模型其他均方根误差的极高均方根误差。

使用 SET1 的决策树（DT）在全部三个参数上的 RMSE 中位数均低于使用 SET2 或 SET3 的决策树（见图17）。此外，随着叶节点数量（即结构复杂度）的增加，使用 SET1 的决策树性能往往会下降。例如，在预测 μ 时， $DT_{L:04}(SET1)$ 的 RMSE 中位数为 1.535，而结构更复杂的 $DT_{L:36}(SET1)$ 的 RMSE 中位数为 2.010。相比之下，使用 SET2 或 SET3 预测任一参数的决策树，在叶节点数量从 4 增加到 36 时，结果均有所改善。决策树的结果高度一致，没有极端离群值。此外，输入集的选择似乎对 RMSE 结果的离散程度影响甚微。

总之，机器学习模型通常优于解析模型。具体而言，使用 SET1 的机器学习模型表现最佳。使用 SET2 的机器学习模型的 RMSE中位数也略低于使用 SET3 的模型，但这些神经网络模型产生的极端离群值数量要多得多。在机器学习模型中，神经网络模型和决策树模型都展现出良好的预测效用。总体而言，神经网络模型结果的一致性低于决策树模型。然而，没有任何单一模型在预测某个特定雨滴谱参数时都始终表现更好。同样，改变模型结构有时会显著提升模型性能，有时则会明显降低其预测效用。因此，不能断定某种结构或模型类型更优。例如，具有 36 个叶节点的决策树在使用 SET3 准确获取D_0 方面表现最佳，而神经网络则更擅长使用 SET1 找到 N_w 。与 Bringi等人（2003年）和 Vulpiani 等人（2006年）的既往研究一致， μ 比N_W 和D_0更难准确预测。

5. 讨论

近年来，已有多项研究采用解析关系、神经网络和随机森林等不同技术，开发用于反演雨滴谱参数的模型。然而，尚无研究对这些不同方法进行全面比较。此外，也尚无研究如此全面地评估用于雨滴谱反演的各类雷达产品和模型结构，或通过重复分析对模型变异性开展如此严格的分析。

就本次讨论而言，我们选取了图17所示分析中性能最佳的模型，以便与文献中的模型（表1）进行比较。我们将忽略使用特征集2（SET2）的模型，因为K_{dp} 似乎仅以显著牺牲模型一致性为代价来提升性能。

	SET1模型	Q1	Q2	Q3
$\boldsymbol{log_{10}N_w}$ (毫米$^{-1}$米$^{-3}$)	$NN_{L3,N50}$	0.070	0.072	0.074
$\boldsymbol{D_0}$ (mm)	DT_{L4}	0.065	0.067	0.068
μ （无量纲）	DT_{L4}	1.505	1.535	1.560
	SET3 模型	Q1	Q2	Q3
$\boldsymbol{log_{10}N_w}$ (毫米$^{-1}$米$^{-3}$)	$NN_{L1,N30}$	0.184	0.185	0.185
$\boldsymbol{D_0}$ (mm)	DT_{L36}	0.109	0.109	0.112
μ (无量纲)	$NN_{L2,N30}$	2.978	2.993	3.008

表1：使用特征集（FS）1 和 3 的性能最佳模型的 RMSE 第 1、2 和 3 四分位数。

Vulpiani 等人（2006）开发的神经网络使用了Z_h、Z_{dr}和K_{dp} 的测量值来反演D_0 和N_w 参数。还使用了 Z_{dr} 的测量值以及（先前估计的）D_0 参数来反演形状参数μ。研究人员所使用的模型具有单个隐藏层，其中用于D_0 和N_w的神经元为6个，用于μ的神经元为12个。在论文中，研究人员指出，其模型“相对于文献中所述的其他估计雨滴谱参数的方法已显示出改进”。Vulpiani的模型对D_0 （0.29mm）和N_w （0.28 mm^{-1}m^{-3}）的均方根误差高于表1中的相应模型，但对μ （1.27）的均方根误差更低。本文的研究结果表明，

Vulpiani（$NN_{L1,N6\ or\ 12}$）所使用的神经网络并非最优结构。此外，Vulpiani使用了K_{dp}，而这似乎并未提升性能；相反，该模型本可以从使用更广泛的特征中获益，类似于SET1中的那些特征。最后，均方根误差表明，Vulpiani所采用的利用Z_{dr} 和D_0 预测μ 的技术可能是一种更优的方法。

Pasarescu（2020）也开发了一种基于神经网络的模型，用于根据天气雷达测量数据（Z_h、Z_{dr} 和 K_{dp}）反演雨滴谱参数。她采用了一个单隐藏层神经网络，隐藏层中有 22 个神经元。asarescu 的模型对 D_0, 0 的均方根误差为 0.09（mm），对 N_w 的均方根误差为 0.26（$mm^{-1}m^{-3}$），对 μ 的均方根误差为 1.57。本研究的结果表明，在估算 D_0，方面，asarescu 的模型优于 Vulpiani 的模型，其均方根误差几乎小了三倍。然而，μ 的均方根误差大于 Vulpiani 的模型。尽管如此，本文显示，当输入参数经过迭代随机化时，均方根误差往往会出现显著变化。因此，并非总能严谨地断定某个模型优于另一个模型。如果无法使用 SET1，Pasarescu 的方法本可通过采用 SET1 或 SET3 得到改进。此外，对于（例如）D_0，的预测，相较于决策树（DT），她采用神经网络模型的方法似乎不够可靠。

Conrick 等（2020）使用解析方法和决策树集成（即随机森林），从极化天气雷达测量数据（Z_h、Z_{dr}）中反演中值体积直径D_0 及其他参数。在反演 D_0, 时，解析方法所得结果的均方根误差为1.75（mm），而随机森林则降至0.15（mm）。本文支持Conrick的论断，即决策树在预测雨滴谱方面展现出颇具前景的能力，因为用于预测D_0 的两个表现最佳的模型均被确定为决策树。

在选择模型时，一个更广泛的考量因素是可用的计算能力。例如，复杂的神经网络所需的训练时间和数据量显著多于解析回归模型。

6. 结论

总之，在获取雨滴谱参数方面，机器学习方法通常比解析方法更准确。然而，没有一种机器学习模型在预测雨滴谱参数时始终表现出更高的能力。相反，机器学习模型的选择及其结构应取决于可获取的雷达数据范围以及模型所要估算的具体参数。例如，DT_{L4} 在使用 SET1 数据时，已展现出相当强的预测D_0 能力。用于训练模型的数据被发现会显著影响模型准确预测雨滴谱参数的能力。通常，SET1 的表现最佳。然而，如果无法获得 SET1 数据，SET3 得出的结果与 SET2 相近，但一致性更高。

7. 局限性与未来工作建议

本研究采用的机器学习方法假设雷达测量不存在误差。然而，已有充分文献表明，雷达反射率参数（包括Z_{dr} 和K_{dp}）的测量值存在噪声。未来工作可向导出的雷达数据中引入噪声，以模拟这一现象。此外，用于训练模型的数据来自英国的某一特定地点。本研究可通过提供来自不同气候区的数据加以扩展，从而可能提高泛化能力。最后，文献将受益于一项类似研究：该研究检验相当范围的结构与参数，但此次使用来自广泛且多样地点的业务天气雷达，以检验不同模型的普适性。

8. 参考文献

Adirosi, E., Gorgucci, E., Baldini, L. 和 Tokay, A.（2014）《通过比较实测降雨率与雷达等效伽马雨滴尺寸分布的降雨率，评估伽马雨滴尺寸分布假设》，应用气象学与气候学杂志，53(6)，第1618-1635页。美国气象学会（2012）翻斗式雨量计：美国气象学会。可访问：https://glossary.ametsoc.org/wiki/Tipping- bucket_rain‾gauge。Atlas, D., Srivastava, R. C. 和 Sekhon, R. S.（1973）《垂直入射降水的多普勒雷达特征》，地球物理学评论，11(1)，第1-35页。Beven, K.（2006）《等效性论点宣言》，水文学杂志，320(1-2)，第18-36页。Brandes, E. A., Zhang, G. 和 Vivekanandan, J.（2004）《极化雷达雨滴尺寸分布反演算法的比较》，大气与海洋技术杂志，21(4)，第584-598页。Bringi, V., Chandrasekar, V., Hubbert, J., Gorgucci, E., Randeu, W. 和 Schoenhuber, M.（2003）《基于雨滴谱仪和双偏振雷达分析的不同气候区雨滴粒径分布》，《大气科学杂志》，60(2)，第354-365页。Bringi, V. N. 和 C handrasekar, V.（2001）《极化多普勒天气雷达：原理与应用》。剑桥大学出版社。Bringi, V. N., Huang, G.-J., Chandrasekar, V. 和 Gorgucci, E.（2002）《利用极化雷达数据估算伽马雨滴尺寸分布模型参数的方法：应用于TRMM/巴西观测活动中的飑线事件》，大气与海洋技术杂志，19(5)，第633-645页。Bringi, V. N., Rico-Ramirez, M. A. 和 Thurai, M.（2011）《利用英国业务型极化C波段雷达估算降雨：与雨量计网络的比较及误差分析》，水文气象学杂志，12(5)，第935-954页。Caracciolo, C., Napoli, M., Porcù, F., Prodi, F., Dietrich, S., Zanchi, C. 和 Orlandini, S.（2012）《雨滴粒径分布与土壤侵蚀》，灌溉与排水工程杂志，138(5)，第461-469页。

Ciach, G. J.（2003）《翻斗式雨量计测量中的局部随机误差》，大气与海洋技术杂志，20(5)，第752-759页。Cluckie, I. D. 和 Rico-Ramirez, M. A.（2004）《天气雷达技术及未来发展》，国际水文科学协会出版物——会议录与报告系列，289，第11-22页。Conrick, R., Zagrodnik, J. P. 和 Mass, C. F.（2020）《利用随机森林回归对太平洋西北沿海地区雨滴粒径分布参数进行双偏振雷达反演》，大气与海洋技术杂志，37(2)，第229-242页。Cybenko, G.（1989）《通过S形函数的叠加进行逼近》，《控制、信号与系统数学》，2(4)，第303-314页。Dai, Q., Zhu, J., Zhang, S., Zhu, S., Han, D. 和 Lv, G.（2020）《基于WRF推导的雨滴粒径分布估算降雨侵蚀力》，*Hydrol. Earth Syst. Sci.*，24(11)，第5407-5422页。Duchon, C. E. 和 Biddle, C. J.（2010）《高降雨强度事件中翻斗式雨量计的少计问题》，*Adv. Geosci.*，25，第11-15页。英国环境署（2021）英国环境署降雨*API*。可访问：https://environment.data.gov.uk/flood-monitoring/doc/rainfall。Goodfellow, I., Bengio, Y. 和 Courville, A.（2016）《深度学习》。麻省理工学院出版社，第190页。Gorgucci, E., Chandrasekar, V., Bringi, V. N. 和 Scarchilli, G.（2002）《利用极化雷达测量估算雨滴粒径分布参数》，《大气科学杂志》，59(15)，第2373-2384页。Gorgucci, E., Scarchilli, G., Chandrasekar, V. 和 Bringi, V. N.（2001）《基于极化雷达测量的降雨估算：不受雨滴形状—尺寸关系变化影响的复合算法》，大气与海洋技术杂志，18(11)，第1773-1786页。Gu, J.-Y., Ryzhkov, A., Zhang, P., Neilley, P., Knight, M., Wolf, B. 和 Lee, D.-I.（2011）《C波段强降雨中的极化衰减订正》，应用气象学与气候学杂志，50(1)，第39-58页。Habib, E., Krajewski, W. F. 和 Kruger, A.（2001）《翻斗式雨量计测量的抽样误差》，《水文工程杂志》，6(2)，第159-166页。Hornik, K., Stinchcombe, M. 和 White, H.（1989）《多层前馈网络是通用逼近器》，神经网络，2(5)，第359-366页。Humphrey, M. D., Istok, J. D., Lee, J. Y., Hevesi, J. A. 和 Flint, A. L.（1997）《翻斗式雨量计自动动态校准的新方法》，大气与海洋技术杂志，14(6)，第1513-1519页。Kathiravelu, G., Lucke, T. 和 Nichols, P.（2016）《雨滴测量技术：综述》，*Water*，8(1)，第29页。Löffler-Mang, M. 和 Joss, J.（2000）《用于测量水凝物尺寸和速度的光学雨滴谱仪》，大气与海洋技术杂志，17(2)，第130-139页。Luyckx, G. 和 Berlamont, J.（2001）《修正翻斗式雨量计降雨测量值的简化方法》，城市排水建模，第767-776页。Marshall, J. S. 和 Palmer, W. M. K.（1948）《雨滴粒径分布》，《气象学杂志》，5(4)，第165-166页。Montopoli, M., Marzano, F. S. 和 Vulpiani, G.（2008）《基于雨滴谱仪数据分析与合成雨滴粒径分布时间序列》，*IEEE*地球科学与遥感汇刊，46(2)，第466-478页。Nemec, J.（1969）《降水与蒸发的仪器、观测方法及基础数据处理。非洲水文气象仪器、观测与网络，DA Davies 编》，世界气象组织，第11-61页。Nemeth, K.《OTT Pluvio2：称重式降水计及降水测量技术的进展》。世界气象组织气象与环境仪器及观测方法技术会议。Nešpor, .和Sevruk, .（1999）‘利用数值模拟对雨量计测量进行感应误差估计’，大气与海洋技术杂志，16(4)，第450–464页。Nystuen, J. A.、Proni, J. R.、Black, P. G.和Wilkerson, J. C.（1996）‘自动雨量计的比较’，大气与海洋技术杂志，13(1)，第62–73页。Pasarescu, A. C.（2020）利用神经网络进行雨滴粒径分布估计。布里斯托大学，布里斯托，英国。Seliga, T. A.和Bringi, V. N.（1976）‘正交极化雷达差分反射率测量在降水测量中的潜在应用’，应用气象学与气候学杂志，15(1)，第69–76页。Sokol, Z.、Szturc, J.、Orellana-Alvear, J.、Popová, J.、Jurczyk, A.和Célleri, R.（2021）‘天气雷达在降雨估算中的作用及其在气象与水文建模中的应用——综述’，遥感，13(3)，第351页。Strangeways, I.（2000）‘降水’，载于Strangeways, I.（编）自然环境测量。剑桥：剑桥大学出版社，第120–160页。Tang, Q.、Xiao, H.、Guo, C.和Feng, L.（2014）‘中国北方和南方雨滴粒径分布特征及其反演的极化雷达参数’，大气研究，135–136，第59–75页。Testud, J.、Oury, S.、Black, R. A.、Amayenc, P.和Dou, X.（2001）‘用于描述雨滴谱的“归一化”分布概念：云物理学与云遥感的工具’，应用气象学杂志，40(6)，第1118–1140页。Tilg, A. M.、Skrzypiński, W. R.、Hannesdóttir, Á.和Hasager, C. B.（2022）‘考虑前缘侵蚀时雨滴粒径参数化与降雨量对叶片寿命计算的影响’，风能，25(5)，第952–967页。Tokay, A.和Short, D. A.（1996）‘来自热带雨滴谱的证据：层状云与对流云降雨的起源’，应用气象学与气候学杂志，35(3)，第355–371页。Tokay, A.、Wolff, D. B.和Petersen, W. A.（2014）‘新型激光光学雨滴谱仪OTT Parsivel 2的评估’，大气与海洋技术杂志，31(6)，第1276–1288页。Tokay, A.、Wolff, D. B.、Wolff, K. R.和Bashor, P.（2003年）‘基斯地区微物理项目（KAMP）期间的雨量计与雨滴谱仪测量’，大气与海洋技术杂志，20(11)，第1460–1477页。Ulbrich, C. W.（1983）‘雨滴粒径分布解析形式的自然变化’，气候与应用气象学杂志，22(10)，第1764–1775页。Vulpiani, G.、Giangrande, S.和Marzano, F. S.（2009）‘基于极化S波段雷达测量的降雨估算：神经网络方法的验证’，应用气象学与气候学杂志，48(10)，第2022- 2036页。Vulpiani, G.、Marzano, F. S.、Chandrasekar, V.、Berne, A.和Uijlenhoet, R.（2006年）‘利用正则化人工神经网络进行极化天气雷达反演雨滴粒径分布’，*IEEE*地球科学与遥感汇刊，44(11)，第3262–3275页。Zhang, G.、Vivekanandan, J.和Brandes, E.（2001）‘一种根据极化雷达测量估计降雨率与雨滴粒径分布的方法’，*IEEE*地球科学与遥感汇刊，39(4)，第830–841页。