

基于机器学习的双参数反演：利用固定偏移GPR数据估算雪层含水量与密度

Zohaib AKBAR^{1,2}, 姜元俊^{1,2*}, Ryan WEBB³, Anja KLOTZSCHE⁴, 朱元甲^{1,2}, Aftab ANWAR^{1,2}, Muhammad Mudassar REHMAN^{1,2}

1. 山地自然灾害与工程安全重点实验室/中国科学院、水利部成都山地灾害与环境研究所，成都 610213

2. 中国科学院大学，北京 100049

3. University of Wyoming, Laramie, WY 82071, USA

4. Agrosphere (IBG-3), Institute of Bio- and Geosciences, Forschungszentrum Jülich GmbH, Jülich 52425, Germany

* 通讯作者，E-mail: yjjiang@imde.ac.cn

2025-04-13收稿，2025-09-24修回，2025-11-12接受，2025-12-17网络版发表

国家重点研发计划项目(2023YFC3008300、2023YFC3008305)、国家自然科学基金项目(42172320)、山地自然灾害与工程安全重点实验室/中国科学院、水利部成都山地灾害与环境研究所自主部署项目(KLMHER-Z06、KLMHER-T07)和中国科学院、水利部成都山地灾害与环境研究所自主部署科研项目(IMHE-CXTD.04)资助

摘要 准确评估雪层体积含水量和体积密度对于理解雪的水文过程、降低雪崩风险以及冰冻圈监测具有重要意义。本文提出一种新型双参数反演框架，该框架集成了合成电磁建模、降维方法和机器学习算法，用于从探地雷达(GPR)数据中提取相对介电常数和对数电阻率。传统雪层测量方法具有侵入性、劳动强度大且仅限于点位观测等局限性。为克服上述局限，建立了一种非侵入性、可扩展且数据驱动的框架，利用合成GPR数据集来表示具有不同含水量和密度分布的多样化雪层条件。使用先进电磁模拟器gprMax，通过有限差分时域模拟生成合成的一维时序反射(A扫描)。随后采用主成分分析(PCA)将每个A扫描进行压缩，得到低维且信息保真的特征集，从而显著提升模型训练效率。基于经过主成分分析(PCA)降维处理的特征集，训练了随机森林、神经网络、支持向量机和极限梯度提升四种机器学习模型。其中，神经网络模型性能最佳，介电常数为 $R^2>0.97$ ，电阻率为 $R^2>0.92$ 。合成数据中引入高斯噪声(信噪比约为6 dB)，并通过针对特定领域进行改进，以提高其在实地数据的泛化能力。模型在中国阿尔泰山脉的两条典型GPR剖面(湿雪T750和饱和雪G125)上进行了验证。神经网络模型预测结果与GPR反演、Snowfork测量及人工雪坑数据高度一致，体积含水量偏差不超过1.5%，体积密度误差在30–84 kg m⁻³范围内。结果表明，基于机器学习的反演方法在结合真实模拟与数据增强的条件下，能够实现可扩展、非侵入式的雪层特性反演，在水文预报、雪情监测及气候敏感型水资源管理中具有重要应用价值。

关键词 雪层，GPR，gprMax，机器学习，反演

1 引言

季节性积雪是地球能量和水平衡的关键组成部分，直接影响雪崩风险、水资源和生态系统动态(Mankin等, 2015)。在气候变化的背景下，积雪范围、持续时间和组成的变化对雪物理性质(如液态含水量LWC和体积密度

ρ_s)的表征提出了更高要求，因此，亟需改进观测和建模工具以精确描述雪层状态(Luce等, 2013; Sturm等, 2017; Coppola等, 2018; Mote等, 2018)。

在过渡期和湿雪条件下测量雪层特性仍然具有技术挑战性(McGrath等, 2019)。基于现场的测量方法通常依赖雪传感器，如Snowfork(Singh等, 2016; Hao等,

中文引用格式：Akbar Z, 姜元俊, Webb R, Klotzsche A, 朱元甲, Anwar A, Rehman M M. 2026. 基于机器学习的双参数反演：利用固定偏移GPR数据估算雪层含水量与密度. 中国科学：地球科学, 56(2): 584–603, doi: 10.1360/N072025-0134

英文引用格式：Akbar Z, Jiang Y, Webb R, Klotzsche A, Zhu Y, Anwar A, Rehman M M. 2026. Machine learning-based dual-parameter inversion for estimating snowpack liquid water content and density using common offset GPR data. Science China Earth Sciences, 69(2): 564–581, <https://doi.org/10.1007/s11430-025-1764-7>

2021)、Denoth设备(Denoth, 1997)和A2光子传感器(WISe)(Webb等, 2021),通过介电常数(ϵ')估计来推断雪的介电行为。然而,这些方法具有侵入性、劳动强度大且空间覆盖有限。为了克服这些限制,遥感平台已成为一种有前景的替代方案,但其空间分辨率较低,难以捕捉雪层的精细变化,且仍需在现场进行测量以校准和验证(Rott等, 2003; Techel和Pielmeier, 2011; Mortimer等, 2020; Jacobs等, 2021; Dharmadasa等, 2024)。

探地雷达(GPR)为雪层评估提供了一种可扩展且非侵入性的替代方法,可实现高分辨率测量、快速数据采集和可靠的地下剖面分析。其有效性已在雪地和冰川环境中的大量研究中得到验证(Tiuri等, 1984; Marshall, 2005; Granlund等, 2009; Bradford等, 2010; Sundström等, 2012; Holbrook等, 2016; McGrath等, 2022; Sidorenko等, 2023)。近期,通过将GPR与无人机(UAV)结合,其应用能力得到进一步扩展,可显著提升空间覆盖范围,实现快速调查,并在偏远及雪崩多发地区保证数据采集的安全性(Jenssen和Jacobsen, 2020; Vergnano等, 2022; Frolov等, 2024; Tjoelker等, 2024)。

在干雪中,可以使用经验模型根据雷达速度合理估算雪密度(Kovacs等, 1995; Lundberg等, 2000; Bonnell等, 2023)。然而,由于介电常数(ϵ')、对数电阻率(ρ)和电磁波衰减之间的非线性耦合,在湿雪环境(空气、冰、水)中解译GPR信号仍然非常复杂。介电常数决定电磁波速度,而电阻率控制传导损耗,两者共同导致信号衰减(Granlund等, 2009; Dupuy等, 2024)。经验模型,如广泛使用的Denoth和Tiuri方程(Denoth等, 1984; Tiuri等, 1984)以及三相混合公式(Heilig等, 2015; Schmid等, 2015),都依赖于简化的、特定站点假设,难以推广至不同雪层类型或含水状态。基于频率衰减技术对噪声高度敏感,需要干净、宽带信号,而这些标准在异质湿雪层中很少得到满足(Turner和Siggins, 1994; Webb等, 2021)。偏移方法,包括Kirchhoff(Moran等, 2000)、频率-波数域(Zhang等, 2021)和逆时偏移(Zhu等, 2020),可以恢复形状和层位等几何信息,但通常忽略振幅保真度,并且无法解析电磁参数(Sun等, 2022)。全波形反演(FWI)作为较先进的解译方式,可利用完整波场同时估计 ϵ' 和 $\log(\rho)$,但计算量大且由于反演本身的不确定性在噪声环境下易陷入局部最优解,稳定性不足(Klotzsche等, 2019; Qin等, 2023)。

这些挑战促使数据驱动的反演技术受到广泛关注。近年来,机器学习的进展为解决地球物理分析中的非线性问题提供了有力工具(Alzubaidi等, 2024; Liu等,

2024; Stephan等, 2024; Namdari等, 2025; Sharma等, 2025; Shehab等, 2025)。尽管如此,机器学习在季节性雪层中的应用仍然有限。Meehan等(2024)将探地雷达(GPR)、激光雷达(LiDAR)与机器学习相结合来估算雪的物理特性,但主要集中于体积密度,未能解决液态水含量(LWC)的估算,而LWC对于水文建模和雪崩预报至关重要。类似地,Webb等(2022)应用反演技术时,仅关注介电常数,忽略了电导率(σ)或对数电阻率(ρ)在湿雪中的关键作用,而其对信号衰减具有显著的影响。Xue等(2024)提出了一种基于深度学习的土壤GPR双参数反演框架,但该研究具有领域特异性,未能解决雪的复杂性。

实现雪层反演的高效监督式机器学习训练依赖于大规模、多样化且带标注的数据集。但受制于环境不确定性、传感器局限以及野外作业条件,仅凭实地测量难以获取此类数据。为克服这一限制,利用FDTD模拟器gprMax(Warren等, 2021)构建了涵盖多种雪层特性的合成数据集。该模拟能够在受控条件下,较准确模拟不同介电常数、电导率和层类的雪层中的雷达波传播行为。此外,主成分分析(PCA)已被证明在机器学习应用中对GPR数据的预处理具有重要价值(Giannakis等, 2024),可在保留重要特征的同时降低数据维度,是雪层研究中一项有价值的技术。Küçükdemirci和Sarris(2022)详细概述了人工智能在地球物理领域的应用,强调了机器学习在处理复杂地下数据中的日益重要性。然而,现有大多数雪层反演框架主要侧重于介电常数(ϵ')估算,往往忽略了对数电阻率(ρ)的重要性,尤其是在湿雪条件下。

这一研究空白凸显了开发能够联合估算介电常数(ϵ')和对数电阻率(ρ)的先进方法的必要性。为此,提出了一个基于机器学习(ML)的双参数反演框架,将物理上真实的合成模拟与数据驱动的建模相结合,实现对湿雪层的更准确、更全面地表征,从而突破传统方法的局限性,为基于探地雷达(GPR)的雪层分析提供可靠的解决方案。

本研究开发并评估了一种双参数数据驱动算法,用于从固定偏移GPR数据中反演雪层介电常数(ϵ')和对数电阻率(ρ)。研究目的:(1)生成合成雪层剖面以模拟不同雪层条件下的电磁特性;(2)在训练机器学习模型前对数据进行主成分分析(PCA);(3)利用实证GPR数据和现场测量数据验证机器学习模型的有效性。基于机器学习的算法可准确估算雪层的体积含水量和体积密度,为雪层特性表征提供重要的水文信息。

2 材料和方法

本研究遵循结构化的工作流程。首先,合成的雪层

模型可利用二维gprMax FDTD模拟器生成(Warren等, 2016), 随后采用主成分分析(PCA)对该数据精简, 以实现高效的机器学习(ML)训练。评估了一组非线性监督机器学习模型, 包括随机森林(RF)、神经网络(NN)、极限梯度提升(XGB)和支持向量机(SVM)。模型性能采用标准回归指标进行评估: 判定系数(R^2)、均方根误差(RMSE)、平均绝对百分比误差(MAPE)。为了提高模型的通用性并弥合合成数据与实际现场数据之间的域差距, 引入高斯噪声(信噪比约为6 dB), 并采用了数据增强技术。最终, 将表现最优的模型应用于实测数据进行验证。通过频域衰减分析、雪叉数据和人工雪坑观测获得的雪层物理数据为精准评估提供了基准。所提出的反演和测试框架概述如图1所示。

2.1 研究区域和现场数据收集

中国新疆北部的阿尔泰山山区(图2A)被认为是中国三大主要稳定雪区之一(Huang等, 2016)。该地区具有高纬度大陆性气候和崎岖地形, 积雪量受海拔、风向重分布和植被覆盖影响表现出剧烈空间变异(Ran等, 2012; Huang等, 2019)。新疆禾木(48°N, 87°E)地处中国、蒙古、哈萨克斯坦和俄罗斯四国交界处, 平均海拔1200 m, 冬季漫长且严寒, 平均气温在-10 °C至-15 °C之间, 积雪覆盖面积广泛(Yang等, 2020), 最大雪层深度约为1.5–2.0 m, 因此成为雪层研究的理想区域。

在野外采集期间采集的多条GPR剖面中, 选择两条特征性强的剖面(T750和G125)进行详细分析, 以覆盖不同海拔、降雪条件及雪层厚度(图2B)。T750剖面位于相对较低的海拔(约1118 m), 朝北阴坡, 显示中等雪密度和残余水分, 属于湿雪条件(0%–3%)。相比之下, G125横剖面位于较高海拔(约1320 m)的南向坡地, 太阳照射充足, 融化速度快, 雪层明显更湿润(3%–8%)。这种分层选址有助于在早春融雪期间对不同雪层条件下模型的泛化能力进行稳健测试。

雪层数据使用瑞典Impulse CO4080雷达获取, 雷达配备800 MHz地面耦合屏蔽天线。天线频率的选择旨在平衡浅层非均匀雪层的穿透深度和空间分辨率。系统采用实时采样(RTS)模式, 以确保信号的高保真度。发射源和接收器在屏蔽雪橇中相距9 cm, 以减少环境电磁干扰, 设备以人工牵引方式沿雪面滑动, 速度约1 m s⁻¹, 保持仪器与积雪接触面耦合(Webb等, 2018)。探地雷达数据采集速率为每道250个样本, 时间帧为15 ns, 用于捕捉预期雪层厚度, 每道间隔为0.45 s。通过机载和外置便携式GPS接收器记录位置, 从而对雷达图进行精确的地理配准。为了进行交叉验证并获得更可靠的雪层估算,

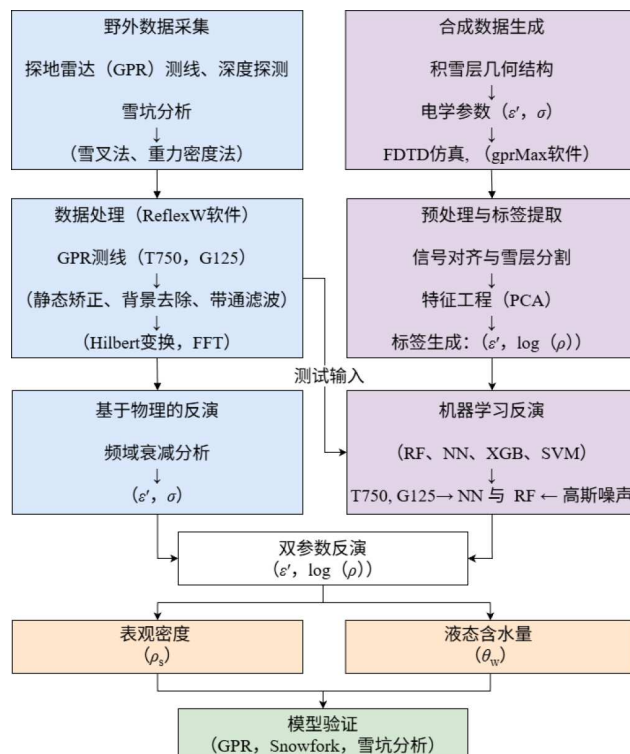


图1 雪层分析混合反演方法概述

除了GPR采集外, 还采用了人工技术。在垂直间隔10 cm的雪坑中获取雪叉介电探针读数, 以估算含水量和密度。此外, 还使用经过校准的800 cm³圆柱形雪采样器进行重量法密度测定, 并与雪叉估算值进行并行分层(图2E)。此外, 还记录了深度探测器读数的数量, 以估算沿探地雷达剖面的雪层厚度。这一综合策略为评估反演模型的预测结果提供了可靠的替代参考数据集。

2.2 GPR数据处理及衰减分析

利用ReflexW软件对GPR现场数据进行如下处理: (1) 时间零点校正, 以校准反射到达时间; (2) 背景去除; (3) 在400 MHz至1600 MHz之间进行带通滤波, 以提高信噪比(SNR), 同时保留相关的频谱内容(图3)。

雷达波速度(v)的测量方法是将双向传播时间(TWTT)与沿剖面的手动雪探头深度(h)关联起来。双向传播时间是通过在ReflexW软件处理后的雷达图中手动选取反射探地雷达信号的第一个到达时间获得的。速度计算公式为 $v=2h/(TWTT)$ (Marshall等, 2005)。然后根据速度计算介电常数的实部(ϵ') $\epsilon'=(c/v)^2$, 其中 c 表示真空中的光速(Neal, 2004)。

雪层复介电常数表示为 $\epsilon^*=\epsilon'+j\epsilon''$, 其中 ϵ' 和 ϵ'' 分别表示介电常数的存储和损耗分量。虚部(ϵ'')决定介电损

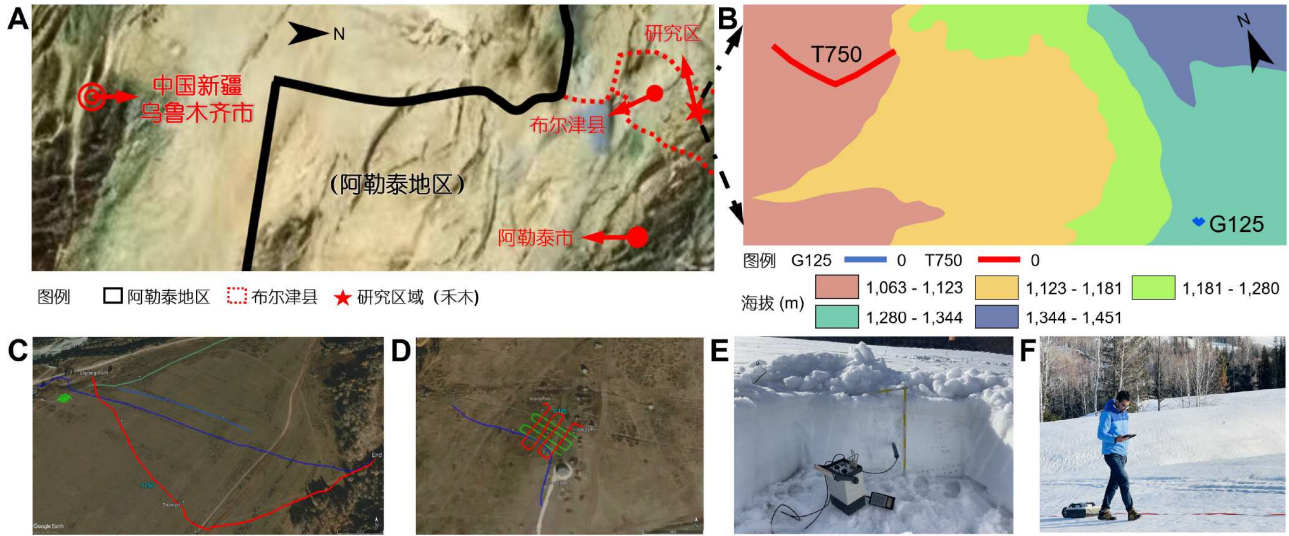


图2 研究区域概览。(A) 研究区域位于中国新疆阿尔泰地区；(B) 数字高程模型，标注有两条探地雷达测线(T750和G125)；(C, D) 谷歌地球影像，红色部分为目标线；(E) 使用Snowfork测量雪坑；(F) GPR剖面测量

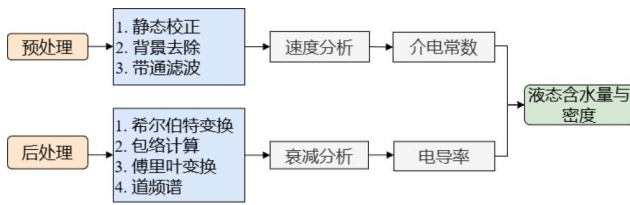


图3 GPR现场数据处理工作流程

耗，在衰减驱动的LWC估算中起着核心作用。雪的电导率(σ)定义为 $\sigma = \sigma' + \epsilon''\omega$ ，其中 σ' 是电导率的实部， ω 是信号的角频率。衰减系数(α)主要由 ϵ'' 控制，并随频率的增加而迅速增加。这种关系可以用公式(1)来描述(Bradford, 2007)。

$$\alpha \approx \sqrt{\frac{\mu_0}{\epsilon'}} \frac{\epsilon''}{2} \omega \quad (1)$$

其中， α 表示衰减系数， ϵ' 和 ϵ'' 分别表示介电常数的实部和虚部， μ_0 表示自由空间的磁导率。对于10 MHz至1 GHz范围内的探地雷达(GPR)频率，通常假设 ϵ' 与频率无关，而更精确的衰减建模方法是结合低频衰减项(α_0)和品质因数(Q^*)(Turner和Siggins, 1994)。由于GPR调查是在该频带内的800 MHz下进行的，因此雪层条件下的衰减可以用公式(2)表示。

$$\alpha = \alpha_0 + \frac{\sqrt{\mu_0 \epsilon'}}{2Q^*} \omega \quad (2)$$

其中， α_0 为低频衰减项， Q^* 为衰减品质因数，它在1–1500 MHz频率范围内保持经验常数。 Q^* 可通过公式(3)估算(Bradford等, 2010)。

$$Q^* = \frac{\epsilon'_s}{2\epsilon''_s} \quad (3)$$

频移法(Valence等, 2022)将空气-雪和雪-地面界面之间的偏移与衰减联系起来，从而可以使用公式(4)检索 Q^* 。

$$\frac{1}{Q^*} = \frac{2(\omega_0^2 - \omega_t^2)}{\pi t \omega_0^2 \omega_t} \quad (4)$$

其中， $\omega_0 = 2\pi f_0$ 和 $\omega_t = 2\pi f_t$ 分别是雪面和雪地界面处的角频率。

在识别每条记录中的雪地界面反射后，使用Re-flexW软件进行频域衰减分析。后处理流程包括几个关键步骤：(1) 希尔伯特变换提取解析信号成分；(2) 瞬时幅度包络计算；(3) 应用傅里叶变换分离频谱成分；(4) 逐道峰值频率分析(图3)。

对连续反射进行频谱分析，并使用中值滤波器在每个窗口上获取峰值频率，以抑制噪声引起的变化。利用这些峰值频率，分别根据公式(3)和(4)估算雪介电常数虚部(ϵ''_s)和角频率(Q^*)。

体积含水量(θ_w)采用公式(5)中的经验公式估算(Sihvola和Tiuri, 2007)。

$$\epsilon''_s = (0.10\theta_w + 0.80\theta_w^2) \epsilon''_w \quad (5)$$

其中， ϵ''_w 是水的虚部介电常数，因此雪介电常数实部 ϵ'_s 可通过公式(6)估算。

$$\epsilon'_s = (0.10\theta_w + 0.80\theta_w^2) \epsilon'_w + \epsilon'_d \quad (6)$$

其中， ϵ'_d 是干雪介电常数，根据公式(7)中的干雪密度

(ρ_d)估算。

$$\varepsilon'_d = (1 + 1.7\rho_d + 0.7\rho_d^2) \quad (7)$$

使用Debye或Cole-Cole色散模型(Zeng等, 2000)对在湿雪条件下主导GPR信号衰减的水介电响应进行建模, 具体取决于弛豫频率。

最后, 利用公式(8)估算雪层体积密度(ρ_s)。

$$\rho_s = \rho_d + \theta_w \rho_w \quad (8)$$

其中, ρ_w 是液态水的密度。所有 θ_w 和 ρ_s 的估计值随后与雪叉法和重量法雪坑测量值进行比较, 以进行现场验证。公式(1)–(8)中使用的变量和单位的完整命名法见网络版附录C。

2.3 使用gprMax生成合成数据

基于经验模型对GPR雪层数据进行人工解译以估算雪物理属性复杂且耗时。为实现可扩展的自动化处理, 引入数据驱动的机器学习(ML)方法(Namdari等, 2025)。然而, 机器学习模型需要大规模、多样化且带标签的数据集, 而仅依赖实地观测难以满足这一需求。实地获取的GPR数据受环境、噪声和传感器限制影响, 不适合作为监督式机器学习模型的训练数据。

为解决数据稀缺问题, 采用开源电磁模拟器gprMax生成大量合成反射阵列。模拟在受控环境下再现真实雪层条件, 通过有限差分法(FDTD)(Yee, 1966), 可精确模拟波在非均匀介质中的传播。所有模拟均在二维空间进行, 采用800 MHz中心频率、0.001 m网格单元尺寸的Ricker小波源。相应的电磁模拟参数汇总于网络版附录A(附表A1)中。

为捕捉雪层的自然变异性, 设计了考虑可变层、介电常数、电导率和地形粗糙度的模拟模型, 并假设磁导率为1。这些模型包括随机介电包裹体(例如冰透镜、地壳层)以及融化-再冻结的地层结构, 以再现过渡期雪层的实际复杂性。完整的模拟参数和合成雪层模型示例见附录A, 典型雪层几何形状如图4所示。

模拟模型代表了覆盖在三层土壤上的五层不同的雪层。每层都被赋予了不同的相对介电常数(ε')和电导率(σ)值, 具体数值见表1, 包括湿雪、干雪、融雪层以及不同的土壤饱和度和条件。编码过程中采用了德拜色散模型, 以融入对湿雪建模至关重要的频率相关介电行为(Liu等, 2019)。模拟中加入了分形表面粗糙度, 以复制自然的异质性, 并加入了完美匹配层(PML)边界条件, 以防止人为的边缘反射。

模拟系统配置了一台800 MHz系统, 与现场部署的探地雷达(GPR)相匹配, 并在10 m剖面上以共偏移距剖面(COP)模式运行, 生成205个A扫描, 空间分辨率为4.8 cm, 与现场采集结果一致(第2.1节)。图5显示了一个合成雪层模型的雷达图(B扫描)输出示例。图中清晰地显示了雪层内部的反射和明显的雪土界面。

为验证合成信号的物理真实性, 对其与现场采集的探地雷达(GPR)轨迹进行了时域和频域的对比。在时域中, 评估了反射振幅和走时模式, 结果显示信号的起始和衰减行为得到准确再现。空气-雪界面和雪-地面界面清晰可见, 凸显了介电常数的垂直差异, 这与模拟模型几何形状(表1)相符。这些界面反射可作为参考点, 用于在时间上对齐合成和真实的探地雷达信号(图6)。

在基于FFT的频谱对比中, 真实频谱在800–

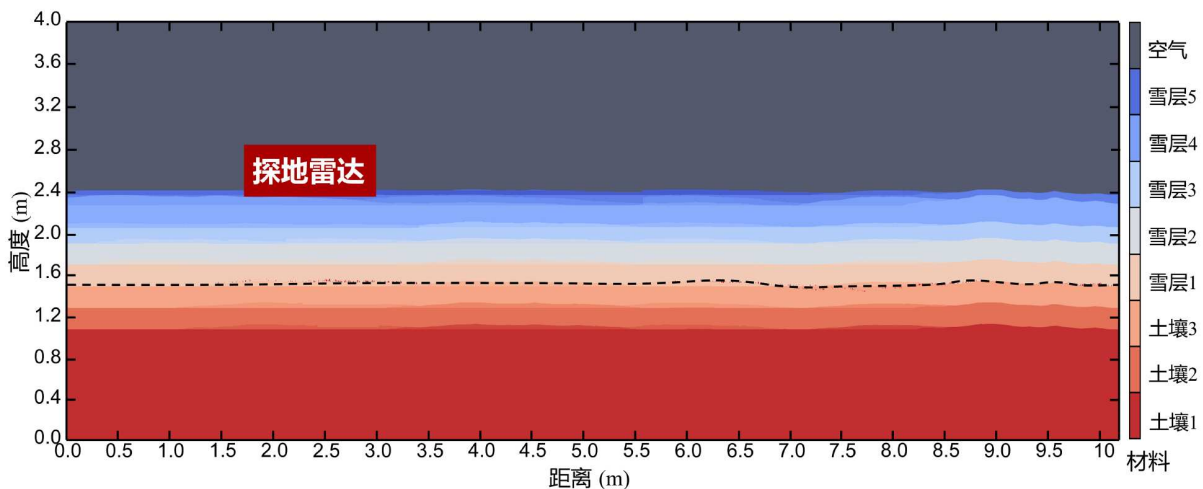


图4 雪层模型几何形状(.vti文件)已在gprMax中配置, 用于合成模拟。红色方框标示了探地雷达天线的位置, 其中发射器(Tx)和接收器(Rx)的固定偏移量为0.09 m

表1 模拟雪层层序的概述^{a)}

层数	层类	厚度(cm)	介电常数(ϵ')	电导率(σ)(S m^{-1})	$\text{Log}(\rho)(\Omega \cdot \text{m})$
雪层: 5	冰壳	10	3.18	0.00001	5.0
雪层: 4	融化层	20	2.50	0.001	3.0
雪层: 3	干雪	15	1.90	0.00001	5.0
雪层: 2	干雪	20	1.70	0.000063	4.2
雪层: 1	湿层	20	2.44	0.0012	2.9
土壤: 3	饱和土	20	11.0	0.2	0.63
土壤: 2	部分饱和土	20	9.0	0.15	0.85
土壤: 1	非饱和土	110	8.0	0.1	1.0

a) 通过电导率(σ)推导电阻率(ρ)值, 关系式为 $\log_{10}(\rho)=\log_{10}(1/\sigma)$ 。

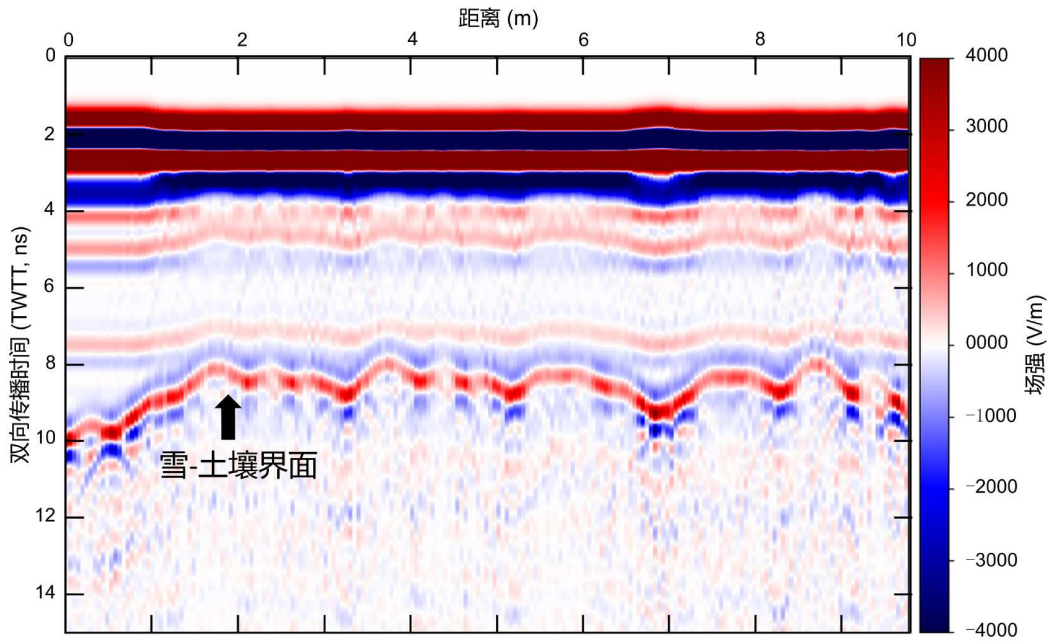


图5 基于积雪模型(图4)模拟生成的合成雷达图谱, 展示了探地雷达信号响应

1200 MHz之间呈现出多个主峰, 表明雪层内存在强烈的反射和不均匀的地下特征。尽管CO4080天线的标称中心频率为800 MHz, 但其宽带脉冲源覆盖更宽的频谱范围, 使峰值频率会随雪层衰减特性发生偏移。在低损耗雪层条件下(图7), 高频成分得以保留, 使谱峰略高于标称天线频率。这验证了合成数据集的物理真实性, 支持其在机器学习模型训练和反演工作流程中的应用。实际信号的峰值频率为920 MHz, 与合成信号的峰值999 MHz高度接近, 这证实了模型可捕捉主要传播行为。3 dB带宽分析显示, 真实数据具有更窄的频率内容, 其带宽减小主要源于真实环境中的更高衰减、散射和能量损耗, 而这些机制在合成建模中并不存在。

该合成数据集构成了基于机器学习的双参数反演算法的基础。A扫描输出与已知的真实标签(ϵ' , $\log(\rho)$),

相结合, 作为训练和评估多个机器学习模型的输入输出对。完整的机器学习工作流程包括信号预处理、主成分分析(PCA)降维以及使用成对电磁属性标签进行监督训练。后续章节将详细介绍每个部分。

2.3.1 基于PCA的输入特征生成

主成分分析(PCA)是一种广泛使用的技术, 可将高维数据转换为一组紧凑的正交分量, 同时保留重要信息(Bengio等, 2013)。本研究采用PCA压缩A扫描信号, 同时保留电磁参数估计所需的关键结构信息(Kaczmarek和Pietrasinski, 2014)。

训练数据集包含5600个合成A扫描, 每个A扫描代表gprMax中模拟的一个独特的雪层结构。每个原始A扫描最初包含6371个时间样本, 捕捉雪层和地下层的反射。

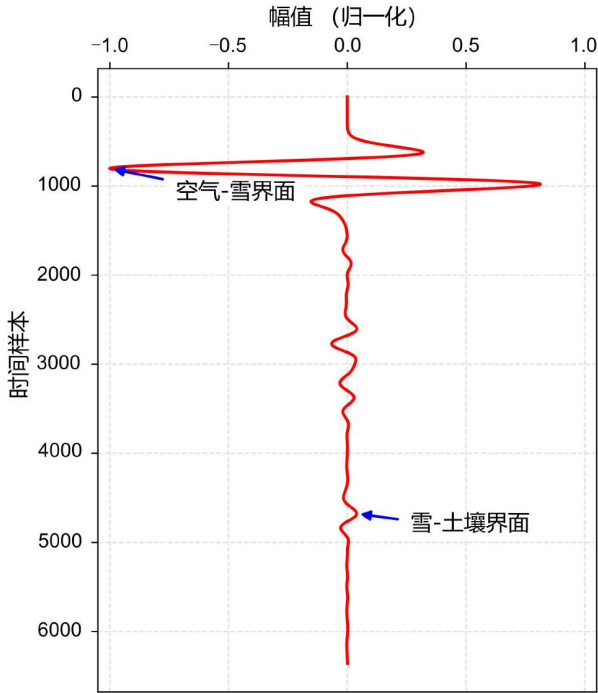


图6 展示来自关键电介界面的振幅反射的A扫描雷达迹线图

为了提高所有轨迹的时间一致性,通过对齐信号起始点进行静校正。随后,通过滤除1.6 GHz以上的频率,将轨迹下采样至936个时间步长,以匹配现场部署中使用的800 MHz GPR系统的工作带宽。为了分离雪层响应,采用局部峰值阈值算法去除了来自地下土壤的反射。将得到的雪层片段(450–600个样本)用零填充至677个时间步长的固定长度,以确保所有轨迹的特征提取均匀一致。

处理后的数据集由以下矩阵表示:

$$\mathbf{A} = [a_1, a_2 \dots a_m] = \mathbf{A}_{i,j} \in \mathbb{R}^{m \times n}$$

其中, $m=5600$ 表示A扫描总数, $n=677$ 表示每条轨迹的时间步数。 a_i 是欠采样和预处理轨迹 $a_i \in \mathbb{R}^{677}$ 。

将PCA应用于 $\mathbf{A}$,得到一组 n 个正交特征向量,作为主轴,表示为

$$\mathbf{P} = \mathbf{P}_{i,j} \in \mathbb{R}^{n \times n}$$

其中, $\mathbf{P}$ 的每一列对应一个主成分(PC)。通过分析累积解释方差曲线确定保留主成分(PC)的数量(Jolliffe和Cadima, 2016)。图8A显示,前40个PC保留了超过99%的总信号方差。图8B展示了重建精度,表明平均绝对重建误差在40个PC之后趋于稳定,表明信息损失最小。

因此,构建了一个简化基矩阵 $\mathbf{Q} \in \mathbb{R}^{677 \times 40}$,并使用最小二乘解将每个A扫描投影到这个40维子空间中(公式(9))。

$$\mathbf{x}_i = (\mathbf{Q}^T \mathbf{Q})^{-1} \mathbf{Q}^T a_i \quad (9)$$

由此为每条轨迹生成了一个压缩的输入特征向量 $\mathbf{x}_i \in \mathbb{R}^{40}$ 。完整的PCA变换输入矩阵如下所示:

$$\mathbf{X} = [x_1, x_2 \dots x_m] = \mathbf{X}_{i,j} \in \mathbb{R}^{m \times 40}$$

这种低维表示显著降低了计算负荷,减轻了过度拟合,并保持了波形完整性。

为进一步理解主成分分析(PCA)成分及其物理相关性,对40个主成分(PC)与电磁地面真实目标(ϵ' , $\log(\rho)$)进行了皮尔逊相关性分析。如附录B(附图B1、附表B1)所述,几个主成分(PC9、PC17和PC12)与目标表现出弱到中等的线性相关性。尽管PCA是一种无监督且由数学驱动变换,但这种经验关联表明某些成分保留了与电

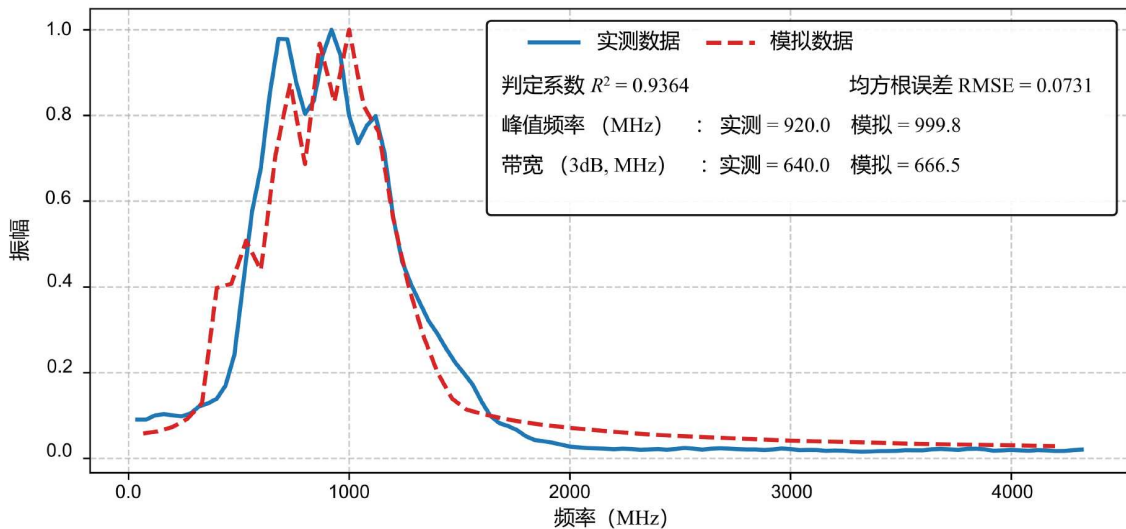


图7 雪层覆盖区域的合成GPR数据与真实GPR数据的频域比较。频谱比对结果显示峰值频率和带宽高度匹配, R^2 为0.9364, RMSE为0.0731

磁行为相关的具有物理意义的信号变化。因此，这种事后分析提供了可解释性的见解，同时又不损害PCA框架的抽象优势。

图9显示了最主要的六个PC。图10表示40个PC如何

有效地近似完整波形，比较原始波形和PCA压缩重建。

2.3.2 地面实况标签生成

为了训练监督机器学习模型，每个预处理的A扫描

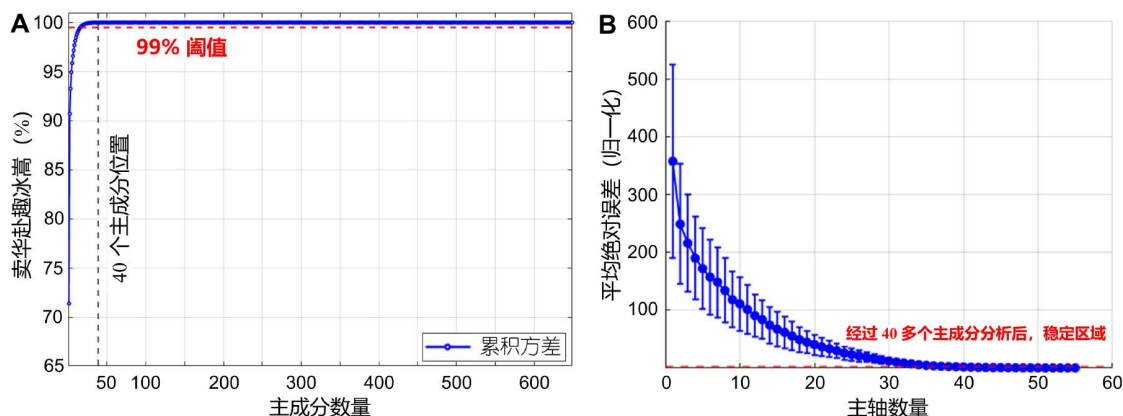


图8 (A) 累积解释方差曲线。红色虚线表示99%阈值，垂直虚线表示40个PC的截止值；(B) 原始A扫描与重建A扫描之间的平均绝对误差。蓝色误差线代表 ± 1 个标准差，突显了重建性能的一致性

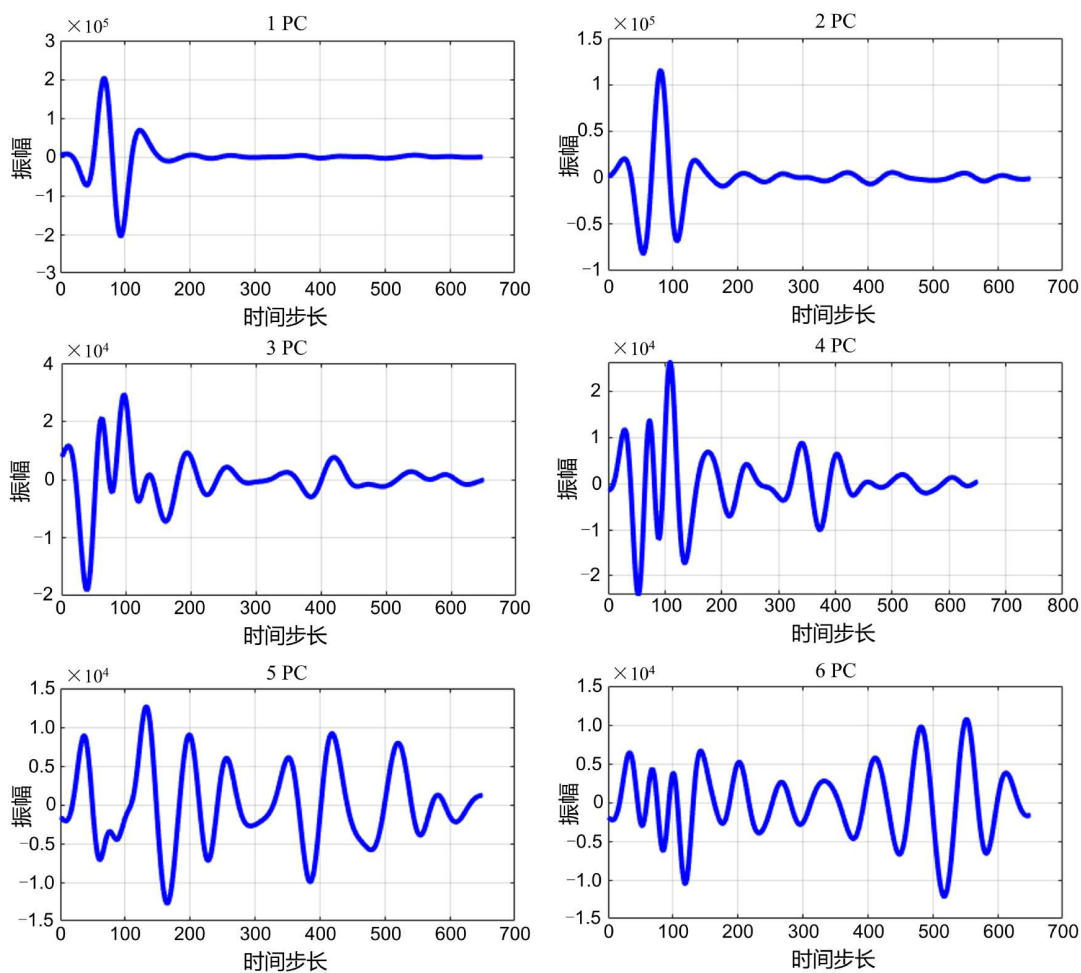


图9 训练集中最主要的六个主成分

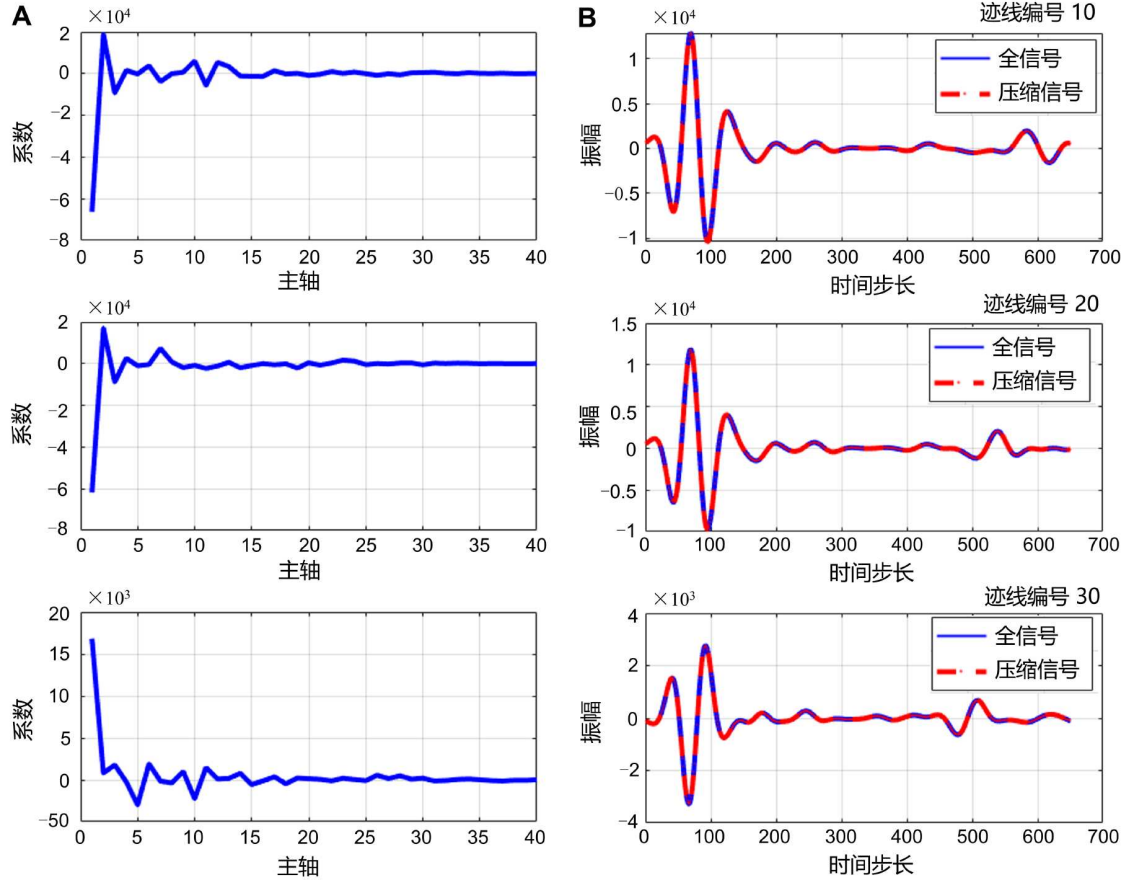


图 10 (A) 用于信号表示的40个主轴; (B) 使用FDTD随机生成的三个原始完整A扫描及其相应的PCA重建

都需要与相应的真实标签配对。此处，电磁属性(介电常数和电阻率)作为目标输出。这些值是从gprMax中执行的全波模拟中提取的，gprMax将空间域编码为一个基于整数的材料索引数组，该数组对应于预定义的材料列表(Warren等, 2019)。通过(#geometry_objects_write)指令将材料索引导出为HDF5格式，并使用查找表将其解码为相关的电磁属性值。为了提高可视化效果和学习稳定性，利用公式(10)将电导率(σ)转换为对数电阻率(ρ)，从而压缩了数值的动态范围、稳定了反演过程，使两个参数置于可比尺度上训练。

$$\log_{10}(\rho) = -\log_{10}(\sigma) \quad (10)$$

初始模型覆盖了整个模拟区域，矩阵大小表示为($\mathbf{M}_i \in \mathbb{R}^{10200 \times 4000}$)。该矩阵代表一个10.2 m×4 m的空间模型，采样分辨率为1 mm($\Delta x = \Delta y = 0.1$ cm)。为了将雪层与上方的自由空间隔离，移除前1.65 m的垂直数据，从而得到了如图11所示的修剪区域，并将调整大小后的矩阵表示为($\mathbf{M}_i \in \mathbb{R}^{10200 \times 2350}$)。

为了在保留关键结构的同时减少输入大小，采用最

近邻插值对矩阵进行重采样，得到按时间步长和轨迹数量缩放的缩放图，如相关图12所示(Giannakis等, 2024)，矩阵的新大小为($\mathbf{I}_i \in \mathbb{R}^{205 \times 936}$, $\Delta x = 4.98$ cm, $\Delta y = 0.25$ cm)，其中， $\mathbf{I}_i$ 是包含雪和土壤反射的插值矩阵。

为分离仅包含雪的属性，应用基于阈值的分割算法，去除了自由空间和土壤层。剩余的雪层区域沿每条轨迹(列)进行深度平均，从而生成每次A扫描的单一介电常数和电阻率值，如下图所示，分别对应两个不同的目标：

$$y^{(\varepsilon')} = [\varepsilon'_1, \varepsilon'_2, \dots, \varepsilon'_m]$$

$$y^{(\log \rho)} = [\log \rho_1, \log \rho_2, \dots, \log \rho_m]$$

其中， m 为迹线数。在示例中， $m=205$ ，如图13A、13C、13E所示。

最后，这两个输出向量被连接成一个标签矩阵：

$$\mathbf{Y} \in \mathbb{R}^{5600 \times 2}$$

$\mathbf{Y}$ 的每一行包含一对基本真值(ε' , $\log(\rho)$)。因此，完整的监督学习任务可以表述为一个多元回归模型：

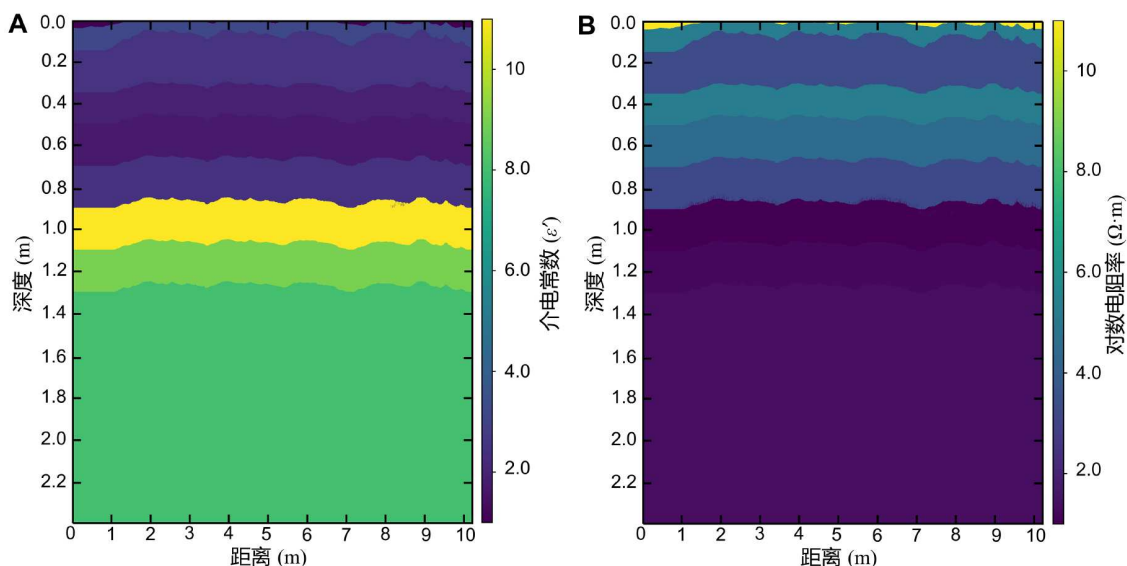


图11 电磁特性的空间分布。(A) 相对介电常数(ϵ')；(B) 对数电阻率(ρ)

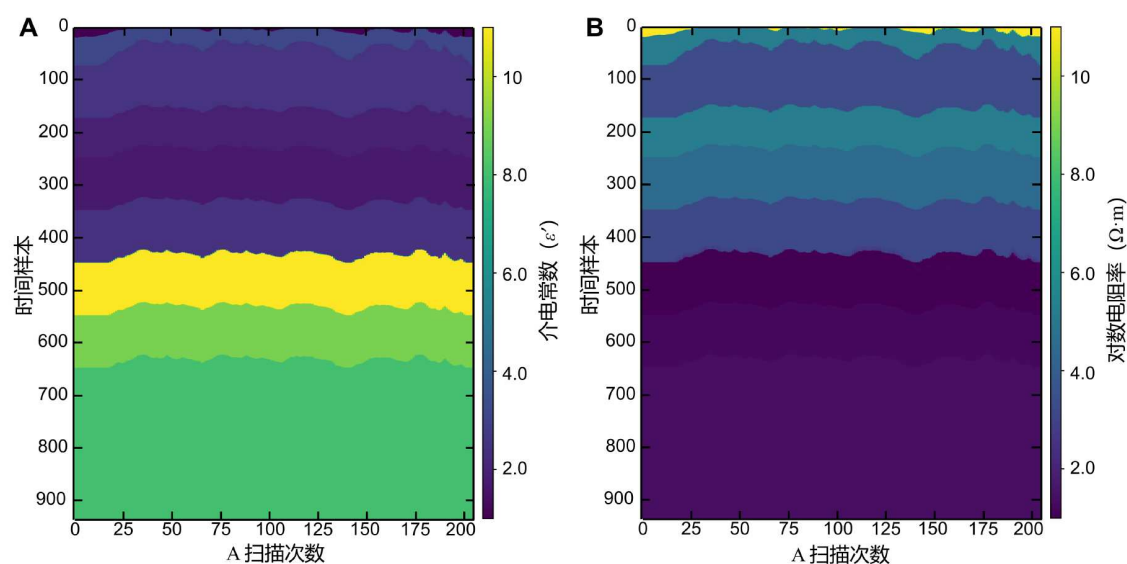


图12 电磁特性插值图。(A) 相对介电常数(ϵ')；(B) 对数电阻率(ρ)。插值结果描绘了雪层和下层土壤的垂直和横向变化

$$f: X \in \mathbb{R}^{5600 \times 40} \rightarrow Y \in \mathbb{R}^{5600 \times 2}$$

其中, X 是每个A扫描的40维PCA特征矩阵, Y 包含相应的EM属性标签。

2.4 机器学习(ML)模型选择

本研究采用了四种机器学习(ML)模型, 包括随机森林(Biau和Scornet, 2016)、神经网络(Yegnanarayana, 2009)、极端梯度提升(Chen和Guestrin, 2016)和支持向量机(Hearst等, 1998)。选择这些模型的依据在于其在多输出回归任务中已被证实的有效性, 以及在地球物理

信号分析中捕捉复杂非线性模式的能力(Filardi等, 2023)。

对于RF、XGB和SVM, 使用随机交叉验证和网格搜索的混合方法调整树深度、学习率和核函数等超参数, 以确保通用性并减少过度拟合(Belete和Huchaiah, 2022)。对于神经网络, 模型结构及训练超参数通过系统的经验调优进行优化。评估的配置包括不同的深度(4–5层)、每层神经元数量(128、256、512)、dropout率(0.2–0.4)、学习率(0.01–0.0001)以及激活函数ReLU和SiLU。最终的模型架构包含五个完全连接的层, 带有SiLU激

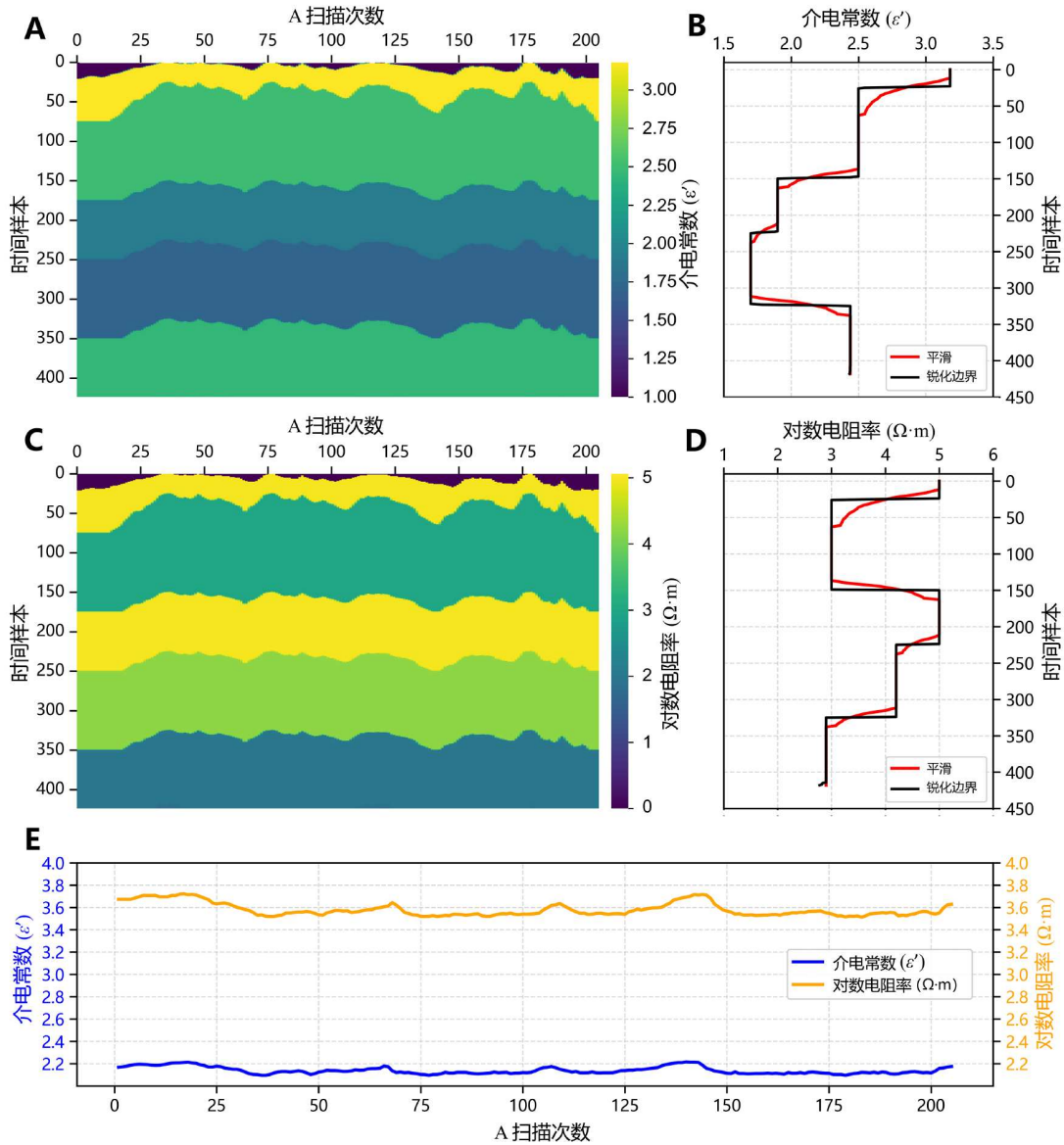


图 13 分段雪层的电磁特性的空间和垂直分布。(A) 相对介电常数(ϵ')的二维图;(B) 随机选择的A扫描的精确和平滑垂直介电常数曲线的比较;(C) 对数电阻率(ρ)的二维图;(D) 随机选择的A扫描的精确和平滑垂直对数电阻率曲线的比较;(E) 整个雪层的深度空间趋势,显示所有迹线中 ϵ' 和 $\log(\rho)$ 的平均变化

活、批量归一化、L2正则化($\lambda=0.0001$)和dropout(0.3)。该模型使用Nadam优化器进行训练,Nadam优化器是随机梯度下降(SGD)的一种变体(Liao等, 2022),以其高效的收敛特性而闻名,特别是在非凸优化任务中。

2.4.1 数据增强和领域适应

为提升模型在实际应用中的稳健性,使用噪声增强的合成数据对两个最有效的模型(NN和RF)进行了进一步微调。将高斯噪声迭代添加到输入特征中,信噪比(SNR)范围为1至10 dB。这种增强模拟了GPR信号中典

型的场畸变,例如天线耦合、地表粗糙度散射和环境干扰,这与近期地球物理反演中的机器学习实践(Xue等, 2024)一致。

神经网络模型在训练过程中通过自定义的批量生成器引入噪声,并动态地将噪声添加到每个小批次中。这不仅提高了模型的抗噪能力,还促进了鲁棒特征表征的学习。此外,还采用了一种领域自适应策略,其中神经网络的较深层被选择性地解冻,并使用降低的学习率进行微调。这种方法能够保留广义的低级时间特征,同时使高级表示适应噪声引起的变化。相比之下,RF模型在

添加噪声后需要从头开始重新训练。这种方法确保了两个模型都能有效地分析含噪声的野外探地雷达数据。随后对它们的泛化能力进行了评估(见3.2节)。

2.4.2 性能评估指标

机器学习模型的性能采用三个回归指标评估：判定系数、均方根误差和平均绝对百分比误差(Anwar等, 2023)。这些指标针对每个预测参数(ε' , $\log(\rho)$)分别计算, 以检验其学习效率和未知数据的泛化能力。这些矩阵提供了绝对误差和相对误差, 以及模型的整体拟合度。令 $y_i^{(j)}$ 表示第 j 个参数的真实值(j th), 其中对于第 i 个样本(i th), $j \in \{\varepsilon', \log(\rho)\}$ 。性能指标定义为

$$\text{RMSE}^{(j)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i^{(j)} - \hat{y}_i^{(j)})^2} \quad (11)$$

$$\text{MAPE}^{(j)} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i^{(j)} - \hat{y}_i^{(j)}}{y_i^{(j)}} \right| \times 100 \quad (12)$$

$$R_{(j)}^2 = 1 - \frac{\sum_{i=1}^n (y_i^{(j)} - \hat{y}_i^{(j)})^2}{\sum_{i=1}^n (y_i^{(j)} - \bar{y}^{(j)})^2} \quad (13)$$

其中, $\hat{y}_i^{(j)}$ 和 $\bar{y}^{(j)}$ 分别为第 j 个参数的模型预测值和平均地面真值, n 为评估样本数。这些指标将在第3.1节和3.2节中用于检验模型在合成数据和现场数据上的性能, 以确定最适合实际GPR反演的架构。

为了实现简洁的可视化效果, 并方便直观地比较多个评估指标, 利用雷达图对介电常数和对数电阻率进行预测。由于指标的尺度和可解释性各不相同, 采用了min-max归一化方法, 将所有值映射到统一的(0, 1)范围内, 这是Bedi和Toshniwal(2018)提出的建议。归一化计算如下:

$$\text{Norm}(X_i) = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (14)$$

其中, X_i 表示第 i 个模型的性能指标值, $X_{\min}$ 和 $X_{\max}$ 分别是所有模型的最小和最大指标值。

对于判定系数 R^2 , 其值越高表示性能越好, 因此直接使用了标准化值。相比之下, 对于基于误差的指标(例如RMSE和MAPE), 其值越低表示性能越好, 因此进行了反演步骤, 以保持整个雷达图的视觉一致性, 如下所示。

$$\text{Norm}_{\text{inv}}(X_i) = 1 - \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (15)$$

这种反演确保雷达图中更高的标准化值始终对应于所有轴上的卓越模型性能。

3 结果

3.1 模型在合成数据上的表现

首先使用模拟数据集对四种机器学习模型(NN、RF、XGB和SVM)进行评估。针对相对介电常数(ε')和对数电阻率(ρ), 使用RMSE、MAPE和 R^2 对训练集和测试集的性能进行评估。结果汇总于表2。

图14展示了训练集-测试集的对比结果, 而图15则展示了归一化指标的雷达图。NN模型对 ε' 和 $\log(\rho)$ 均始终保持最高的预测精度和泛化能力。RF模型也表现良好, 在误差幅度和稳定性方面取得了平衡。相比之下, SVM和XGB模型的结果较弱, 尤其是在电阻率方面。

3.2 现场数据测试

3.2.1 噪声增强条件下的性能

现场采集的探地雷达信号本身会受到环境噪声和系统噪声的影响。为了模拟这种变化, 将对应于6 dB (25%)信噪比的高斯噪声注入到合成数据集中。该噪声水平反映了现实世界中雪层调查中常见的退化情况, 并为评估模型在退化输入条件下的稳健性提供了基础。

图16比较了噪声添加前后 ε' 和 $\log(\rho)$ 的性能指标

表2 机器学习模型在合成数据集上的介电常数(ε')和对数电阻率(ρ)的性能指标

模型		RMSE(训练/测试)	MAPE%(训练/测试)	R^2 (训练/测试)
介电常数(ε')	NN	0.0258/0.0442	0.64/1.50	0.9926/0.9784
	RF	0.0313/0.0721	0.65/1.51	0.9891/0.9426
	XGB	0.0094/0.1061	0.30/2.21	0.9900/0.8756
	SVM	0.0712/0.0991	2.82/3.44	0.9438/0.8915
对数电阻率(ρ)	NN	0.0603/0.0976	0.91/2.23	0.9900/0.9241
	RF	0.0470/0.1133	0.92/2.24	0.9818/0.8979
	XGB	0.0272/0.1267	0.50/2.40	0.9930/0.8722
	SVM	0.0877/0.1431	2.40/3.56	0.9365/0.8370

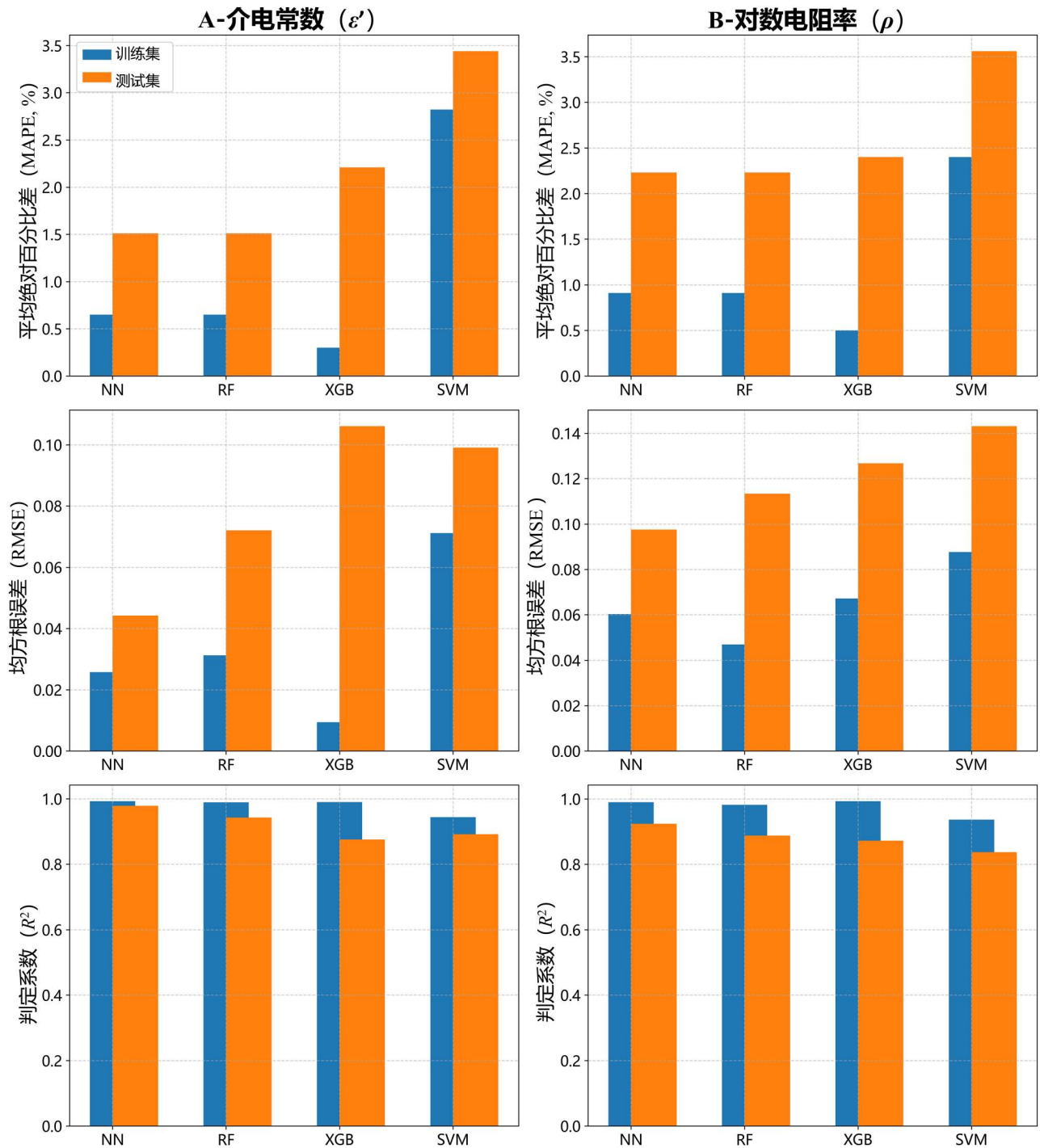


图 14 使用不同ML模型对(A)介电常数(ϵ')和(B)对数电阻率(ρ)预测的训练测试性能指标(MAPE、RMSE、 R^2)进行条形比较

RMSE、MAPE和 R^2 。正如预期的那样，所有指标的噪声后性能均略有下降，表明精度降低但鲁棒性增强。此外，图17展示了噪声增强前后的 $\log_{10}(\text{MAPE})$ 分布。红色条形图(添加噪声)向右移动表示预测误差增加，但分布形状保持不变。这种一致性表明模型在劣化输入条件下仍保持了泛化能力。NN模型表现出卓越的鲁棒性，因为它在

训练期间学习了泛化特征，并且对噪声引起的失真不太敏感。相比之下，RF表现出更高的可变性。

3.2.2 使用GPR衍生的电磁特性进行机器学习(ML)模型验证

为了评估模型在真实雪层条件下的性能，在两个对

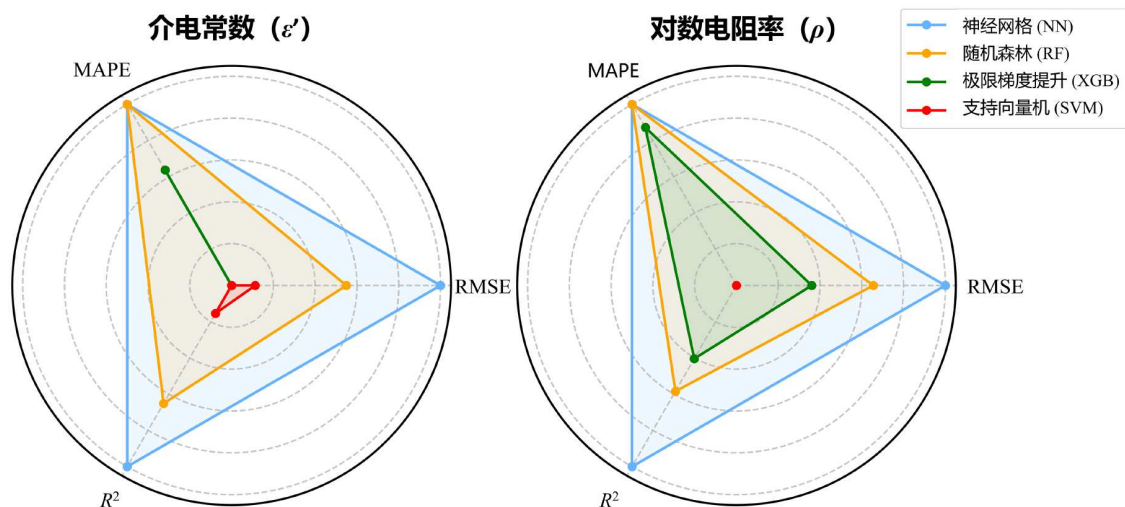


图15 基于雷达图的机器学习模型性能评估。用于估计 ϵ' 和 $\log(\rho)$ 。归一化值越大，模型性能越好。神经网络模型在这两个参数上都表现出色，而随机森林模型也表现出了相当高的准确率

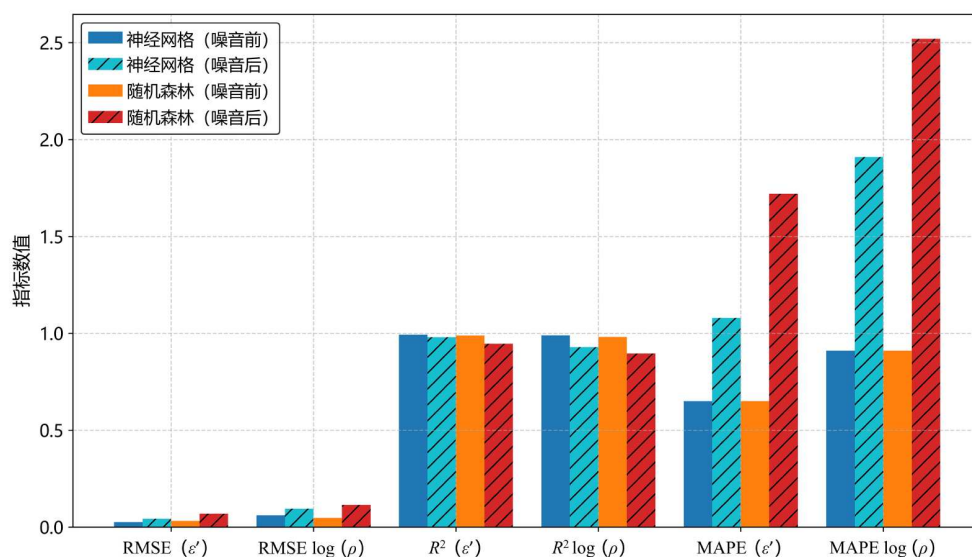


图16 添加噪声前后性能指标比较

比鲜明的GPR野外样地上评估了NN和RF模型：T750(潮湿，阴凉，约1118 m)和G125(潮湿，阳光直射，约1320 m)。使用频域GPR衰减分析(见2.2节)得出地面真实电磁参数。

图18显示了ML预测值与GPR得出的 ϵ' 和 $\log(\rho)$ 值的比较。NN在所有 ϵ' 估计值中表现出更接近的一致性：T750上为2.186 vs. 1.96，G125上为2.80 vs. 2.52。RF表现出更大的偏差，尤其是在潮湿条件下。对于 $\log(\rho)$ ，两个模型都表现良好。NN和RF对T750的预测值均略有低估(分别为3.067 vs. 3.35 $\Omega \cdot m$)和(3.042 vs. 3.35 $\Omega \cdot m$)。在G125

上，两个模型都在接近的范围内，NN略微接近(3.194 vs. 3.24 $\Omega \cdot m$)。

3.2.3 使用物理特性进行验证：LWC和体积密度

为了评估ML反演框架的水文相关性和实际现场精度，将ML得出的体积含水量(θ_w)和雪层体积密度(ρ_s)估计值与基于GPR的衰减反演、雪叉探测器和重力雪坑分析获得的参考测量值进行了比较。表3总结了T750(湿雪层)和G125(湿雪层)两个对比样地的完整结果。此外，还分别报告了相对于雪叉法和手动重力法参考的LWC和

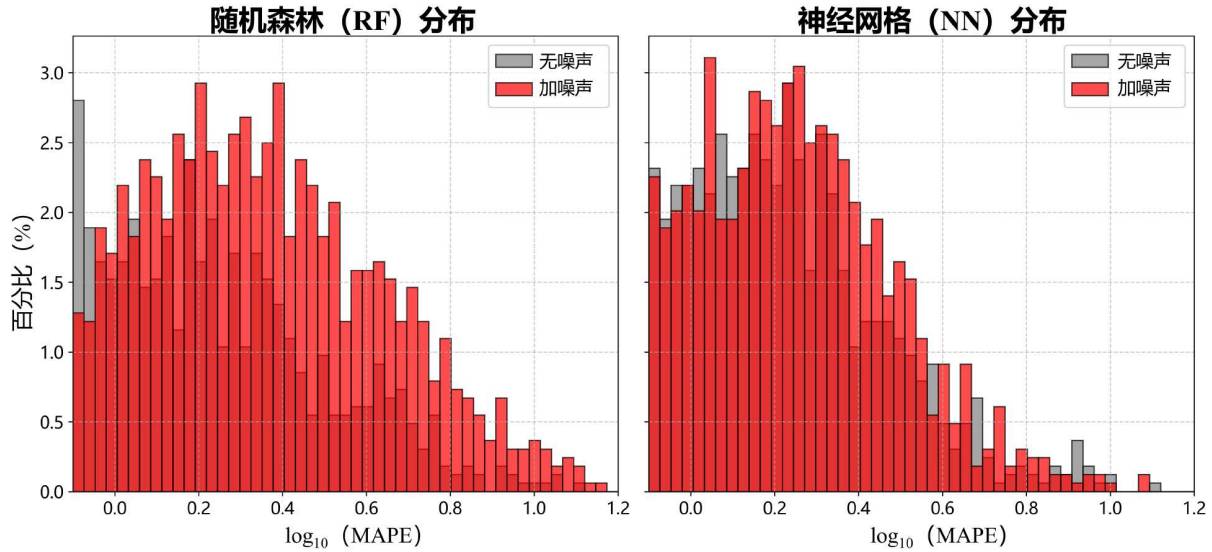


图 17 在无噪声和噪声增强条件下,使用RF和NN模型计算的介电常数和电阻率的 $\log_{10}(\text{MAPE})$ 值分布。无噪声分布范围较窄,而噪声增加则表现出较高的变异性,尤其是在RF模型中

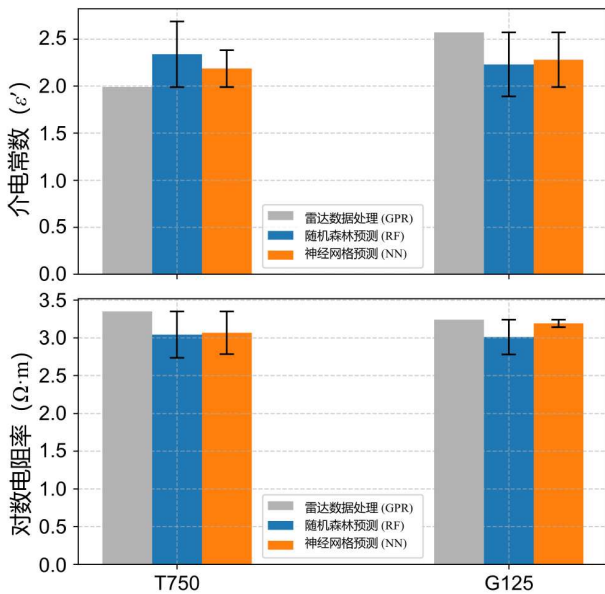


图 18 ML预测与GPR衍生的电磁特性

密度的绝对误差。

在图19中,NN模型在LWC和密度方面与地面测量值始终保持着更好的一致性。在T750时,NN的LWC估计值(3.8%)与Snowfork值(3.2%)和GPR值(2.7%)的误差在1%以内。在G125时,NN几乎与Snowfork值(6.14% vs. 6.0%)相符,并且与GPR参考值保持接近。NN的密度预测值与人工参考值的误差在 $\pm 0.08 \text{ g cm}^{-3}$ 以内,且在两条剖面上保持了高度一致性。相比之下,RF低估了LWC,尤其是在湿润的G125剖面上,但保持了稳定的密度估计

值。RF的T750密度预测值(0.486 g cm^{-3})略微高估了重量法(0.404 g cm^{-3})和GPR法(0.383 g cm^{-3})的值。

总体而言,神经网络(NN)模型在综合评估和真实GPR现场数据中始终表现出色,能够准确预测不同条件下雪层的电学特性(ϵ' 和 $\log(\rho)$)和物理特性(θ_w , ρ_s)。该模型在噪声增强输入下仍保持稳健,并在从潮湿阴凉坡到潮湿阳光照射地形的各种湿度条件下保持了普遍性。随机森林(RF)在密度预测方面表现出稳定性,但对LWC变化的敏感性有限,尤其是在较湿的雪层中。这些结果证实,将综合模拟、噪声增强和多源验证相结合,可以为使用GPR进行非侵入式雪层监测提供一个可靠且可扩展的框架。

4 讨论

本研究开发并评估了一个数据驱动的反演框架,该框架集成了合成GPR模拟、基于主成分的降维方法和机器学习,用于根据共偏移距GPR数据反演雪层介电常数和电阻率。基于合成数据集和现场数据集的验证表明,该方法能够有效估算关键的电磁特性,并进而推导出体积含水量(θ_w)和体积密度(ρ_s)。

4.1 弥合合成到现场泛化差距

基于机器学习的反演面临的一个主要挑战是,受控的合成训练数据集与包含环境干扰和传感器变异的嘈杂真实现场测量数据之间的域转移会导致性能下降。为了缓解这个问题,在合成数据集中注入迭代高斯噪声(信噪

表3 ML预测值与参考LWC和体积密度值的比较

横剖面	模型	$\theta_w(\%)$	$\rho_s(\text{g cm}^{-3})$	$\Delta\theta_w(\%)$	$\Delta\rho_s(\text{g cm}^{-3})$
T750	RF	1.6	0.486	-1.6	0.082
	NN	3.8	0.437	0.6	0.033
	GPR	2.7	0.383	-0.5	-0.021
	雪叉	3.2	0.420	参考	/
	重量法	/	0.404	/	参考
G125	RF	2.9	0.491	-3.1	-0.004
	NN	6.14	0.579	0.14	0.084
	GPR	5.1	0.537	-0.9	0.042
	雪叉	6.0	0.517	参考	/
	重量法	/	0.495	/	参考

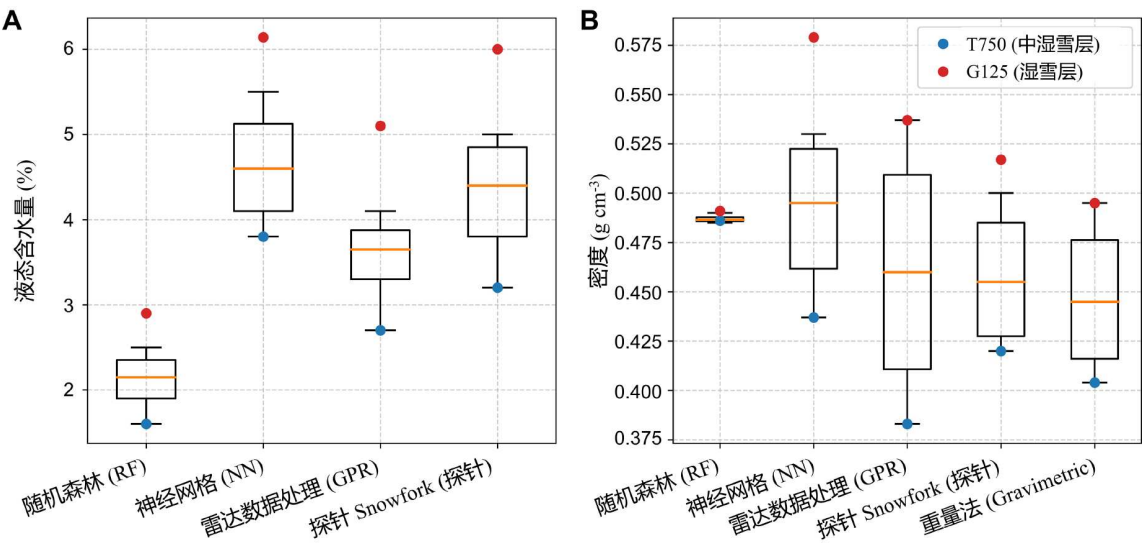


图19 图(A)和图(B)分别比较了通过机器学习模型(随机森林、神经网络)、野外测量(GPR)和原位测量(探针法和重量法)估算的液态水含量和密度值。箱线图分别用蓝色和红色标记显示了T750和G125样带的分布情况。密度的重力估算是指人工采集雪芯样本并称重

比约为6 dB)，并应用主成分分析(PCA)进行特征空间自适应。这种策略减少了过拟合，并提高了对现场条件的可迁移性。

在所有测试模型中，神经网络(NN)在噪声增强条件下表现出最可靠的泛化能力。近期针对冰冻圈应用的GPR-ML研究也强调了类似的迁移能力有限的问题。例如，Xiao等(2025)的研究表明，将GPR横剖面数据纳入ML分类器，可显著提高青藏高原多年冻土(冻结超过2年的地面)测绘模型的泛化能力，这凸显了提升模型从有限或有偏差数据集迁移能力的策略的重要性。框架的独特之处在于明确解决了雪层反演中合成数据与实地数据之间的差距，从而扩展了其在不同雪况下的适用性。

4.2 比较评估和准确性

使用从两个对比鲜明的雪层横剖面(T750和G125)收

集的现场数据对ML反演模型进行评估。电磁估计值(ϵ' 和 $\log(\rho)$)用于推导体积LWC和体积密度(ρ)，并与来自雪叉探测器、雪坑密度和基于经验衰减的反演的独立地面真实测量值进行交叉验证。这些辅助数据集仅用于验证，不用作模型输入。与参考测量值相比，NN模型实现了LWC偏差 $\leq 1.5\%$ 和密度误差 $\pm 0.03\text{--}0.084\text{ g cm}^{-3}$ ，在潮湿和湿润雪层条件下均表现出卓越的稳健性。为了将这些结果具体化，表4总结了主要基于GPR的雪层反演研究的性能。

比较结果凸显了框架的准确性和实用效率。Meehan等(2024)报告称，使用GPR结合LiDAR测得的密度均方根误差(RMSE)为 27 kg m^{-3} ，但这种对辅助数据的依赖增加了运营成本并限制了可扩展性。Sundström等(2012)，使用基于衰减的GPR方法，报告称LWC误差为1.8%至3.0%，低估率高达50%，这凸显了经验方法在不

表4 基于GPR的雪层研究中LWC和密度估计的精度比较

来源	GPR类型	辅助数据	ML使用	体积含水量(%) 准确度	体积密度(kg m^{-3}) 准确度
本研究	合成+现场 GPR	雪坑、雪叉、经验衰减 (仅验证)	是	$\leq 1.5\%$	$\pm 30\text{--}80 \text{ kg m}^{-3}$
Bellamy(2025)	多偏移 GPR	雪芯	无	RMSE=0.96%–4.86%	RMSE=34–73 kg m^{-3}
Meehan等(2024)	场地 GPR	LiDAR, 雪坑	是	N/A	RMSE=27 kg m^{-3}
Sundström等(2012)	场地 GPR	雪坑	无	RMSE=1.8%–3.0%; 低估高达50%	N/A

同雪况下的局限性。Bellamy(2025)提出了多偏移距GPR技术,用于估算密度和LWC,其均方根误差(RMSE)分别为34–73 kg m^{-3} 和0.96%–4.86%,但需要进行侵入式雪芯校准。相比之下,框架仅使用合成和共偏移距GPR数据即可实现更高的精度,并已通过独立的实地测量进行了验证。通过减少对辅助采样和破坏性采样的依赖,该方法为大规模雪层监测提供了更高的便携性和操作效率。

4.3 物理可解释性和水文相关性

该框架的核心创新在于能够同时从探地雷达数据中反演双电磁参数(ϵ' 和 $\log(\rho)$)。介电常数(ϵ')控制波速,反映雪的密度和液态水含量,电阻率(ρ)则表征传导路径,控制湿雪中的衰减。与单参数方法相比,两者的联合反演能够提供更符合物理规律、更全面的雪层水文表征。

其水文意义重大。在潮湿和湿润条件下准确反演体积LWC和容重,可提高融雪预报和水平衡模型的保真度,尤其是在过渡融化阶段。与依赖简化衰减定律或几何重建的经验或基于迁移的技术相比,神经网络在雷达轨迹模式和电磁特性之间建立了广义关系。这绕过了特定地点的校准,并增强了在不同雪况下的稳健性。

冰冻圈探地雷达反演的最新进展进一步佐证了这一贡献。Meehan等(2024)将探地雷达和激光雷达与机器学习回归相结合,估算了雪层密度, Bellamy(2025)使用多偏移距探地雷达进行密度和液态水含量的反演, Xiao等(2025)开发了一种用于多年冻土测绘的探地雷达增强机器学习框架。这些研究表明机器学习和先进反演技术的作用日益增强,但大多数研究仍然局限于单参数反演或需要侵入式辅助校准。相比之下,该框架在保持物理可解释性的同时,提供了一种可扩展且非侵入式的解决方案,可同时解析两个电参数。这种双重能力确保了对雪层密度和液态水含量的一致估算,使该框架非常适合在不同条件下进行雪层监测。

4.4 未来验证的局限性和范围

本研究聚焦于阿尔泰山脉四月初的两条横剖面(T750和G125),代表了中度湿润的雪层。这些条件可能无法完全捕捉不同海拔、植被覆盖或融雪阶段雪层特性的巨大差异。未来的研究应致力于更广泛的空间和季节验证,以增强模型的通用性。

在极湿雪(8%–15%LWC)至泥雪(>15%LWC)条件下的性能是另一个限制因素,因为严重的衰减可能会降低反演精度。需要在这种极端条件下获取更多现场数据来定义模型极限并优化学习策略。虽然神经网络模型的浅层架构提供了可解释性和抗过拟合能力,但探索更深层的架构,例如卷积神经网络(CNN)或Transformer,可以提高对复杂雪层结构和雷达信号模式细微变化的灵敏度。这些架构对于高密度雪层或极端雪层条件尤其有用,而传统模型在这些条件下可能难以应对。

虽然雪当量(SWE)并非直接估算,但准确反演LWC、密度和深度为雪当量(SWE)评估奠定了坚实的基础。由于雪当量(SWE)是密度和深度的函数,因此精确的估算对于预测融雪动态、水资源可用性和气候变化影响至关重要。未来的工作还应考虑雪层特征,例如冰透镜和内部湿层,这些特征会影响雪崩风险和雪当量(SWE)的估算。

5 结论

本研究引入并验证了一个基于机器学习的反演框架,用于同时估算探地雷达(GPR)数据中的介电常数(ϵ')和对数电阻率(ρ)。该框架结合了gprMax中FDTD模拟生成的合成数据集、主成分分析(PCA)降维以及多种机器学习(ML)算法。在湿雪和湿雪条件下对对比样带(T750和G125)进行的实地评估证实了神经网络(NN)模型的稳健性,并与雪坑和探针测量结果高度一致。该框架准确检索了雪层的关键特性,包括体积含水量(θ_w)和体积密度

(ρ)，展现了从合成训练到实际条件的可靠泛化能力。

其核心创新在于其双参数反演技术，能够同步估算共偏移距探地雷达的介电常数和电阻率。该方法能够以物理一致性的方式推导雪层特性，同时避免特定地点的校准，从而为传统反演技术提供了一种非侵入式且可扩展的替代方案。随着该框架在更广泛的雪层类型和条件下得到进一步验证，其在雪层水文学、雪崩预报和气候敏感型水资源管理等领域具有巨大的实际应用潜力。

然而，由于有限的灵敏度以及相关天线效应，利用探地雷达实验数据准确估算电阻率仍然具有挑战性。未来的研究应进一步优化反演策略，并结合改进的天线模型，以确保获得可靠的定量结果。模拟中使用了Ricker子波，但若采用实测的天线小波，可提高与实际数据的吻合度。

补充材料 补充材料为作者提供的原始数据，作者对其学术质量和内容负责，详见网络版(<http://earthcn.scichina.com>)。

参考文献

- Alzubaidi L, Chlaib H K, Fadhel M A, et al. 2024. Reliable deep learning framework for the ground penetrating radar data to locate the horizontal variation in levee soil compaction. *Eng Appl Artif Intelligence*, 129: 107627
- Anwar A, Wenyi Y, Jing L, et al. 2023. Predicting the compressive strength of cellulose nanofibers reinforced concrete using regression machine learning models. *Cogent Eng*, 10: 2225278
- Bedi J, Toshniwal D. 2018. Attribute selection based on correlation analysis. In: Balas V E, Jain L C, Balas M M, eds. *Advances in Big Data and Cloud Computing*. Singapore: Springer. 45–55
- Belete D M, Huchaiah M D. 2022. Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. *Int J Comput Appl*, 44: 875–886
- Bellamy K M. 2025. Spatial and temporal variability of temperate snowpack properties in northern Maine and southeast Alaska with multi-offset ground-penetrating radar. Doctoral Dissertation. Orono: University of Maine
- Bengio Y, Courville A, Vincent P. 2013. Representation learning: A review and new perspectives. *IEEE Trans Pattern Anal Mach Intell*, 35: 1798–1828
- Biau G, Scornet E. 2016. A random forest guided tour. *TEST*, 25: 197–227
- Bonnell R, McGrath D, Hedrick A R, et al. 2023. Snowpack relative permittivity and density derived from near-coincident lidar and ground-penetrating radar. *Hydrol Process*, 37: e14996
- Bradford J H. 2007. Frequency-dependent attenuation analysis of ground-penetrating radar data. *Geophysics*, 72: J7–J16
- Bradford J H, Harper J T, Brown J. 2010. Correction to “Complex dielectric permittivity measurements from ground-penetrating radar data to estimate snow liquid water content in the pendular regime”. *Water Resour Res*, 46: 2009WR008959
- Chen T, Guestrin C. 2016. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD Int Conf Knowledge Discovery and Data Mining*. 785–794
- Coppola E, Raffaele F, Giorgi F. 2018. Impact of climate change on snow melt driven runoff timing over the Alpine region. *Clim Dyn*, 51: 1259–1273
- Denoth A. 1997. The monopole-antenna: A practical snow and soil wetness sensor. *IEEE Trans Geosci Remote Sens*, 35: 1371–1375
- Denoth A, Foglar A, Weiland P, et al. 1984. A comparative study of instruments for measuring the liquid water content of snow. *J Appl Phys*, 56: 2154–2160
- Dharmadasa V, Kinnard C, Baraër M. 2024. A new interpolation method to resolve under-sampling of UAV-lidar snow depth observations in coniferous forests. *Cold Regions Sci Tech*, 220: 104134
- Dupuy B, Lee M, Chatel M, et al. 2024. Snow measurements with UAV GPR tools. In: *Proceedings of NSG 2024 4th Conf Airborne, Drone and Robotic Geophysics*
- Filardi V, Cheung A, Khan R, et al. 2023. Data-driven soil water content estimation at multiple depths using SFCW GPR. In: *Proceedings of 2023 IEEE Int Opportunity Research Scholars Symp (ORSS)*
- Frolov D M, Seliverstov Y G, Koshurnikov A V, et al. 2024. GPR snow cover investigation to determine its spatial and temporal variability. *BIO Web Conf*, 113: 04003
- Giannakis I, Warren C, Giannopoulos A, et al. 2024. Investigating the capabilities of deep learning for processing and interpreting one-shot multi-offset GPR data: A numerical case study for lunar and Martian environments, arXiv: 2410.14386
- Granlund N, Lundberg A, Feiccabrino J, et al. 2009. Laboratory test of snow wetness influence on electrical conductivity measured with ground penetrating radar. *Hydrol Res*, 40: 33–44
- Hao J, Mind'je R, Feng T, et al. 2021. Performance of snow density measurement systems in snow stratigraphies. *Hydrol Res*, 52: 834–846
- Hearst M A, Dumais S T, Osuna E, et al. 1998. Support vector machines. *IEEE Intell Syst Their Appl*, 13: 18–28
- Heilig A, Mitterer C, Schmid L, et al. 2015. Seasonal and diurnal cycles of liquid water in snow—Measurements and modeling. *J Geophys Res-Earth Surf*, 120: 2139–2154
- Holbrook W S, Miller S N, Provart M A. 2016. Estimating snow water equivalent over long mountain transects using snowmobile-mounted ground-penetrating radar. *Geophysics*, 81: WA183–WA193
- Huang X, Deng J, Ma X, et al. 2016. Spatiotemporal dynamics of snow cover based on multi-source remote sensing data in China. *Cryosphere*, 10: 2453–2463
- Huang X, Liu C, Wang Y, et al. 2019. Snow cover variations across China from 1952–2012. *Cryosphere Discuss*, 2019: 1–21
- Jacobs J M, Hunsaker A G, Sullivan F B, et al. 2021. Snow depth mapping with unpiloted aerial system lidar observations: A case study in Durham, New Hampshire, United States. *Cryosphere*, 15: 1485–1500
- Jenssen R O R, Jacobsen S. 2020. Drone-mounted UWB snow radar: Technical improvements and field results. *J ElectroMagn Waves Appl*, 34: 1930–1954
- Jolliffe I T, Cadima J. 2016. Principal component analysis: A review and recent developments. *Phil Trans R Soc A*, 374: 20150202
- Kaczmarek P, Pietrasinski J. 2014. Principal component analysis in interpretation of A-scan measurements in GPR system. In: *Proceedings of 2014 15th Int Radar Symp (IRS)*
- Klotzsche A, Vereecken H, van der Kruk J. 2019. Review of crosshole ground-penetrating radar full-waveform inversion of experimental data: Recent developments, challenges, and pitfalls. *Geophysics*, 84: H13–H28
- Kovacs A, Gow A J, Morey R M. 1995. The *in-situ* dielectric constant of polar firm revisited. *Cold Regions Sci Tech*, 23: 245–256
- Küçükdemirci M, Sarris A. 2022. GPR data processing and interpretation based on artificial intelligence approaches: Future perspectives for archaeological prospection. *Remote Sens*, 14: 3377
- Liao L, Li H, Shang W, et al. 2022. An empirical study of the impact of

- hyperparameter tuning and model optimization on the performance properties of deep neural networks. *ACM Trans Softw Eng Methodol*, 31: 1–40
- Liu H, Xing B, Wang H, et al. 2019. Simulation of ground penetrating radar on dispersive media by a finite element time domain algorithm. *J Appl Geophys*, 170: 103821
- Liu Z, Gu X, Li J, et al. 2024. Deep learning-enhanced numerical simulation of ground penetrating radar and image detection of road cracks. *Chin J Geophys*, 67: 2455–2471
- Luce C H, Abatzoglou J T, Holden Z A. 2013. The missing mountain water: Slower westerlies decrease orographic enhancement in the Pacific Northwest USA. *Science*, 342: 1360–1364
- Lundberg A, Thunehed H, Bergström J. 2000. Impulse radar snow surveys —Influence of snow density. *Hydrol Res*, 31: 1–14
- Mankin J S, Viviroli D, Singh D, et al. 2015. The potential for snow to supply human water demand in the present and future. *Environ Res Lett*, 10: 114016
- Marshall H P. 2005. Snowpack spatial variability: Towards understanding its effect on remote sensing measurements and snow slope stability. Doctoral Dissertation. Boulder: University of Colorado at Boulder
- Marshall H P, Koh G, Forster R R. 2005. Estimating alpine snowpack properties using FMCW radar. *Ann Glaciol*, 40: 157–162
- McGrath D, Webb R, Shean D, et al. 2019. Spatially extensive ground-penetrating radar snow depth observations during NASA's 2017 SnowEx campaign: Comparison with *in situ*, airborne, and satellite observations. *Water Resour Res*, 55: 10026–10036
- McGrath D, Bonnell R, Zeller L, et al. 2022. A time series of snow density and snow water equivalent observations derived from the integration of GPR and UAV SfM observations. *Front Remote Sens*, 3: 886747
- Meehan T G, Hojatimalekshah A, Marshall H P, et al. 2024. Spatially distributed snow depth, bulk density, and snow water equivalent from ground-based and airborne sensor integration at Grand Mesa, Colorado, USA. *Cryosphere*, 18: 3253–3276
- Moran M L, Greenfield R J, Arcone S A, et al. 2000. Multidimensional GPR array processing using Kirchhoff migration. *J Appl Geophys*, 43: 281–295
- Mortimer C, Mudryk L, Derksen C, et al. 2020. Evaluation of long-term Northern Hemisphere snow water equivalent products. *Cryosphere*, 14: 1579–1594
- Mote P W, Li S, Lettenmaier D P, et al. 2018. Dramatic declines in snowpack in the western US. *NPJ Clim Atmos Sci*, 1: 2
- Namdari H, Moradikia M, Zekavat S, et al. 2025. Advancing precision agriculture: Machine learning-enhanced GPR analysis for root-zone soil moisture assessment in mega farms. *IEEE Trans Agri Elect*, 3: 98–109
- Neal A. 2004. Ground-penetrating radar and its use in sedimentology: Principles, problems and progress. *Earth-Sci Rev*, 66: 261–330
- Qin T, Bohlen T, Allroggen N. 2023. Full-waveform inversion of ground-penetrating radar data in frequency-dependent media involving permittivity attenuation. *Geophys J Int*, 232: 504–522
- Ran Y H, Li X, Lu L, 等. 2012. Large-scale land cover mapping with the integration of multi-source information based on the Dempster-Shafer theory. *Int J Geogr Inf Sci*, 26: 169–191
- Rott H, Nagler T, Scheiber R. 2003. Snow mass retrieval by means of SAR interferometry. In: Proceedings of FRINGE 2003 Workshop. Frascati
- Schmid L, Koch F, Heilig A, et al. 2015. A novel sensor combination (upGPR-GPS) to continuously and nondestructively derive snow cover properties. *Geophys Res Lett*, 42: 3397–3405
- Sharma S, Garg A, Shukla B K. 2025. Machine learning-based prediction of flexural strength in graphene-enhanced RC beams using ANN, GPR, REP tree, and AR_M5P models. *Asian J Civ Eng*, 26: 2991–3006
- Shهاب M, Al-Kaltakchi M T, Dukhan A, et al. 2025. Enhancing ground penetrating radar (GPR) data analysis utilizing machine learning. *J eng sustain dev*, 29: 321–330
- Sidorenko A I, Bobrova D A, Pletnev D, et al. 2023. GPR results for indicating layers of snow strata. *E3S Web Conf*, 389: 02012
- Sihvola A, Tiuri M. 2007. Snow fork for field determination of the density and wetness profiles of a snow pack. *IEEE Trans Geosci Remote Sens*, 45: 717–721
- Singh K K, Kumar A, Kulkarni A V, et al. 2016. Experimental investigation of dielectric properties of seasonal snow at field observatories in the northwest Himalaya. *Ann Glaciol*, 57: 319–327
- Stephan S M, Allroggen N, Tronicke J. 2024. Adding realistic noise models to synthetic ground-penetrating radar data. *Near Surf Geophys*, 22: 127–139
- Sturm M, Goldstein M A, Parr C. 2017. Water and life from snow: A trillion dollar science question. *Water Resour Res*, 53: 3534–3544
- Sun H, Yang X, Gong J, et al. 2022. Joint physics and data driven full-waveform inversion for underground dielectric targets imaging. *IEEE Trans Geosci Remote Sens*, 60: 1–11
- Sundström N, Kruglyak A, Friborg J. 2012. Modeling and simulation of GPR wave propagation through wet snowpacks: Testing the sensitivity of a method for snow water equivalent estimation. *Cold Regions Sci Tech*, 74–75: 11–20
- Tecler F, Pielmeier C. 2011. Point observations of liquid water content in wet snow-investigating methodical, spatial and temporal aspects. *Cryosphere*, 5: 405–418
- Tiuri M, Sihvola A, Nyfors E, et al. 1984. The complex dielectric constant of snow at microwave frequencies. *IEEE J Ocean Eng*, 9: 377–382
- Tjoelker A R, Baraër M, Valence E, et al. 2024. Drone-based ground-penetrating radar with manual transects for improved field surveys of buried ice. *Remote Sens*, 16: 2461
- Turner G, Siggins A F. 1994. Constant Q attenuation of subsurface radar pulses. *Geophysics*, 59: 1192–1200
- Valence E, Baraer M, Rosa E, et al. 2022. Drone-based ground-penetrating radar (GPR) application to snow hydrology. *Cryosphere*, 16: 3843–3860
- Vergnano A, Franco D, Godio A. 2022. Drone-borne ground-penetrating radar for snow cover mapping. *Remote Sens*, 14: 1763
- Warren C, Giannopoulos A, Giannakis I. 2016. gprMax: Open source software to simulate electromagnetic wave propagation for ground penetrating radar. *Comput Phys Commun*, 209: 163–170
- Warren C, Giannopoulos A, Gray A, et al. 2019. A CUDA-based GPU engine for gprMax: Open source FDTD electromagnetic simulation software. *Comput Phys Commun*, 237: 208–218
- Warren C, Giannopoulos A, Giannakis I, et al. 2021. gprMax: Open source software to simulate electromagnetic wave propagation for ground penetrating radar. *Eng Min Geophys* 2021
- Webb R W, Jennings K S, Fend M, et al. 2018. Combining ground-penetrating radar with terrestrial LiDAR scanning to estimate the spatial distribution of liquid water content in seasonal snowpacks. *Water Resour Res*, 54, doi: 10.1029/2018WR022680
- Webb R W, Marziliano A, McGrath D, et al. 2021. *In situ* determination of dry and wet snow permittivity: Improving equations for low frequency radar applications. *Remote Sens*, 13: 4617
- Webb R W, Musselman K N, Ciafone S, et al. 2022. Extending the vadose zone: Characterizing the role of snow for liquid water storage and transmission in streamflow generation. *Hydrol Process*, 36: e14541
- Xiao Y, Liu G, Hu G, 等. 2025. Mapping mountain permafrost via GPR-augmented machine learning in the northeastern Qinghai-Tibet Plateau. *Remote Sens*, 17: 2015
- Xue J, Huang Q, Wu S, et al. 2024. Real-time dual-parameter full-waveform inversion of GPR data based on robust deep learning. *Geophys J Int*, 238: 1755–1771
- Yang J, Jiang L, Luo J K, et al. 2020. Snow depth estimation and historical data reconstruction over China based on a random forest machine

- learning approach. *Cryosphere*, 14: 1763–1778
- Yee K. 1966. Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. *IEEE Trans Antennas Propagat*, 14: 302–307
- Yegnanarayana B. 2009. Artificial Neural Networks. New Delhi: PHI Learning Pvt. Ltd
- Zeng X, McMechan G A, Xu T. 2000. Synthesis of amplitude-versus-offset variations in ground-penetrating radar data. *Geophysics*, 65: 113–125
- Zhang L, Xu Y, Zeng Z, et al. 2021. Simulation of Martian near-surface structure and imaging of future GPR data from Mars. *IEEE Trans Geosci Remote Sens*, 60: 1–12
- Zhu W, Huang Q, Liu L, et al. 2020. Three-dimensional reverse time migration of ground-penetrating radar signals. *Pure Appl Geophys*, 177: 853–865

(责任编辑：岳天祥)