

基于电磁感应数据的电导率反演模型研究

吴家林¹, 彭杰^{1*}, 白建铎¹, 王佳文¹, 纪文君², 王楠³

(1.塔里木大学 植物科学学院, 新疆 阿拉尔 843300; 2.中国农业大学 土地科学与技术学院, 北京 100083; 3.浙江大学 环境与资源学院, 杭州 310058)

摘要:【目的】土壤盐渍化是限制新疆南部棉花高产的主要因子, 准确获取区域尺度土壤剖面盐分信息。【方法】以南疆阿拉尔垦区为研究区, 以田间尺度采集的 30 个不同盐渍化程度棉田的 540 个样点的 0~0.375、0~0.750、0~1.000 m 的土壤剖面电导率数据和对应的电磁感应数据为数据源, 采用线性模型和非线性模型分别构建了田间尺度和区域尺度的土壤剖面电导率的电磁感应反演模型, 并采用缩减建模样本量方法进一步检验了区域尺度模型的可靠性和稳定性。【结果】多元线性回归 (MLR)、偏最小二乘回归 (PLSR) 和主成分回归 (PCR) 建模方法的田间尺度模型 R^2 在 0.88~0.95, 而对应的区域尺度模型 R^2 在 0.34~0.53。基于随机森林 (RF)、神经网络 (NN) 和支持向量机 (SVM) 非线性建模方法构建的土壤剖面电导率的区域尺度电磁感应反演模型 R^2 在 0.60~0.85, 其中 RF 模型的精度最高。0~0.375、0~0.750、0~1.000 m 土壤剖面电导率的 RF 反演模型 R^2 分别为 0.80、0.85 和 0.84, 相较于线性建模方法的区域尺度模型精度有明显的提高。RF 区域尺度模型的样本数量由 540 个缩减到 240 个, 模型精度没有明显变化, 表明采用区域尺度模型, 可大幅度降低土壤剖面样本采集数量, 从而可显著提高采样效率和降低采样成本。【结论】区域尺度下构建土壤剖面电导率反演模型时, 随机森林建模方法效果较优, 模型预测能力具有较高的可靠性。

关键词: 电磁感应; 区域尺度; 土壤电导率; 反演模型

中图分类号: S156.41

文献标志码: A

doi: 10.13522/j.cnki.ggps.2020503

OSID:



吴家林, 彭杰, 白建铎, 等. 基于电磁感应数据的电导率反演模型研究[J]. 灌溉排水学报, 2021, 40(4): 80-87.

WU Jialin, PENG Jie, BAI Jianduo, et al. Calculating Electrical Conductivity of Soil Using Electromagnetic Induction Data[J]. Journal of Irrigation and Drainage, 2021, 40(4): 80-87.

0 引言

【研究意义】土壤盐渍化是现今世界干旱地区的主要农业问题^[1-3]。新疆地处西北干旱地区, 2018 年棉花种植面积已经占全国棉花种植面积的 4/5, 同时作为我国最大的盐土区, 其盐渍土面积达 1 100 万 hm^2 ^[4-5]。南疆作为棉花的主产区, 棉花产业长期受高蒸降比和灌溉排水不畅导致的土壤盐渍化以及次生盐渍化危害, 土壤盐渍化已成为新疆棉花产量和品质提升及持续发展的重大限制条件和障碍因素。表层土壤盐渍化信息由于时空变异性强烈, 不利于土壤盐渍化程度的客观诊断, 从而很难为土壤盐渍化改良和精准灌溉提供可靠的支撑依据, 而土壤剖面盐分信息相对于表层土壤盐分信息来讲, 其时空变异性相对较弱, 信息相对较稳定。掌握棉田土壤盐分空间变异性, 前提是大面积、准确、及时地掌握棉田盐渍化土壤剖面的盐分空间分布特征信息。传统的土壤剖面盐渍化调

查方法为野外定点采样结合室内分析测定, 多以挖剖面、土钻钻孔等侵入式为主, 且相关研究多局限于田间小尺度。在费时耗力和代表性差的双重影响下, 难以满足目前针对区域尺度的农田土壤盐渍化管理所急需的大面积土壤剖面盐分信息获取的需求^[6-9]。

【研究进展】电磁感应技术可非接触直接获取土壤剖面表观电导率信息, 具有快速、代表性强和精度高等优点, 适用于大面积土壤盐渍化的监测, 已被广泛应用于土壤盐渍化的调查与研究中^[10-12]。近年来, 大地电导率仪在土壤属性(土壤盐分、有机质、土壤含水率、阳离子交换量等)监测方面发挥着重要作用^[13-18]。杨劲松等^[19]以黄河三角洲海涂区为研究对象, 基于电磁感应数据并结合 GIS 与地统计学确定了最佳的空间插值方式, 分析了不同植被类型下土壤盐渍剖面分布特征。吴亚坤等^[20]利用线性建模方法构建了不同季节及不同土层土壤含盐量的解译模型, 从三维角度评估了不同季节土壤盐分变异特征。国内外学者同时期也进行了相关研究^[21-27]。【切入点】以往基于电磁感应数据所建立的土壤盐分解译模型研究区域多以田间尺度为主, 建模方法主要以线性模型居多。而不同条田的土壤盐分、土壤水分、阳离子交换量、土壤有机质以及土壤质地等均具有较大差异, 造成模型参数难以统一。因此, 基于田间尺度建立的电磁感

收稿日期: 2020-09-10

基金项目: 兵团中青年创新领军人才项目 (2020CB032); 国家重点研发计划项目 (2018YFE0107000)

作者简介: 吴家林 (1995-), 男, 硕士研究生, 主要从事土壤盐渍化遥感监测与数字制图研究。E-mail: jlwzky@163.com

通信作者: 彭杰 (1977-), 男, 教授, 主要从事资源与环境高光谱遥感监测机理与方法的研究。E-mail: pjzky@163.com

应线性解译模型能否外延至区域尺度,有待于进一步研究。此外,将非线性建模方法应用于区域尺度构建基于表观电导率的土壤电导率反演模型的报道甚少。在区域尺度上,探明表观电导率与土壤电导率之间是否满足线性关系,不同线性建模方法能否适用于区域尺度的土壤电导率反演模型研究,以及非线性建模方法相较于线性建模方法在区域尺度上是否有更好的应用前景,是电磁感应技术应用于大面积农田土壤剖面盐分信息获取所需解决的关键问题。【拟解决的关键问题】利用新疆南疆阿拉尔垦区内 30 个不同地点的棉田剖面表观电导率与土壤电导率数据,通过区域尺度模型与田间尺度模型的模型精度对比,不同线性模型和不同非线性模型在区域尺度下的模型精度对比,提出构建棉田土壤电导率反演模型的思路和方法,研究结果为 EM38-MK2 在区域尺度棉田土壤盐渍化监测中提供理论参考,为发展精准农业提供新思路。

1 材料与方法

1.1 研究区概况

研究区为新疆南部阿拉尔垦区(图 1),地理位置介于(80°30'—81°58'E, 40°22'—40°57'N),全垦区东西相距 281 km,南北相距 180 km。平均海拔高度 1 011 m,北部为冲积扇平原,南部为塔克拉玛干沙漠,地势北高南低,属典型暖温带极端大陆性干旱荒漠气候。且水土资源丰富,年均气温 10.7 °C,年日照 2 556.3~2 991.8 h,无霜期 220 d,年均蒸发量为 1 988 mm,年均降水量 50 mm,蒸降比约为 40 倍。土壤质地类型主要为壤土、沙壤土以及沙土。土壤盐分主要由硫酸盐、亚硫酸盐和氯化盐构成。地下水矿化度 0.6~6 g/L,地下水埋深 1~3 m^[28],土壤 pH 值在 7.26~9.23 之间变化。主要种植作物包括棉花、红枣、苹果和水稻等,其中棉花种植面积最大,约 1.55×10⁵hm²,土壤样品采集与表观电导率数据采集均在棉田中。受地形、强蒸发、降雨稀少以及高蒸降比、土地开垦、灌溉排盐系统不配套的影响,垦区内土壤盐渍化现象明显且盐渍化土壤分布广泛,致使农作物受到不同程度的盐渍化危害。

1.2 表观电导率数据和土壤样品采集

EM38-MK2 测量点和土样采集点如图 2 所示,本研究使用加拿大 GEONICS 公司生产的新式大地电导率仪 EM38-MK2 采集土壤表观电导率数据(EC_a , mS/m),采用美国犀牛钻机公司生产的 Rhino S1 土壤采集系统采集土壤剖面样品。EM38-MK2 工作时提供 2 种测定模式 EMH 和 EMV,每种模式可测定 2 个不同深度土壤剖面的表观电导率,EMH 模式下的测定深度为 0~0.375、0~0.750 m,表观电导率数据分别记录为 $EMH_{0.375}$ 和 $EMH_{0.75}$,EMV 模式下的测定深度为 0~0.750 和 0~1.500 m,表观电导率数据分别记录为 $EMV_{0.75}$ 和 $EMV_{1.5}$ 。

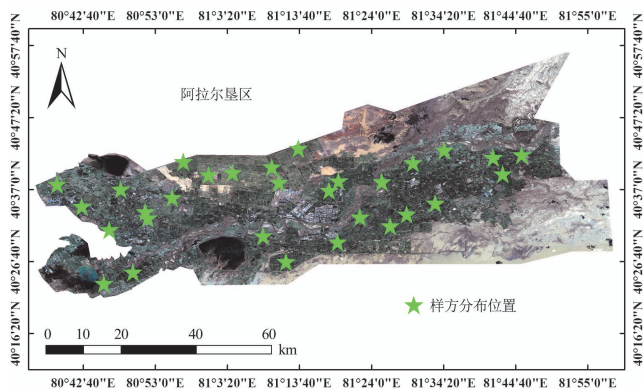


图 1 研究区地理位置与调查样方分布

Fig.1 The geographical location of the study area and the distribution of survey samples

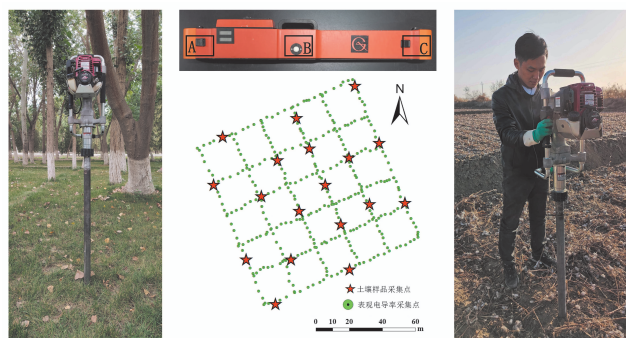


图 2 EM38-MK2 测量点和土样采集点分布

Fig.2 Distribution of EM38-MK2 measuring points and soil sample collection points

Rhino S1 土壤采集系统取样管长度为 1 m,内置中空 PVC 取芯管,取芯管直径 36 mm,可完整保存土壤样品的原状性。田间作业时,单管取样时间约 30~60 s。完成后取出取芯管获得深度为 1 m 的完整土壤剖面样品,用密封橡皮塞分别将取芯管底部和上部密封(底红顶黑)以防止水分蒸发和盐分流失。

综合考虑垦区面积大小、土壤盐渍化程度、棉花种植集中度、交通道路状况等因素,在阿拉尔垦区内棉花种植区域选取 30 个棉田布设样方,采集土壤表观电导率数据和土壤剖面样品。采样时间为 2019 年 3 月中下旬,如图 1 所示,在研究区中选取 30 个不同盐渍化程度棉田(面积均大于 10 hm²),棉田内布设大小为 100 m×100 m 的样方。

如图 2 所示,采集表观电导率数据时,为了了解样方内表观电导率数据的阈值范围,便于采样点位的选取,采集 6 条平行棉花种植方向和 6 条垂直棉花种植方向的 EMH 和 EMV 模式下连续线状表观电导率数据,形状为网格状,数据间距约 3 m。根据连续线状表观电导率数据采集时观测的阈值范围,样方内遵循代表性原则在连续线状表观电导率数据行走路线上采集 18 个特征样点表观电导率数据,每个样点 EMH 和 EMV 模式各测定 3 次,获取的表观电导率数据取平均值。并在 18 个表观电导率数据测定

位上使用 Rhino S1 土壤采集系统采集 18 管深度为 0~1 m 的土壤剖面样本数据, 30 个样方共采集 540 组特征样点的表观电导率数据, 每组包含 4 种表观电导率数据, 共计 2 160 个样点表观电导率数据。采集 540 管土壤样品数据, 用手持 GPS 记录每个采样点的纬度和经度 (误差精度小于 8 m), 每管土壤样品实验室处理为 3 个深度 (0~0.375、0~0.750、0~1.000 m), 共计 1 620 个土壤样品。在去除石子、植物根和地膜等残留物后将土壤样本置于牛皮纸上自然风干, 取风干土样研磨过 2 mm 筛备用, 每个深度样本数据保留 300 g 土壤, 按土水比为 (1:1) 比例制备滤液, 测定其土壤电导率, 用于构建电导率与表观电导率之间的反演模型。

1.3 建模方法及模型评价指标

本研究以 Rhino S1 土壤采集系统采集、处理、测定的 540 组土壤电导率 and 对应 EM38-MK2 测定的 540 组土壤表观电导率数据为样本总体, 以 3 个不同深度土壤剖面 (0~0.375、0~0.750、0~1.000 m) 电导率为因变量, 4 种测定位下的土壤表观电导率 ($EMH_{0.375}$ 、 $EMH_{0.75}$ 、 $EMV_{0.75}$ 、 $EMV_{1.5}$) 为自变量, 建立研究区棉田不同深度土壤剖面 (0~0.375、0~0.750、0~1.000 m) 电导率的反演模型。

建模方法为多元线性回归 (MLR)、主成分回归 (PCR)、偏最小二乘回归 (PLSR)、支持向量机 (SVM)、随机森林 (RF)、神经网络 (NN) 6 种。其中, RF 是一种构建神经网络, 利用多棵决策树对样本数据进行训练、分类和预测的机器学习方法, 各决策树单元间相互无关联且随机排列, 有效降低了单一分类器的分析误差, 其计算量大, 但提升了分类准确度和模型预测精度, 适用于大规模数据的高效处理^[29]。NN 是基于大量样本经过反复训练的学习过程, 能被赋予完成某项任务的具体功能, 在拥有传统数学模型中能处理海量多维数据的优点的同时, 还具有高度自由的非线性、泛化性等优良特性^[30]。建模思路为区域尺度模型与田间尺度模型 2 种, 区域尺度模型是将 30 个样方的

表观电导率数据和土壤电导率数据汇总, 建立统一的反演模型, 田间尺度模型针对单一样方的表观电导率数据和土壤电导率数据建立单独的反演模型。

模型预测精度和稳定性的评价指标为决定系数 (R^2)、均方根误差 (RMSE)、平均绝对百分比误差 (MAPE)、相对分析误差 (RPD)。 R^2 反映观测值与预测值间的相关性强度; RMSE 检验模型的预测能力; MAPE 检验模型的精度。相对于 RPD 而言, 预测精度分为 5 个等级, 当 $RPD < 1.5$ 时表明模型无法进行预测, 当 $1.5 < RPD < 2$ 时表明模型可进行粗略估计, 当 $2 \leq RPD < 2.5$ 时表明模型有较好的预测能力, 当 $2.5 \leq RPD < 3.0$ 时模型具有很好的预测能力, 当 $RPD \geq 3.0$ 时模型具有极好的预测能力^[4]。 R^2 和 RPD 大、RMSE 和 MAPE 小的模型, 其预测能力好, 稳定性高。

2 结果与分析

2.1 剖面土壤电导率统计特征

土壤剖面电导率统计特征见表 1。由表 1 可知, 不同深度土壤剖面的电导率介于 0.120~11.015 dS/m 之间, 变幅较大, 最小值分布于 0~0.375 m 土壤剖面中, 最大值分布于 0~0.750 m 土壤剖面中, 其中, 0~0.375 m 土壤剖面电导率范围为 0.120~8.963 dS/m, 0~0.750 m 土壤剖面的为 0.185~11.015 dS/m, 0~1.000 m 土壤剖面的为 0.354~10.650 dS/m。不同深度土壤剖面的电导率平均值在 2.080~3.769 dS/m, 0~0.375 m 的电导率最低, 0~0.750 m 的次之, 0~1.000 m 的最高, 具有随土壤剖面深度增加而增加的趋势, 这是棉花冬季灌水将表层土壤盐分淋溶到深层土壤中的结果。从变异系数来看, 各层土壤电导率的变异系数差异明显, 0~0.375、0~0.750、0~1.500 m 土壤剖面变异系数分别为 65.10%、57.00%、53.90%。通常情况下, 变异系数小于 10% 为弱变异, 变异系数 10%~100% 为中等变异, 变异系数大于 100% 为强变异^[31], 不同土壤深度下土壤电导率均呈中等变异强度。

表 1 剖面土壤电导率的统计特征值

Table 1 Statistical characteristic value of profile soil conductivity

土层深度/m	样品数	最小值/(dS m ⁻¹)	最大值/(dS m ⁻¹)	平均值/(dS m ⁻¹)	标准差/(dS m ⁻¹)	变异系数/%
0~0.375	540	0.120	8.963	2.080	1.355	65.10
0~0.750	540	0.185	11.015	3.252	1.853	57.00
0~1.000	540	0.354	10.650	3.769	2.031	53.90

2.2 基于线性建模方法的田间尺度模型精度

为验证线性建模方法在田间尺度模型中的可靠性, 选用 MLR、PLSR 和 PCR 等线性建模方法建立田间尺度模型。以每个棉田样方采集的 18 个土壤样本为总体, 采用分层抽样的方法按 2:1 比例分为建模集与预测集, 其中建模集 12 组样本, 预测集 6 组样本, 将 30 个棉田总体样本数累加。

不同线性模型反演精度见表 2。由表 2 可知, 在建模集中, 3 种线性建模方法的不同深度土壤剖面模型精度差异较小, 0~0.375 m 土壤深度 MLR 模型精度低于 PLSR 和 PCR, R^2 为 0.88, RMSE、MAPE 分别为 0.50 dS/m 和 0.21, 0~0.750 m 和 0~1.000 m 土壤深度下, PLSR 和 PCR 基本保持一致, 0~0.750 m 的 PLSR 模型精度最高, R^2 为 0.95, RMSE、MAPE 分别

为 0.41 dS/m 和 0.12。在验证集中, 3 种不同深度土壤剖面电导率反演模型的各指标相较于建模集模型精度均有一定程度下降, 但不同建模方法 R^2 在 0.83 以上, 表明模型稳定不同模型预测集中 RPD 最高值为 2.78, 最低值为 2.35, 可见模型具有较好或很好的

预测能力。不同深度土壤剖面电导率模型反演效果优劣排序依次为 $PLSR > PCR > MLR$ 。在田间尺度模型中表观电导率与土壤电导率满足线性关系, 所选用的 3 种线性建模方法在区域尺度棉田中适用性较好。

表 2 不同线性建模方法的田间尺度模型精度

Table 2 Field scale model accuracy of different linear modeling methods

建模方法	土层深度/m	建模			验证			
		R^2	$RMSE/(dS\ m^{-1})$	$MAPE$	R^2	$RMSE/(dS\ m^{-1})$	$MAPE$	RPD
MLR	0~0.375	0.88	0.50	0.21	0.84	0.56	0.24	2.36
	0~0.750	0.94	0.48	0.13	0.87	0.67	0.18	2.68
	0~1.000	0.92	0.59	0.14	0.87	0.78	0.18	2.55
PLSR	0~0.375	0.91	0.42	0.19	0.84	0.54	0.23	2.42
	0~0.750	0.95	0.41	0.12	0.86	0.73	0.16	2.47
	0~1.000	0.95	0.46	0.12	0.88	0.72	0.17	2.78
PCR	0~0.375	0.91	0.42	0.19	0.83	0.56	0.23	2.35
	0~0.750	0.95	0.42	0.12	0.88	0.67	0.16	2.68
	0~1.000	0.95	0.47	0.12	0.88	0.71	0.17	2.78

2.3 基于线性建模方法的区域尺度模型精度

选取 MLR、PLSR 和 PCR 3 种建模方法对土壤电导率进行预测并进行模型精度检验, 将田间尺度模型的建模集与预测集分别汇总后得到区域尺度模型的

建模集与预测集, 其中建模集 360 组土壤样本数据, 预测集 180 组土壤样本数据。表观电导率数据为自变量, 土壤电导率为因变量, 构建基于线性建模方法的区域尺度模型。

表 3 不同线性建模方法的区域尺度模型精度

Table 3 Regional scale model accuracy of different linear modeling methods

建模方法	土层深度/m	建模			验证			
		R^2	$RMSE/(dS\ m^{-1})$	$MAPE$	R^2	$RMSE/(dS\ m^{-1})$	$MAPE$	RPD
MLR	0~0.375	0.34	1.08	0.57	0.33	1.15	0.55	1.22
	0~0.750	0.53	1.26	0.41	0.51	1.33	0.44	1.42
	0~1.000	0.49	1.44	0.42	0.48	1.48	0.43	1.38
PLSR	0~0.375	0.35	1.08	0.57	0.33	1.15	0.55	1.22
	0~0.750	0.51	1.28	0.42	0.48	1.34	0.46	1.37
	0~1.000	0.48	1.45	0.43	0.47	1.49	0.43	1.37
PCR	0~0.375	0.34	1.08	0.57	0.33	1.15	0.55	1.22
	0~0.750	0.51	1.28	0.42	0.46	1.38	0.46	1.37
	0~1.000	0.48	1.46	0.43	0.47	1.49	0.43	1.37

不同线性建模方法的区域尺度模型精度评价结果如表 3 所示, 在建模集中, 3 种建模方法的模型精度差异不明显, 0~0.375 m 土壤深度 PLSR 模型精度略优于 MLR 和 PCR, R^2 为 0.35, $RMSE$ 、 $MAPE$ 分别为 1.08 dS/m 和 0.57, 0~0.750 m 和 0~1.000 m 土壤深度下, PLSR 和 PCR 模型精度基本保持一致, 且均低于 MLR, 在 0~0.750 m 的 MLR 模型精度最高, R^2 为 0.53, $RMSE$ 、 $MAPE$ 分别为 1.26 dS/m 和 0.41。在验证集中, 3 种模型的各指标相较于建模集均没有明显下降, 表明模型比较稳定, MLR 模型的各项指标略优于 PLSR 和 PCR。MLR、PLSR 和 PCR 建模方法的 RPD 均 < 1.5 , 说明 3 种线性模型不具备对全区样本数据进行预测的能力。由于土壤中大多数溶解盐类均以离子状态存在, EM38-MK2 测定的表观电导率数

据反映的是土壤空间范围内游离态电解质量, 同一棉田样方中土壤质地和土壤含水率等影响表观电导率数据的因素变化基本稳定, 田间尺度模型对样本有较好或很好的预测能力。区域尺度模型考虑到不同棉田样方的地域跨度较大, 土壤质地、地下水位和土壤含水率等因素变化强烈, 降低了区域尺度模型的精度, 导致区域尺度模型的预测能力大幅下降。

2.4 基于非线性建模方法的区域尺度模型精度

选用 RF、NN 和 SVM 3 种非线性建模方法构建基于表观电导率数据的土壤剖面电导率反演模型。结果如表 4 所示, 相较于 MLR、PLSR 和 PCR 线性建模方法, RF、NN 和 PCR 区域尺度模型精度均有明显提高。在建模集中, 3 种建模方法的精度差异较大, 0~0.750 m 深度 RF 模型精度最高, R^2 为 0.85, $RMSE$ 和 $MAPE$

分别为 0.78 dS/m 和 0.27, 0~0.375 m 深度剖面的 R^2 、 $RMSE$ 和 $MAPE$ 分别为 0.80、0.61 dS/m 和 0.33, 为 RF 模型中最低值。0~1.000 m 剖面的模型精度介于二者之间, 略低于 0~0.750 m 深度下模型精度。NN 和 SVM 模型精度有不同程度的降低, 但不同深度的模型精度变化趋势与 RF 保持一致, 具有随土壤剖面电导率增加而升高的趋势。综上所述, 建模集中, RF 模型性能明显优于 NN 和 SVM 模型, 反演效果优劣排序依次为 RF 模型>NN 模型>SVM 模型。在验证集中, 不同深度的 RF 模型精度相较于建模集均没有明显差

异, 且 RPD 指标均>2.0, 说明模型比较稳定, 能在区域尺度下较好地对本样数据进行预测。NN 模型的 R^2 高于 SVM, $RMSE$ 和 $MAPE$ 低于 SVM。不同深度下 NN 和 SVM 的 RPD 指标均介于 1.5~2.0 之间, R^2 指标介于 0.54~0.72 之间, 说明 NN 和 SVM 两模型均只能对样本数据进行粗略估计。综合考察各模型建模集与预测集的评价指标, 3 种模型的预测性能和稳定性从高到低排序依次为 RF 模型>NN 模型>SVM 模型, 表明非线性建模方法可有效改善基于土壤剖面电导率的区域尺度模型的预测精度。

表 4 不同非线性建模方法的区域尺度模型精度对比

Table 4 Comparison of regional scale model accuracy of different nonlinear modeling methods

建模方法	土层深度/m	建模			验证			
		R^2	$RMSE/(dS\ m^{-1})$	$MAPE$	R^2	$RMSE/(dS\ m^{-1})$	$MAPE$	RPD
RF	0~0.375	0.80	0.61	0.33	0.79	0.64	0.35	2.13
	0~0.750	0.85	0.78	0.27	0.83	0.86	0.27	2.47
	0~1.000	0.84	0.90	0.26	0.81	0.91	0.29	2.34
NN	0~0.375	0.61	1.23	0.4	0.58	1.20	0.44	1.71
	0~0.750	0.76	1.32	0.31	0.72	1.10	0.34	1.95
	0~1.000	0.73	1.47	0.35	0.66	1.28	0.38	1.86
SVM	0~0.375	0.60	1.18	0.61	0.54	1.26	0.65	1.61
	0~0.750	0.74	1.4	0.47	0.67	1.48	0.53	1.86
	0~1.000	0.71	1.58	0.49	0.64	1.61	0.56	1.66

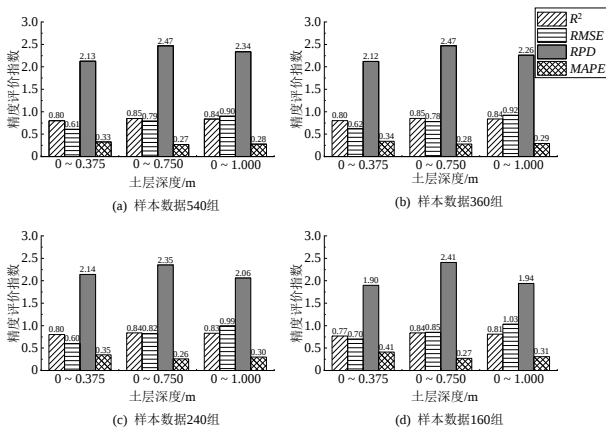


图 3 RF 模型不同样本数据量下的精度指标

Fig.3 RF model accuracy index under different sample data sizes

2.5 基于 RF 模型的不同样本数据量精度指标比较

区域尺度模型中应用效果最优的为 RF 模型, 为探究其能否在缩小数据量的条件下获得同样的模型精度, 以期达到减少样本数据获取量, 减少人力、物力以及时间成本的投入, 对模型可靠性进行验证的目的, 随机抽取总样本数量分别设置 540、360、240、160 共 4 个样本数据量梯度, 总样本数据量为 160 时, 无法按照 2:1 的比例进行建模和预测, 调整建模集数据量为 107, 预测集数据量为 53。结果如图 3 所示, 从 0~0.375 m 深度来看, 540、360 和 240 样本数据量梯度下模型精度差异不大, R^2 均为 0.80, $RMSE$ 在 0.60~0.62 dS/m 波动, $MAPE$ 在 0.33~0.35 之间, $2.0 < RPD < 2.5$, 不同样本数据量下模型均有较好

的预测能力。当样本数据量为 160 时, R^2 从 0.80 降低至 0.77, $RMSE$ 和 $MAPE$ 数值上升, $1.50 < RPD < 2.0$, 说明模型对样本数据具有粗略估计的能力。

0~0.750 m 深度下, 4 种不同数据量的 RF 模型 R^2 稳定, $RMSE$ 在 0.78~0.85 dS/m 波动, $MAPE$ 在 0.26~0.28, RPD 保持在 2.35~2.47 之间, 表明样本数据量减少模型仍具有较好的预测能力。在 0~1.000 m 深度下, 随样本数据量的减少 R^2 由 0.84 降低至 0.81, $RMSE$ 和 $MAPE$ 呈上升趋势, RPD 由 2.34 降低为 1.94, 模型预测能力由较好预测等级降低至粗略估计等级。总体分析, 样本数据量 540、360、240 共 3 个梯度时, RF 模型保留了原有对样本数据的预测能力。在样本数据量为 160 时, 模型精度有所下降, 但能对样本数据进行粗略估计。4 个不同样本数据梯度中, $RMSE$ 、 $MAPE$ 和 RPD 等精度评价指标出现随样本数据量的减少反而略微升高的现象, 造成该现象的原因可能是在缩减样本数据时剔除了异常样本数据。

3 讨论

本研究表明, 采用线性建模方法时, 区域尺度模型反演精度明显低于田间尺度模型, 二者具有明显差异性, 不同线性建模方法无法有效提高区域尺度模型的反演精度, 且各线性模型的预测能力均较低。采用非线性建模方法时, 区域尺度模型的模型反演精度显著提高, 最佳模型预测能力提升至对样本数据有较好的预测能力。根据田间实地采样考察, 实验室分析和

EM38-MK2 工作原理,其原因可能有以下几点:①土壤含水率不同,表观电导率受土壤含水率影响^[4,32],研究区棉田冬季和春季灌溉用水主要引于塔里木河和阿克苏河,灌溉顺序大致为由西向东,部分地区由于灌溉配套系统不完善和灌水量不足,不进行冬季和春季灌水或少灌,导致各棉田土壤含水率不同,造成了区域尺度模型和田间尺度模型的精度差异。②土壤质地类型不同,在土壤温度、土壤含水率相对一致的前提下,质地越轻,表观电导率读数越小。而全垦区内各棉田地地域跨度较大,棉田土壤质地类型主要有壤土、沙壤土和沙土 3 种,不同土壤质地类型对线性建模方法在区域尺度下模型的精度造成了影响。③土壤有机质不同,土壤中固体土粒由矿物质和有机质组成,而形成有机质腐殖质具有带电性,其在土壤中量的多少可引起大地电导率仪磁场的变化。在土壤质地和含水率相对一致的前提下,土壤有机质量的多少会造成表观电导率测定值的高低,二者正相关。不同棉田由于本身肥力和施用有机肥量各不相同,造成了线性模型无法对区域尺度的土壤电导率进行反演^[33]。④建模方法不同,非线性建模方法相较于线性建模方法具有各决策树单元互无关联且随机排列,能有效降低区域尺度模型的分析误差,提高了模型的精度。综上,在同一棉田中,土壤含水率、土壤质地类型和土壤有机质相对保持一致,模型参数稳定,表观电导率和土壤电导率二者保持线性关系,线性建模方法在田间尺度高度适用,而不同棉田中,土壤含水率、土壤质地类型和土壤有机质量各不相同,导致土壤电导率反演模型参数不同,从而造成了在区域尺度下线性建模方法无法对样本进行预测,而非线性建模方法由于其本身具备的高度非线性可有效提高区域尺度土壤电导率反演模型精度。

如何有效筛选出能进一步提高区域尺度模型反演精度的建模方法,对利用 EM38-MK2 获取区域尺度农田土壤盐渍化信息有重要意义。面对区域尺度更大,种植作物类型更复杂的土壤盐渍化信息获取的研究,还有待于进一步探索。

4 结 论

适用于田间尺度模型的 MLR、PLSR 和 PCR 线性建模方法无法对区域尺度模型的样本数据进行预测。在田间尺度模型中不同线性建模方法 R^2 在 0.88~0.95 之间,在区域尺度模型中不同线性建模方法 R^2 降低至 0.34~0.53 之间,无法对样本数据进行有效预测。将 RF、NN 和 SVM 等非线性建模方法应用于区域尺度模型中,相较于线性建模方法,区域尺度模型的模型精度明显提高,不同非线性建模方法的 R^2 在 0.60~0.85 之间,其中,NN 和 SVM 建模方法只能对样本数据进行粗略估计,模型精度最高、预测能力最好的 RF 模型在不同深度下均保持了对样本数据

较好的预测能力。通过缩减 RF 区域尺度模型的样本数量,验证了 RF 区域尺度模型可大幅降低采集土壤剖面的样本数量,从而提高采样效率和降低采样成本。

参考文献:

- [1] PENG J, BISWAS A, JIANG Q, et al. Estimating soil salinity from remote sensing and terrain data in southern Xinjiang Province, China[J]. *Geoderma*, 2019, (337): 1 309-1 319.
- [2] LI J, PU L, HAN M, et al. Soil salinization research in China: advances and prospects[J]. *Journal of Geographical Sciences*, 2014, 24(5): 943-960.
- [3] 李艳菊, 丁建丽, 米热古力·艾尼瓦尔. 渭-库绿洲土壤剖面盐分分布特征及驱动因子分析[J]. *灌溉排水学报*, 2019, 38(6): 58-65.
- LI Y J, DING J L, AINIWAR M. Distribution characteristics and driving factors of soil salinity in Wei-Ku Oasis[J]. *Journal of Irrigation and Drainage*, 2019, 38(6): 58-65.
- [4] 刘新路, 彭杰, 冯春晖, 等. 基于电磁感应仪数据的南疆棉田土壤电导率反演模型研究[J]. *土壤学报*, 2020, 57 (3): 646-655.
- LIU X L, PENG J, FENG C H, et al. Inversion model for soil conductivity in cotton field in South Xinjiang based on EM38-MK2 data[J]. *Acta Pedologica Sinica*, 2020, 57 (3): 646-655.
- [5] 刘新路, 彭杰, 冯春晖, 等. 基于电磁感应数据棉田土壤盐分剖面时空变异特征研究[J]. *中国土壤与肥料*, 2020 (1): 17-23.
- LIU X L, PENG J, FENG C H, et al. Study on spatial and temporal variation characteristics of soil profile salt in cotton field based on EM38-MK2 data[J]. *Soil and Fertilizer Sciences in China*, 2020(1): 17-23.
- [6] 范未华, 轩俊伟, 李保国, 等. 长期滴灌棉田表层土壤盐分时空变化特征[J]. *灌溉排水学报*, 2020, 39 (11): 83-89.
- FAN W H, XUAN J W, LI B G, et al. Temporal and spatial changes of surface soil salinity in cotton fields with long-term drip irrigation[J]. *Journal of Irrigation and Drainage*, 2020, 39(11): 83-89.
- [7] 彭杰, 刘焕军, 史舟, 等. 盐渍化土壤光谱特征的区域异质性及盐分反演[J]. *农业工程学报*, 2014, 30(17): 167-174.
- PENG J, LIU H J, SHI Z, et al. Regional heterogeneity of hyperspectral characteristics of salt-affected soil and salinity inversion[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2014, 30(17): 167-174.
- [8] 彭杰, 王家强, 向红英, 等. 土壤含盐量与电导率的高光谱反演精度对比研究[J]. *光谱学与光谱分析*, 2014, 34(2): 510-514.
- PENG J, WANG J Q, XIANG H Y, et al. Comparative study on hyperspectral inversion accuracy of soil salt content and electrical conductivity[J]. *Spectroscopy and Spectral Analysis*, 2014, 34(2): 510-514.
- [9] SHEPHERD K D, WALSH M G. Development of reflectance spectral libraries for characterization of soil properties[J]. *Soil science society of America journal*, 2002, 66(3): 988- 998 .
- [10] ZARE E, ARSHAD M, ZHAO D, et al. Two-dimensional time-lapse imaging of soil wetting and drying cycle using EM38 data across a flood irrigation cotton field[J]. *Agricultural Water Management*, 2020(241): 106383.
- [11] DOOLITTLE J A, BREVIK E C. The use of electromagnetic induction techniques in soils studies[J]. *Geoderma*, 2014(223): 33-45.
- [12] BREVIK E C, FENTON T E, HORTON R. Effect of daily soil temperature fluctuations on soil electrical conductivity as measured with the Geonics® EM-38[J]. *Precision Agriculture*, 2004, 5(2): 145-152.
- [13] FARZAMIAN M, SANTOS F A M, KHALIL M A. Application of EM38 and ERT methods in estimation of saturated hydraulic

- conductivity in unsaturated soil[J]. *Journal of applied geophysics*, 2015, (112): 175-189.
- [14] ZHAO D, LI N, ZARE E, et al. Mapping cation exchange capacity using a quasi-3d joint inversion of EM38 and EM31 data[J]. *Soil and Tillage Research*, 2020(200): 104618.
- [15] 杨劲松, 姚荣江. 基于磁感式土壤表观电导率空间变异性的插值方法比较[J]. *农业工程学报*, 2007, 23(9): 50-57.
- YANG J S, YAO R J. Comparison of interpolation approaches based on spatial variability of apparent soil electrical conductivity with an electromagnetic induction[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2007, 23(9): 50-57.
- [16] LI X M, YANG J S, LIU M X, et al. Spatio-temporal changes of soil salinity in arid areas of south Xinjiang using electromagnetic induction[J]. *Journal of Integrative Agriculture*, 2012, 11(8): 1 365-1 376.
- [17] DOOLITTLE J A, BREVIK E C. The use of electromagnetic induction techniques in soils studies[J]. *Geoderma*, 2014(223): 33-45.
- [18] RODRIGUES JR F A, BRAMLEY R G V, GOBBETT D L. Proximal soil sensing for precision agriculture: Simultaneous use of electromagnetic induction and gamma radiometrics in contrasting soils[J]. *Geoderma*, 2015(243): 183-195.
- [19] 杨劲松, 姚荣江, 刘广明. 电磁感应仪用于土壤盐分空间变异性的指示克立格分析评价[J]. *土壤学报*, 2008, 45(4): 585-593.
- YANG J S, YAO R J, LIU G M. Analysis and evaluation on spatial variability of soil salinity by indicator Kriging with an electromagnetic induction gauge[J]. *Acta Pedologica Sinica*, 2008, 45(4): 585-593.
- [20] 吴亚坤, 刘广明, 苏里坦, 等. 基于电磁感应研究新疆土壤盐分三维空间变异对季节的响应[J]. *农业工程学报*, 2017, 33(14): 173-178.
- WU Y K, LIU G M, SU L T, et al. Response of three-dimensional spatial variability of soil salinity to change of season of Xinjiang based on electromagnetic induction[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2017, 33(14): 173-178.
- [21] MOGHADAS D, TAGHIZADEH MEHRJARDI R, TRIANTAFILIS J. Probabilistic inversion of EM38 data for 3D soil mapping in central Iran[J]. *Geoderma Regional*, 2016, 7(2): 230 -238.
- [22] GRELLIER S, FLORSCH N, CAMERLYNCK C, et al. The use of Slingram EM38 data for topsoil and subsoil geoelectrical characterization with a Bayesian inversion[J]. *Geoderma*, 2013(200): 140-155.
- [23] HEIL K, SCHMIDHALTER U. Comparison of the EM38 and EM38-MK2 electromagnetic induction-based sensors for spatial soil analysis at field scale[J]. *Computers and Electronics in Agriculture*, 2015(110): 267-280.
- [24] WATSON H D, NEELY H L, MORGAN C L S, et al. Identifying subsoil variation associated with gilgai using electromagnetic induction [J]. *Geoderma*, 2017(295): 34-40.
- [25] YAN G, YIN Z, ZHOU L, et al. Using proximal sensor data for soil salinity management and mapping[J]. *Journal of Integrative Agriculture*, 2019, 18(2): 340-349.
- [26] DING J, YU D. Monitoring and evaluating spatial variability of soil salinity in dry and wet seasons in the Werigan-Kuqa Oasis, China, using remote sensing and electromagnetic induction instruments[J]. *Geoderma*, 2014(235): 316-322.
- [27] RHOADES J D, RAATS P A C, PRATHER R J. Effects of liquid - phase electrical conductivity, water content, and surface conductivity on bulk soil electrical conductivity[J]. *Soil Science Society of America Journal*, 1976, 40(5): 651-655.
- [28] 赵凯, 罗东海, 郭超, 等. 新疆阿拉尔地区土壤盐渍化特征[J]. *中国煤炭地质*, 2018, 30(10): 46-49.
- ZHAO K, LUO D H, GUO C, et al. Soil Salinization Features in Aral Area, Xinjiang[J]. *Coal geology of China*, 2018, 30(10): 46-49.
- [29] 庄巧蕙. 基于改进随机森林算法的研究与应用[D]. 厦门: 华侨大学, 2019.
- ZHUANG Q H. Research and application of improved random forest algorithms[D]. Xiamen: Huaqiao University, 2019.
- [30] JIN P, LU L, TANG Y, et al. Quantifying the generalization error in deep learning in terms of data distribution and neural network smoothness[J]. *Neural Networks*, 2020(130): 85-99.
- [31] 李晓明, 杨劲松, 李冬顺. 基于电磁感应(EM38)典型半干旱区土壤盐分空间变异研究[J]. *土壤通报*, 2010, 41(3): 695-699.
- LI X M, YANG J S, LI D S. Spatial variability of soil salinity in typical semiarid area based on electromagnetic induction(EM38)[J]. *Chinese Journal of Soil Science*, 2010, 41(3): 695-699.
- [32] 吴家林, 冯春晖, 刘新路, 等. 基于电磁感式大地电导率仪的南疆干旱区土壤电导率反演研究[J]. *土壤通报*, 2019, 50(6): 1 278-1 284.
- WU J L, FENG C H, LIU X L, et al. Inversion of soil electrical conductivity based on electromagnetic induction data in the arid region of Southern Xinjiang[J]. *Chinese Journal of Soil Science*, 2019, 50(6): 1 278-1 284.
- [33] 赵军伟, 蒋平安, 盛建东, 等. EM38 电磁发生仪测定结果的影响因素分析[J]. *干旱区地理*, 2005, 28(3): 362-366.
- ZHAO J W, JIANG Q A, SHENG J D, et al. Analysis on the factors affecting the soil salt contents measured with electromagnetism generator Em38[J]. *Arid Land Geography*, 2005, 28(3): 362-366.

Calculating Electrical Conductivity of Soil Using Electromagnetic Induction Data

WU Jialin¹, PENG Jie^{1*}, BAI Jianduo¹, WANG Jiawen¹, JI Wenjun², WANG Nan³

(1.College of Plant Sciences, Tarim University, Alar 843300, China;

2.College of Land Science and Technology, China Agricultural University, Beijing 100083, China;

3.College of Environment and Resources, Zhejiang University, Hangzhou 310058, China)

Abstract: 【Background】 Along with drought, soil salinization is one of the most important abiotic stresses facing agricultural production in arid and semi-arid regions like Xinjiang in northwest China. Saline soil is estimated to have reached 11 million hm² in Xinjiang, and is a major limiting factor in cotton industry in this region. Accurately measuring soil salinity is hence important to safeguard cotton production in Xinjiang. **【Objective】** The purpose of this

paper is to present a new method to estimate salinity distribution at field and regional scales to help improve irrigation and cultivation management. **【Method】** The experiments were conducted at Alar reclamation area in southern Xinjiang. Electrical conductivity of 540 soil samples taken from 30 cotton fields with different salinization levels were measured in soil profile at 0~0.375, 0~0.750, and 0~1.000 m depth. Using the electromagnetic induction data, linear and nonlinear models were constructed to inversely calculate the electrical conductivity of soil at field and regional scales respectively. Stability and reliability of the models was verified against ground-true data using the sample-size-reduction method. **【Result】** At field scale, the coefficient of determination R^2 associated with the multiple linear regression model (MLR), the partial least square regression (PLSR) model and the principal component regression model (PCR) varied from 0.88 to 0.95, while their associated R^2 for regional scale was from 0.34 to 0.53. The R^2 of the nonlinear model built on the random forest (RF), neural network (NN) and support vector machine (SVM) varied between 0.60 and 0.85, with the RF most accurate. The R^2 of the RF model for calculating the electrical conductivity of the soil profile at 0~0.375, 0~0.750, 0~1.000 m was 0.80, 0.85 and 0.84, respectively. Compared with the linear model, the model for regional scale significantly improved the accuracy. The sample size in the RF model for regional scale was reduced from 540 to 240, while the accuracy remained almost unchanged, indicating the regional-scale model can reduce the number of soil profile without compromising modelling accuracy.

【Conclusion】 For constructing the inversion model for estimating soil-profile electrical conductivity at regional scale, the random forest method is most accurate.

Key words: electromagnetic induction; regional scale; soil electrical conductivity; inversion model

责任编辑：白芳芳

关于评选优秀论文的公告

本刊已开启优秀论文评选活动，每年评选优秀论文 10 篇，每篇奖励 800 元，并颁发获奖证书，届时将在期刊网站首页展示，同时微信公众号推送。欢迎广大读者、作者积极向我刊投稿。

《灌溉排水学报》编辑部