doi:10.6041/j.issn.1000-1298.2018.05.028

# 基于岭回归的土壤含水率高光谱反演研究

 

 张智韬<sup>1,2</sup>
 王海峰<sup>1,2</sup>
 KARNIELI Arnon<sup>3</sup>
 陈俊英<sup>1,2</sup>
 韩文霆<sup>2</sup>

 (1. 西北农林科技大学水利与建筑工程学院,陕西杨凌 712100; 2. 西北农林科技大学中国旱区节水农业研究院, 陕西杨凌 712100; 3. 本古里安大学 Jocob Blaustein 沙漠研究所,思德博克 84990)

**摘要**: 以以色列南部 Seder Boker 地区采集的粘壤土样品为研究对象。在室内利用 ASD FieldSpec 3 型高光谱仪获 取土壤的原始光谱,在进行数据预处理和不同数学变换后,共获取了 4 种光谱指标:光谱反射率(REF)、倒数之对 数(LR)、一阶微分(FDR)和去包络线(CR)。采用偏最小二乘回归法(PLSR)、逐步回归法(SR)和岭回归法(RR) 构建了基于不同指标的土壤含水率高光谱反演模型,并对反演结果进行精度验证与比较。结果表明:REF - PLSR 模型在所有回归模型中的反演与预测效果均为最优( $R_e^2$  = 0.990, $R_p^2$  = 0.987),在逐步回归模型和岭回归模型中,LR -SR( $R_e^2$  = 0.981, $R_p^2$  = 0.971)、LR - RR( $R_e^2$  = 0.975, $R_p^2$  = 0.979)均为最佳模型。对于其他 3 种指标,虽然逐步回归法 和岭回归法的建模效果较偏最小二乘回归法略有下降,但  $R_e^2$ 均大于 0.9, $R_p^2$ 均大于 0.8,RPD 均大于 2.5,RMSE 均 小于 0.03,模型仍具有较好的反演效果;逐步回归法和岭回归法均实现了模型的简化,但岭回归法采用有偏估计从 而提高了模型的稳健性,且实现了波段的优选(用于建模的波段数仅为全光谱的 0.3%)。粘壤土土壤含水率 LR -RR 高光谱反演模型的建立为高光谱模型的优化、土壤含水率的快速测定提供了途径。

关键词:土壤含水率;高光谱遥感;回归分析;岭回归;波段选择

中图分类号: S127; S152.7 文献标识码: A 文章编号: 1000-1298(2018)05-0240-09

# Inversion of Soil Moisture Content from Hyperspectra Based on Ridge Regression

ZHANG Zhitao<sup>1,2</sup> WANG Haifeng<sup>1,2</sup> KARNIELI Arnon<sup>3</sup> CHEN Junying<sup>1,2</sup> HAN Wenting<sup>2</sup>
(1. College of Water Resources and Architectural Engineering, Northwest A&F University, Yangling, Shaanxi 712100, China
2. Institute of Water Saving Agriculture in Arid Areas of China, Northwest A&F University, Yangling, Shaanxi 712100, China
3. Jocob Blaustein Institute for Desert Research, Ben-Gurion University of the Negev, Sede Boker 84990, Israel)

Abstract: Obtaining soil moisture quickly and timely can grasp the needs of water of the crops, which is very important for the agricultural production. Soil spectral reflectance provides an alternative method to classical physical and chemical analysis of soil in laboratory for the estimation of a large range of key soil properties. Therefore, the soil moisture was quickly achieved by using hyperspectral technology and the application of ridge regression was explored in the optimization and quantitative analysis of hyperspectral bands. Totally 91 soil samples were collected from the soil depth of 0 ~5 cm in Seder Boker area in the southern Israeli. These soil samples were analyzed in the process of physical and chemical properties in laboratory. After that, the raw hyperspectral reflectance of soil samples was measured by an ASD FieldSpec 3 instrument equipped with a high intensity contact probe under the darkroom conditions. Next, the raw spectral reflectance (REF) was transformed to three spectral indices, i. e. inverse-log reflectance (LR), the first order differential reflectance (FDR) and continuum removal reflectance (CR). Regression models of soil moisture with different indices were established by three methods: partial least squares regression (PLSR), stepwise regression (SR) and ridge regression (RR). The inversion results of the model were validated and compared with each other. The results showed that the method of LR transform can eliminate the interference of external factors much better, and it appeared to be the optimal spectral index in stepwise regression model and ridge regression model ( $R_c^2$  were 0.981 and 0.975, and  $R_p^2$  were 0.971 and 0.979). For the three spectral indices about REF, FDR and CR,

收稿日期: 2017-10-27 修回日期: 2017-11-29

基金项目:国家重点研发计划项目(2017YFC0403302、2016YFD0200700)和杨凌示范区科技计划项目(2016NY-26)

作者简介:张智韬(1976—),男,副教授,博士,主要从事遥感技术在节水灌溉及水资源中的应用研究,E-mail: zhitaozhang@ 126. com

although the modeling effect of SR and RR was slightly lower than that of PLSR, the coefficient of modeling determination was above 0.9. Both SR and RR had simplified and optimized the model, but RR had better validation results and the number of bands used for modeling was only 0.3% of the full spectrum (400 ~ 2 400 nm). After comparing the three regression models established with the four spectral indices, the LR - RR model not only had the characteristics of simple model and less calculation, but also improved the robustness of the model better by using biased estimation at the cost of losing the part accuracy. The result indicated that ridge regression method can not only achieve the efficient selection of hyperspectral bands, but also use the LR - RR hyperspectral inversion model for the reference of monitoring the aerospace hyperspectral remote sensing of regional soil moisture in the future. **Key words**: soil moisture content; hyperspectral remote sensing; regression analysis; ridge regression; band selection

# 0 引言

农业生产中,对土壤含水率进行实时、准确的监 测对提升农作物灌溉管理水平、作物长势状况和产 量预测能力等具有十分重要的指导作用<sup>[1-2]</sup>。但传 统的土壤含水率测定方法如干燥法、中子仪法、γ射 线透射法、电测法等在土壤有损度、测定周期方面有 弊端。因此,亟需一种高效无损的土壤含水率估算 方法。

近年来,具有信息量大、无破坏、非接触、零污 染、简便易行等特点的高光谱技术得以快速发展,为 区域农田墒情的快速监测提供了一种新的技术手 段<sup>[3]</sup>。国内外学者在利用高光谱技术估测土壤含 水率方面作了大量的研究<sup>[4-11]</sup>。虽然通过高光谱 技术可以实现土壤含水率较高精度的反演,但近红 外光谱存在吸收较弱,谱带复杂且重叠度高的缺点, 在土壤高光谱数据中必然存在与含水率不相关的冗 余信息。所以,在定量分析高光谱数据时进行变量 的优选就尤为关键。目前,常用的变量优选方法主 要有:基于竞争性自适应重加权采样<sup>[12]</sup>、连续投影 算法<sup>[13]</sup>、多元逐步线性回归<sup>[14-15]</sup>、变量投影重要性 分析<sup>[16]</sup>、遗传算法<sup>[17]</sup>等。岭回归分析作为一种改 进的最小二乘法,在病态数据处理及特征提取方面 有较好的效果,并可实现模型的简化和鲁棒性的提 高<sup>[18]</sup>。祝鹏等<sup>[19]</sup>基于岭回归方法对太湖 15 个陆 源水样 CRS 与紫外吸光度进行回归分析,从190~ 400 nm 的紫外光谱吸光度中优选出 5 个波段建模, 在 k = 0.5 时的决定系数  $R^2 = 0.69$ 。张曼等<sup>[20]</sup>应用 岭回归方法优选出了原光谱3%的波段,由此建立 的近红外--岭回归模型较好地反演了小麦蛋白质含 量。但岭回归在高光谱遥感反演土壤含水率方面的 应用研究还未见报道,缺乏对土壤含水率的高光谱-岭回归反演预测模型。

本文以以色列南部 Seder Boker 地区采集到的 粘壤土为研究对象,经过室内理化性质分析及光谱 测定处理等工作,尝试用岭回归法对不同含水率梯 度下的土壤光谱反射率(REF)、倒数之对数(LR)、 一阶微分(FDR)和去包络线(CR)等指标进行优选 与建模,同时建立偏最小二乘回归模型和逐步回归 模型。通过模型的精度评定及验证结果比较后,综 合评价高光谱--岭回归模型在反演土壤含水率方面 的效果,以期得到最佳的光谱反演指标。

## 1 试验材料与方法

#### 1.1 研究区概况

研究区位于以色列南部的 Seder Boker 地区 (34°47′E,30°52′N,海拔 640 m),该地区为亚热带 半干旱气候,夏季炎热干旱,冬季温暖湿润,年降水 量介于 100~200 mm 之间,土壤类型为粘壤土。以 色列与我国均面临"总量少、时空分布不均"的水资 源问题<sup>[21]</sup>。但以色列拥有世界最发达的精准农业 技术水平,其先进的节水灌溉技术使得本国的水资 源利用率达 98%,居世界之首。

### 1.2 土壤采集与制备

本次试验在以色列 Seder Boker 地区共采集了 91 个不同含水率的土样,含水率在 0.28% ~ 31.41%之间,平均值为 7.44%,由于该地区较为干 旱,表层土壤水分蒸发强烈,导致大部分土样含水率 较低。用直径和深度分别为 7.5、5.5 cm 的环刀采 集土壤表层以下 5 cm 处的土壤,取得原状土样。采 集时剔除浸入体,再将环刀置于塑料盒中封存、编号 并称量后带回实验室。

#### 1.3 土壤含水率的测定

从环刀内取 20 g 左右有代表性的土样放入铝 盒内,盖上盒盖称量,记录下铝盒的编号和质量,打 开盒盖将铝盒置于干燥箱内,在 105 °C、24 h 恒温条 件下用干燥法测得土样质量含水率  $\theta_m$ , $\theta_m$ 的计算公 式为

$$\theta_m = \frac{W_1 - W_2}{W_2 - W_3} \times 100\% \tag{1}$$

式中 W<sub>1</sub> —— 原状土样(含铝盒)质量 W<sub>2</sub> —— 干燥后原状土样(含铝盒)质量 W<sub>3</sub> —— 空铝盒质量

#### 1.4 光谱测定

土壤样品的高光谱数据采用 ASD FieldSpec 3 型地物光谱仪测得。波谱范围为350~2500 nm,采 样间隔为1.4 nm(350~1000 nm)和2 nm(1000~ 2500 nm), 重采样间隔为1 nm。光谱测量在暗室中 进行,光源为能够提供平行光线的50W卤素灯。 使用 5°视场角光纤探头,光源到土壤表面的距离 L、 光源入射角度 A 和探头距土壤表面高度 H 采用洪 永胜等[22]研究得出的室内较为理想的土壤高光谱 几何测试参数组合:L = 50 cm、 $A = 30^{\circ}$ 、H = 15 cm。 从环刀中取土样置于直径 10 cm、深 2 cm 的黑色盛 样皿中,装满后用直尺将土壤表面刮平,在暗室内进 行光谱测定。每次测定光谱前,先进行暗电流的去 除和白板定标。每个土样在4个方向上(转动3次, 每次90°)进行测量,每个方向上保存5条光谱曲 线,共20条,利用 ViewSpec Pro V6.0.11 软件进行 算术平均后得到土样的实际反射光谱数据。

#### 1.5 光谱数据预处理与光谱指标提取

每个土样光谱去除噪声较大的边缘波段 350~ 399 nm 和 2 401~2 500 nm,采用 Savitzky – Golay 滤

波进行平滑处理,一定程度上消除了由光谱测试环 境、高频随机噪声和杂散光等干扰因素引起的噪声, 提高数据的信噪比<sup>[10,23]</sup>。在土壤原始光谱反射率 (Raw spectral reflectance, REF)基础上,计算其倒数 之对数(Inverse-log reflectance, LR)、一阶微分(First order differential reflectance, FDR)和去包络线 (Continuum removal, CR)3种指标。光谱经LR变 换增强了相似光谱之间的差异,适当减少了随机误 差<sup>[24]</sup>;FDR处理可以消除背景噪声的干扰,改善多 重共线性,提高光谱分辨率和灵敏度,易找到相关性 高的波段<sup>[25]</sup>;CR处理能够突出光谱曲线的吸收和 反射特征,增强光谱曲线各波段之间的对比性<sup>[26]</sup>; 指标 REF、LR、FDR 在 ViewSpec Pro V6.0.11软件 中处理获得,指标 CR 利用 ENVI 5.1 的 Continuum Removed 模块处理得到。

#### 1.6 建模集和验证集的划分

将 91 个土样按测得的质量含水率 θ<sub>m</sub>从大到小 排序,每隔 2 个样本取出 1 个作为验证集样本,共取 得 30 个(33%),其余 61 个(67%)作为建模集样 本,最终建模集和验证集的样本比例为 2:1,可保证 建模样本与验证样本范围一致且分布均匀。土壤含 水率的特征描述见表 1,3 个样本集中,变异系数均 在 100% 以上,为强变异强度。

表 1 土壤含水率统计特征 Tab.1 Statistical characteristics of soil moisture content

样本类型	样本数	最小值/%	最大值/%	平均值/%	标准偏差/%	变异系数/%	峰度系数	偏度系数
总体样本	91	0.28	31.41	7.44	8.21	110.22	0.65	1.24
建模样本	61	0.28	31.41	7.57	8.41	111.04	0.71	1.26
验证样本	30	0.31	28.50	7.19	7.91	110.09	0.74	1.24

#### 1.7 模型建立与验证

采用 3 种回归方法:偏最小二乘回归法(Partial least squares regression, PLSR)、逐步回归法 (Stepwise regression, SR)和岭回归法(Ridge regression, RR)建立高光谱遥感对土壤含水率的反 演模型。其中,偏最小二乘回归法在高光谱遥感模 型中得到了广泛的研究与应用<sup>[10,12,25]</sup>,可较好地解 决自变量之间存在的共线性问题。而逐步回归法是 一种便捷高效的模型优化方法,对高光谱数据的 "降维"具有较好的作用<sup>[14-15]</sup>。岭回归分析作为一 种专门用于共线性数据分析的有偏估计方法,实际 上是一种改良的最小二乘法,通过放弃最小二乘的 无偏性和部分精度来获得效果稍差但稳定性更好的 回归模型,兼具"抗共线性"和自变量筛选的作用。

通过建模决定系数 $(R_c^2)$ 、验证决定系数 $(R_p^2)$ 、 均方根误差(RMSE)和相对分析误差(RPD)来综合 评价模型的效果。其中,当0.66  $\leq R^2 \leq 0.80$ 时,模 型拟合效果较好,当0.81  $\leq R^2 \leq 0.90$ 时,模型拟合 结果很好,当 $R^2 \geq 0.90$ 时,模型拟合效果极好<sup>[27]</sup>。 而 RMSE 越接近于0,表征模型的预测精度越高,预 测能力越强。均方根误差(RMSE)与相对分析误差 (RPD)的计算公式为

$$R_{\rm MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \overline{y_i})}$$
(2)

$$R_{\rm PD} = \frac{\sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-1}}}{\frac{y_{\rm RMSE}}{y_{\rm RMSE}}}$$
(3)

式中 y<sub>i</sub>——土壤含水率模型预测值

- *y<sub>i</sub>*——土壤含水率模型实测值
- ŷ<sub>i</sub>——土壤含水率模型预测值的平均值
- n——模型检验样本的个数

Y<sub>RMSE</sub>——土壤含水率模型的均方根误差

当相对分析误差在 2.5 以上时,表明模型具有 极好的预测能力;当相对分析误差在2.0~2.5之间 时,表明模型具有很好的定量预测能力;当相对分析 误差在 1.8~2.0 之间时,表明模型具有定量预测能 力;当相对分析误差在1.4~1.8之间时,表明模型 具有一般的定量预测能力;当相对分析误差在1.0~ 1.4 之间时,表明模型只有区别高值和低值的能力; 当相对分析误差小于1.0时,表明模型不具备预测 能力<sup>[28]</sup>。

#### 结果与分析 2

2.1



不同含水率土样的高光谱曲线特征分析 图 1 为挑诜的 12 条较典型的土壤样本原始及

> Fig. 1 Raw and pretreated spectral reflectance curves of soil samples with different soil moisture contents

去包络线(CR)4种光谱指标作为 PLSR 分析的自变 量,以土壤含水率为因变量,采用 Leave-one-out 交 叉验证法来确定回归模型中的最佳因子数<sup>[25]</sup>,建立 PLSR 模型。建模及验证结果如表 2 所示。

	表 2	土壤含水	×率的]	PLSR 模型	민
Tab. 2	PLSI	R model	of soil	moisture	conten

光谱	建模决定	验证决定	均方根误差	相对分析
指标	系数 $R_c^2$	系数 $R_p^2$	RMSE	误差 RPD
REF	0.990	0.987	0.011	6.60
LR	0.971	0.971	0.013	5.68
FDR	0.987	0.965	0.015	4.93
CR	0.981	0.977	0.012	6.12

结果表明,基于4种光谱指标建立的全波段土

反射光谱曲线,可以发现,12条光谱曲线波形基本 相似,同时土壤光谱反射率随含水率的增加而降低, 当含水率超过某一阈值(一般认为是田间持水率) 时,土壤反射率随含水率的增加而增加<sup>[29]</sup>。而在 1450、1950 nm 附近存在明显的水分吸收峰。 图 1b、1c、1d 分别是土壤原始光谱经倒数之对数 (LR)、一阶微分(FDR)和去包络线(CR)3种预处 理后的反射率图。图 1d 中,土壤经 CR 处理后的光 谱曲线都归一化到0~1之间,光谱吸收带变得更加 明显:除明显的2个水分吸收峰之外,还存在500、 2 200 nm 2 个较微弱的水分吸收峰。

预处理后的光谱反射率曲线,其中图 1a 为土样原始

# 2.2 偏最小二乘回归模型的建立与分析

运用土壤全波段(400~2400 nm)原始光谱反

壤含水率的 PLSR 模型均具有很好的效果,建模集 的  $R_{e}^{2}$  均在 0.97 以上,验证集的  $R_{p}^{2}$  均在 0.96 以上,  $R_{c}^{2}$  和  $R_{b}^{2}$  相差很小且非常接近于 1,其中 REF -PLSR 模型的  $R_{e}^{2}$  和  $R_{e}^{2}$  均为最高,分别为 0.990 和 0.987。比较各种光谱指标所建立的模型发现,LR-PLSR 模型具有最好的"鲁棒性"( $R_{a}^{2}/R_{c}^{2}$  接近于 100%),其次为 REF - PLSR(99.71%)、CR - PLSR (99.50%)、FDR - PLSR(97.70%)。比较4种模型 的均方根误差 RMSE 发现, REF - PLSR 模型的预测 精度最高(RMSE 为 0.011),经过不同的数学变换, 模型的误差略微增加。而4种模型的相对分析误差 RPD 均大于 2.5, 说明模型均有极好的预测能力, 其 中尤以 REF - PLSR 和 CR - PLSR 模型的 RPD 最 高,均达到 6.0 以上。以上分析表明 PLSR 模型在 拟合和预测方面均有很好的效果。

#### 2.3 逐步回归模型的建立与分析

逐步回归是高光谱反演研究中常用的研究方法,根据自变量对因变量的作用或显著程度,由大到 小逐个引入回归方程,剔除对因变量作用不显著的 自变量,从而极大地简化模型,建立最优回归方 程<sup>[15]</sup>。

本文运用全波段4种光谱指标(REF、LR、FDR、 CR)作为自变量,土壤含水率为因变量。变量入选 和剔除的显著水平分别设为0.10和0.15,由"最 优"回归子集所建立模型的结果见表3。

	表 3	土壤含水	×率的 SR	模型
Tab. 3	SR	model of	soil moist	ture content

* 读	波段数	中心波长/nm		验证决定	拘方相误	相对分析
北屆	W tX XX			五血(C)C 系数 R <sup>2</sup>	均力很快 主 BMSF	唱AD 70 误差 RPD
1H WN	1	833	0.288	0.637	<u> 2 1010日</u>	<u> </u>
	2	033 222 0 207	0. 200	0.037	0.034	2.25
REF	2	784 822 2 227	0.925	0.822	0.034	2.35
	5	784 822 2011 2287	0.939	0.801	0.029	2.03
	4	1070	0.944	0.808	0.028	2.70
	1	1 970	0.838	0.878	0.042	1. 14
LR	2	1464,1970	0.959	0.922	0.022	3. 59
	3	1 464 ,1 970 ,2 149	0.978	0.959	0.016	4.95
	4	1 464 ,1 928 ,1 970 ,2 149	0. 981	0.971	0.013	5.72
	1	1 351	0.932	0.903	0.026	2.99
	2	1 351 ,1 788	0.951	0.903	0.025	3.17
	3	1 351 ,1 713 ,1 788	0.964	0.918	0.024	3.39
	4	1 351 ,1 354 ,1 713 ,1 788	0.968	0.878	0.044	3.45
	5	1 351 ,1 354 ,1 713 ,1 788 ,1 920	0.973	0.878	0.044	1.23
	6	1 351 ,1 354 ,1 713 ,1 788 ,1 920 ,2 246	0.978	0.896	0.026	3.10
	7	1 351 ,1 354 ,1 713 ,1 723 ,1 788 ,1 920 ,2 246	0.981	0.903	0.024	3.12
FDR	8	1 351 ,1 354 ,1 703 ,1 713 ,1 723 ,1 788 ,1 920 ,2 246	0.986	0.910	0.023	3.29
	9	1 351 ,1 354 ,1 703 ,1 713 ,1 723 ,1 788 ,1 791 ,1 920 ,2 246	0.987	0.913	0.023	3.36
	10	1 351 ,1 354 ,1 703 ,1 713 ,1 723 ,1 786 ,1 788 ,1 791 ,1 920 ,2 246	0.989	0.925	0.021	3.64
	11	1 351 ,1 354 ,1 703 ,1 713 ,1 723 ,1 786 ,1 788 ,1 791 ,1 920 ,2 246 ,2 315	0.990	0.923	0.022	3.57
	12	1 351 ,1 354 ,1 703 ,1 713 ,1 723 ,1 786 ,1 788 ,1 791 ,1 920 ,2 222 ,2 246 , 2 315	0. 991	0.916	0.023	3.45
	13	1 351 ,1 354 ,1 703 ,1 713 ,1 723 ,1 767 ,1 786 ,1 788 ,1 791 ,1 920 ,2 222 , 2 246 ,2 315	0. 992	0.900	0.025	3.18
	14	1 351 ,1 354 ,1 652 ,1 703 ,1 713 ,1 723 ,1 767 ,1 786 ,1 788 ,1 791 ,1 920 , 2 222 ,2 246 ,2 315	0. 993	0. 899	0.025	3.17
	15	1 351 ,1 354 ,1 652 ,1 703 ,1 713 ,1 723 ,1 767 ,1 786 ,1 788 ,1 791 ,1 920 , 2 166 ,2 222 ,2 246 ,2 315	0. 996	0.877	0.028	2.86
	1	2 239	0.459	0.607	0.077	0.01
$\mathbf{CR}$	2	1 802 , 2 239	0.910	0.880	0.028	2.37
	3	944,1 802,2 239	0.920	0.886	0.027	2.55

由表 3 可以看出,逐步回归在高光谱数据的 "降维"过程中,对模型的简化有显著的效果。通过 波段的"筛选"过程,去除大量的冗余数据(约 99%)而保留下与含水率相关性最高的几个变量。 建模决定系数  $R^2_{\epsilon}$  随建模波段数量的增加而增加,从 单个波段开始"骤增"后渐趋平缓。不同光谱指标 中通过筛选的波段数目也不同。其中,通过一阶微 分 FDR 处理后的光谱数据,较大程度地提高了光谱 与含水率之间的相关性,通过逐步回归筛选的波段 数最多达 15 个。但根据逐步回归模型算法简捷、运 算量小等特点综合考虑当波段数 m = 3 时, FDR -SR 模型效果最佳,此时建模决定系数  $R_e^2$  和验证决 定系数  $R_p^2$ 分别为 0.964 和 0.918,建模及验证效果 均较佳,并且预测精度较高(RMSE 为 0.024)。当 波段数  $m = 4 \cdot 5 \cdot 6$  时,  $R_e^2$  虽略有增加,但  $R_p^2$  下降到 0.90 以下,而 RMSE 均大于 0.024,此时模型较差。 当  $m \ge 7$  时,模型的效果才得以"改善",但此时模型 复杂度提高,运算量增加,因此不建议使用。当波段 数 m 增加到 10 个以上时,逐步回归模型建模效果 不断提高,预测效果不断降低,这与肖雪梦等<sup>[30]</sup>的 研究成果一致。

因为 REF - SR、LR - SR、CR - SR 3 个模型的波 段数均较少,所以取最大波段数 *m* 时的模型为最佳 模型。比较 4 个模型发现,LR - SR 模型的  $R_c^2$  最高 为 0.981,  $R_p^2$  最高为 0.971, RMSE 最低为 0.013, RPD 为 5.72(大于 2.5),表明该模型具有极好的预 测能力。剩余 3 种模型效果依次为: FDR - SR、 REF - SR、CR - SR 模型,其中,虽然 CR - SR 模型效 果最差,但其筛选剩余的波段数目最少(*m* = 3)且 RPD 为 2.55(大于 2.5),所以仍具有极好的预测能 力。通过  $R_p^2/R_c^2$  来比较模型的稳定性发现:LR - SR 最高达 99.01%,而其余 3 种模型分别为:CR - SR (96.29%)、FDR - SR (95.24%)、REF - SR (92.00%),均低于最差的偏最小二乘回归模型 (FDR - PLSR),说明逐步回归法对模型的简化也会 造成稳定性一定程度的下降。

#### 2.4 岭回归模型的建立与分析

岭回归方法是 1970 年 HOERL 等<sup>[31-32]</sup>针对复 共线性数据分析的有偏估计方法,以放弃最小二乘 法的无偏性和部分精度为代价来获得效果稍差但更 稳定且符合实际的回归方程,是一种改进的最小二 乘估计。在岭回归估计中,通过选择不同的岭参数 *k*值可以得到不同的回归系数,由岭迹法选择变量, 消减波段之间的相关性。通常选择 *k*值的一般原则<sup>[33]</sup>是:①各回归系数的岭估计基本稳定。②采用 岭回归法得到的各自变量符号更具有实际意义。 ③回归系数没有不合乎经济意义的绝对值。④残差 平方和增加不太多。

而岭回归用于选择波长的原则是:①剔除掉标 准化岭回归系数比较稳定且绝对值很小的自变量。 ②剔除标准化岭回归系数不稳定,但随着 k 的增加 振动趋于零的波长。③剔除标准化岭回归系数很不 稳定的波长。④根据去掉波长后重新进行岭回归分 析的结果,去掉一个或若干个回归系数不稳定的 波长。

以全波段4种光谱指标(REF、LR、FDR、CR)作 为自变量,土壤含水率为因变量,根据 k 值选择原则 和波长选择原则,确定岭参数 k 和最优波长,建立 REF - RR、LR - RR、FDR - RR、CR - RR 4 种岭回归 模型(表4)并比较分析。

表 4 土壤含水率的 RR 模型 Tab. 4 RR model of soil moisture content

光谱	波段数		岭参数	建模决定	验证决定	均方根误	相对分析
指标	m	中心波卡/nm	k	系数 $R_c^2$	系数 $R_p^2$	差 RMSE	误差 RPD
REF	6	985,1441,1503,2043,2295,2334	0.0002	0.958	0.956	0.017	4.45
LR	5	1 401 ,1 415 ,1 488 ,1 799 ,1 863	0.0002	0.975	0.979	0.012	5.89
FDR	7	1 087 ,1 151 ,1 317 ,1 629 ,1 724 ,2 125 ,2 152	0.3	0.919	0.921	0.025	2.58
CR	6	1 359 , 1 408 , 1 556 , 1 885 , 1 995 , 2 154	0.02	0.910	0.940	0.020	3.67

由表 4 可见,利用不同光谱指标(REF、LR、 FDR、CR)所建立的 4 种岭回归模型中,岭参数 k 的 取值浮动较大(k=0.000 2~0.3),而筛选出的波段 数(m=5~7)较逐步回归法(m=3~4)略微增加, 但模型的复杂程度和运算量相差不大。对经过数学 变换后的 3 种光谱指标(LR、FDR、CR)建立的模型, 模型的验证决定系数  $R_p^2$ 较建模决定系数  $R_e^2$  分别提 高 0.41%、0.25%、3.33%, REF – RR 虽略有降低, 但是  $R_p^2$  比  $R_e^2$  仅下降 0.11%,所以岭回归模型较为 稳健。

与对应的偏最小二乘回归模型进行比较:LR -RR 模型的建模决定系数  $R_c^2 = 0.975$ ,提高 0.42%, 其他 3 个模型均有不同程度的下降(降幅为 3.25%、6.95%、7.27%),而在验证决定系数  $R_p^2$ 方 面,LR - RR 模型  $R_p^2$ 增加了 0.84%, REF - RR、FDR -RR、CR - RR 分别下降了 3.07%、4.52%、3.71%。 由此可以看出,只有 5 个特征波段的 LR - RR 模型 与 2001 个波段的 LR - PLSR 模型对土壤含水率均 有极好的拟合效果。从模型预测方面来看,LR - RR 模型(RMSE 为 0.012, RPD 为 5.89)比 LR - PLSR 模型(RMSE 为 0.013, RPD 为 5.68)略有提高,故具 有极强的预测精度和预测能力,然后依次为 REF -RR、CR - RR、FDR - RR 模型,其中最差的 FDR -RR 模型 RMSE 为 0.025、RPD 为 2.58,虽然降幅最 大,但仍具有很强的定量预测能力。

与逐步回归模型比较:REF - RR、LR - RR、FDR -RR、CR - RR 4 种岭回归模型的建模决定系数  $R_c^2$  变 幅分别为:1.43%、-0.61%、-4.69%、-1.08%, 验证决定系数  $R_p^2$  变幅分别为 10.13%、0.79%、 0.32%、6.15%。结合该对比发现,在模型拟合度方 面,岭回归模型 REF - RR、CR - RR 较优于对应的 逐步回归模型,而 LR - RR 模型与 LR - SR 模型拟 合度相仿,但 LR - RR 模型的  $R_c^2$ 和  $R_p^2$ 最高分别为 0.975和0.979。在模型预测方面,LR - RR 模型的 RMSE和 RPD(0.012,5.89)均优于 LR - SR 模型 (0.013,5.72),其余3种模型的 RMSE 均在允许范 围内,且 RPD 均大于2.5,具有极好的预测能力。总 之,LR - RR 模型不仅对波段具有较好的优选作用, 还具有较高的稳定性和预测精度,为最佳模型,其次 为 REF - RR 和 CR - RR,FDR - RR 模型相对较差。

# 3 讨论

高光谱遥感与宽波段遥感相比具有光谱分辨率 高和波段连续性强的特点,能够获得更为精细的光 谱信息,是定量分析浅层土壤水分与光谱特征参数 关系的重要工具,可以更为精确地获得土壤含水率 的敏感波段,而对光谱进行不同的数学变换也能在 一定程度上消除光谱测量中的某些人为和自然因素 的干扰,增强信噪比<sup>[11]</sup>,以期提高土壤含水率的预 测精度。

运用3种不同的回归方法对4种光谱指标进行 建模,发现倒数之对数处理后的光谱数据,在逐步回 归建模和岭回归建模中具有较好的效果,证明对数 据进行 LR 预处理可以更为有效地处理非线性问 题,达到增强相似光谱之间的差异,减小随机误差的 作用<sup>[24,34]</sup>。而在3种回归模型中,反演精度最高的 是基于全波段的 PLSR 模型:建模集和验证集的决 定系数 R<sup>2</sup>最高分别达到 0.990 和 0.987,进一步验 证了偏最小二乘-高光谱遥感模型在定量预测土壤 浅层含水率方面的可行性[8,10,12]。但基于全波段 (400~2400 nm)的偏最小二乘回归模型复杂、计算 量大,难以在实际生产中运用。为实现模型的简化, 本文采用逐步回归和岭回归2种方法建模并进行比 较分析。结果表明,2种方法从具有"维数灾难"及 多重共线性的全波段高光谱数据中,有效挖掘了与 土壤含水率相关的波段信息。而岭回归法作为一种 改进的参数估计方法,通过损失部分精度为代价从 而获得对病态数据具有更强耐受性的回归模型(表 现在大部分岭回归模型的建模决定系数  $R^2$  较低,但 验证决定系数  $R_a^2$  较高)。

以往土壤含水率的高光谱定量反演研究主要利 用土壤原始光谱反射率及相对应的倒数之对数,一、 二阶微分,去包络线等光谱预处理方法,结合偏最小 二乘、逐步回归、主成分等方法构建回归模型。而岭 回归分析作为一种"新型"的多元回归方法,在高光 谱领域的研究较少:祝鹏等<sup>[19]</sup>在对太湖水样 CRS 的分析和张曼等<sup>[20]</sup>对小麦蛋白质的光谱测定中均 用到了岭回归方法,并取得了较好的效果。由此也 在一定程度上说明了岭回归法在高光谱的定量分析 方面具有一定的可行性,又由于岭回归在高光谱遥 感反演土壤含水率方面的研究还未见报道,故本文 对此进行了一定的研究分析。而基于不同光谱指标 所建立的岭回归-高光谱遥感反演模型,具有逐步回 归模型简便的特点,兼具偏最小二乘回归模型精度 高( $R^2$ 均在 0.90 以上),预测能力较强(RPD 均在 2.5 以上)的长处,说明岭回归法在高光谱定量反演 土壤含水率方面同样具有可行性。但本文所用到的 光谱指标较少,光谱预处理方法单一,有待于进一步 尝试更多不同的光谱指标变换在土壤含水率岭回 归反演模型中的应用,同时,未来还应尝试利用岭 回归法来充分挖掘土壤光谱信息,实现对土壤不 同成分(如有机质、盐分、重金属等)的快速而准确 的预测。

# 4 结论

(1)逐步回归模型在应用方面具有简便性和快速性,筛选了与水分相关的显著性波段,但 *R<sup>2</sup><sub>p</sub>*大部分小于 0.9,验证效果较差。岭回归模型的 *R<sup>2</sup><sub>c</sub>*、*R<sup>2</sup><sub>p</sub>*均大于 0.9, RMSE 均在 0.03 以内, RPD 均大于 2.5, 故拟合优度、预测效果极好,同时考虑到岭回归 法建模仅使用了优选后的 5~7 个波段(全波段的 0.3%), 故岭回归法在波段筛选方面具有"少而精"的特点。

(2)基于不同光谱指标建立的岭回归模型所选择的岭参数 k 值不同,筛选出波段也不同。但根据预测结果( $R_p^2$ 均高于  $R_e^2$ ,RMSE 很小且在误差允许的范围内,RPD 较高,模型具有极强的预测能力)分析看,波段间的相关性已大大减弱,模型具有较强的稳健性和预测能力。

(3) 基于倒数之对数变换指标建立的逐步回归 模型与岭回归模型,其决定系数最高,故该指标在 2 种回归模型中具有更为显著的作用。而只使用 5 个 波段建立的 LR - RR 模型,其 *R*<sup>2</sup><sub>c</sub> 和 *R*<sup>2</sup><sub>p</sub> 最高分别为 0.975 和 0.979, RMSE 为 0.012, RPD 为 5.89, 是进 行土壤含水率高光谱反演的最优模型。

#### 参考文献

- 1 张小超,吴静珠,徐云.近红外光谱分析技术及其在现代农业中的应用[M].北京:电子工业出版社,2012.
- 2 陈仲新,任建强,唐华俊,等.农业遥感研究应用进展与展望[J].遥感学报,2016,20(5):748-767.

3 宋韬,鲍一丹,何勇.利用光谱数据快速检测土壤含水量的方法研究[J].光谱学与光谱分析,2009,29(3):675-677.

CHEN Zhongxin, REN Jianqiang, TANG Huajun, et al. Progress and perspectives on agricultural remote sensing research and applications in China[J]. Journal of Remote Sensing, 2016, 20(5): 748-767. (in Chinese)

SONG Tao, BAO Yidan, HE Yong. Research on the method for rapid detection of soil moisture content using spectral data [J]. Spectroscopy and Spectral Analysis, 2009, 29(3): 675-677. (in Chinese)

- 4 BOWERS S A, HANKS R J. Reflection of radiant energy from soil[J]. Soil Science, 1971, 100(3): 130-138.
- 5 BOWERS S A, SMITH S J. Spectrophotometric determination of soil water content [C] // SSSA Proceedings, 1972, 36: 978-980.
- 6 HUMMEL J W, SUDDUTH K A, HOLLINGER S E. Soil moisture and organic matter prediction of surface and subsurface soils using an NIR soil sensor[J]. Computers and Electronics in Agriculture, 2001, 32(2): 149-165.
- 7 HASSAN-ESFAHANI L, TORRES-RUA A, JENSEN A, et al. Assessment of surface soil moisture using high-resolution multispectral imagery and artificial neural networks [J]. Remote Sensing, 2015, 7(3): 2627 - 2646.
- 8 王海江,张花玲,任少亭,等. 基于高光谱反射特性的土壤水盐状况预测模型研究[J/OL]. 农业机械学报, 2014, 45(7): 133-138. http://www.j-csam.org/jcsam/ch/reader/view\_abstract.aspx? flag = 1&file\_no = 20140721&journal\_id = jcsam. DOI:10.6041/j.issn.1000-1298.2014.07.021.

WANG Haijiang, ZHANG Hualing, REN Shaoting, et al. Prediction model of soil water-salt based on hyperspectral reflectance characteristics [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(7): 133-138. (in Chinese) 刘洋,丁潇,刘焕军,等. 黑土土壤水分反射光谱特征定量分析与预测[J]. 土壤学报, 2014, 51(5): 1021-1026.

- LIU Yang, DING Xiao, LIU Huanjun, et al. Quantitative analysis of reflectance spectrum of black soil as affected by soil moisture for prediction of soil moisture in black soil[J]. Acta Pedologica Sinica, 2014, 51(5): 1021 1026. (in Chinese)
- 10 朱亚星,周桢津,洪永胜,等. 耦合高光谱数据估算土壤含水率的方法[J]. 华中师范大学学报:自然科学版, 2017, 51(1): 123-129.

ZHU Yaxing, ZHOU Zhenjin, HONG Yongsheng, et al. Prediction of soil moisture content based on coupled hyperspectral data [J]. Journal of Central China Normal University: Natural Sciences, 2017, 51(1): 123 - 129. (in Chinese)

- 11 陆婉珍.现代近红外光谱分析技术[M].2版.北京:中国石化出版社,2006.
- 12 于雷,朱亚星,洪永胜,等. 高光谱技术结合 CARS 算法预测土壤水分含量[J]. 农业工程学报, 2016,32(22): 138-145. YU Lei, ZHU Yaxing, HONG Yongsheng, et al. Determination of soil moisture content by hyperspectral technology with CARS algorithm[J]. Transactions of the CSAE, 2016,32(22): 138-145. (in Chinese)
- 13 成忠,张立庆,刘赫扬,等. 连续投影算法及其在小麦近红外光谱波长选择中的应用[J]. 光谱学与光谱分析, 2010, 30(4):949-952.

CHENG Zhong, ZHANG Liqing, LIU Heyang, et al. Successive projections algorithm and its application to selecting the wheat near-infrared spectral variables [J]. Spectroscopy and Spectral Analysis, 2010, 30(4): 949-952. (in Chinese)

14 张德虎,田海清,武士钥,等.河套蜜瓜糖度可见近红外光谱特征波长提取方法研究[J].光谱学与光谱分析,2015, 35(9):2505-2509.

ZHANG Dehu, TIAN Haiqing, WU Shiyao, et al. Study on extraction methods of characteristic wavelength of visible near infrared spectroscopy used for sugar content of Hetao muskmelon[J]. Spectroscopy and Spectral Analysis, 2015, 35(9): 2505 - 2509. (in Chinese)

15 叶勤,姜雪芹,李西灿,等. 基于高光谱数据的土壤有机质含量反演模型比较[J/OL]. 农业机械学报,2017,48(3):164-172. http://www.j-csam.org/jcsam/ch/reader/view\_abstract.aspx? flag = 1&file\_no = 20170321&journal\_id = jcsam. DOI:10. 6041/j.issn.1000-1298.2017.03.021.

YE Qin, JIANG Xueqin, LI Xican, et al. Comparison on inversion model of soil organic matter content based on hyperspectral data[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(3): 164 - 172. (in Chinese)

16 贾生尧,唐旭,杨祥龙,等.可见-近红外光谱技术结合递归变量选择算法对土壤全氮与有机质含量测定研究[J].光谱学与光谱分析,2014,34(8):2070-2075. JIA Shengyao, TANG Xu, YANG Xianglong, et al. Visible and near infrared spectroscopy combined with recursive variable

selection to quantitatively determine soil total nitrogen and organic matter[J]. Spectroscopy and Spectral Analysis, 2014, 34(8): 2070 - 2075. (in Chinese)

17 孙俊,路心资,张晓东,等. 基于高光谱图像的红豆品种 GA - PNN 神经网络鉴别[J/OL]. 农业机械学报, 2016, 47(6):
 215 - 221. http://www.j-csam.org/jcsam/ch/reader/view\_abstract.aspx? flag = 1&file\_no = 20160628&journal\_id = jcsam.
 DOI:10.6041/j.issn.1000-1298.2016.06.028.

SUN Jun, LU Xinzi, ZHANG Xiaodong, et al. Identification of red bean variety with probabilistic GA – PNN based on hyperspectral imaging[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(6): 215 – 221. (in Chinese)

- 18 MARYAM I, HASSAN G. Ridge regression-based feature extraction for hyperspectral data[J]. International Journal of Remote Sensing, 2015, 36(6): 1728-1742.
- 19 祝鹏,华祖林,李惠民,等. 基于岭回归分析法的太湖 CRS 紫外光谱回归模型[J]. 光谱实验室, 2011, 28(6): 2748 2752.

ZHU Peng, HUA Zulin, LI Huimin, et al. Regression model for ultraviolet spectroscopy of CRS in Taihu lake based on ridge regression [J]. Chinese Journal of Spectroscopy Laboratory, 2011, 28(6): 2748-2752. (in Chinese)

20 张曼,刘旭华,何雄奎,等. 岭回归在近红外光谱定量分析及最优波长选择中的应用研究[J]. 光谱学与光谱分析, 2010,

30(5): 1214 - 1217.

ZHANG Man, LIU Xuhua, HE Xiongkui, et al. Study on the application of ridge regression to near-infrared spectroscopy quantitative analysis and optimum wavelength selection[J]. Spectroscopy and Spectral Analysis, 2010, 30(5): 1214 - 1217. (in Chinese)

21 王参民. 以色列水资源问题研究[D]. 开封:河南大学, 2016.

WANG Canmin. Study on issues of water resources in Israel[D]. Kaifeng: Henan University, 2016. (in Chinese)

22 洪永胜,于雷,耿雷,等.应用 DS 算法消除室内几何测试条件对土壤高光谱数据波动性的影响[J].华中师范大学学报: 自然科学版,2016,50(2):303-308.

HONG Yongsheng, YU Lei, GENG Lei, et al. Using direct standardization algorithm to eliminate the effect of laboratory geometric parameters on soil hyperspectral data fluctuate characteristic [J]. Journal of Central China Normal University: Natural Sciences, 2016, 50(2): 303 - 308. (in Chinese)

- 23 SAVITZKY A, GOLAY M J E. Smoothing and differentiation of data by simplified least squares procedures [J]. Analytical Chemistry, 1964, 36(8): 1627 - 1639.
- 24 吴明珠,李小梅,沙晋明.亚热带土壤铬元素的高光谱响应和反演模型[J].光谱学与光谱分析,2014,34(6):1660-1666.

WU Mingzhu, LI Xiaomei, SHA Jinming. Spectral inversion models for prediction of total chromium content in subtropical soil [J]. Spectroscopy and Spectral Analysis, 2014, 34(6): 1660 - 1666. (in Chinese)

- 25 于雷,洪永胜,耿雷,等. 基于偏最小二乘回归的土壤有机质含量高光谱估算[J]. 农业工程学报, 2015, 31(14): 103 109. YU Lei, HONG Yongsheng, GENG Lei, et al. Hyperspectral estimation of soil organic matter content based on partial least squares regression[J]. Transactions of the CSAE, 2015, 31(14): 103 - 109. (in Chinese)
- 26 彭小婷,高文秀,王俊杰.基于包络线去除和偏最小二乘的土壤参数光谱反演[J].武汉大学学报:信息科学版,2014, 39(7):862-866.

PENG Xiaoting, GAO Wenxiu, WANG Junjie. Inversion of soil parameters from hyperspectra based on continuum removal and partial least squares regression [J]. Geomatics and Information Science of Wuhan University, 2014, 39(7): 862 - 866. (in Chinese)

27 殷哲,雷廷武,陈展鹏,等.近红外传感器测量不同种类土壤含水率的适应性研究[J/OL].农业机械学报,2014,45(3): 148-151.http://www.j-csam.org/jcsam/ch/reader/view\_abstract.aspx?flag=1&file\_no=20140325&journal\_id=jcsam. DOI:10.6041/j.issn.1000-1298.2014.03.025.

YIN Zhe, LEI Tingwu, CHEN Zhanpeng, et al. Adaptability of near-infrared sensor for moisture measurement of different soils [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2014, 45(3): 148-151. (in Chinese)

28 张秋霞,张合兵,张会娟,等. 粮食主产区耕地土壤重金属高光谱综合反演模型[J/OL]. 农业机械学报, 2017, 48(3): 148-155. http://www.j-csam.org/jcsam/ch/reader/view\_abstract.aspx? flag = 1&file\_no = 20170319&journal\_id = jcsam. DOI:10.6041/j.issn.1000-1298.2017.03.019.

ZHANG Qiuxia, ZHANG Hebing, ZHANG Huijuan, et al. Hybrid inversion model of heavy metals with hyperspectral reflectance in cultivated soils of main grain producing areas[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(3): 148 - 155. (in Chinese)

29 刘秀英,王力,宋荣杰,等.黄绵土风干过程中土壤含水率的光谱预测[J/OL].农业机械学报,2015,46(4):266-272. http://www.j-csam.org/jcsam/ch/reader/view\_abstract.aspx? flag = 1&file\_no = 20150439&journal\_id = jcsam. DOI:10. 6041/j.issn.1000-1298.2015.04.039.

LIU Xiuying, WANG Li, SONG Rongjie, et al. Prediction of soil moisture content in air-drying loess using spectral data[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(4): 266 - 272. (in Chinese)

- 30 肖雪梦,张应应. 三种回归方法在消除多重共线性及预测结果的比较[J]. 统计与决策, 2015(24): 75-78.
- 31 HOERL A E, KENNARD R W. Ridge regression: biased estimation for nonorthogonal problems [J]. Technometrics, 1970, 12(1): 55-67.
- 32 HOERL A E, KENNARD R W. Ridge regression: applications to nonorthogonal problems [J]. Technometrics, 1970, 12(1): 69-82.
- 33 何晓群,刘文卿.应用回归分析[M].4版.北京:中国人民大学出版社,2015.
- 34 王敬哲,塔西甫拉提·特依拜,张东.基于分数阶微分的荒漠土壤铬含量高光谱检测[J/OL]. 农业机械学报,2017,48(5): 152-158. http://www.j-csam.org/jcsam/ch/reader/view\_abstract.aspx? flag = 1&file\_no = 20170518&journal\_id = jcsam. DOI:10.6041/j.issn.1000-1298.2017.05.018.

WANG Jingzhe, TASHPOLAT Tiyip, ZHANG Dong. Spectral detection of chromium content in desert soil based on fractional differential [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(5): 152 - 158. (in Chinese)