

胡艳培, 白铁成, 陈好斌, 等. 不同近红外光谱预处理方法对胡杨叶片含水量检测模型的影响[J]. 江苏农业科学, 2018, 46(19): 253–257.
doi:10.15889/j.issn.1002-1302.2018.19.066

不同近红外光谱预处理方法 对胡杨叶片含水量检测模型的影响

胡艳培¹, 白铁成¹, 陈好斌², 姚江河¹, 刘冠华¹, 杨洪坤¹

(1. 塔里木大学信息工程学院, 新疆阿拉尔 843300; 2. 塔里木大学经济与管理学院, 新疆阿拉尔 843300)

摘要:近红外光谱预处理是胡杨叶片含水量光谱检测的前提, 然而样本背景、电噪声和杂散射等会干扰近红外光谱的测量结果。为了避免以上因素对近红外光谱产生的影响, 采用连续投影算法 (successive projections algorithm, 简称 SPA)、数据中心化、归一化和标准正态变量变换方法 (standard normal variate transformation, 简称 SNV) 对原始光谱数据进行预处理, 使用偏最小二乘法 (partial least squares, 简称 PLS) 建立胡杨叶片水分含量检测的校正集和预测集模型。结果表明, 直接使用原始光谱, 利用 SPA 算法筛选变量数为 7 个, 模型预测精度为 0.971 44, RMSPCV 为 0.046 132, 相关系数 $r=0.674\ 24$, RMSEP = 0.021 434; 使用原始光谱 + 标准正态变量变换方法, 利用 SPA 算法筛选变量数为 6 个, 模型预测精度为 0.976 63, RMSPCV 为 0.045 642, 相关系数 $r=0.774\ 72$, RMSEP = 0.018 24。SNV + SPA 预处理方法能够有效地消除噪声和散射, 提高模型的预测精度, 相关性明显增加, 降低数据维数和误差, 可用于胡杨叶片水分含量的快速、无损检测, 同时对其他作物叶片水分含量光谱预处理检测具有一定的参考意义。

关键词:近红外光谱预处理方法; 叶片含水量; 标准正态变量变换; SPA; PLS

中图分类号: S718.3; S127 **文献标志码:** A **文章编号:** 1002-1302(2018)19-0253-04

在塔里木河周边生长着大片的天然胡杨林, 这道天然林带绵延数百公里, 可以减缓和防止世界上第二大沙漠——塔克拉玛干沙漠北移, 且具有调节当地气候、防止水土流失导致土地沙化和保障绿洲农业生产等优点。胡杨树通过多年进化长出不同大小的叶片, 小叶片可有效减少水分蒸发, 大叶片为自身的生长提供养分与能量, 因此胡杨叶片能够合理地利用每一滴水。近年来, 胡杨正在面临虫害和缺水等新的问题, 导致塔里木河沿岸大片的胡杨林濒临死亡。因此对胡杨的健康状况实施有效保护的前提条件是及时有效地进行监测, 而叶片的含水量是判断胡杨生长状况的重要依据。传统的人工对叶片检测的方法不但费时而且费力, 效果不是很好, 所以研究一种快速、有效、无损的检测方法对胡杨林的有效保护具有重要的现实意义。

近红外光谱 (near infrared, 简称 NIR) 技术检测已被广泛地应用在食品、石化、烟草、医药、农业等行业, 此技术是一种低成本、快速和无损的检测方法^[1-5], 其主要应用于农产品品质鉴定 (小麦、花生、番茄、玉米、红枣、苹果和大豆等)^[6-13]、植物叶片水分含量检测^[14]、土壤性质分析^[15]和早期病虫害诊断^[16]等, 另外近红外光谱技术在指纹、血迹、火药等刑事犯罪检验方面具有可行性^[17], 但有关胡杨叶片含水量的光谱预处理方法研究的相关报道很少。

在近红外光谱应用时, 不仅可以检测到样品中有用的信息, 还可以检测到其他的信息和噪声, 这是因为受到样本背景、电噪声、杂散射等因素的干扰。在建立模型前对原始的光谱数据进行预处理是分析样品中成分的前提条件, 不同光谱的预处理方法、校正样本的选择方法和建模方法有很多, 比如归一化、标准化、均值中心化、标准正态变量变换、多元散射校正和去趋势等算法是常用的光谱预处理方法。禁忌搜索算法、连续投影算法 (successive projections algorithm, 简称 SPA)、粒子群优化算法、遗传算法 (genetic algorithm, 简称 GA)、相关系数法等是常用的特征波长的筛选方法。模型建立方面, 主要采用竞争性自适应重加权算法 (competitive adaptive reweighted least squares, 简称 CARS)、偏最小二乘法 (partial least squares, 简称 PLS)、误差反向传输 (back propagation, 简称 BP) 神经网络、支持向量机 (support vector machine, 简称 SVM) 等方法。SPA 算法不仅可以从严重重叠的光谱信息中提取有用的信息, 还可以消除波长变量之间的共线性干扰, 提高建模的精度, 在火龙果总酸含量^[18]、葡萄果皮花色苷含量^[19]、油菜叶片氨基酸总量^[20]、羊肉品质^[21]无损检测中的应用也验证了连续投影算法的有效性。虽然各种光谱的预处理方法、建模方法得到了普遍的应用, 但不同作物的光谱响应会有所不同, 胡杨叶片含水量的建模方法和光谱预处理方法需要进一步去验证。

本研究以胡杨叶片为研究对象, 比较多种不同的光谱预处理方法对胡杨叶片水分含量模型的影响, 进而确定适合于近红外光谱分析胡杨叶片含水量的预处理方法, 以此来简化模型的运算过程, 为大面积遥感监测以及进一步研究胡杨叶片的含水量提供理论依据和数据支撑。

收稿日期: 2017-05-30

基金项目: 国家自然科学基金 (编号: 61362026)。

作者简介: 胡艳培 (1987—), 女, 河南周口人, 硕士研究生, 主要从事农业遥感研究。E-mail: 2239015656@qq.com。

通信作者: 白铁成, 博士研究生, 副教授, 主要从事遥感与数字农业研究。E-mail: baitiecheng1983@163.com。

1 材料与方法

1.1 样品采集

试验用的胡杨叶片样品为采集于新疆第一师十四团胡杨林中的 300 张胡杨叶片,从中选择完好无损的 100 张胡杨叶片,用自封袋包装并放置在冰箱中储藏,然后集中一个时间点进行光谱采集和水分测定。

1.2 光谱仪器及数据采集

试验中采用 Zolix GaiaSorter 近红外成像高光谱仪,光谱数据的采集操作和数据存储采用配套的软件获得。样本在室内恒温下进行扫描,光谱分辨率为 5 nm,采样点为 4 nm,测定范围为 900 ~ 1 700 nm,获取一维影像和光谱信息,通过 ENVI (Research System Inc, 美国) 软件获取每张叶片的平均光谱值作为原始光谱,每个样本数据测量 4 次取平均值,导出 Excel 2010 文件。标准正态变量变换、数据中心化、归一化、SPA 和偏最小二乘回归算法在 MATLAB 2014a (The MathWorks Inc, 美国) 中实现,叶片水分含量采用远近红外快速干燥箱进行烘干法测量。

1.3 光谱预处理方法

本研究通过 4 种方法对胡杨叶片原始光谱数据进行预处理与分析,分别采用数据中心化、归一化 (normalization) 和标准正态变量变换 (standard normal variate transformation, 简称 SNV) 3 种预处理方法。

1.4 校正集样本划分

近红外光谱分析要从复杂的信息中有效地提取具有代表性的信息建立模型,不仅可以提高建模精度,还可以减少模型的储存空间。还有就是采集到的样本不含化学成分数据的,如果不进行筛选而直接进行分析,费用也是昂贵的。目前,样本的选择方法主要包括遥感技术 (random sampling, 简称 RS)、基于欧氏距离的样本选择方法 (kennard - stone, 简称 KS)、基于欧式距离和浓度的样本选择方法 (sample set partitioning based on joint x - y distance, 简称 SPXY) 剔除方式、连续投影算法 (successive ections algorithm, 简称 SPA) 等。本研究基于光谱主成分和样品含水量分析,选用 SPXY 方法进行校正集样本的选择,从 100 个胡杨叶片样本中选择 65 个校正集样本,35 个预测集样本。

1.5 特征波长的提取方法

校正模型可通过波长选择使模型简化、消除非线性或无关变量从而使模型具有更强的预测能力和更好的稳健性。

SPA 算法可以从严重重叠的光谱信息中提取有用的信息、消除变量之间的共线性干扰、简化建模,在波长变量选择上得到广泛的应用。SPA 算法在初始状态时挑取 1 个波长,向前运行时采用循环的方法,每循环 1 次计算出它在未选入波长上的投影,进而选取在投影向量中的最大波长,然后将投影向量引入到波长组合中,直到循环结束。设 $X_{K(0)}$ 表示初始的迭代向量, $X_{M \times K}$ 表示光谱的吸收矩阵, M 表示样本数, K 表示波长数,用 $N(N < M - 1)$ 表示提取的波长数, X_j 记为列变量,最后提出 $[X_{K(0)} 0, \dots, N - 1]$ 的变量。

因此,建立了 $N \times K$ 波长组合对,模型的建立通过提取波长数和迭代向量的组合来实现,使用 RMSE 确定模型的差异,从中选出变量的个数、相对应的波长、最小均方根误差。为了

简化模型,防止出现过度拟合,建立 MIR (多元线性回归) 的模型,剔除各个波长的残差平方和,得到用于建模的特征波长。

1.6 建模方法

使用偏最小二乘法 (partial least squares, 简称 PLS) 建立模型,采用相关系数 r 、预测精度 (precision)、预测均方根误差 (root mean square erroe of prediction, 简称 RMSEP) 和交叉验证预测均方差 (root mean square prediction of cross validation, 简称 RMSPCV) 等参数来评价各个预处理方法对所建立模型的影响。对于同一样品集的同组分,预测均方根误差越小,表示所建模型的预测能力越强,结果越准确,预测越精确。

2 结果与分析

2.1 胡杨叶片水分含量

由胡杨叶片的水分含量测定统计结果可知,水分含量最大值为 0.678 6,最小值为 0.446 7,平均值为 0.605 2,标准偏差为 0.042 5。胡杨叶片样品数按照 3 : 1 的比例分为校正集和预测集。

2.2 样品光谱及处理后光谱

图 1 - a 为采集到的 100 张胡杨叶片的原始光谱。可以看出,光谱内存在大量的散射基线漂移和噪声,因此需要对原始的光谱数据进行预处理,建立和优化模型。如图 1 - b 和图 1 - c 所示,预处理后的光谱有效地去除了噪声、背景干扰、散射、基线漂移影响,光谱特征增强,有利于特征波长的选择。图 1 - d 显示,归一化后的光谱波峰和波谷区分明显,消除了部分噪声、散射的影响。对比分析可知,图 1 - b 和图 1 - c 效果更佳,更能说明原始光谱预处理之后的效果。

2.3 SPA 处理结果

对原始光谱进行 SNV、数据中心化和归一化处理后的光谱通过连续投影算法进行变量的选择,最终变量数采用标准误差也称均方根误差 (root mean square error, 简称 RMSE) 选定,如图 2 为 SPA 处理后的筛选变量个数与波长效果,具体波长选择结果如表 1 所示。

由表 1 可以看出,从原始光谱 + SPA 处理后共选出 7 个光谱特征波长,考虑到 908 nm 在光谱仪边缘位置,所以建模时要剔除 908 nm 特征变量,最后选择 6 个特征波长。从原始光谱 + SNV + SPA 初步筛选出 7 个光谱特征波长,由于 1 758 nm 为边缘位置,要剔除,最终获得 6 个特征波长。从原始光谱 + 数据中心化 + SPA 处理后初步筛选出 6 个光谱特征波长,1 755 nm 为边缘位置,要剔除,最终获得 5 个特征波长。从原始光谱 + 归一化 + SPA 处理后初步筛选出 8 个光谱特征波长,由于 905 nm 和 908 nm 在边缘位置,所以要剔除,最终获得 6 个特征波长。从选择的变量波长可以看出,胡杨叶片水分含量在 1 000 至 1 700 nm 波段有较高的相关性光谱特征,可用于胡杨叶片水分的快速无损检测,采用 SPA 算法选择光谱变量不仅可以减少模型建立的复杂度,还降低了光谱数据维数。

2.4 预测模型建立

本研究对胡杨叶片光谱数据进行优化采用原始光谱 + SPA、原始光谱 + 归一化 + SPA、原始光谱 + 数据中心化 + SPA、原始光谱 + SNV + SPA 4 种光谱预处理方法,然后建立

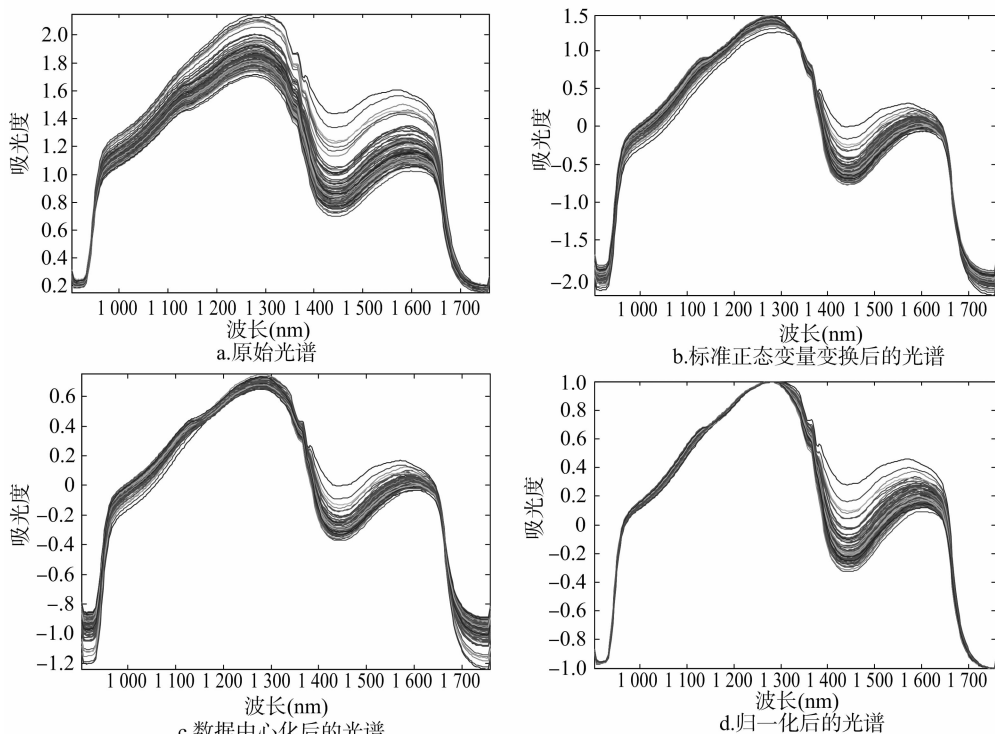


图1 原始光谱和预处理后的光谱

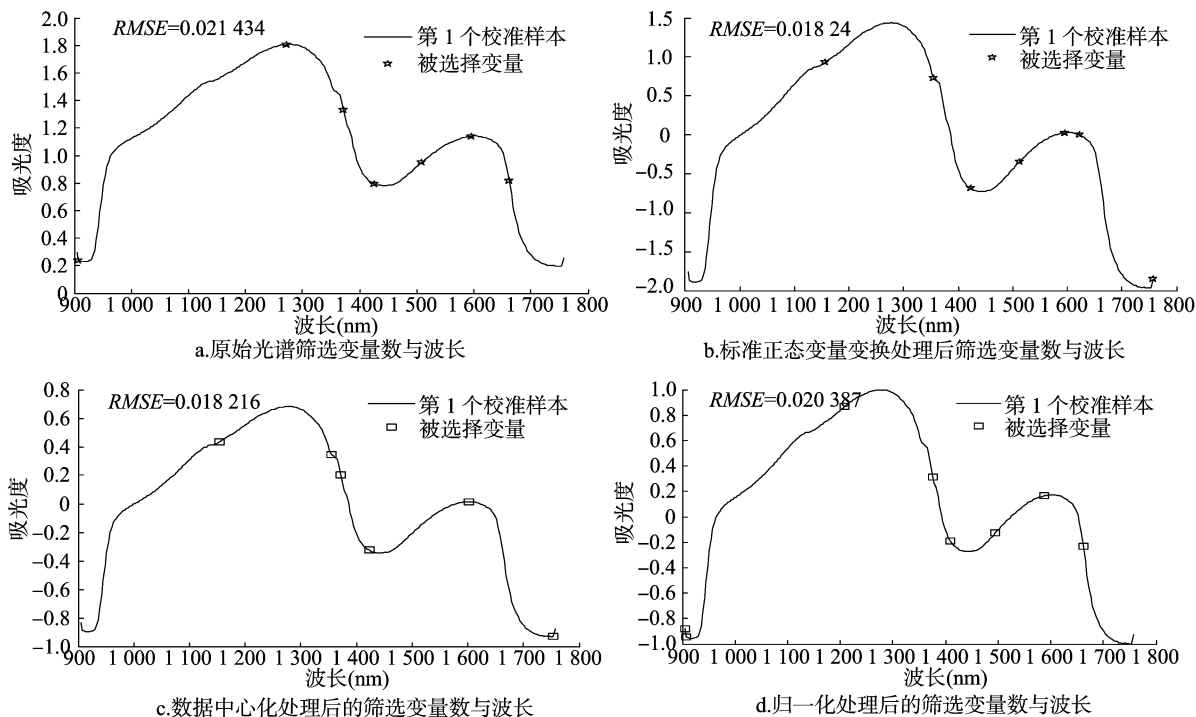


图2 SPA 处理后的筛选变量数与波长效果

PLS 模型,预测的散点图如图 3 所示。

2.5 光谱建模精度及预处理分析

表 2 为采用 4 种光谱的数据预处理方法对胡杨叶片建模优化后的结果。由表 2 可以看出,不同的光谱预处理方法对 PLS 建立的模型参数存在一定的影响。在 4 种光谱预处理方法中, r 最大值为 0.774 72,最小值为 0.674 24;RMSEP 最小值为 0.018 216,最大值为 0.021 434;Precision 最大值为

0.976 63,最小值为 0.971 44。综合对比分析得出,通过采用原始光谱 + SNV + SPA 的方法预处理后, r 由 0.674 24 增加到 0.774 72,Precision 由 0.971 44 增加到 0.976 63,RMSEP 由 0.021 434 降低到 0.018 24,RMSPCV 由 0.046 132 降低到 0.045 642。结果表明,原始光谱 + SNV + SPA 降低了噪声的干扰,提高了建模精度和相关性,因此所建立的模型具有较好的预测效果。

表 1 不同预处理方法波长的筛选结果

光谱预处理方法	初步筛选特征波长 (nm)	最后筛选特征波长 (nm)	剔除波长 (nm)
原始光谱 + SPA	908、1 271、1 372、1 426、1 508、1 595、1 662	1 271、1 372、1 426、1 508、1 595、1 662	908
原始光谱 + 数据 中心化 + SPA	1 755、1 423、1 153、1 602、1 372、1 355	1 423、1 153、1 602、1 372、1 355	1 755
原始光谱 + 归一 化 + SPA	905、908、1 209、1 375、1 409、1 495、1 588、1 662	1 209、1 375、1 409、1 495、1 588、1 662	905、908
原始光谱 + SNV + SPA	1 156、1 355、1 423、1 512、1 595、1 623、1 758	1 156、1 355、1 423、1 512、1 595、1 623	1 758

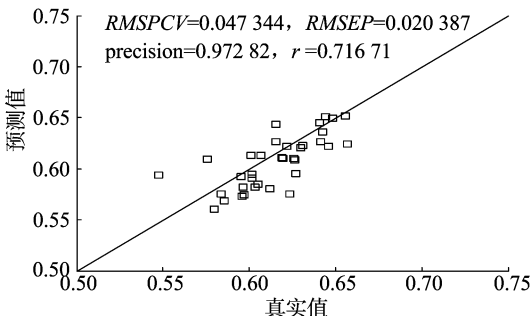
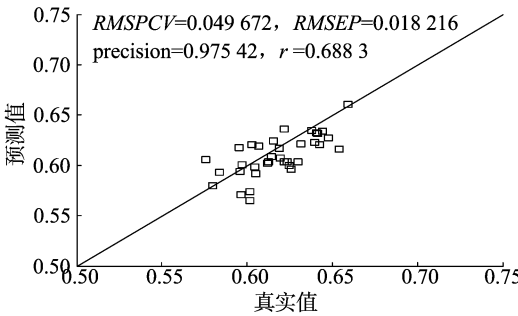
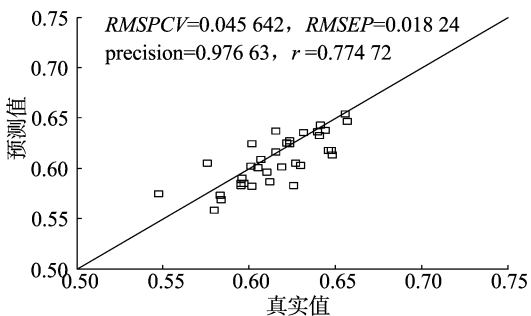
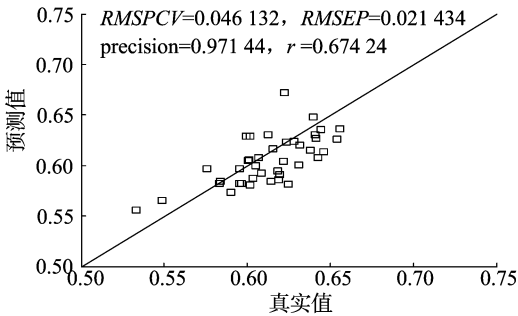


图3 光谱的预测散点

表 2 不同光谱预处理方法及参数比较

光谱预处理方法	RMSE	RMSPCV	RMSEP	Precision	r
原始光谱 + SPA	0.021 434	0.046 132	0.021 434	0.971 44	0.674 24
原始光谱 + 数据中心化 + SPA	0.018 216	0.049 672	0.018 216	0.975 43	0.688 30
原始光谱 + 归一化 + SPA	0.020 387	0.047 344	0.020 387	0.972 82	0.716 71
原始光谱 + SNV + SPA	0.018 240	0.045 642	0.018 24	0.976 63	0.774 72

3 结论

本研究利用近红外光谱技术来检测胡杨叶片水分含量，并且建立了胡杨叶片的预测模型，同时对比了 4 种不同光谱预处理方法。综合分析得出，原始光谱 + 标准正态变量变换 + SPA 预处理后相关系数最大，所建模型的预测精度最高，交叉验证预测均方差最小，分别为 0.774 72、0.976 63、0.045 642。由此可以看出，SNV + SPA 近红外光谱预处理方法可用于胡杨叶片水分含量的检测。

参考文献：

[1] 孙 通,徐惠荣,应义斌. 近红外光谱分析技术在农产品/食品品质在线无损检测中的应用研究进展[J]. 光谱学与光谱分析, 2009,29(1):122-126.
[2] 刘多强,李长征,孙建章,等. 近红外光谱技术的发展及在石化领

域中的应用[J]. 化工时刊,2011,25(6):32-34.
[3] 李瑞丽,张保林,王建民. 近红外光谱检测技术在烟草分析中的应用及发展趋势[J]. 河南农业科学,2013,42(6):1-6.
[4] 肖 雪,梁琼麟,王义明,等. 近红外光谱技术在医药领域中的应用进展[J]. 现代仪器,2011,17(5):9-12.
[5] 王艳龙,王胜宝,李新生. 近红外光谱技术在农业领域的应用及展望[J]. 安徽农学通报(上半月刊),2012,18(15):33-35.
[6] 谭正林,吴谋成,傅廷栋. 近红外光谱技术在农产品品质检测中的应用[J]. 中国油料作物学报,2012,34(4):455-460.
[7] 姬玉梅. 近红外光谱技术在小麦品质育种中的应用研究[J]. 湖北农业科学,2012,51(10):2096-2097.
[8] 史文青,方晓璞,崔 诚,等. 近红外光谱技术在花生品质分析中的应用[J]. 粮油加工(电子版),2014(6):30-33.
[9] 李天华. 成熟期番茄品质近红外光谱无损检测技术研究[D]. 济南:山东大学.
[10] 黄华军,严衍禄,申兵辉,等. 鉴别玉米杂交种纯度的近红外光

史 勇,肉孜·阿木提.一种利用太阳能提升温室灌溉用井水温度装置的设计与试验[J].江苏农业科学,2018,46(19):257-260.
doi:10.15889/j.issn.1002-1302.2018.19.067

一种利用太阳能提升温室灌溉用井水温度装置的设计与试验

史 勇,肉孜·阿木提

(新疆农业大学机电工程学院/新疆农业工程装备创新设计重点实验室,新疆乌鲁木齐 830052)

摘要:针对新疆地区温室大棚灌溉用井水温度过低而影响植物正常生长等问题,设计并研制一种温室灌溉用井水太阳能增温装置。设计一种适合快速加热的吸热板和蛇形太阳能热水器,研制便于移动的太阳能集热器的底架装置。对太阳能井水增温装置进行深入理论分析,试验结果,单个集热器使用后可提高水温约 3~5℃,采用 2 个集热器串联方式使用,可提高水温约 6~8℃。表明温室灌溉用井水太阳能增温装置可有效满足秋冬季温室灌溉用水的温度要求,保证作物的正常生长,具有良好的发展前景。

关键词:温室;井水;灌溉;太阳能集热器

中图分类号: S214.3 **文献标志码:** A **文章编号:** 1002-1302(2018)19-0257-04

新疆乌鲁木齐地区的冬季和初春时节气温较低,最低可达 -20℃^[1],导致温室大棚内灌溉用井水温度也较低,低于温室内直接灌溉用水温度 14~16℃ 的要求。灌溉用水温度过低会导致作物产生寒害,影响作物的正常生长,从而造成作物减产,影响农民增收。

太阳能是一种清洁的可再生能源^[2],采用太阳能作为能源用于温室大棚的增温系统,没有安全隐患,不污染环境^[3]。目前,国内外众多学者也从事温室太阳能增温技术的研究。山本雄二郎最早将太阳能技术用于温室地温的提升,取得较好效果^[4]。Bargach 等研究,利用太阳能热水器获得热水,并

将其传递到温室的地下软管,用以提升温室的地温^[5]。张海莲等利用太阳能加热水结合填埋钢管的方式进行温室地温提升研究^[6]。熊培桂等在青藏高原的温室大棚中研究太阳能储热系统^[7]。刘圣勇等利用太阳能热水器和土壤埋管的方式设计并研制了温室加热系统^[8]。戴巧利等利用太阳能空气集热器和土壤蓄热的方式进行了温室大棚的增温试验^[9]。Wang 等设计的主动式太阳能温室蓄热系统具有夏天储热冬天用,白天储热夜间用的效果^[10]。新疆农业大学冯前前等研制的太阳能温室地温提升装置利用蛇形太阳能空气集热器结合土壤蓄热的方式,在乌鲁木齐南郊水西沟村进行了试验,该装置可使温室土壤 10~20 cm 深度的温度平均升高 1.5~3.0℃^[11]。上述研究证明,利用太阳能作为能源用来提高温室温度具有应用广泛、增温效果显著等优点,并且节约能源,保证作物的正常生长。

目前,现有温室用太阳能增温系统主要应用于温室地温的提升,而针对温室灌溉用井水增温的装置和系统研究较少。因此,为了适应市场的需求,设计一种温室大棚专用的太阳能井水增温装置,对现有的平板式太阳能集热器进行改进设计,研制便于移动的太阳能集热器底架装置,试验证明,该装置可

收稿日期:2017-03-21

基金项目:新疆维吾尔自治区农机化新技术新机具研制开发项目(编号: XNJY2012-08)。

作者简介:史 勇(1986—),男,新疆哈密人,硕士,讲师,从事农产品加工机械、机电一体化技术的研究。E-mail: 280974136@qq.com。

通信作者:肉孜·阿木提,硕士,教授,从事农产品加工机械、农产品干燥技术的研究。E-mail: roziyamat@163.com。

谱分析技术研究[J]. 光谱学与光谱分析,2014,22(5):125.

[11] 彭云发. 近红外光谱技术在南疆红枣品质快速无损检测中的应用研究[D]. 阿拉尔:塔里木大学,2015.

[12] 马玉娟,赵见军,孟永宏,等. 基于 MATLAB 软件的苹果内部品质系统化无损检测[J]. 食品工业科技,2014,35(13):61-62.

[13] 张 鹏,李江阔,陈绍慧. 苹果品质近红外光谱无损检测技术研究进展[J]. 保鲜与加工,2013,13(3):1-7.

[14] 李光君. 热成像技术与近红外光谱技术结合无损检测西拉葡萄叶片水分含量[J]. 山西农业科学,2016,44(10):1467-1475.

[15] 丁海泉,卢启鹏. 近红外光谱技术在土壤养分分析中的研究进展及应用前景[J]. 光谱学与光谱分析,2012,32(1):88-91.

[16] 王晓周,周国民. 基于近红外光谱技术的农作物病害诊断[J]. 农机化研究,2010,32(6):171-174.

[17] 齐敏璐,陈奕桦,王新全,等. 近红外光谱成像技术在现场物证搜索中的应用研究[J]. 刑事技术,2017,42(1):15-20.

[18] 罗 霞,洪添胜,罗 阔,等. 小波变换和连续投影算法在火龙果总酸中无损检测中的应用[J]. 光谱学与光谱分析,2016,36(5):1345-1351.

[19] 吴 迪,宁纪锋,刘 旭,等. 基于高光谱成像技术和连续投影算法检测葡萄果皮花色苷含量[J]. 食品科学,2014,35(8):57-61.

[20] 刘 飞,张 帆,方 慧,等. 连续投影算法在油菜叶片氨基酸总量无损检测中的应用[J]. 光谱学与光谱分析,2009,29(11):3079-3083.

[21] 杨菊梅,贺晓光,王松磊,等. 冷鲜羊肉品质的高光谱成像无损检测[J]. 食品工业科技,2016,37(22):84-89.