

遥感与气象数据结合预测小麦灌浆期白粉病

马慧琴^{1,2}, 黄文江^{2*}, 景元书¹

(1. 南京信息工程大学, 应用气象学院, 气象灾害预报预警与评估协同创新中心, 南京 210044;

2. 中国科学院遥感与数字地球研究所, 数字地球重点实验室, 北京 100094)

摘要: 利用多源数据对区域尺度上小麦白粉病的发生状况准确及时地预报能为农业服务和农业植保等部门提供重要信息, 实现小麦白粉病的有效预防。研究利用一景 2014 年 5 月 6 日的 landsat8 遥感影像提取出植被指数、地表温度(land surface temperature, LST)和影像中各波段反射率特征, 同时用 2014 年 3 月—5 月份的站点逐日地面气象资料计算获得各气象特征, 并经过 GIS 空间插值分析得到相应的空间气象特征。通过 Relief 算法和泊松相关系数相结合的方式对遥感和气象特征的筛选, 最终得出改进的简单比值指数(modified simple ratio index, MSR)、重归一化植被指数(re-normalized difference vegetation index, RDVI)、3 月 21 日—4 月 20 日总日照时数和 4 月 11 日—5 月 10 日大于 0.1 mm 降雨日数。采用相关向量机(relevance vector machine, RVM)的方法分别用筛选出的遥感、气象数据特征及 2 种数据特征相结合的方式构建了河北省石家庄市藁城、晋州和赵县 3 地区小麦灌浆期白粉病的发生预测模型, 并对 3 种不同数据模型进行了验证与评估。试验结果表明, 遥感气象数据模型的总体精度达到 84.2%, 优于遥感数据模型的 80.0%和气象数据模型的 74.7%。进而得出, 相比于单站点准确和空间不连续的气象数据和类型单一的遥感数据, 遥感气象数据更适合于区域尺度范围内的作物病虫害发生发展状况的预测研究。

关键词: 遥感; 气象; 预测; 相关向量机; 气象数据; 小麦白粉病;

doi: 10.11975/j.issn.1002-6819.2016.09.023

中图分类号: S4; TP79

文献标志码: A

文章编号: 1002-6819(2016)-09-0165-08

马慧琴, 黄文江, 景元书. 遥感与气象数据结合预测小麦灌浆期白粉病[J]. 农业工程学报, 2016, 32(9): 165—172.

doi: 10.11975/j.issn.1002-6819.2016.09.023 http://www.tcsae.org

Ma Huiqin, Huang Wenjiang, Jing Yuanshu. Wheat powdery mildew forecasting in filling stage based on remote sensing and meteorological data[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2016, 32(9): 165—172. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2016.09.023 http://www.tcsae.org

0 引言

小麦白粉病是影响小麦生长发育的主要病害之一。近年来随着气候变化等因素的影响, 中国小麦白粉病的发生发展和分布特征发生了变化, 对其研究工作也越来越多, 研究方法与内容也在不断改进、加深^[1]。小麦白粉病常在小麦生育后期爆发, 受害后一般可减产 5%~10%, 重病田达 20%以上。准确及时地预测小麦白粉病对病害的有效预防具有重要意义。

目前, 关于小麦白粉病的预测研究主要集中在气象因子对小麦白粉病影响的研究, 多基于田间观测, 通过与气象资料对比分析, 建立小麦白粉病发生发展与各气象因子及因子组合间的定量关系^[2]。国内外学者利用气象数据进行了作物病虫害预报研究。居为民等^[3]根据小麦白粉病始病期、严重度和同期气象资料, 采用滑动相关普

查法分析了气象条件对白粉病发生的影响作用, 并在此基础上建立了白粉病的预报模式。钱拴等^[4]采用影响我国小麦白粉病发病流行的关键环流特征因子距平作为预测因子, 建立了白粉病长期气象预测模式。姚树然等^[5]用气象资料和白粉病观测调查资料通过灾变规律关键时方法建立了白粉病发生程度的预测指标模型。Stansbury 等^[6]利用湿热指数模型和降雨模式 2 种气象建模方法对西澳大利亚州地区小麦印度腥黑穗病的易感时期进行了预报, 并分析了小麦易染该病的气候原因。然而, 利用气象数据预测白粉病的方法具有一定的局限性。尽管气象条件是小麦白粉病发生流程度度的决定因素^[3], 但其他因子如小麦长势信息、农田环境(土壤含水量, 地表温度)、作物品种和施肥措施等条件与小麦白粉病的发生也息息相关。另外, 用于建立气象数据预测模型的气象站点数据具有单点准确客观、因其数量有限而空间信息不连续的特点, 从而使得模型虽能够给出较大范围的病害发生概率, 但其在面上并不连续, 无法给出一定区域内地块级的空间病害发生概率。

相比于气象数据, 遥感数据可以给出连续的空间信息, 同时可实现对作物的生长和生境信息的实时反馈。目前采用遥感数据预测病害以及将遥感和气象信息结合用于病害预测的研究较少。沈文颖等^[7]通过大田小区人工接种试验, 采用不同白粉病危害程度下的常规光谱特征

收稿日期: 2015-12-29 修订日期: 2016-01-22

基金项目: 中国科学院国际合作局对外合作重点项目: 主要作物主要病虫害遥感监测与预警研究(131211KYBS20150034); 国家自然科学基金项目“多源数据小麦病害遥感识别与监测方法研究”(41271412); 江苏省普通高校自然科学基金资助项目(15KJA170003)。

作者简介: 马慧琴, 女, 甘肃人, 现主要从事农业气象和植被定量遥感研究。南京 南京信息工程大学应用气象学院, 210044。Email: 1033513161@qq.com
*通信作者: 黄文江, 博士, 研究员, 博士生导师。北京 中国科学院遥感与数字地球研究所数字地球重点实验室, 100094。Email: huangwj@radi.ac.cn

参数、比值指数和归一化指数, 利用因子分析 (factor analysis, FA) 与 BP 神经网络 (back propagation neural network, BPNN) 相结合的方法对小麦叶片白粉病严重度进行了模型模拟, 结果表明反演模型对小麦白粉病的整个病症期均具有很好的适用性。张竞成等^[8]采用环境星影像数据反演的各种植被指数, 地表温度及作物干旱指数, 通过 logistic 回归方法构建了小麦白粉病的发生预测模型。张竞成等^[9]还利用气象因子结合环境星数据, 通过 logistic 回归方法来构建小麦白粉病发生概率预测模型。罗菊花等^[10]利用 TM (landsat5 TM) 影像反演的修订后的归一化水分植被指数 (modified normalized difference water index, MNDWI) 和地表温度 (land surface temperature, LST) 构建了二维特征空间来预测小麦蚜虫。Jonas Franke 等^[11]探讨了多光谱遥感对作物疾病多时相分析的潜力, 通过 3 景遥感影像分析病害的时空变化, 并采用决策树、混合调谐匹配滤波结果和归一化植被指数来区分不同病害程度的小麦区域。罗菊花等^[12]利用环境星数据反演的地表温度、垂直干旱指数 (perpendicular drought index, PDI) 和归一化植被指数 (NDVI) 结合温度、降水、风速等气象数据, 通过 Logistic 回归方法建立了小麦蚜虫发生预测模型。上述研究表明作物长势信息和生境特征与病虫害的发生发展状况密切相关。

当前对作物病虫害的预报方法以数理统计应用为主, 同时还有专家评估及系统模拟的方法。这些方法依次具有普适性差、人为主观性和构建困难的缺陷。相关向量机 (relevance vector machine, RVM) 是一种机器学习和数据挖掘的常用方法。因其具有优越的学习能力, 常被用于商业决策的改善、疾病检测及天气预测等^[13-17]。唐翠翠等^[18]利用该方法对北京郊区的冬小麦蚜虫的发生概率进行了预测, 并取得了较高的预测精度。

基于上述分析, 考虑影响小麦白粉病发生因子的多样性, 本文尝试通过气象数据和遥感数据以及 2 种数据相结合, 利用相关向量机的方法对区域尺度的小麦白粉病的发病情况开展预测研究, 并对 3 种数据所建立的预测模型的优越性进行评价。

1 材料与方 法

1.1 研究区概况

研究区位于河北省石家庄市的藁城 (114°39'~114°59'E, 37°51'~38°18'N)、晋州 (114°58'~115°12'E, 37°48'~38°10'N) 和赵县 (114°37'~115°4'E, 37°37'~37°53'N) (如图 1)。该区域在气候分区中属黄河流域白粉病易发气候区^[19]。统计资料显示, 该地区白粉病发生时间连续性强, 爆发特征明显, 受灾面积范围大。

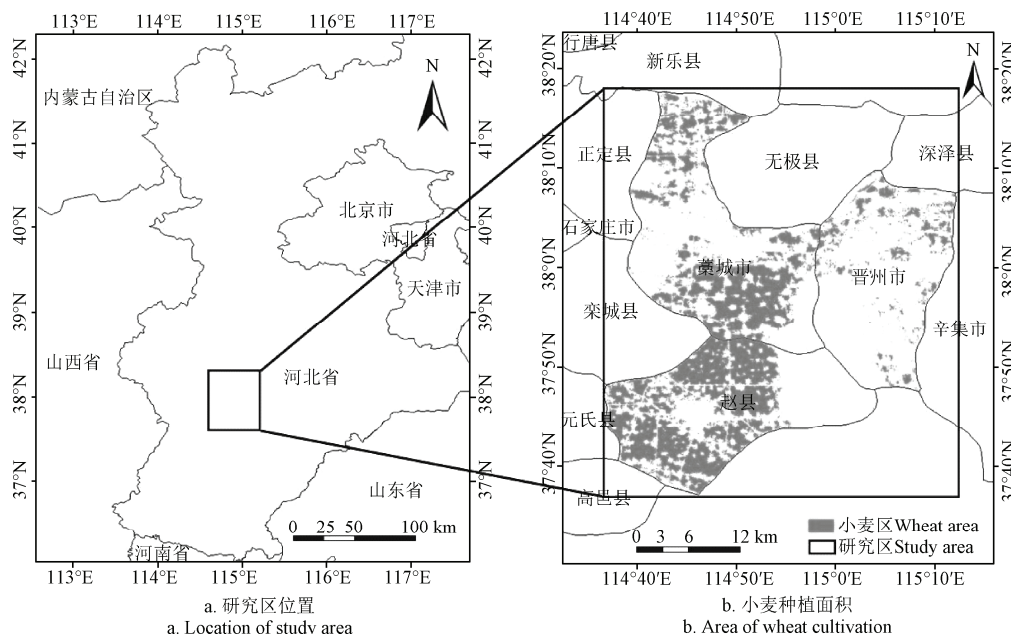


图 1 研究区概况

Fig.1 General situation of study area

1.2 数据获取

研究所用数据主要包括遥感数据、气象数据和小麦白粉病地面调查数据。遥感数据为由 Landsat8 传感器获取的 2014 年 5 月 6 日、2014 年 5 月 22 日和 2014 年 6 月 5 日 3 景影像, 其基本参数见表 1。气象数据来源于河北省气象局各站点气象资料库。区域范围包括石家庄市藁城、晋州、赵县 3 地区及其周边的高邑、栾城、深泽、石家庄市、无极、辛集、元氏、正定共 11 个市县。重点获取了 2014 年 3 月到 5 月的逐日平均气温、降雨量、日

照时数和相对湿度等地面气象数据。小麦白粉病病害地面调查数据于 2014 年 5 月下旬 (小麦灌浆期) 在藁城、晋州和赵县调查获得。获取时根据 Landsat8 数据的空间分辨率, 确定每个调查样点为 30 m×30 m 的正方形区域, 用手持式 GPS 定位仪记录样点区域的中心经纬度信息。尽量选取范围内全为小麦, 地势平坦且远离建筑物的地块。每块样地中采用分布均匀的 9 点调查法调查, 每点调查面积为 1 m²。调查内容包括小麦发病严重度、小麦长势、株高等。最终获得 95 个样点数据。

表 1 landsat8 卫星的基本参数

Table 1 Basic parameters of landsat8 satellite

主要载荷 Main load	波段号 Band NO.	波段名称 Band name	光谱范围 Spectral range/ μm	空间分辨率 Spatial resolution/m
运行陆地成像仪 Operational land imager, OLI	1	New deep blue	0.433~0.453	30
	2	Blue	0.450~0.515	30
	3	Green	0.525~0.600	30
	4	Red	0.630~0.680	30
	5	NIR	0.845~0.885	30
	6	SWIR2	1.560~1.660	30
	7	SWIR3	2.100~2.300	30
	8	PAN	0.500~0.680	15
	9	SWIR	1.360~1.390	30
热红外传感器 Thermal infrared sensor, TIRS	10	TIR	10.30~11.30	100
	11	TIR	11.50~12.50	100

注：辐射分辨率为 12 bit，重访周期为 16 d。
Note: Radiometric resolution is 12 bit and revisit cycle is 16 d.

1.3 数据处理

1.3.1 遥感数据处理

获取研究需要的数据后，首先需要对影像进行辐射定标和大气校正及影像裁剪处理。辐射定标的公式如下

$$L = ML \times Q_{cal} + AL \quad (1)$$

式中 L 为辐射亮度，其单位为 $\text{W}/(\text{m}^2 \cdot \text{Sr} \cdot \mu\text{m})$ ， ML (RADIANCE_MULT_BAND_x) 为卫星传感器的增益值，

AL (RADIANCE_ADD_BAND_x) 为卫星传感器补偿值 (偏置)， Q_{cal} 为影像像元灰度值。其中 ML 和 AL 可直接在影像头文件中获取。热红外波段的辐射亮度公式如下

$$L = 3.3420 \times 10^{-4} \times Q_{cal} + 0.1 \quad (2)$$

辐射定标完成后，采用 ENVI5.1 软件中的 FLAASH 大气校正模块将影像的辐射亮度转为反射率。

遥感数据预处理完成后，根据当地的作物类型及特点，首先利用 2014 年 5 月 22 日的影像通过 NDVI、数字高程模型 (digital elevation model, DEM) 的阈值设置初步剔除非植被地物和部分林地、再根据 2014 年 5 月 6 日和 2014 年 6 月 5 日 2 景影像的近红外波段反射率的阈值设置剔除草地，最后用 3 景影像的近红外波段合成的假彩色图像结合监督分类中最大似然分类的方法最终实现对研究区小麦种植面积的提取，总体精度达到 94%。最终提取的小麦种植面积分布见图 1b。

之后，利用预处理后的 2014 年 5 月 6 日的遥感影像进行了遥感数据预测特征的提取。本文选取出 landsat8 各波段反射率和对小麦白粉病比较敏感的几种植被指数^[20]以及表征生境因子的地表温度 (LST) 共 15 个特征作为白粉病预测模型的初选遥感特征因子。各植被指数的具体名称及计算方法详见表 2。采用常用的单通道算法利用热红外 11 波段对研究区的地表温度进行反演^[21-22]。

表 2 宽波段植被指数计算公式
Table 2 Vegetation index of wide band range

植被指数 Vegetation Index	缩写 Abbreviation	计算公式 Formula	参考文献 References
归一化植被指数 Normalized difference vegetation index	NDVI	$(R_{NIR} - R_R) / (R_{NIR} + R_R)$	Rouse et al, 1973
改进的简单比值指数 Modified simple ratio index	MSR	$(R_{NIR} / R_R - 1) / (\sqrt{R_{NIR} / R_R} + 1)$	Chen, 1996
重归一化植被指数 Re-normalized difference vegetation index	RDVI	$(R_{NIR} - R_R) / \sqrt{R_{NIR} / R_R}$	Roujean & Breon, 1995
三角植被指数 Triangular vegetation index	TVI	$60 \times (R_{NIR} - R_G) - 100 \times (R_R - R_G)$	Broge et al, 2000
优化的土壤调节植被指数 Optimized soil adjusted vegetation index	OSAVI	$(R_{NIR} - R_R) / (R_{NIR} + R_R + 0.16)$	Rondeaux et al, 1996
绿波段归一化植被指数 Green normalized difference vegetation index	GNDVI	$(R_{NIR} - R_G) / (R_{NIR} + R_G)$	Gitelson et al, 1994

注： R_{NIR} 为近红外波段的反射率； R_R 为红波段的反射率； R_G 为绿波段的反射率。
Note: R_{NIR} means near-infrared reflectance; R_R means red band reflectance; R_G means green band reflectance.

1.3.2 气象数据处理

大量研究表明小麦白粉病的发生发展与气候条件密切相关^[2-5]。因此，气象因子是病害发生发展的重要驱动因子，尤其是气温、湿度、降雨和风速。在温湿度适宜的条件下，病菌的繁衍生育速度最快；降雨是通过影响大气的相对湿度来间接影响病菌的产生和消失；寡照则有利于白粉病的流行。

考虑到研究区小麦白粉病的发病时间及特点，并结合文献调研结果，将通过气象站点的逐日地面气象资料计算所得的 3 月和 4 月的月平均气温，月大于 0.1mm 降雨日数，月总日照时数、月平均相对湿度、月温雨系数 (温雨系数=某时段降水总量/该时段平均气温) 和月降雨系数 (降雨系数=(降雨量×降雨日数)^{1/2})^[23]；3 月下旬至 5 月上旬的旬平均气温、旬大于 0.1 mm 降雨日数、旬总日照时数、旬平均相对湿度、旬温雨系数和旬降雨系数；3 月 21 日~4 月 20 日平均气温、大于 0.1 mm 降雨日数、总日照时数、平均相对湿度、温雨系数和降雨系

数；以及 4 月 11 日~5 月 10 日平均气温、大于 0.1 mm 降雨日数、总日照时数、平均相对湿度、温雨系数和降雨系数 54 个特征因子作为白粉病预测模型的初选气象特征。

因有限的气象站点无法满足气象因子的时空分布精度，虽然测量精度较高，但对样本点的代表性差。为获取调查样点相对较为准确的各样本点的气象预测特征，利用 GIS 软件对气象特征进行空间插值分析。有学者研究发现对于同一种插值方法，参与插值的气象站点数目不同，插值结果不同，只有选择合适的站点数目时才会有很高的精度^[24-25]。基于此，在插值分析的站点选择时，不仅采用了藁城、晋州和赵县这 3 个研究区站点的数据，同时还结合了研究区周边的市县总共 11 个气象站点的数据，具体的气象站点分布见图 2。研究通过 arcGIS10.2 软件空间分析中的反距离加权法 (inverse distance weighted, IDW) 对研究区的各气象特征进行空间插值分析。

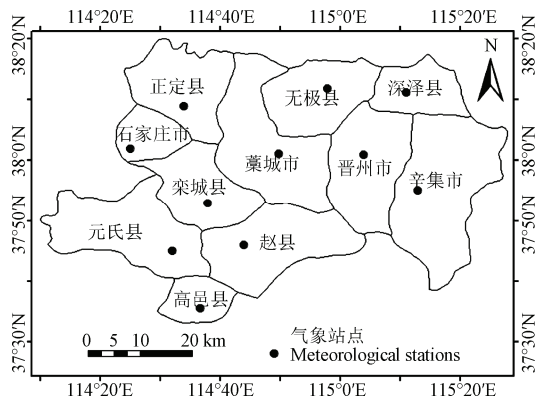


图2 气象站点分布

Fig.2 Distribution of meteorological stations

1.4 模型特征的选取及模型的构建

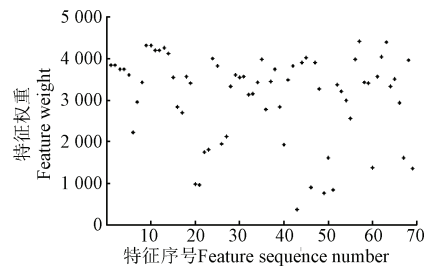
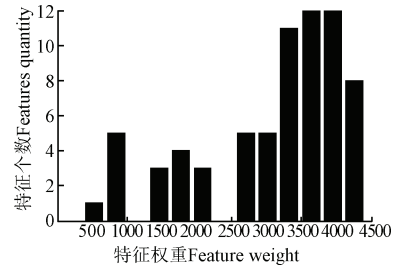
1.4.1 模型特征的选取

特征选择是统计学领域的经典问题，也是机器学习领域的重要问题，对机器学习领域的所有问题都有重大意义。合适的特征选择可以有效去除不相关和冗余特征，提高学习算法的效率^[26]。很多学习算法受不相关或冗余特征的负面影响，而选择好的特征既可以减小计算复杂度，提高分类准确度，又有助于寻找更精简更易理解的算法模型；另外，不断出现的数目庞大或特征维数高的大规模数据处理问题对特征选择算法提出了严峻的挑战^[27]。

Relief 算法最早由 Kira^[28]提出，是公认的性能较好的特征评估方法，主要是根据特征对近距离样本的区分能力来评估特征。该算法核心思想为：好的特征应使同类样本接近，而使不同类样本远离。Relief 算法具有运行效率高，不限制数据类型及对特征间关系不敏感的优点，能较好地去除无关特征，但不能去除冗余特征^[29]。为利用 Relief 算法的优点，同时解决其不能去除冗余的缺点，通过采用 Relief 与泊松相关系数 (poisson correlation) 相结合的方法来进行预测特征选取^[29]。首先，通过 Relief 算法计算出共 69 个遥感及气象特征在小麦健康与发生白粉病分类中的权重，图 3 为计算所得的各特征的权重分布图。之后，结合计算获得的特征权重分布图中各特征变量的权重分布情况，为使得筛选出的特征变量的数量不至于过多，将权重阈值设置为 4000，筛选出权重最大的 11 个特征变量。然后将筛选出的特征变量两两计算其泊松相关系数，在冗余度大的 2 个特征中，保留分类权重较大的特征。最终，筛选出了 MSR、RDVI、3 月 21 日~4 月 20 日总日照时数和 4 月 11 日~5 月 10 日大于 0.1 mm 降雨日数 4 个特征用于小麦白粉病发生预测模型的构建。

1.4.2 模型的构建

获取的地面调查点中共有 95 个点与遥感影像同步。调查点的病害严重度分 5 个等级，0 (无病害)、1 (轻度)，2 (中度)，3 (重度)，4 (特重)。考虑到轻度病害与无病害小麦较难区分，研究将无病害与轻度 2 个等级合并为健康，用 -1 表示，而将其余中度、重度、特重 3 个等级合并为发病，用 1 表示。

a. 基于 Relief 算法得到的不同特征的权重值
a. Different features weight values based on Relief algorithmb. 特征权重值分布
b. Distribution of feature weight values图3 Relief 算法特征权重计算结果图
Fig.3 Calculation results chart of feature weight by Relief algorithm

相关向量机是一种新的基于贝叶斯统计学习理论的学习方法，与支持向量机 (support vector machine, SVM) 相比，具有概率型输出、更稀疏和核函数选择更自由等优点^[30]。目前相关向量机国内外主要应用于医学影像处理、数字图像处理和高光谱图像分类等领域^[31,32]。本研究中小麦白粉病的发生预测实质为二分类问题，而相关向量机用于二分类预测的模型形式为

$$P(t|\lambda) = \prod_{i=1}^n \sigma(y(x_i; \lambda))^{t_i} (1 - \sigma(y(x_i; \lambda)))^{1-t_i} \quad (3)$$

式中 λ 为样本的权重向量， n 为样本数， σ 为逻辑函数， x_i 为训练样本向量， t_i 为训练样本标号， $(x_i, t_i) (i=1, 2, \dots, n)$ 为训练样本集合， y 为分类函数。若假设 t 等于 1 代表小麦发病， t 等于 -1 表示病害未发生，则预测规则为：若 $P \leq 0.5$ ，则白粉病不发生；若 $P > 0.5$ ，则发生白粉病。

研究采用由 Relief 算法结合泊松相关系数的方法筛选获得 4 个特征变量，利用 RVM 方法，构建仅以遥感数据 MSR、RDVI 为输入变量，仅以气象数据 3 月 21 日~4 月 20 日总日照时数和 4 月 11 日~5 月 10 日大于 0.1 mm 降雨日数为输入变量，以及以遥感气象数据 MSR、RDVI、3 月 21 日~4 月 20 日总日照时数和 4 月 11 日~5 月 10 日大于 0.1 mm 降雨日数为输入变量的 3 个不同数据类型的小麦白粉病发生预测模型。

在 95 个地面调查样本点中，均匀地选取出总样本数的 2/3 即 63 个样本数据作为训练样本用于模型构建，而剩余的 3/1 即 32 个样本数据用于模型的验证。

2 结果与分析

2.1 研究区小麦白粉病发生预测

采用研究区 2014 年 5 月 6 日遥感影像数据以及 2014

年 3~5 月份气象数据插值结果数据，以单个像元为基本处理单元，用 Relief 算法与泊松相关系数相结合的方式筛选出 4 个特征量 MSR、RDVI、3 月 21 日~4 月 20 日总日照时数和 4 月 11 日~5 月 10 日大于 0.1 mm 降雨日数，利用 RVM 方法建立遥感数据、气象数据和遥感气象数据模型，得到了 2014 年 5 月下旬小麦灌浆期的白粉病发生空间分布图（图 4）。从预测结果分布图可以看出，2014 年河北省石家庄市藁城、晋州、赵县这 3 个地区白粉病发生严重。3 种数据模型相对比发现，不同数据模型预测的结果差异较大，气象数据预测结果中赵县麦区均发生白粉病，藁城和晋州也全面发生，健康

小麦分布为极少部分；遥感数据预测结果中 3 地区均有健康麦区与病害发生麦区，从图中（图 4b）可以看出赵县发生病害面积最大，藁城次之，晋州病害面积最少；遥感气象数据预测结果介于二者之间。但 3 种数据预测的结果大体趋势一致，均为赵县白粉病发生最为严重，藁城次之，晋州发病最轻。另外，比较气象数据和遥感数据模型发现，遥感数据模型预测结果中，小麦健康地块与病害地块之间的分布较为均匀；而气象数据模型预测结果则表现为整块的病害区域，且健康麦区为极少的一部分，这一结果间接表明气象数据模型的拟合优度不高。

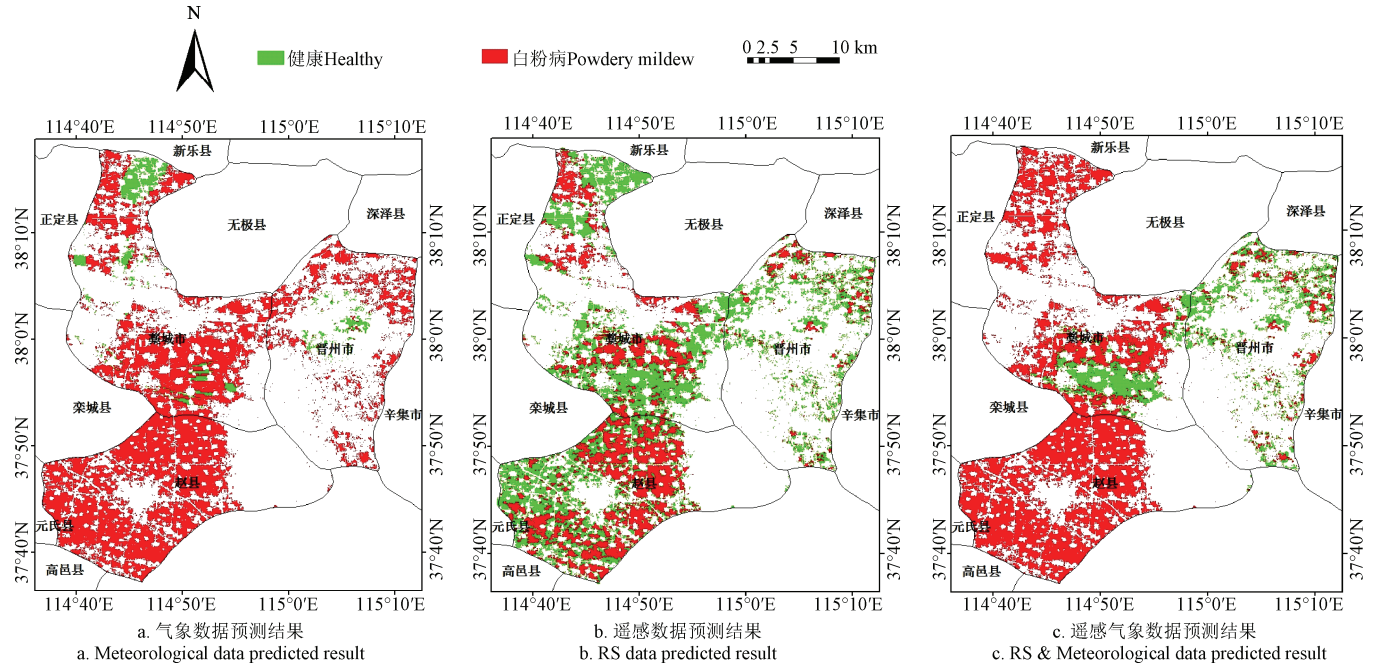


图 4 RVM 模型预测小麦白粉病发生空间分布图

Fig.4 Predicted spatial distribution map of wheat powdery mildew by relevance vector machine

2.2 模型的评估与验证

对 3 种数据建立的模型进行 Spearman 拟合优度检验，并通过配对样本检验和有序变量的相关性分析获取统计量参数 Somers'D、kendall's Tau-c 和 Goodman-Kruskal Gama^[33-35]。3 种参数的取值范围分别在[-1,+1]之间，值越大表明模型精度越高。从表 3 可以看出，3 种数据模型的 Spearman 相关性值为气象数据 0.511，遥感数据 0.609，遥感气象数据 0.693，且均达到极显著相关。

表 3 RVM 模型的拟合优度评价

Table 3 Evaluation of goodness of fit of RVM models

数据类型 Data types	统计量参数 Statistic parameters			
	Spearman relevance	Somers'D	Kendall's Tau-c	Goodman- Kruskal gamma
气象数据 Meteorological data	0.511***	0.511	0.506	0.827
遥感数据 RS data	0.609***	0.609	0.605	0.897
遥感气象数据 RS & Meteorological data	0.693***	0.693	0.689	0.946

注：***表示显著性达到 0.001 显著水平。

Note: *** Means the significance reaches 0.001 significant level.

3 种模型的 Somers'D、kendall's Tau-c 和 Goodman-Kruskal

Gama 值均表现为遥感气象数据模型最大，遥感数据模型次之，气象数据模型最小，说明 3 种数据模型中遥感气象数据模型的拟合优度最好。

独立样本数据对模型的稳定性和可靠性进行验证更能反映模型实际的预测精度^[18]。采用 2014 年 5 月下旬小麦灌浆期的地面调查数据对模型的结果进行评价。表 4 列举了小麦白粉病模型的预测样本、训练样本及总样本的漏分、错分信息及模型的预测精度、训练精度、总体精度及其 Kappa 系数。从 3 种数据模型的预测情况来看，在 3 种数据模型中气象数据模型的预测精度最低，遥感数据对小麦白粉病发生的预测精度最高，遥感气象数据模型则介于二者之间，且 3 种数据预测模型均可取得较为满意的预测效果。从模型的训练情况可以看出，遥感气象数据模型的训练精度最高，气象数据模型最低，遥感数据模型则介于二者之间。对比 3 种数据模型的训练情况和预测情况发现，气象数据模型和遥感数据模型均表现为训练精度低于预测精度；而遥感气象数据模型则表现为训练精度高于预测精度，尽管其训练精度高于遥感数据模型的训练精度，但其预测精度却低于遥感数据

模型。考虑导致上述结果的原因可能是, 遥感气象数据模型相对于另 2 种模型变量较多, 模型相对较复杂, 同时遥感变量与气象变量间存在一定的相关性, 导致该模型在训练过程中出现过拟合现象。考虑模型的总体情况发现, 遥感气象数据模型的总体精度最高, 为 84.2%, 且 Kappa 系数也最大, 为 0.686; 遥感数据模型的总体精度 (80.0%) 和 Kappa 系数 (0.602) 次之, 而气象数据模型

的总体精度 (74.7%) 和 Kappa 系数 (0.500) 均最小。从模型总的漏分及错分情况来看, 3 种数据模型均表现为将病害地块错分为健康地块的情况较为严重; 遥感气象数据模型无论是对健康地块还是病害地块, 其漏分和错分均为最小, 且明显低于气象数据模型和遥感数据模型。以上结果表明, 与单一类型数据源相比, 多源数据更适合于区域空间尺度范围内小麦病害预测模型的构建。

表 4 RVM 预测模型的总体验证结果
Table 4 Overall verification results of RVM model

数据类型 Data types	实际样本 The actual samples	预测样本 Prediction samples				训练样本 Training samples				总体样本 Overall samples				
		健康 Healthy	白粉病 PM	总和 Sum	精度 Accuracy/%	健康 Healthy	白粉病 PM	总和 Sum	精度 Accuracy/%	健康 Healthy	白粉病 PM	总和 Sum	精度 Accuracy/%	Kappa 系数 Kappa coefficient
气象 Meteorological	健康 Healthy	13	6	19		23	11	34		36	17	53		
	白粉病 PM	1	12	13	78.1	6	23	29	73.0	7	35	42	74.7	0.500
	总和 Sum	14	18	32		29	34	63		43	52	95		
遥感 RS	健康 Healthy	14	5	19		23	8	31		37	13	50		
	白粉病 PM	0	13	13	84.4	6	26	32	77.8	6	39	45	80.0	0.602
	总和 Sum	14	18	32		29	34	63		43	52	95		
遥感结合气象 RS & Meteorological	健康 Healthy	11	3	14		28	8	36		39	11	50		
	白粉病 PM	3	15	18	81.3	1	26	27	85.7	4	41	45	84.2	0.686
	总和 Sum	14	18	32		29	34	63		43	52	95		

3 结 论

本文利用 landsat8 遥感数据和气象数据建立了小麦白粉病的发生预测模型, 选取了表征小麦生长状况的 MSR 和 RDVI 以及对白粉病的发生有决定作用的 3 月 21 日~4 月 20 日的总日照时数以及 4 月 11 日~5 月 10 日大于 0.1 mm 降雨日数, 采用相关向量机方法对区域尺度的小麦白粉病发生概率进行预测, 并对 3 种数据模型的结果进行了比较分析。结果表明:

1) 在区域尺度范围内的小麦白粉病发生情况预测中, 遥感气象数据模型与气象数据模型和遥感数据模型相比, 其对白粉病的预测总体精度分别提高了 9.5 和 4.2 个百分点。说明遥感气象数据模型的拟合效果更佳, 更适合于空间区域尺度范围内小麦病害发生情况的预测。

2) 采用遥感数据和气象数据作为输入变量的 RVM 模型所得的效果最好, 精度达到了 84.2%。与单一数据类型的模型相比, 遥感气象数据结合的多源数据更为全面地反映了病害信息, 使模型在输入参数中融合了更多信息, 提高了小麦白粉病发生概率预测的精度。

训练样本的数量及质量均会影响模型的精度, 该研究开展时考虑到成本问题, 采集样本数量较少, 仅为河北小部分地区一年的数据, 模型的通用性有待进一步验证。在今后的研究中考虑用多年、多地区的数据来验证模型的通用性。考虑到小麦白粉病的发生受多种因素的共同作用, 模型构建时选取了影响病害发生的重要气象因子, 对小麦白粉病比较敏感的植被指数和生境因子, 但是仍然有一些重要因子因未能获取到, 没有进入模型, 如农田管理信息等。在今后的研究中, 尽可能周全地考虑影响气象数据空间插值的各个因素, 采用更好的获取空间范围内气象数据的方法, 融合更多数据源, 同时寻

找更优的特征筛选方法, 并结合本文方法来构建预测模型, 以便更为准确地进行小麦白粉病的空间分布、发病严重度及发展趋势的预测。

[参 考 文 献]

- [1] 李彤霄. 我国小麦白粉病预报方法研究进展[J]. 气象与环境科学, 2013, 36(3): 44-48.
Li Tongxiao. Research progress of wheat powdery mildew forecasting method[J]. Meteorological and Environmental Sciences, 2013, 36(3): 44-48. (in Chinese with English abstract)
- [2] 郑秋红, 杨霏云, 朱玉洁. 小麦白粉病发生气象条件和气象预报研究进展[J]. 中国农业气象, 2013, 34(3): 358-365.
Zheng QiuHong, Yang Feiyun, Zhu Yujie. Meteorological factors and its forecast on wheat powdery mildew occurrence in China[J]. Chinese Journal of Agrometeorology, 2013, 34(3): 358-365. (in Chinese with English abstract)
- [3] 居为民, 高苹. 气象条件对小麦白粉病发生影响的研究[J]. 气象, 2006, 26(2): 50-53.
Ju Weimin, Gao Ping. The study of influence of meteorological condition on winter wheat powdery mildew[J]. Meteorological Monthly, 2006, 26(2): 50-53. (in Chinese with English abstract)
- [4] 钱拴, 霍治国, 叶彩玲. 我国小麦白粉病发生流行的长期气象预测研究[J]. 自然灾害学报, 2005, 14(4): 56-63.
Qian Shuan, Huo Zhiguo, Ye Cailing. Long-term meteorological prediction research on epidemic of wheat powdery mildew in China[J]. Journal of Natural Disasters, 2005, 14(4): 56-63. (in Chinese with English abstract)
- [5] 姚树然, 李春强, 霍治国, 等. 河北小麦白粉病流行的气象指标研究[J]. 自然灾害学报, 2008, 17(4): 38-43.
Yao Shuran, Li Chunqiang, Huo Zhiguo, et al. A research on climatic indices for epidemic of wheat powdery mildew[J]. Journal of Natural Disasters, 2008, 17(4): 38-43. (in Chinese with English abstract)
- [6] Stansbury C D, Mckirdy S J. Forecasting climate suitability for Karnal bunt of wheat: a comparison of two

- meteorological methods[J]. *Australasian Plant Pathology*, 2002, 31(1): 81—92.
- [7] 沈文颖, 李映雪, 冯伟, 等. 基于因子分析-BP 神经网络的小麦叶片白粉病反演模型[J]. *农业工程学报*, 2015, 31(22): 183—190.
Shen Wenying, Li Yingxue, Feng Wei, et al. Inversion model for severity of powdery mildew in wheat leaves based on factor analysis-BP neural network[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2015, 31(22): 183—190. (in Chinese with English abstract)
- [8] Zhang Jingcheng, Yuan Lin, Nie Chenwei, et al. Forecasting of powdery mildew disease with multi-sources of remote sensing information[C]// *International Conference on Agro-Geoinformatics. IEEE*, 2014:1-5.
- [9] Zhang Jingcheng, Pu Ruiliang, Yuan Lin, et al. Integrating remotely sensed and meteorological observations to forecast wheat powdery mildew at a regional scale[J]. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 2014, 7(11):4328-4339.
- [10] Luo juhua, Zhao Chunjianag, Huang Wenjiang, et al. Discriminating wheat aphid damage degree using 2-dimensional feature sapce derived from landsat 5 TM[J]. *Sensor Letters*, 2012, 10(1/2): 608—614.
- [11] Jonas Franke, Gunter Menz. Multi-temporal wheat disease detection by multi-spectral remote sensing[J]. *Precision Agriculture*, 2007, 8(3): 161—172.
- [12] Luo Juhua, Huang Wenjiang, Zhao Jingling, et al. Predicting the probability of wheat aphid occurrence using satellite remote sensing and meteorological data[J]. *Optik-International Journal for Light and Electron Optics*, 2014, 125(19): 5660—5665.
- [13] Hossain A, Nasser M. Recurrent support and relevance vector machines based model with application to forecasting volatility of financial returns[J]. *Journal of Intelligent Learning Systems and Applications*, 2011, 3(4): 230.
- [14] Majumder S K, Nirmalya G, Gupta P K. Relevance vector machine for optical diagnosis of cancer[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2005, 36(4): 323—333.
- [15] 祝志慧, 刘婷, 马美湖. 基于高光谱信息融合和相关向量机的种蛋无损检测[J]. *农业工程学报*, 2015, 31(15): 285—292.
Zhu Zhihui, Liu Ting, Ma Meihu. Hatching eggs nondestructive detection based on hyperspectral-imaging information and RVM[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2015, 31(15): 285—292. (in Chinese with English abstract)
- [16] Liu X, Hu Z Q, Yang G. Wind speed forecasting based on relevance vector machine and empirical mode decomposition[J]. *Advances in Information Sciences and Service Sciences*, 2013, 5(10): 279—287.
- [17] 李慧杰, 刘亚南, 卫志农, 等. 基于相关向量机的短期风速预测模型[J]. *电力自动化设备*, 2013, 33(10): 28—32.
Li Huijie, Liu Yanan, Wei Zhinong, et al. Short-term wind speed forecasting model based on relevance vector machine[J]. *Electric Power Automation Equipment*, 2013, 33(10): 28—32. (in Chinese with English abstract)
- [18] 唐翠翠, 黄文江, 罗菊花, 等. 基于相关向量机的冬小麦蚜虫遥感预测[J]. *农业工程学报*, 2015, 31(6): 201—207.
Tang Cuicui, Huang Wenjiang, Luo Juhua, et al. Forecasting wheat aphid with remote sensing based on relevance vector machine[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2015, 31(6): 201—207. (in Chinese with English abstract)
- [19] 霍治国, 陈林, 刘万才, 等. 中国小麦白粉病发生地域分布的气候分区[J]. *生态学报*, 2002, 22(11): 1873—1880.
Huo Zhiguo, Chen Lin, Liu Wancai, et al. Climatic zonation of wheat powdery mildew in China[J]. *Acta Ecologica Sinica*, 2002, 22(11): 1873—1880. (in Chinese with English abstract)
- [20] 张竞成. 多源遥感数据小麦病害信息提取方法研究[D]. 浙江大学, 2012.
Zhang Jingcheng. *Methods for Information Extraction of Wheat Disease Based on Multi-source Remote Sensing Data*[D]. Zhejiang University, 2012. (in Chinese with English abstract)
- [21] Jiménez-Muñoz J C, Sobrino J A. A generalized single-channel method for retrieving land surface temperature from remote sensing data[J]. *Journal of Geophysical Research: Atmospheres (1984—2012)*, 2003, 108(D22): 2015—2023.
- [22] 徐涵秋. 新型 Landsat8 卫星影像的反射率和地表温度反演[J]. *地球物理学报*, 2015(3): 741—747.
Xu Hanqiu. Retrieval of the reflectance and land surface temperature of newly-launched Landsat 8 satellite[J]. *Chinese Journal of Geophysics*, 2015(3): 741—747. (in Chinese with English abstract)
- [23] 余卫东, 朱晓东, 杨君健, 等. 商丘市小麦白粉病预测模式研究[J]. *气象与环境科学*, 2008, 31(1): 20—23.
Yu Weidong, Zhu Xiaodong, Yang Junjian, et al. Study on prediction model for wheat powdery mildew in Shangqiu city[J]. *Meteorological and Environmental Sciences*, 2008, 31(1): 20—23. (in Chinese with English abstract)
- [24] 李军龙, 张剑, 张丛, 等. 气象要素空间插值方法的比较分析[J]. *草业科学*, 2006, 23(8): 6—11.
Li Junlong, Zhang Jian, Zhang Cong, et al. Analyze and compare the spatial interpolation methods for climate factor[J]. *Pratacultural Science*, 2006, 23(8): 6—11. (in Chinese with English abstract)
- [25] 马轩龙, 李春娥, 陈全功. 基于 GIS 的气象要素空间插值方法研究[J]. *草业科学*, 2008, 25(11): 13—19.
Ma Xuanlong, Li Chun'e, Chen Quangong. Study on the method of GIS based spatial interpolation of climate factors in China[J]. *Pratacultural Science*, 2008, 25(11): 13—19. (in Chinese with English abstract)
- [26] Kohavi R, John G. Wrappers for feature subset selection[J]. *Artificial Intelligence*, 1997, 97(1/2): 273—324.
- [27] 张丽新. 高维数据的特征选择及基于特征选择的集成学习研究[D]. 北京: 清华大学, 2004.
Zhang Lixin. *Study on Feature Selection and Ensemble Learning Based on Feature Selection for High-Dimensional Datasets*[D]. Beijing: Tsinghua University, 2004. (in Chinese with English abstract)
- [28] Kira K, Rendell L. The feature selection problem: Traditional methods and a new algorithm[C]//*Proceedings of the Ninth National conference on Artificial Intelligence*. New Orleans: AAAI Press, 1992: 129—134.
- [29] 张丽新, 王家威, 赵雁南, 等. 基于 Relief 的组合式特征选择[J]. *复旦学报: 自然科学版*, 2004, 43(5): 893—898.
Zhang Lixin, Wang Jiawen, Zhao Yannan, et al. Combination feature selection based on Relief[J]. *Journal of Fudan University: Natural Science*, 2004, 43(5): 893—898. (in Chinese with English abstract)
- [30] 杨树仁, 沈洪远. 基于相关向量机的机器学习算法研究与应用[J]. *计算技术与自动化*, 2010, 29(1): 43—47.
Yang Shuren, Shen Hongyuan. Research and application of machine learning algorithm based on relevance vector machine[J]. *Computing Technology and Automation*, 2010, 29(1): 43—47. (in Chinese with English abstract)
- [31] 杨国鹏, 余旭初, 周欣, 等. 基于相关向量机的高光谱影像分类研究[J]. *测绘学报*, 2010, 39(6): 572—578.
Yang Guopeng, Yu Xuchu, Zhou Xin, et al. Research on

- relevance vector machine for hyperspectral imagery classification[J]. *Acta Geodaetica et Cartographica Sinica*, 2010, 39(6): 572—578. (in Chinese with English abstract)
- [32] 赵春晖, 张毅. 相关向量机分类方法的研究进展与分析[J]. *智能系统学报*, 2012(4): 294—301.
Zhao Chunhui, Zhang Yi. Research progress and analysis on methods for classification of RVM[J]. *CAAI Transactions on Intelligent Systems*, 2012(4): 294—301. (in Chinese with English abstract)
- [33] Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*[M]. New York: Springer, 2009.
- [34] David H A, Edwards A W F. *Annotated Readings in the History of Statistics*[M]. New York: Springer Science & Business Media, 2001.
- [35] Lemeshow S, Hosmer D W. A review of goodness of fit statistics for use in the development of logistic regression models[J]. *American Journal of Epidemiology*, 1982, 115(1): 92—106.

Wheat powdery mildew forecasting in filling stage based on remote sensing and meteorological data

Ma Huiqin^{1,2}, Huang Wenjiang^{2*}, Jing Yuanshu¹

(1. Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, School of Applied Meteorology, Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China)

Abstract: Powdery mildew is one of the main serious diseases for winter wheat. An accurate and timely forecasting of the wheat powdery mildew occurrence at the regional scale by using multi-source data can provide important information for crop protection decision making, and achieving effective prevention of wheat powdery mildew. In this study, the Landsat8 remote sensing image was used to extract the land surface temperature (LST), the vegetation indices which included normalized difference vegetation index (NDVI), modified simple ratio index (MSR), re-normalized difference vegetation index (RDVI), triangular vegetation index (TVI), optimized soil adjusted vegetation index (OSAVI), green normalized difference vegetation index (GNDVI), and the band reflectance features. Then we obtained the parameters of wheat growth environment condition such as air temperature, number of rainy days with more than 0.1 mm rainfall, total sunshine hour, average relative humidity, temperature-rain coefficient (the ratio of total rainfall in a period of time to average temperature of the same period) and rainfall coefficient (the square root of the product of rainfall and number of rainy days) in different time steps (including month, 10 days and sensitive period) with the site daily meteorological data; and then we got the corresponding space meteorological features by using the inverse distance weighted (IDW) method in GIS (geographic information system) spatial interpolation analysis. Next, we implemented screening features with the combination of relief algorithm and Poisson's correlation coefficient, and finally got the MSR, the RDVI, the total sunshine hour from March 21st to April 20th, and the number of rainy days with more than 0.1 mm rainfall from April 11th to May 10th, which were as optimal explanatory variables for developing the powdery mildew forecasting model. The relevance vector machine (RVM) model was used to improve business decisions, detect disease, and forecast weather. And then we used it to predict the probability of powdery mildew occurrence in filling stage of wheat in Gaocheng, Jinzhou and Zhaoxian County, Shijiazhuang City, Hebei Province through remote sensing and meteorological data. The model combining remote sensing and meteorological data produced a higher Spearman relevance value than the single remote sensing data or the meteorological data model, and moreover, the values of Somers'D, Goodman-Kruskal Gamma, and Kendal's Tau-c of the remote sensing and meteorological data model were all higher than those of the other 2 models. They all indicated that the remote sensing and meteorological data model had a better performance than the other 2 models. The results showed that: the overall accuracy of the remote sensing and meteorological data model was the highest among the 3 methods, with lower omission and wrong judgement than the other 2 models. Furthermore, the overall accuracy and the kappa coefficient of the remote sensing and meteorological data model were 84.2% and 0.686 respectively, which showed better performance over the remote sensing data model (80.0% and 0.602) and the meteorological data model (74.7% and 0.500). These results reveal that compared with the single meteorological data or remote sensing data, the combination of remote sensing and meteorological data is more suitable for the prediction of crop disease occurrence situation in the regional scale.

Keywords: remote sensing; meteorology; forecasting; relevance vector machine; meteorological data; wheat powdery mildew