

# 基于 PCA 和多元回归算法构建大棚黄瓜霜霉病预警系统

李映, 葛喜珍\* (北京联合大学生物化学工程学院, 北京 100023)

**摘要** 采用大棚气象条件检测系统记录了大棚黄瓜在不同种植季的环境条件变化, 将其与不同季节的黄瓜霜霉病发病率的数据进行了相关研究。利用机器学习领域的主成分分析(PCA), 使用 50% 的数据作为训练集, 找出了 2 种权重较高的特征值, 并针对其与霜霉病发病率进行了多元回归分析, 计算出最佳的回归模型。利用该模型, 对剩余 50% 的数据进行预测。结果表明, 该降维分析和多项式回归得到的预警系统可根据环境检测数据, 有效预测霜霉病的发病率, 准确度达到 85%, 为黄瓜霜霉病的预防提供了重要的预警信息。

**关键词** 大棚黄瓜; 霜霉病; 预警系统; 机器学习

中图分类号 S-058 文献标识码 A

文章编号 0517-6611(2022)21-0232-03

doi: 10.3969/j.issn.0517-6611.2022.21.058

开放科学(资源服务)标识码(OSID):



## Construction of Early-warning System of Greenhouse Cucumber Against Downy Mildew with PCA and Multiple Regression Algorithm

LI Ying, GE Xi-zhen (College of Biochemical Engineering, Beijing Union University, Beijing 100023)

**Abstract** By using the environmental condition record system, data of greenhouse cucumber planting were obtained. Its relationship with the plant disease rate of downy mildew was analyzed. PCA dimension reduction algorithm in machine learning was used to treat 50% of the total data as training data to construct an early-warning system. As a result, 2-dimension matrix accompanied with a regression equation was confirmed effective to predict the disease rate at similar conditions. The predicting of the remaining infection rate indicated this method had accuracy over 85% compared to the recorded data. In general, this research provided a novel method on predicting the infection rate of downy mildew of greenhouse cucumber.

**Key words** Greenhouse cucumber; Downy mildew; Early-warning system; Machine learning

黄瓜霜霉病(downy mildew)是一种由真菌引起的病害, 是大棚黄瓜种植中常年发生的最普遍、最难防治的病害<sup>[1-3]</sup>。黄瓜霜霉病主要由病原真菌 *Pseudoperonospora cubensis* 引起, 从幼苗到收获阶段均可发生, 以成株的危害较重<sup>[4-5]</sup>。该种病原真菌主要侵染黄瓜的叶片, 由叶片基部向上部发展, 在叶面形成浅黄色或白色病斑, 有时可延伸至叶片背面<sup>[6]</sup>。病斑会导致叶片枯黄, 使得植物失去正常的生理功能和作用, 最后导致植物的死亡。由于霜霉病的传染性较强且引起的损失惨重, 因此对霜霉病的防治一直以来是大棚黄瓜种植的重要工作<sup>[7]</sup>。用于防治霜霉病的化学农药主要有波尔多液、石硫合剂等, 生物农药主要有苦参碱、芽孢杆菌制剂等<sup>[8-10]</sup>。但喷洒农药只能抑制霜霉病的蔓延, 对霜霉病的预防和预测则是更有效的控制手段。

大棚黄瓜霜霉病的发病规律已经较为清楚。在一年中的不同种植季, 霜霉病往往出现在温度较低、湿度较大、光照强度较小的种植季<sup>[11]</sup>。在该气象条件下, 黄瓜霜霉病的病原菌可以实现高效的繁殖, 并产生大量的孢子蔓延到大棚内其他区域<sup>[12]</sup>。目前, 已有使用孢子收集器或图像识别系统研究黄瓜各类病害的发生规律和预警系统<sup>[13-14]</sup>, 然而这些装置价格和维护成本较高, 且后续的孢子成分分析需要专业设备的人员进行, 因此限制了其广泛的使用。相比之下, 使用气象条件监控系统进行预警是一种可行性较高的选择。鉴于此, 笔者针对河北邢台地区的气温特征, 利用不同大棚内

的气候监控系统, 研究气象条件与霜霉病发生的规律; 使用机器学习中的 PCA 降维算法提高计算速度, 并给出回归方程进行预测。

### 1 材料与方法

**1.1 仪器与试验场地** 大棚气象监测装置由石家庄圣启科技有限公司提供, 每小时取大棚内温度、湿度、光照、土壤温度等气象信息。大棚所在地为河北省隆尧怡东农牧有限公司黄瓜大棚(37°40'N, 114°92'E)。共使用 5 个大棚, 在不同的种植季节记录大棚内的气象条件, 同时统计相应的霜霉病发生率。

**1.2 田间试验设计** 温室长 60.0 m, 宽 7.50 m, 株距 0.40 m, 行距 0.70 m。自移栽之日起 60 d 记录数据。霜霉病的发病率按照 5 点法取样, 每采样点选取 10 株黄瓜, 每株选取上、中、下 3 个部位进行霜霉病发病率的统计。发病率(%) = 病株数/调查总株数 × 100。发病率大于 60.0% 时, 将该试验区域的植物清除。

**1.3 数据分析和预警系统的建立** 数据的分析和统计在 python 中进行, 使用 Sci-Learn 包进行数据分析。根据获取的数据将大棚内的空气温度、湿度和光照强度进行日均值的计算, 将不同种植季节的数据进行合并后进行预测。根据所有数据的不同, 使用归一化的方法将不同单位的数据进行归一以便计算降维分析的权重。在降维分析中使用置信度为 85% 计算特征值数量, 得到降维分析的数据后, 使用多元多项式回归的方法进行回归方程的计算。

### 2 结果与分析

**2.1 不同种植季数据采集** 在每年的单数月份采集邢台地区不同大棚的气象条件和霜霉病的发病率, 并从每个大棚中随机选择 50% 的数据进行组合, 统计和分析霜霉病的发病

**基金项目** 河北省科技成果转化重点项目(19026517Z); 北京市教委科技计划一般项目(KM202011417006)。

**作者简介** 李映(1989—), 男, 河南许昌人, 副教授, 博士, 从事微生物学研究。\* 通信作者, 教授, 博士, 从事生物农药研究。

**收稿日期** 2021-12-31

率,统计结果见表 1。从表 1 可以看出,在不同的季节,霜霉病的发病率出现了明显的不同。在空气温度低于 20 ℃ 的季节,霜霉病的发生率较高。相反的是,在温度较高、光照强度

较高的种植季,霜霉病的发生率较低,甚至在 10.0% 以下。因此,霜霉病的发生率与环境因素存在非常明显的相关性,使用数学模型可预期达到良好的预测效果。

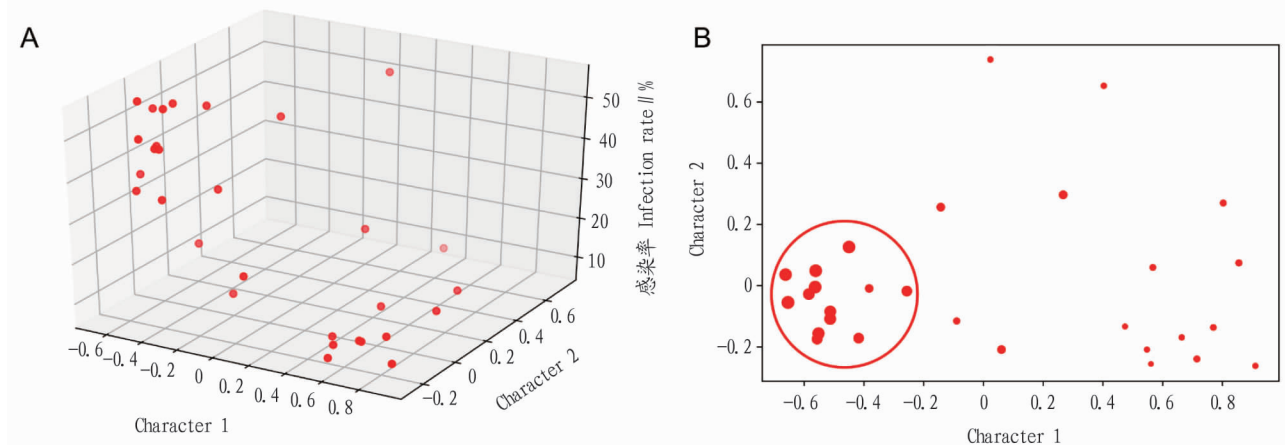
表 1 不同大棚不同种植季节的气象条件和霜霉病发生率

Table 1 Infection rate of downy mildew at different environmental conditions

大棚编号 Greenhouse code	月份 Month	发病率 Disease incidence %	空气温度 Air temperature ℃	空气湿度 Air humidity %	土壤温度 Soil temperature ℃	土壤含水量 Soil moisture content//%	光照强度 Illumination intensity//lx
1	1	48.3	15.5	87.5	13.0	17.9	1 161.9
2	1	36.3	17.4	86.9	11.3	17.9	1 206.8
3	1	35.2	18.5	86.0	13.9	18.0	1 225.9
4	1	44.6	14.9	89.6	14.4	18.2	1 333.8
5	1	45.3	15.2	89.1	14.2	18.1	1 323.3
1	3	51.0	13.3	91.4	15.0	19.3	1 859.7
2	3	52.3	11.6	92.1	13.9	18.8	1 963.8
3	3	50.6	10.2	93.3	13.1	18.6	1 807.5
4	3	35.6	18.8	89.8	16.6	18.8	2 031.0
5	3	43.6	12.5	90.7	12.4	18.5	2 095.1
1	5	16.9	26.8	74.3	23.8	21.9	5 212.0
2	5	19.3	22.7	68.2	22.8	21.3	7 242.0
3	5	12.1	25.4	78.5	21.9	20.5	5 242.0
4	5	11.6	26.6	65.2	22.7	20.1	5 242.0
5	5	13.8	23.3	62.6	21.2	19.7	6 212.0
1	7	9.6	30.8	65.3	25.4	19.5	4 943.7
2	7	10.2	29.1	72.1	25.1	19.4	3 507.8
3	7	8.7	25.3	87.3	23.7	19.1	5 773.0
4	7	9.3	26.5	83.6	24.0	18.9	5 753.9
5	7	13.4	27.8	74.9	24.4	18.8	3 847.2
1	9	23.6	19.2	83.9	17.3	18.7	2 912.7
2	9	25.9	15.5	96.4	15.7	18.6	3 165.4
3	9	24.7	19.9	83.9	17.7	18.5	4 734.4
4	9	32.9	20.9	82.8	18.6	21.3	4 331.6
5	9	18.5	25.7	88.2	20.3	23.0	3 312.3
1	11	36.8	10.6	91.9	15.8	18.0	669.6
2	11	46.9	17.0	92.1	19.5	20.2	1 297.6
3	11	49.8	17.2	86.5	19.0	22.9	1 100.4
4	11	54.3	10.6	91.9	15.8	18.0	669.6
5	11	52.1	13.3	93.4	16.5	18.3	695.9

2.2 主成分分析和回归计算 将上述数据在 python 中的 Sci-Learn 包中进行数据主成分分析,设置主成分的包含度为 85%,得到了降维后的二维数据,计算出的特征值对应的

贡献度分别为 73.6%和 13.7%,合计超过 85.0%。将降维后的数据与实际的发病率进行对比作图,结果如图 1 所示。图 1A 中显示为特征向量与发病率的三维数据图,可看到在



注:图 B 中圆形面积代表感染率

Note: The area of circle in B indicated infection rate

图 1 数据降维后特征值与发病率的三维作图(A)和二维映射图(B)

Fig. 1 Data visualization(3D for A and 2D for B) after dimension reduction

发病率较高的情况下,特征值集中于一个较小的区域。随后将该三维图进行影射至二维坐标系,并以圆点大小表示发病率的高低(图1B)。由图中可以看出,霜霉病发病率较高的区域集中在2个特征值均较小的区域。针对该种情况,为了增加预测精度,使用多元多项式的回归进行回归方程的预测。回归后的公式如下:

$$F=0.195-3.11x_1+3.08x_2-0.337x_1x_2+0.271x_1^2+1.24x_2^2-1.46x_2^3+0.189x_1^2x_2-1.58x_1x_2^2$$

式中, $F$ 为霜霉病的发病率, $x_1$ 和 $x_2$ 分别为数据降维之后的2项特征指标。

**2.3 预警精确度的计算** 针对上述得出的预警模型进行精确度的计算,使用训练数据与实际数据的对比(图2),并使用方程预测剩余的50%数据,预测结果和准确率如表2所示。从表2可以看出,该方程整体的预测准确率大于85.0%。在感染率大于25.0%的情况下,模型的预测准确率均较高;而在感染率较低的区间,预测准确率有所降低,为85.0%~90.0%。在实际种植过程中,低温高湿度的环境会导致霜霉病的快速发生,因此对于该种情况下的准确预测显得更为重要。因此,该方程作为大棚黄瓜霜霉病的预测模型可有效进行霜霉病的监控和预警。

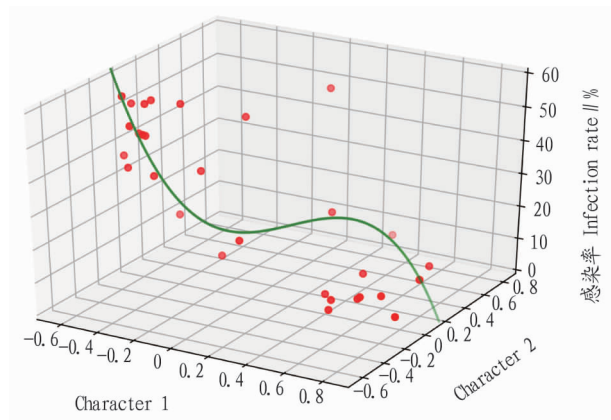


图2 数据模型与实际数据的结果对比

Fig. 2 Comparison between predicted data and investigated data

### 3 结论与讨论

黄瓜霜霉病是大棚黄瓜种植中的一种重要病害,然而一直以来缺少安全有效的农药对其进行防控。前人的研究表明,在温度低、湿度大的环境下,霜霉病的发生率会大幅度提高,严重者会导致绝收。然而有些地区因担心霜霉病的发生,提前施洒大量农药,这不仅造成了农药的大量浪费,违背了农业农村部“双减”的政策导向<sup>[15]</sup>,还会降低农产品品质并污染空气和地下水源。因此,研究霜霉病的发病规律并在高风险的时段进行针对性的提前施药,是较可行的防控方案。

根据以往的气象数据和霜霉病的发生率,该研究模拟出了预警方程用于预测霜霉病的发生规律。与实际数据对比可知,该方法可准确预测霜霉病的发生率,且在高风险条件下预测的准确率可达90.0%,为施药提供有效的指导。

表2 利用数据降维和多元回归对实际发病率的结果预测分析

Table 2 Infection rate prediction by using dimension reduction and regression %

月份 Month	实际发病率 Actual incidence rate	预测发病率 Predicted incidence rate	准确率 Accuracy rate
1	48.0	51.3	93.1
	36.3	36.6	99.1
	39.9	38.4	96.2
	40.2	41.5	96.8
	43.5	40.1	92.2
3	55.6	55.3	99.5
	51.1	53.9	94.4
	51.9	55.5	93.0
	34.1	33.7	98.7
	38.7	35.2	91.0
5	10.8	9.9	91.6
	13.6	15.5	86.1
	7.1	6.6	92.8
	13.2	13.6	96.6
	11.0	9.8	88.3
7	10.1	10.7	95.0
	10.5	9.8	93.3
	8.0	9.0	87.2
	10.0	8.8	87.2
	4.7	4.8	97.1
9	8.9	9.3	96.0
	18.6	19.3	96.7
	22.9	20.3	88.5
	19.0	20.8	90.6
	12.2	13.9	86.2
11	39.3	41.7	93.9
	51.5	51.3	99.5
	50.2	52.8	94.8
	58.6	62.3	93.8
	56.4	56.7	99.3

### 参考文献

- [1] 熊艳,王鹤冰,向华丰,等. 黄瓜霜霉病研究进展[J]. 中国农学通报, 2016,32(1):130-135.
- [2] 陈永明,谷莉莉,林双喜,等. 黄瓜霜霉病的研究进展及登记防治农药的分析[J]. 农学学报,2018,8(8):9-15,100.
- [3] 吴凤兰. 黄瓜霜霉病危害症状及防治方法探讨[J]. 农业灾害研究, 2016,6(5):1-3.
- [4] 鲍宇. 棚室黄瓜霜霉病和角斑病的区别与防治[J]. 现代化农业,2015(1):16-17.
- [5] 李昊. 日光温室黄瓜霜霉病及其防治[J]. 蔬菜,2016(7):65-66.
- [6] 段显德. 黄瓜霜霉病斑生长规律的初步研究[J]. 辽东学院学报(自然科学版),2015,22(2):99-101.
- [7] 董海涛. 怎样有效防治黄瓜霜霉病[J]. 农民致富之友,2017(22):78.
- [8] 陈颖潇,何胥,施洁君,等. 黄瓜霜霉病生防菌株的筛选及防病促生研究[J]. 安徽农业科学,2015,43(23):121-124.
- [9] 王永存,李聪晓,刘桂芳. 10% 氟噻唑吡乙酮对黄瓜霜霉病防治效果研究[J]. 农业灾害研究,2016,6(6):60-61.
- [10] 冯佳新. 怎样有效防治黄瓜霜霉病[J]. 农民致富之友,2015(9):148.
- [11] 王琳. 黄瓜霜霉病的发生及防治[J]. 现代园艺,2017(9):137-138.
- [12] 郭继民,张苗,管明乐. 黄瓜霜霉病的药剂防治研究[J]. 中国果菜, 2016,36(10):65-66.
- [13] 马浚诚,杜克明,郑飞翔,等. 可见光光谱和支持向量机的温室黄瓜霜霉病图像分割[J]. 光谱学与光谱分析,2018,38(6):1863-1868.
- [14] 杜慧,蒋晶晶,王春明,等. 天水地区葡萄霜霉病田间病情、孢子囊数量动态及病害始发关键因子分析[J]. 草业学报,2020,29(5):191-197.
- [15] 瞿剑. 40 项共性技术促蔬菜大棚化肥农药双减超 30%[J]. 农民文摘, 2020(7):64.