

# 基于机器学习与气象灾害指标的苹果相对气象产量预测

罗琦<sup>1,2</sup> 茹晓雅<sup>1,2</sup> 姜元<sup>1,2</sup> 冯浩<sup>2,3</sup> 于强<sup>3,4</sup> 何建强<sup>1,2</sup>

(1. 西北农林科技大学旱区农业水土工程教育部重点实验室, 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. 中国科学院水利部水土保持研究所黄土高原土壤侵蚀与旱地农业国家重点实验室, 陕西杨凌 712100;

4. 陕西省气象局秦岭和黄土高原生态环境气象重点实验室, 西安 710016)

**摘要:** 为及时准确地预测我国黄土高原苹果产量, 首先选取黄土高原苹果产区 86 个基地县的气象观测数据, 分别提取出苹果生长季内不同月份的气温、降水量、太阳辐射等气象特征变量, 花期冻害时间、连阴雨时间和标准化降水蒸发指数 (Standardized precipitation evapotranspiration index, SPEI) 等气象灾害特征变量, 以及气象站点经度、纬度和高程等空间特征变量, 再根据斯皮尔曼相关性分析确定影响苹果产量的最重要气象特征变量。然后, 采用梯度提升树 (Gradient boosting decision tree, GBDT)、支持向量机 (Support vector machine, SVM)、贝叶斯正则化反向传播神经网络 (Bayesian regularization back propagation artificial neural network, BRBP) 和多元线性回归 (Multiple linear regression, MLR) 算法, 建立苹果相对气象产量的预测模型, 并确定最佳模型输入特征变量组合。最后, 基于不同生育期和生长季内各月份最佳模型输入特征变量组合, 分析不同模型预测苹果相对气象产量的提前期。结果表明: 影响苹果相对气象产量的最重要气象特征变量为最高气温、最低气温、空气相对湿度、降水量和太阳辐射; 最佳模型输入变量组合为最重要气象特征变量、空间特征变量和灾害特征变量; 基于最佳模型输入变量组合, GBDT 和 BRBP 模型精度较好 (相关系数  $r$  为 0.77, 均方根误差 (RMSE) 为 0.44;  $r$  为 0.70, RMSE 为 0.44), 而 MLR 模型表现最差 ( $r$  为 0.63, RMSE 为 0.49)。在苹果不同生育期内, GBDT 和 BRBP 模型在各个生育期内均能获得相对较高的苹果相对气象产量预测精度, SVM 和 MLR 模型可在果实膨大期获取较为理想的苹果相对气象产量模拟结果。在苹果生长季内各月份, GBDT、SVM、BRBP 和 MLR 模型可在苹果成熟期前 1~2 个月实现对苹果相对气象产量的早期预测。本研究可为黄土高原苹果产量早期产量预测提供科学依据和技术参考。

**关键词:** 苹果; 产量预测; 机器学习; 气象灾害; 黄土高原

中图分类号: S165.27

文献标识码: A

文章编号: 1000-1298(2023)09-0352-13

OSID:



## Prediction of Apple Relative Meteorological Yields Based on Machine Learning and Meteorological Disaster Indices

LUO Qi<sup>1,2</sup> RU Xiaoya<sup>1,2</sup> JIANG Yuan<sup>1,2</sup> FENG Hao<sup>2,3</sup> YU Qiang<sup>3,4</sup> HE Jianqiang<sup>1,2</sup>

(1. Key Laboratory for Agricultural Soil and Water Engineering in Arid Area, Ministry of Education, 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. State Key Laboratory of Soil Erosion and Dryland Farming on the Loess Plateau, Institute of Water and Soil Conservation, Chinese Academy of Science and Ministry of Water Resource, Yangling, Shaanxi 712100, China

4. Key Laboratory of Eco-Environment and Meteorology for the Qinling Mountains and Loess Plateau, Shaanxi Provincial Meteorological Bureau, Xi'an 710016, China)

**Abstract:** Aiming to predict apple production in Loess Plateau in a timely and accurate manner, based on historical weather records in a total of 86 counties in the apple producing areas of the Loess Plateau, the data of different meteorological feature variables (e.g., temperature, precipitation, and radiation in

收稿日期: 2023-03-20 修回日期: 2023-06-14

**基金项目:** 国家重点研发计划项目(2021YFD1900700)、陕西省重点研发计划重点产业创新链(群)-农业领域项目(2019ZDLNY07-03)、西北农林科技大学人才专项资金项目(千人计划项目)和高等学校学科创新引智计划(111计划)项目(B12007)

**作者简介:** 罗琦(1997—),女,博士生,主要从事农业生态系统模拟研究,E-mail: qiluo@nwfu.edu.cn

**通信作者:** 何建强(1977—),男,教授,博士生导师,主要从事农业生态系统模拟研究,E-mail: jianqiang\_he@nwfu.edu.cn

different apple growing months), spatial feature variables (e. g., latitude, longitude and elevation of meteorological stations), and meteorological disaster feature variables (e. g., time of freezing damage at flowering stage, time of continuous rain, and standardized precipitation evapotranspiration index (SPEI)) were extracted at first. The influential feature factors were determined according to Spearman correlation analysis. Next, the prediction models for apple relative meteorological yield were established based on different algorithms (i. e., gradient boosting decision tree, GBDT; support vector machine, SVM; Bayesian regularization back propagation artificial neural network, BRBP; multiple linear regression, MLR). At the same time, the optimal combination of model input feature variables was determined for each of the established yield prediction models. Finally, based on the optimal combinations of input feature variables in different apple growth periods and in different months of apple growing seasons, the prediction leading time were analyzed for different simulation models for apple relative meteorological yield. The results were as follows: the influential meteorological feature variables were the highest temperature, lowest temperature, air relative humidity, precipitation and solar radiation. The best model input variable combination was selected as the influential meteorological, spatial and disaster feature variables. Based on the best combination of model input variables, the GBDT and BRBP models had better prediction accuracy ( $r$  was 0.77, RMSE was 0.44;  $r$  was 0.70, RMSE was 0.44), while the MLR model performed the worst ( $r$  was 0.63, RMSE was 0.49). In different growth periods of apples, the GBDT and BRBP models could obtain relatively high apple yield prediction accuracy in each growth period, while the SVM and MLR models could obtain relatively ideal simulation results in apple fruit expansion period. In each month of the apple growing season, the GBDT, SVM, BRBP and MLR models could realize early prediction of apple relative meteorological yield about one to two months before apple maturity. The research result can provide a scientific foundation and technical reference for apple yield prediction on the Loess Plateau.

**Key words:** apple; yield prediction; machine learning; meteorological disaster; Loess Plateau

## 0 引言

目前,我国黄土高原是全世界最大的苹果 (*Malus pumila* Mill.) 种植区之一,其苹果种植面积和产量分别占全球的 25.2% 和 26.3%<sup>[1-2]</sup>。由于苹果种植业的经济效益较高,黄土高原大部分传统农田都已改建为苹果园,苹果产业已成为当地农村经济的支柱产业之一,是当地农民的主要收入来源。因此,及时准确地预测苹果产量对当地苹果产业的发展具有重要意义。

黄土高原大部分地区光热资源丰富、昼夜温差大,对苹果树的生长发育和结果极为有利<sup>[3]</sup>。降水、气温和太阳辐射是影响苹果生长发育的重要气象要素,但气象灾害也会对苹果产量形成过程造成极大的影响。黄土高原地区干旱少雨,属于典型生态脆弱区,容易发生各种气象灾害。已有相关研究分析了各类气象灾害对苹果产量的影响。例如,RODRIGO<sup>[4]</sup>发现春季花期冻害很少导致果树死亡,但会严重损害苹果花芽,是限制苹果产量的关键因素;韩文静等<sup>[5]</sup>采用连阴雨指标分析发现苹果着色成熟期和采收期受到连阴雨危害的程度高于其他生育期;戴安然等<sup>[6]</sup>研究结果表明,旱灾是限制中国苹果生产发展的主要因素之一,会使果树正常生长受到抑制,导致叶片萎蔫、落果、枝条抽干等。此外,

干旱指标中标准化降水蒸散发指数 (Standardized precipitation evapotranspiration index, SPEI) 对降水和潜在蒸散发同样敏感,1—6 个月尺度的 SPEI 可用于评估农业干旱,且该指标还被广泛应用于不同作物的产量预测模型中<sup>[7]</sup>。

为了量化气象因子和气象灾害因子对苹果产量的影响,有必要建立上述因子和苹果产量之间的数学映射关系。但是不同气象因子和气象灾害因子之间存在较强的共线性,并且这些因子与苹果产量之间通常为非线性关系,将传统的线性模型应用于作物产量预测往往不能获得理想的预测结果。近年来,机器学习算法由于其众多优势而逐渐被广泛应用于作物产量预测研究<sup>[8]</sup>。这类方法从训练数据集中提取信息建立统计模型,随后根据测试数据集评估模型模拟精度,从而在自变量和非自变量之间建立非线性关系<sup>[9-10]</sup>。相关研究表明机器学习算法可在苹果产量预测中取得较好的效果。例如,LI 等<sup>[11]</sup>利用气象数据和支持向量机 (Support vector machine, SVM) 算法对陕西省 28 个基地县的苹果进行了估产,但研究区域只限制于陕西省;KHAN 等<sup>[12]</sup>基于农业生产总值建立了基于列文伯格马夸尔特 (Levenberg - Marquardt optimization, LM)、尺度共轭梯度 (Scale conjugate gradient back propagation, SCG)、贝叶斯正则化反向传播神经网络 (Bayesian

regularization back propagation, BRBP) 算法的苹果产量预报模型,但其选择的模型输入变量并不包括影响苹果产量形成的环境因素;景辉等<sup>[13]</sup>采用气象数据和多元线性回归(Multiple linear regression, MLR)、反向传播神经网络(Back propagation neural network, BPNN)算法建立了苹果产量早期预测模型,但并未考虑气象灾害因子对苹果产量的影响,且只考虑了各生育期内气象因子对苹果产量的影响。总体而言,上述研究的不足在于研究区域往往较小,所选择的苹果产量预测模型输入特征变量较少考虑苹果生长季内不同月份气象因子的影响,以及气象灾害因子对苹果产量的影响。

本文以我国黄土高原苹果产区为研究区域,基于气象、空间和气象灾害3类特征变量与产区苹果单产统计数据,选择梯度提升树(Gradient boosting decision tree, GBDT)、SVM和BRBP神经网络3种机器学习算法和MLR算法,构建黄土高原苹果相对气象产量预测模型。首先,采用斯皮尔曼相关性分析确定影响黄土高原苹果相对气象产量的最重要气象特征变量,然后以最重要气象特征变量、气象灾害特征变量和空间特征变量为输入变量,选择苹果相对气象产量预测模型的最佳模型输入特征变量组合,最后基于不同生育期和生长季内各月份的最佳模型输入特征变量组合,分析不同模型预测苹果产量的提前期,以期黄土高原苹果产量早期预测提供科学依据和技术参考。

## 1 材料与方法

### 1.1 研究区域

研究区域为我国黄土高原苹果产区,主要包括河南、山西、陕西和甘肃4省在内的共86个苹果基地县(34°22′~37°42′N, 104°53′~112°46′E, 海拔84~4216 m, 图1)。黄土高原属于半干旱大陆性季风气候,是我国最大的苹果优生区,其多项气象因素能满足优质苹果生长需求。研究区内年均温为6~16℃,年降水量为201~1010 mm,年日照时数为1294~2900 h。

### 1.2 数据来源与处理

1981—2017年共37年的逐日气象观测数据(包括降水量、平均气温、最高气温、最低气温、日照时数及空气相对湿度等)来源于中国气象科学数据共享网(CMDSSS, <http://cdc.cma.gov.cn/>)。气象观测站空间位置信息(包括经度、纬度、海拔)来自各县市内的气象观测站。河南、山西、陕西和甘肃共86个县级行政区的苹果单产数据来自《河南统计年鉴》<sup>[14]</sup>、《山西统计年鉴》<sup>[15]</sup>、《陕西统计年鉴》<sup>[16]</sup>和

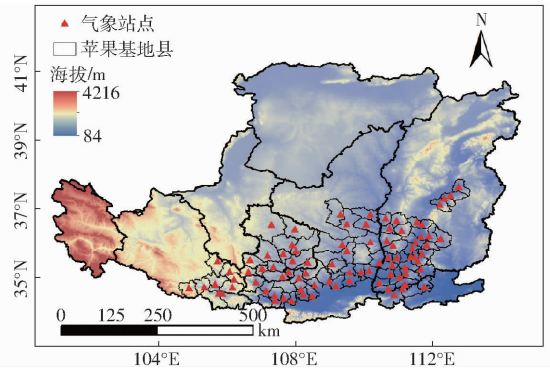


图1 黄土高原86个苹果生产基地县和气象站点分布图

Fig. 1 Distribution of 86 base apple-producing counties and weather stations on Loess Plateau

《甘肃发展年鉴》<sup>[17]</sup>,以及各市人民政府依申请公开数据(表1)。本文剔除了各县市单产缺失年份,共获得1852个有效单产数据。

数据归一化处理可消除不同数据的量纲影响,提升模型各项性能<sup>[18]</sup>。因此,模型建立前统一对输入数据进行归一化处理,最后将模型输出值进行反归一化处理。计算公式为

$$x_n = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

式中  $x_n$ ——归一化后的输出数据

$x$ ——待归一化的输入数据

$x_{\max}$ ——待归一化输入数据最大值

$x_{\min}$ ——待归一化输入数据最小值

### 1.3 模型输入特征变量和目标变量

#### 1.3.1 模型输入特征变量优选

依据文献[19-20],黄土高原产区的苹果全生育期为本年11月至次年10月,可主要划分为落叶期(本年11月)、休眠期(本年12月—次年2月)、萌芽幼果期(次年3—5月)、果实膨大期(次年6—8月)和着色成熟期(次年9—10月)。本文选取不同苹果生育期内的9个气象特征变量、3个空间特征变量和3个气象灾害特征变量作为构建苹果相对气象产量预测模型的输入变量,具体特征变量如表2所示。

太阳辐射数据根据各气象站点观测的日照时数和Angstrom- Prescott经验公式<sup>[21]</sup>进行估算,公式为

$$R_s = \left( a_s + b_s \frac{n}{N} \right) R_a \quad (2)$$

其中  $R_a = (24 \times 60 / \pi) G_{sc} d_r (\omega_s \sin \varphi \sin \delta + \cos \varphi \cos \delta)$  (3)

$$d_r = 1 + 0.033 \cos(2\pi J / 365) \quad (4)$$

$$\delta = 0.409 \sin(2\pi J / 365 - 1.39) \quad (5)$$

$$\omega_s = \arccos(-\tan \varphi \tan \delta) \quad (6)$$

表 1 黄土高原苹果产区 86 个生产基地县产量记录年份

Tab. 1 Years with observed apple yield for 86 base apple-producing counties on Loess Plateau

省份	市	县	年份
河南	三门峡	灵宝、渑池、陕县	1981—2017
	运城	河津、稷山、绛县、临猗、平陆、芮城、万荣、闻喜、夏县、盐湖、永济、垣曲	2007—2017
山西	临猗	安泽、大宁、汾西、浮山、古县、洪洞、侯马、霍州、吉县、蒲县、曲沃、襄汾、乡宁、隰县、尧都、翼城、永和	2007—2017
	晋中	祁县、榆次、平遥、太谷	2007—2016
	延安	安塞、宝塔、延川、延长、宜川、富县、洛川	
	铜川	宜君、耀县	
陕西	渭南	白水、芮城、合阳、富平、韩城、澄城	1981—2017
	宝鸡	陇县、千阳、凤翔、陈仓、岐山、扶风	
	咸阳	旬邑、长武、淳化、永寿、乾县、彬县、礼泉	
	庆阳	合水、华池、环县、宁县、庆城、西峰、镇原、正宁	2000—2016
甘肃	平凉	崇信、华亭、泾川、静宁、崆峒、灵台、庄浪	2000—2017
	天水	甘谷、秦安、秦州、清水、麦积、武山、张家川	2006—2017

表 2 苹果相对气象产量模拟模型输入特征变量

Tab. 2 Input feature variables of prediction models of apple relative meteorological yields

特征变量	变量	计算方法
气象特征变量	平均气温 $T_{mean}$ (°C)	月日平均气温的算术平均值
	最低气温 $T_{min}$ (°C)	月平均气温极小值
	最高气温 $T_{max}$ (°C)	月平均气温极大值
	月平均气温日较差 $T_d$ (°C)	月平均日最高气温与最低气温的差值
	5°C 有效积温 $AT$ (°C)	月日平均气温与 5°C 之差的总和
	负积温 $NT$ (°C)	月低于 0°C 的逐日日平均气温的总和
需水特征变量	空气相对湿度 $RH$ (%)	月日空气相对湿度的算术平均值
供水特征变量	降水量 $P$ (mm)	月日降水量的总和
辐射特征变量	太阳辐射 $R$ (MJ/(m <sup>2</sup> ·d))	Angstrom - Prescott 经验公式
气象灾害特征变量	花期冻害时间 $N_{4 \leq 0^\circ C}$ (d)	4 月日最低气温小于等于 0°C 的时间
	连阴雨时间 $CR$ (d)	9—10 月连续 3 d 降水量大于等于 0.1 mm 的时间
	1 个月尺度标准化降水指数 (SPEI)	Thornthwaite 法
空间特征变量	经度 $Lon$ (°)	气象观测站经度
	纬度 $Lat$ (°)	气象观测站纬度
	海拔 $Ele$ (m)	气象观测站海拔

$$N = 24\omega_s / \pi \quad (7)$$

式中  $N$ ——最大可能日照时数, h

$\omega_s$ ——太阳时角, rad

$\varphi$ ——纬度, rad

$\delta$ ——太阳磁偏角, rad

$J$ ——年内某天的日序数

$d_r$ ——日地间相对距离的倒数

$G_{sc}$ ——太阳常数, 取 0.082 MJ/(m<sup>2</sup>·min)

$R_a$ ——大气外总辐射, MJ/(m<sup>2</sup>·min)

$a_s$ 、 $b_s$ ——经验系数, 取 0.19 和 0.54<sup>[22]</sup>

$n$ ——实际日照时数, h

$R_s$ ——地表总辐射, MJ/(m<sup>2</sup>·min)

1 个月尺度标准化降水指数 (SPEI) 根据 Thornthwaite 法<sup>[23]</sup> 通过 R 语言中的“SPEI”包计算得到, 具体计算过程主要分为 3 个步骤。

(1) 计算潜在蒸散量  $PET$ , 计算式为

$$PET_i = 16 \left( \frac{10T_{mean}}{H} \right)^A \quad (8)$$

其中  $H = \left( \frac{T_{mean}}{5} \right)^{1.514} \quad (9)$

$$A = 6.75 \times 10^{-7} H^3 - 7.71 \times 10^{-5} H^2 + 1.792 \times 10^{-2} H + 0.49 \quad (10)$$

式中  $A$ ——常数

$H$ ——年热量指数

$PET_i$ ——月潜在蒸散量, mm/d

(2) 计算气候水平衡, 计算式为

$$D_i = P_i - PET_i \quad (11)$$

式中  $D_i$ ——月降水量与蒸散量差值, mm/d

$P_i$ ——月降水量, mm/d

(3) 对  $D_i$  数据序列标准化, 采用三参数的 log-logistic 概率分布  $F(x)$  对其进行拟合, 计算出每个  $D_i$  对应的 SPEI 值, 计算式为

$$SPEI = W - \frac{C_0 + C_1 W + C_2 W^2}{1 + d_1 W + d_2 W^2 + d_3 W^3} \quad (12)$$

$$\text{其中 } W = \sqrt{-2 \ln P} \quad (13)$$

$$P = \begin{cases} 1 - F(x) & (P \leq 0.5) \\ 1 - P & (P > 0.5) \end{cases} \quad (14)$$

式中  $SPEI$ ——1 个月尺度标准化降水指数

其中  $C_0$ 、 $C_1$ 、 $C_2$ 、 $d_1$ 、 $d_2$ 、 $d_3$  为常数项,取 2.515 517、0.802 853、0.010 328、1.432 788、0.189 269、0.001 308。

为消除气象变量多年周期性循环的影响,本文将苹果多年生长季内的气象特征变量进行平均,再将其与目标变量(苹果相对气象产量)进行斯皮尔曼相关性分析(Spearman correlation analysis),选择对苹果产量影响最大的重要气象特征变量,以减少输入数据的复杂程度。在每类气象特征变量中,选择与目标变量的相关系数绝对值最大且显著的特征变量。如果同类特征变量之间的相关系数小于 0.5 且其与目标变量显著相关时,则可选择多个特征变量。

### 1.3.2 模型目标变量构建

苹果产量的形成受多种因素的影响,主要包括气象要素、技术措施和其他因素,相应地苹果产量可以分解为气象产量、趋势产量和随机产量。为了更好地反映气象要素造成的苹果产量波动,降低时间和地域对产量影响的限制,采用相对气象产量作为目标变量定量评估气象要素对产量形成的影响。计算公式为

$$y_w = \frac{y_c}{y_t} \quad (15)$$

$$\text{其中 } y_t = a_0 + \sum_{i=1}^M a_i \cos(wit) + \sum_{i=1}^M b_i \sin(wit) \quad (16)$$

$$y_c = y - y_t + y_r \quad (17)$$

式中  $y$ ——实际产量,  $\text{kg}/\text{hm}^2$

$y_w$ ——相对气象产量,正值表示气象要素有利于作物生长发育即产量增加,负值表示产量减少

$y_t$ ——趋势产量,  $\text{kg}/\text{hm}^2$ ,可采用一阶傅里叶拟合来估算<sup>[24]</sup>

$y_c$ ——气象产量,  $\text{kg}/\text{hm}^2$

$y_r$ ——随机产量,  $\text{kg}/\text{hm}^2$ ,本文忽略不计

$a_0$ ——常系数,取  $-2.94 \times 10^{10} \sim 2.17 \times 10^{11}$

$w$ ——常系数,取 0.30 ~ 0.90

$a_i$ ——各级系数,取  $-2.17 \times 10^{11} \sim 2.94 \times 10^{10}$

$b_i$ ——各级系数,取  $-2.09 \times 10^4 \sim 1.22$

$M$ ——傅里叶展开级数,取 1

$t$ ——年份

## 1.4 苹果相对气象产量模拟模型

### 1.4.1 机器学习算法

#### (1) 梯度提升树

梯度提升树(GBDT)是基于 Booting 方法的集成模型。该方法采用加法模型和前向分步算法实现学习优化<sup>[25]</sup>。GBDT 回归算法中以回归树作为弱学习器,学习任务时,利用损失函数的负梯度作为当前学习器的伪残差,根据伪残差拟合回归树,直到满足要求停止重复此过程。GBDT 算法一般比随机森林算法的训练结果更准确<sup>[9]</sup>,收敛速度快且不易出现过拟合。R 语言中提供了 GBDT 算法包“gbm”,所需要的参数包括:分布函数(distribution,选择“gaussian”)、树终节点的最小个数(n.minobsinnode,取 10)、学习率(shrinkage,取 0.01)、回归树数量(n.trees,取 200)以及单棵回归树最大深度(interaction.depth,以 1 为步长在 [0, 10] 内取值)。

#### (2) 支持向量机

支持向量机(SVM)通过非线性映射将自变量映射到高维的特征空间,在高维特征空间中寻找一个最优超平面,使得所有训练样本距离该最优分类面误差最小<sup>[26]</sup>。对于回归应用,SVM 算法主要利用核函数构造线性回归方程求得最优超平面<sup>[27]</sup>。该算法可利用 R 语言提供的“e1071”包来实现,所需参数包括:核函数类型(Kernel,定为高斯径向基核函数)、核函数系数(gamma,以 0.01 为步长在 [0.01, 1.00] 内选择)以及惩罚系数(cost,以 0.1 为步长在 [0.1, 100] 内确定)。

#### (3) 贝叶斯正则化反向传播神经网络

贝叶斯正则化反向传播神经网络(BRBP)融合传统反向传播神经网络算法和贝叶斯理论自动选择最优的正则化参数,能够避免过拟合问题,极大提高模型的泛化能力<sup>[28]</sup>。采用包含输入层、隐含层和输出层的 3 层神经网络拓扑结构,隐含层的节点数根据试凑法,以预测误差最小为原则来确定。Matlab 软件中的“newff”函数可以实现该类算法,所需参数包括:输入节点数  $a$ 、输出节点数  $b$ 、隐含层节点数  $l$ 、节点传递函数(隐含层为 tanh,输出层为 purelin)、训练函数(trainbr)、迭代次数  $k$ 、学习率  $\eta$  以及目标函数误差(goal)。其中,参数  $k$ 、 $\eta$ 、goal 经反复试算后分别取 1 000、0.1、 $10^{-5}$ 。

#### (4) 多元线性回归

多元线性回归(MLR)是一种广泛应用于单个目标变量与 2 个及以上输入特征变量的线性回归方法<sup>[29]</sup>。目标变量通常受多种特征变量的影响,因此

多元线性回归相对于一元线性回归更具适用性。在多元线性回归中,方程的参数通过最小二乘法获得<sup>[30]</sup>。R 语言提供的“lm”函数可实现 MLR 算法。

### 1.4.2 苹果相对气象产量模拟模型构建

如图 2(图中 M、G、D 分别表示最重要气象灾害特征变量、空间特征变量、气象灾害特征变量)所示,以 GBDT 算法为例,简述苹果相对气象产量  $y_w$  预测模型的建立过程。首先,基于苹果全生育期最重要气象特征变量,逐次添加空间特征变量和气象灾害特征变量,形成不同的模型输入特征变量组合。其次,应用 GBDT 算法建立模型估算苹果相对气象产量,以  $y_w$  估算值和实测值之间的误差最小化为优化目标,获得最佳模型输入特征变量组合。然后,逐次添加苹果落叶休眠期、萌芽幼果期、果实膨大期、着色成熟期内的最佳模型输入变量组合,构建不同生育期的苹果相对气象产量预测模型,以模型测试集均方根误差(RMSE)最小化为依据,确定苹果相对气象产量预测的提前生育期。最后,逐次添加苹果生长季内各月份的最佳模型输入特征变量组合,构建苹果生长季内不同月份的苹果相对气象产量预测模型,以模型测试集归一化 RMSE 最小化为依据,确定苹果相对气象产量预测的提前月份。并且在不同

同生育期和生长季内不同月份的模型预测精度分析过程中,将相对气象产量转换为实际产量,比较其与以相对气象产量为目标变量的机器学习模型精度,从而判断苹果相对气象产量模拟模型构建方法精度。

模型模拟前,随机选取 70% 观测数据作为训练数据集,30% 数据作为测试数据集,以 10 折交叉验证误差为原则对模型参数进行寻优。为降低模型精度极大值和极小值的影响,模型的最终精度为 100 次运行误差的 50% 分位数。

### 1.5 模型精度评价

采用皮尔逊相关系数(Pearson correlation coefficient,  $r$ )和均方根误差(Root mean square error, RMSE)2 个统计指标来评价模型模拟精度。

## 2 结果与分析

### 2.1 影响苹果产量的气象特征变量

本文分析了目标变量(苹果相对气象产量)与气象特征变量之间的相关关系(图 3,  $P < 0.05$ ),发现气温特征变量  $T_{mean}$ 、 $T_{min}$ 、 $T_{max}$ 、 $T_d$ 、 $AT$ 、 $NT$  与目标变量之间存在正相关关系,其中  $T_{max}$  与目标变量的相关性最高( $r = 0.26$ )。为了避免输入因子间共线

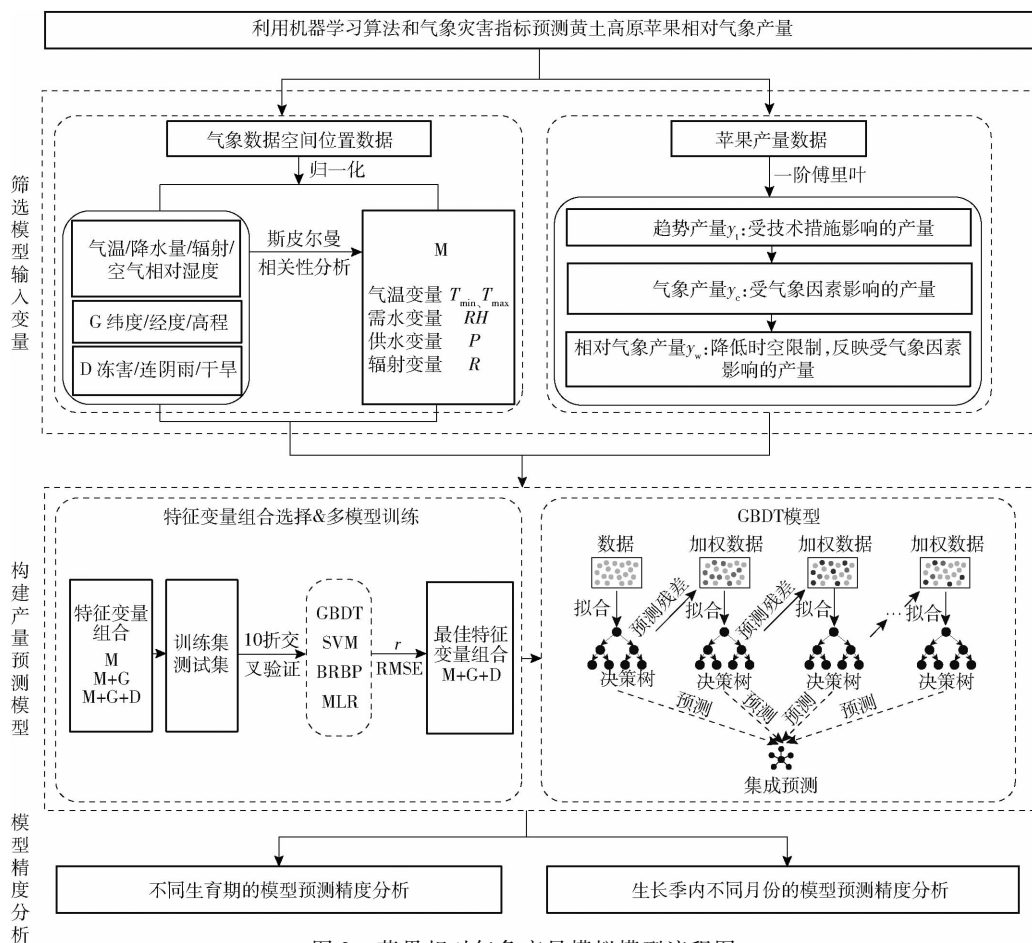


图 2 苹果相对气象产量模拟模型流程图

Fig. 2 Flow chart for establishment of simulation model of relative meteorological yield of apple

性对模型精度的影响,可同时选择与  $T_{\max}$  相关系数小于 0.5 的其他气温特征变量,其中符合要求的两个气温特征变量为  $T_{\min}$  与  $NT$ ,但是它们之间存在较强的相关性,且  $T_{\min}$  与目标变量之间的相关性强于  $NT$ ,因此最终只选择  $T_{\max}$  和  $T_{\min}$  作为模型输入的气温特征变量。需水特征变量  $RH$  与目标变量之间呈负相关关系,而供水特征变量  $P$ 、辐射特征变量  $R$  与目标变量之间的相关性系数较小。因此,本文最终选取  $T_{\max}$ 、 $T_{\min}$ 、 $RH$ 、 $P$ 、 $R$  作为苹果相对气象产量模拟模型输入的最重要气象特征变量。

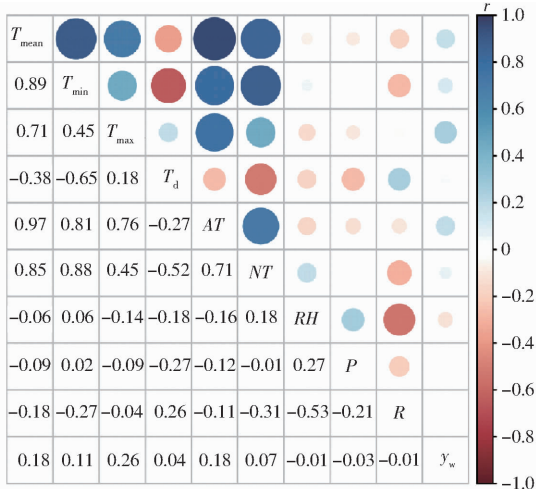


图3 目标变量(苹果相对气象产量)与气象特征变量关联热图

Fig. 3 Heat map of correlation between target variable of apple relative meteorological yield and meteorological feature variables

## 2.2 基于多种特征变量的不同苹果产量模型模拟精度分析

分别使用苹果整个生长季内的最重要气象特征变量组合(M),最重要气象特征变量和空间特征变量组合(M+G),以及最重要气象特征变量、空间特征变量和气象灾害特征变量组合(M+G+D)作为不同的模型输入特征变量组合,分别驱动基于GBDT、SVM、BRBP和MLR算法的苹果相对气象产量模拟模型,并比较不同模型测试数据集的模拟精度(图4,图中虚线表示1:1线;\*\*表示 $P < 0.01$ )。当仅以最重要气象特征变量M作为模型输入变量特征组合时,SVM模型具有最大的 $r$ 值(0.70),而BRBP模型的预测误差最小(RMSE为0.46)且数据点均匀分布在1:1线的两侧(图4a、4d、4g、4j)。当以M+G作为模型输入特征变量集时,与仅以M作为模型输入特征变量的情形相比,GBDT和BRBP的RMSE分别降低2%、2.17%,同时GBDT、BRBP和MLR模型的 $r$ 分别提升4.76%、2.99%和1.72%,但SVM模型的精度无变化(图4b、4e、4h、

4k)。可见在模型输入气象特征变量中再加入空间特征变量,可以提高苹果相对气象产量模拟模型的预测精度。进一步在上述模型输入变量中再添加气象灾害特征变量(M+G+D),结果表明增加气象灾害特征变量可使GBDT、BRBP和MLR模型预测精度进一步提升,其中GBDT模型的模拟精度提升效果最为明显,其 $r$ 增加16.67%,RMSE减少10.20%,而SVM模型的模拟精度则有所下降(图4c、4f、4i、4l)。

总体而言,基于GBDT和BRBP这2种机器学习算法和M+G+D输入特征变量组合建立的苹果相对气象产量预测模型,其模拟精度普遍优于MLR算法模拟。这可能是因为苹果相对气象产量与最重要气象特征变量、空间特征变量和气象灾害特征变量之间存在较强的非线性关系,而机器学习算法比多元线性回归方法更能准确描述这种非线性关系。因此,本文采用M+G+D组合作为最佳的模型输入特征变量组合。

## 2.3 模型模拟误差空间分析

以苹果整个生长季内的M+G+D特征变量组合作为模型输入变量,驱动基于不同算法建立的苹果相对气象产量模拟模型,可进一步分析不同模型在黄土高原各省份的总体模拟精度(图5,图中每个箱体包含各省苹果生产基地县的苹果相对气象产量模拟结果,箱体上、下限为数据的上四分位数和下四分位数,中间实线为数据中位数,箱体外的两条线为上边缘和下边缘,超出上下边缘的值为异常值)。由图5可知,不同模型在不同省份的表现存在较大差异,其中BRBP模型在河南省和陕西省表现最佳,SVM模型在山西省和甘肃省表现较佳。河南、山西、陕西和甘肃各省最佳模型的RMSE均值分别为0.47、0.11、0.39和0.20。在这4个省份中,山西省的模型总体估计误差相对较小,4个模型的RMSE平均值仅为0.24,且误差分布相对较为集中。但河南省RMSE箱型图的箱体最长,这可能由于本文中河南省仅有3个苹果基地县,数据量较少使得箱型图的分布范围较大。

## 2.4 苹果不同生育期输入特征变量组合对模型模拟精度的影响

进一步探究在M+G+D输入变量组合下,不同模型在不同生育期(落叶休眠期、萌芽幼果期、果实膨大期、着色成熟期)的苹果产量预测精度(图6)。当以相对气象产量作为目标变量时,对于GBDT和BRBP模型,各生育期模型的模拟精度无明显变化,其RMSE分别保持在0.44~0.46和0.44~0.45之间,这表明这2种模型在苹果各个生

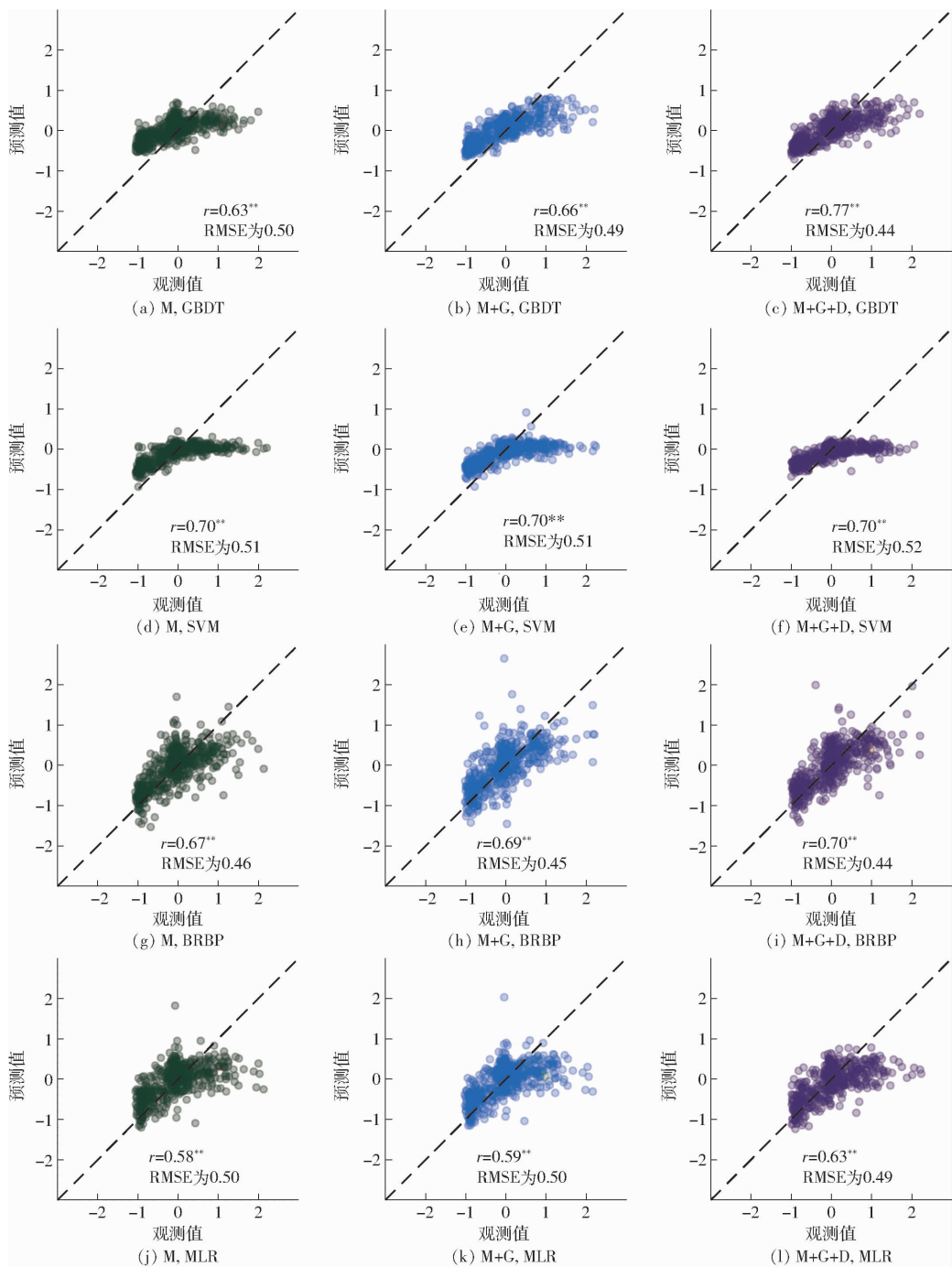


图 4 基于 M、G、D 组合和不同算法构建的苹果相对气象产量模拟模型预测精度对比

Fig. 4 Comparison of prediction accuracies of different simulation models of apple relative meteorological yield based on combination of M, G, D feature variables and different algorithms

育期内均能获得相对较高的相对气象产量预测精度(图 6a)。在不同苹果生育期,SVM 模型的 RMSE 先减小后增加,次年 8 月产量预测的 RMSE 具有最低值(0.51),可见 SVM 模型可在果实膨大期获取最好的预测结果。MLR 模型模拟的 RMSE 随着苹果生长发育的推进而逐渐减小,并在次年 8 月获得最佳的预测结果,与次年 10 月该模型的苹果相对气象产量预测精度仅相差 0.01,表明该模型可在苹果果实膨大期就获取较为理想的苹果相对气象产量预测结果。当将相对气象产量转换为实际产量时,

GBDT、BRBP、MLR 模型在各生育期的模拟精度均呈现减小趋势(图 6b)。具体而言,GBDT 和 BRBP 模型在苹果各个生育期内均能获得相对较高的实际产量预测精度,其 RMSE 分别保持在 4.42 ~ 4.90 kg/hm<sup>2</sup> 和 4.62 ~ 5.20 kg/hm<sup>2</sup> 之间。MLR 模型在果实膨大期的预测精度与着色成熟期仅相差 0.03 kg/hm<sup>2</sup>,表明该模型在果实膨大期也能获取较为理想的苹果产量预测精度。SVM 模型的 RMSE 先减小后增加,并且在果实膨大期获取最好的预测结果(5.04 kg/hm<sup>2</sup>)。



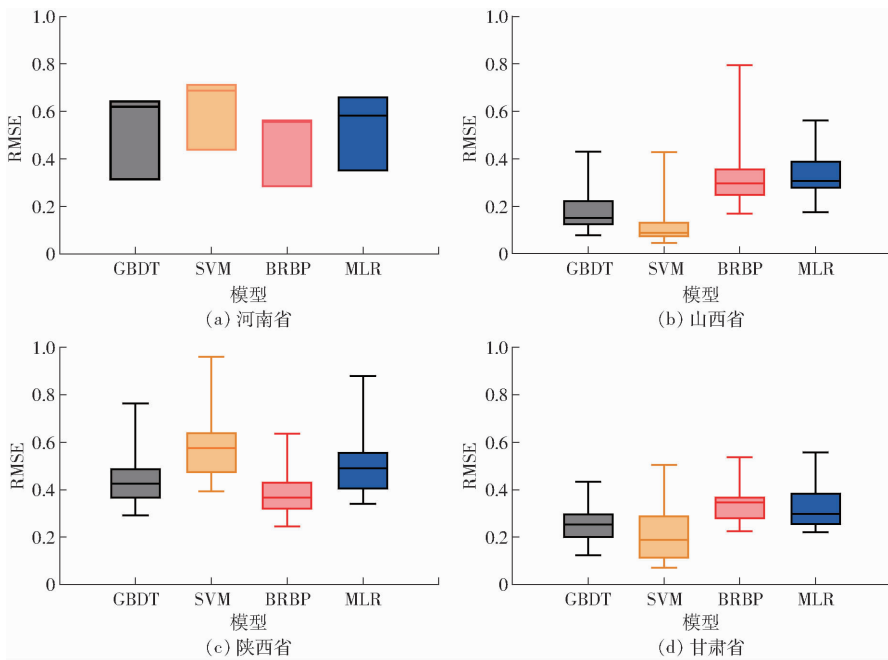


图5 基于最重要气象、空间和气象灾害特征变量组合和不同算法的模拟模型在4个不同省份模拟苹果相对气象产量的RMSE箱型图

Fig. 5 Box plots of RMSE of predictions of apple relative meteorological yields with different models established based on input variable combination of influential meteorological, spatial and meteorological disaster feature variables and different algorithms in four different provinces

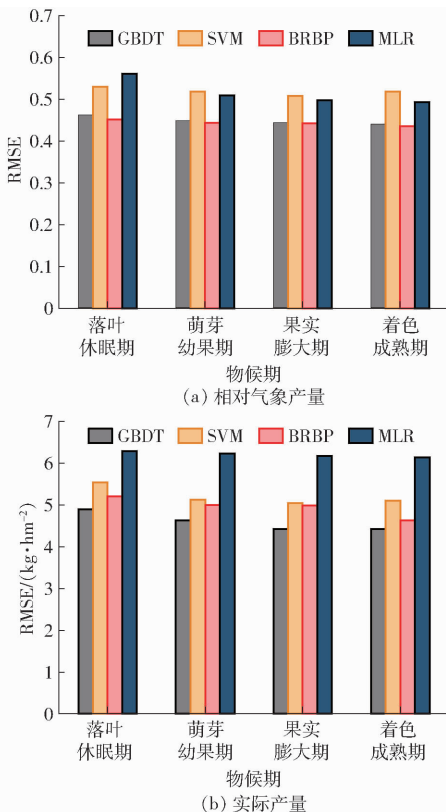


图6 基于GBDT、SVM、BRBP和MLR算法的不同模拟模型在苹果不同生育期模拟黄土高原苹果产量的精度对比

Fig. 6 Comparison of prediction accuracies of apple yield by different models based on GBDT, SVM, BRBP, and MLR algorithms in different growth periods of apple on Loess Plateau

## 2.5 苹果生长季内不同月份的模型预测精度

在生长季内不同月份对苹果相对气象产量进行预测模拟并分析其精度,可确定出苹果产量预测的最佳提前期。根据生长季内不同月份各模型模拟结果归一化RMSE,可知4种模型的模拟精度都呈现出类似的变化趋势(图7a)。即随着苹果生育期的推进,模型输入变量不断增多,各模型的归一化RMSE逐渐减少。当以相对气象产量作为目标变量时,对于GBDT和BRBP算法,基于生长季内所有月份输入特征变量的结果最佳。与基于生长季所有月份输入变量的模型模拟结果归一化RMSE相比,在次年9月这2个模型的归一化RMSE分别与之一仅相差0.62%和1.04%,精度非常相近。SVM算法可在次年8月获取最佳的预测精度。基于MLR算法的苹果相对气象产量预测模型可在次年9月获得最小的归一化RMSE。当将相对气象产量转换为实际产量时,GBDT和BRBP算法在次年9月的产量预测精度分别与次年10月仅相差0.38%和1.34%,而SVM和MLR算法分别在次年8月和次年9月模型预测精度最佳(图7b)。综上所述,本文建立的4种不同模型均可在苹果成熟期前1~2个月实现苹果相对气象产量的相对准确预测。

## 3 讨论

### 3.1 影响苹果产量的特征变量和产量去趋势方法 气象要素和气象灾害对苹果产量具有重要的影

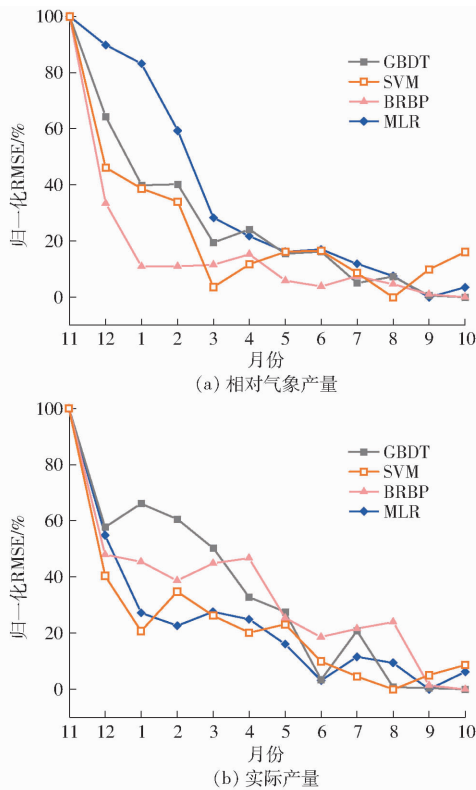


图 7 基于 GBDT、SVM、BRBP 和 MLR 算法的不同模拟模型在苹果生长季内各月份模拟黄土高原苹果产量的归一化 RMSE

Fig. 7 RMSE of predictions of apple yield by different models based on GBDT, SVM, BRBP, and MLR algorithms in each month of growing season of apple on Loess Plateau

响。本文选择了苹果产量形成过程中具有代表性的气象变量和气象灾害变量,包括  $T_{\text{mean}}$ 、 $T_{\text{min}}$ 、 $T_{\text{max}}$ 、 $T_d$ 、 $AT$ 、 $NT$ 、 $RH$ 、 $P$ 、 $R$ 、 $N_{4 \leq 0^\circ\text{C}}$ 、 $CR$  和 SPEI。其次,气象特征变量筛选结果表明  $T_{\text{min}}$ 、 $T_{\text{max}}$ 、 $RH$ 、 $P$ 、 $R$  是影响黄土高原苹果相对气象产量预测模型的精度的重要气象特征变量,这与前人的研究结果相一致。白秀广等<sup>[31]</sup>和张强<sup>[32]</sup>研究表明,气温对苹果产量有正向影响,降水量、空气相对湿度和日照时数对苹果产量有负面影响。本文中降水量和太阳辐射与苹果相对气象产量之间的相关系数较小,这可能是由于黄土高原降水不均匀特征和灌溉等因素影响了降水量与苹果产量间的相关性<sup>[33]</sup>。此外,本文中苹果产量的样本量可能也影响了太阳辐射与苹果产量之间的相关性。随后,在最重要气象特征变量中依次加入空间特征变量和气象灾害特征变量,均可进一步提升基于 GBDT、BRBP 和 MLR 算法的苹果相对气象产量模拟模型的预测精度。当模型输入特征组合为“气象特征变量 M + 空间特征变量 G + 气象灾害特征变量 D”时模型精度最高。刘峻明等<sup>[34]</sup>研究表明,气象站点空间位置会影响作物的积温和物候,但由于气候条件往往存在一定的年际变化,导致空间

位置和气象要素的相关性较弱。所以,作物产量预测模型中需要同时考虑气象特征变量和空间特征变量。FENG 等<sup>[35]</sup>研究发现,需要在作物产量预测模型中加入气象灾害,以此减少气象灾害给作物产量预测结果带来的不确定性。然而,本文中加入了气象灾害特征变量会导致 SVM 模型的预测精度降低,这可能是由于该算法易受数据维度的影响,而气象灾害因子的加入增加了样本维度,从而导致模型性能受损<sup>[36]</sup>。

苹果相对气象产量仅受气象条件的影响,但是该产量不能通过测量直接获取,因此需对实际苹果产量数据进行预处理,而数据处理方法的选择就甚为重要。本文采用了一阶傅里叶对所有基地县的苹果单产去趋势得到苹果相对气象产量,这可以在一定程度上消除其他因素对产量的影响,从而能准确地反映苹果产量随气象因子的非线性变化。然而,产量去趋势其他方法还包括:线性回归、多项式回归、指数回归、对数回归、滑动平均、Logistic 回归、HP 滤波、局部加权回归等<sup>[37-40]</sup>。此外,由于社会、经济发展不平衡,不同区域的作物产量变化趋势也会存在差异性<sup>[41]</sup>。因此,在今后的研究中需要在各苹果生产县采用不同的去趋势方法获得更加准确、客观的苹果相对气象产量,从而提升模型的模拟精度。

### 3.2 基于机器学习算法的苹果相对气象产量模拟模型

机器学习模型能够较好地建立气象灾害因子与产量之间的关系,而这是作物模型普遍无法实现的<sup>[42]</sup>。在模拟黄土高原苹果相对气象产量的过程中,本文发现非线性算法 GBDT、BRBP 优于线性算法 MLR。当以 M + G + D 组合作为模型输入特征变量组合来驱动本文所建立的模型模拟苹果相对气象产量时,机器学习模型 (GBDT、BRBP 模型) 的  $r$  和 RMSE 分别比线性模型 (MLR 模型) 高 11.11% ~ 22.22% 和低 10.20%, 这表明模型输入特征变量与苹果相对气象产量之间存在较强的非线性关系,机器学习算法比线性算法能够更为准确地描述这种非线性关系。前人研究也发现气温、太阳辐射、降水量、空气相对湿度等气象要素、空间要素和气象灾害要素与苹果产量之间存在非线性关系<sup>[11, 24, 34, 43-44]</sup>, 本文结果与这些结论基本相同。然而,基于 SVM 算法建立的苹果相对气象产量预测模型的 RMSE 均高于线性模型,这可能是由于输入变量之间存在较强的线性关系,破坏了 SVM 模型的预测性能,从而导致 SVM 模型预测结果与线性模型相近甚至更差<sup>[45]</sup>。此外,本文还探究了在苹果生长季内不同月份各模型预测苹果相对气象产量的精

度。结果表明,当以 M + G + D 特征变量组合作为模型输入变量时,本文建立的 4 种模型均可在果实成熟前的 1~2 个月获得较高精度的苹果相对气象产量预测结果。本文所构建的苹果相对气象产量模拟模型在苹果生长季内不同月份的预测中模拟精度存在差异,这一方面可能因为不同月份气象特征变量对苹果产量的影响并非同等重要<sup>[13]</sup>。另一方面可能是由于特征变量的输入个数随着月份的递增而增加,机器学习模型对更多数量的输入特征通常模拟精度可能更佳。已有研究表明小麦、玉米、水稻等粮食作物均可在收获前的 1~2 个月实现较为准确的早期产量预测,这些结果也与本文的结果较为一致<sup>[21, 46]</sup>。并且将相对气象产量转换为实际产量后,4 种模型在苹果生长季内不同月份的模拟精度变化趋势基本一致,仅在数值上存在部分差异,可能是苹果实际产量受到技术措施等其他因素影响所导致的。这表明本文所构建的苹果产量模型具有较好的预测精度。

### 3.3 气象特征变量和气象灾害特征变量对苹果相对气象产量模拟精度的影响

本文通过筛选,最终确定了预测苹果相对气象产量的输入特征变量为:气象特征变量  $T_{\max}$ 、 $T_{\min}$ 、 $RH$ 、 $P$ 、 $R$ ,气象灾害特征变量  $N_{4 \leq 0^{\circ}\text{C}}$ 、 $CR$ 、 $SPEI$ ,空间特征变量  $Lat$ 、 $Lon$ 、 $Ele$ 。以机器学习模型(GBDT、SVM、BRBP、MLR 模型)在苹果生长季内不同月份的表现来看,空间特征变量在模型输入中是固定不变的,但是随着苹果生育期从冬季休眠期(本年 11 月)发展到收获期(次年 10 月),输入模型的最重要气象特征变量和气象灾害特征变量的时间序列逐渐递增,而此过程中所有模型的 RMSE 均逐渐减小(图 7)。这表明气象特征变量和气象灾害特征变量在模型性能的后续提升过程中发挥着关键作用。因此,相比于空间特征变量,气象特征变量和气象灾害特征变量对苹果相对气象产量模拟模型精度的影响更大。

### 3.4 苹果相对气象产量模拟模型的不足

本文建立的黄土高原苹果相对气象产量预测模

型也存在一定的不足。首先,本文所建立模型的输入数据较为单一,然而苹果的生长发育过程非常复杂,产量形成过程中受到众多因素的影响,例如土壤<sup>[40, 47]</sup>、灌溉、施肥以及园艺管理措施<sup>[48]</sup>等。因此,在未来的研究中应考虑将上述因子纳入模型的输入特征变量组合,这将会进一步提高苹果相对气象产量模型的模拟精度。其次,本文选用的 3 种机器学习算法均为黑箱模型,并不能定量化解释输入变量对目标变量的影响<sup>[49-50]</sup>。最后,本文还未能将所选的最佳预测模型在其他不同苹果种植区域和不同的气候环境下进行充分验证。例如,LI 等<sup>[11]</sup>建立了陕西省县域产量预测模型,并预测了未来气候条件下 28 个苹果基地县的产量。因此,在随后的研究中拟将本文建立的苹果相对气象产量预测模型推广至整个黄土高原的苹果种植适宜区,并利用其分析未来气候变化条件下黄土高原产区的苹果相对气象产量变化。

## 4 结论

(1) 基于 GBDT 和 BRBP 机器学习算法的苹果相对气象产量模拟模型的精度均优于基于传统线性回归算法的模型,这表明最重要气象特征变量、空间特征变量和气象灾害特征变量与苹果产量之间存在较强的非线性关系,而机器学习算法能够更为准确地描述这种非线性关系。

(2) 以最重要气象特征变量、空间特征变量和气象灾害特征变量作为模型输入变量组合,GBDT 和 BRBP 算法模型在各个生育期内均能获得相对较高的苹果相对气象产量预测精度,SVM 和 MLR 算法模型可在果实膨大期获取较为理想的苹果相对气象产量模拟结果。

(3) 基于最重要气象特征变量、空间特征变量和气象灾害特征变量组合,机器学习算法(GBDT、SVM、BRBP)和线性算法(MLR)的产量模拟模型均可在苹果成熟期前 1~2 个月实现对黄土高原苹果相对气象产量较为准确的预测。

## 参 考 文 献

- [1] CHEN M Y, SHAO M A, WEI X R, et al. Response of the vertical distribution of soil water and nitrogen in the 5 m soil layer to the conversion of cropland to apple orchards in the Loess Plateau, China[J]. Agriculture, Ecosystems & Environment, 2022, 333: 107960.
- [2] ZHAO L H, GAO X D, HE N N, et al. Ecohydrological advantage of young apple tree-based agroforestry and its response to extreme droughts on the semiarid Loess Plateau[J]. Agricultural and Forest Meteorology, 2022, 321: 108969.
- [3] WANG S F, AN J, ZHAO X N, et al. Age- and climate-related water use patterns of apple trees on China's Loess Plateau[J]. Journal of Hydrology, 2020, 582: 124462.
- [4] RODRIGO J. Spring frosts in deciduous fruit trees-morphological damage and flower hardiness[J]. Scientia Horticulturae, 2000, 85(3): 155-173.
- [5] 韩文静,姜会飞,霍治国,等. 中国北方苹果连阴雨灾害时空分布特征[J]. 中国农业资源与区划, 2022, 43(6): 81-91.

- HAN Wenjing, JIANG Huifei, HUO Zhiguo, et al. Spatial-temporal distribution characteristics of apple continuous cloudy-rainy weather disaster in northern China[J]. Chinese Journal of Agricultural Resources and Regional Planning, 2022, 43(6): 81 – 91. (in Chinese)
- [6] 戴安然, 霍治国, 姜会飞. 中国苹果旱涝灾害研究进展[J]. 中国农业气象, 2022, 43(5): 362 – 379.  
DAI Anran, HUO Zhiguo, JIANG Huifei. Review on drought and waterlogging disasters for apple in China[J]. Chinese Journal of Agrometeorology, 2022, 43(5): 362 – 379. (in Chinese)
- [7] SHI L J, FENG P Y, WANG B, et al. Quantifying future drought change and associated uncertainty in southeastern Australia with multiple potential evapotranspiration models[J]. Journal of Hydrology, 2020, 590: 125394.
- [8] 马战林, 文枫, 周颖杰, 等. 基于作物生长模型与机器学习算法的区域冬小麦估产[J]. 农业机械学报, 2023, 54(6): 136 – 147.  
MA Zhanlin, WEN Feng, ZHOU Yingjie, et al. Regional winter-wheat yield estimation based on coupling of machine learning algorithm and crop growth model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(6): 136 – 147. (in Chinese)
- [9] PAUDEL D, BOOGAARD H, DE WIT A, et al. Machine learning for regional crop yield forecasting in Europe[J]. Field Crops Research, 2022, 276: 108377.
- [10] CAI Y P, GUAN K Y, LOBELL D, et al. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches[J]. Agricultural and Forest Meteorology, 2019, 274: 144 – 159.
- [11] LI M R, GUO J P, HE J Q, et al. Possible impact of climate change on apple yield in Northwest China[J]. Theoretical and Applied Climatology, 2020, 139(1 – 2): 191 – 203.
- [12] KHAN T, QIU J T, ALI QURESHI M A, et al. Agricultural fruit prediction using deep neural networks[J]. Procedia Computer Science, 2020, 174: 72 – 78.
- [13] 景辉, 杨华, 赵惠瑾, 等. 基于幼果期气象数据苹果产量早期预测研究——以运城市为例[J]. 中国农学通报, 2021, 37(8): 132 – 136.  
JING Hui, YANG Hua, ZHAO Huijin, et al. Early prediction of apple yield based on meteorological data in young fruit stage: an example of Yuncheng City[J]. Chinese Agricultural Science Bulletin, 2021, 37(8): 132 – 136. (in Chinese)
- [14] 河南省统计局. 《河南统计年鉴》[J]. 北京: 中国统计出版社, 2021.
- [15] 山西省统计局. 《山西统计年鉴》[J]. 北京: 中国统计出版社, 2021.
- [16] 陕西省统计局. 《陕西统计年鉴》[J]. 北京: 中国统计出版社, 2021.
- [17] 甘肃省统计局. 《甘肃发展年鉴》[J]. 北京: 中国统计出版社, 2021.
- [18] BRAY M, HAN D. Identification of support vector machines for runoff modelling[J]. Journal of Hydroinformatics, 2004, 6(4): 265 – 280.
- [19] 邱美娟, 刘布春, 刘园, 等. 中国北方苹果种植需水特征及降水适宜性[J]. 应用气象学报, 2021, 32(2): 175 – 187.  
QIU Meijuan, LIU Buchun, LIU Yuan, et al. Water requirement and precipitation suitability of apple planting in northern China[J]. Journal of Applied Meteorological Science, 2021, 32(2): 175 – 187. (in Chinese)
- [20] LI M R, GUO J P, XU C D, et al. Identifying climatic factors and circulation indices related to apple yield variation in main production areas of China[J]. Global Ecology and Conservation, 2018, 16: e00478.
- [21] LI L C, WANG B, FENG P Y, et al. Crop yield forecasting and associated optimum lead time analysis based on multi-source environmental data across China[J]. Agricultural and Forest Meteorology, 2021, 308 – 309: 108558.
- [22] HE C, LIU J D, XU F, et al. Improving solar radiation estimation in China based on regional optimal combination of meteorological factors with machine learning methods[J]. Energy Conversion and Management, 2020, 220: 113111.
- [23] PRIESTLEY C H B, TAYLOR R J. On the assessment of surface heat flux and evaporation using large scale parameters[J]. Monthly Weather Review, 1972, 100(2): 81 – 92.
- [24] 丁焯毅, 杨栋, 朱佳敏, 等. 基于灾害指数的奉化水蜜桃气象产量模拟[J]. 生态学杂志, 2021, 40(4): 1146 – 1153.  
DING Yeyi, YANG Dong, ZHU Jiamin, et al. Climate yield simulation of juicy peach in Fenghua based on disaster index[J]. Chinese Journal of Ecology, 2021, 40(4): 1146 – 1153. (in Chinese)
- [25] FRIEDMAN J H. Greedy function approximation: a gradient boosting machine[J]. Annals of Statistics, 2001, 29(5): 1189 – 1232.
- [26] PONTIL M, VERRI A. Properties of support vector machines[J]. Neural Computation, 1998, 10(4): 955 – 974.
- [27] SMOLA A J, SCHÖLKOPF B. A tutorial on support vector regression[J]. Statistics and Computing, 2004, 14(3): 199 – 222.
- [28] SUN Z, CHEN Y, LI X Y, et al. A Bayesian regularized artificial neural network for adaptive optics forecasting[J]. Optics Communications, 2017, 382: 519 – 527.
- [29] ROUBAH A, MEISS H, VILLERD J, et al. Predicting the abundances of aphids and their natural enemies in cereal crops: machine-learning versus linear models[J]. Biological Control, 2022, 169: 104866.
- [30] GUO Y H, FU Y S, HAO F H, et al. Integrated phenology and climate in rice yields prediction using machine learning methods[J]. Ecological Indicators, 2021, 120: 106935.
- [31] 白秀广, 李小盼. 气候变化对中国苹果主产区生产布局变迁的影响分析[J]. 北方园艺, 2016(1): 204 – 209.

- BAI Xiuguang, LI Xiaopan. Impact of climate change on the layout of apple production in China[J]. Northern Horticulture, 2016(1): 204–209. (in Chinese)
- [32] 张强. ‘富士’苹果果实品质与土壤养分和气象因子关系的研究[D]. 北京: 中国农业大学, 2018.  
ZHANG Qiang. Study on relationship between fruit quality of ‘Fuji’ apple and soil nutrition, meteorological factors[D]. Beijing: China Agricultural University, 2018. (in Chinese)
- [33] 白秀广, 陈晓楠, 郑少锋. 气候变化对苹果主产区单产及单产增长的贡献研究[J]. 中国农业大学学报, 2015, 20(4): 82–91.  
BAI Xiuguang, CHEN Xiaonan, ZHENG Shaofeng. Impact and contribution of climate change on apple productivity in China [J]. Journal of China Agricultural University, 2015, 20(4): 82–91. (in Chinese)
- [34] 刘峻明, 和晓彤, 王鹏新, 等. 长时间序列气象数据结合随机森林法早期预测冬小麦产量[J]. 农业工程学报, 2019, 35(6): 158–166.  
LIU Junming, HE Xiaotong, WANG Pengxin, et al. Early prediction of winter wheat yield with long time series meteorological data and random forest method[J]. Transactions of the CSAE, 2019, 35(6): 158–166. (in Chinese)
- [35] FENG P Y, WANG B, LIU D L, et al. Dynamic wheat yield forecasts are improved by a hybrid approach using a biophysical model and machine learning technique[J]. Agricultural and Forest Meteorology, 2020, 285–286: 107922.
- [36] KASAMPALIS D S, TSOVALTZIS P, NTOUROS K, et al. The use of digital imaging, chlorophyll fluorescence and Vis/NIR spectroscopy in assessing the ripening stage and freshness status of bell pepper fruit [J]. Computers and Electronics in Agriculture, 2021, 187: 106265.
- [37] YE T, NIE J L, WANG J, et al. Performance of detrending models of crop yield risk assessment: evaluation on real and hypothetical yield data[J]. Stochastic Environmental Research and Risk Assessment, 2015, 29(1): 109–117.
- [38] LU J Y, CARBONE G J, GAO P. Detrending crop yield data for spatial visualization of drought impacts in the United States, 1895–2014[J]. Agricultural and Forest Meteorology, 2017, 237–238: 196–208.
- [39] ZHAO D, WANG Y, REN C, et al. Comparative analysis of three fitting methods of rice trend yield[J]. Chinese Journal of Eco-Agriculture, 2017, 25(3): 345–355.
- [40] ZHANG Q, ZHOU B B, LI M J, et al. Multivariate analysis between meteorological factor and fruit quality of Fuji apple at different locations in China[J]. Journal of Integrative Agriculture, 2018, 17(6): 1338–1347.
- [41] CHEN J D, TIAN H W, HUANG J, et al. Climate-driven yield variability for winter wheat in Henan Province, north China and its relation to large-scale atmospheric circulation indices[J]. International Journal of Plant Production, 2021, 15(1): 79–91.
- [42] EITZINGER J, THALER S, SCHMID E, et al. Sensitivities of crop models to extreme weather conditions during flowering period demonstrated for maize and winter wheat in Austria[J]. Journal of Agricultural Science, 2013, 151(6): 813–835.
- [43] OBSIE E Y, QU H C, DRUMMOND F. Wild blueberry yield prediction using a combination of computer simulation and machine learning algorithms[J]. Computers and Electronics in Agriculture, 2020, 178: 105778.
- [44] DE SOUZA A V, NETO A B, PIAZENTIN J C, et al. Artificial neural network modelling in the prediction of bananas’ harvest [J]. Scientia Horticulturae, 2019, 257: 108724.
- [45] KOK Z H, SHARIFF A, ALFATNI M, et al. Support vector machine in precision agriculture: a review[J]. Computers and Electronics in Agriculture, 2021, 191: 106546.
- [46] FENG P Y, WANG B, LUO J J, et al. Using large-scale climate drivers to forecast meteorological drought condition in growing season across the Australian wheatbelt[J]. Science of the Total Environment, 2020, 724: 138162.
- [47] HUANG X, CHEN T, ZHOU P Y, et al. Prediction and optimization of fruit quality of peach based on artificial neural network [J]. Journal of Food Composition and Analysis, 2022, 111: 104604.
- [48] SHAN Q J, WANG Z K, LING H B, et al. Unreasonable human disturbance shifts the positive effect of climate change on tree-ring growth of *Malus sieversii* in the origin area of world cultivated apples[J]. Journal of Cleaner Production, 2021, 287: 125008.
- [49] SU X, WU Y Y, PEI H J, et al. Prediction of coke yield of FCC unit using different artificial neural network models[J]. China Petroleum Processing & Petrochemical Technology, 2016, 18(3): 102–109.
- [50] LISCHIED G, WEBBER H, SOMMER M, et al. Machine learning in crop yield modelling: a powerful tool, but no surrogate for science[J]. Agricultural and Forest Meteorology, 2022, 312: 108698.