

doi:10.6041/j.issn.1000-1298.2021.10.026

基于 LAI 和 VTCI 及 Copula 函数的冬小麦单产估测

王鹏新¹ 陈 弛¹ 张树誉² 张 悦¹ 李红梅²

(1. 中国农业大学信息与电气工程学院, 北京 100083; 2. 陕西省气象局, 西安 710014)

摘要: 受全球变暖影响,近年来干旱事件发生的频率和强度均呈显著增加的趋势,严重影响了农作物的产量。因此,选择合适的监测指标、构建准确的产量估测模型,对保障国家粮食安全具有十分重要的意义。以关中平原为研究区域,基于与作物长势密切相关的条件植被温度指数(VTCI)和叶面积指数(LAI),采用主成分分析法(PCA)结合 Copula 函数分别构建县域尺度单变量(VTCI 或 LAI)、双变量(VTCI 和 LAI)的冬小麦单产估测模型。结果表明,基于 PCA-Copula 构建的综合 LAI 与冬小麦的单产模型精度最高($R^2 = 0.567, P < 0.001$),但综合 VTCI 和 LAI 与冬小麦间的单产模型($R^2 = 0.524, P < 0.001$)用于研究区域 2012—2017 年各县(区)的冬小麦单产估测时误差分布更为集中、最大误差更小,比基于综合 VTCI、综合 LAI 建立的估产模型的估测结果更可靠,表明应用 PCA-Copula 构建的双变量估产模型更适合大范围的冬小麦单产估测。

关键词: 冬小麦; 估产; 主成分分析; Copula 函数; 条件植被温度指数; 叶面积指数

中图分类号: TP79 **文献标识码:** A **文章编号:** 1000-1298(2021)10-0255-09

OSID:

Winter Wheat Yield Estimation Based on Copula Function and Remotely Sensed LAI and VTCI

WANG Pengxin¹ CHEN Chi¹ ZHANG Shuyu² ZHANG Yue¹ LI Hongmei²

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. Shaanxi Provincial Meteorological Bureau, Xi'an 710014, China)

Abstract: Affected by the global warming, the frequency and intensity of drought events have shown a significant increase in recent years, which has seriously affected crop yields. Therefore, selecting appropriate monitoring indicators and constructing accurate yield estimation models are of great significance to ensure the country's food security. The Guanzhong Plain in Shaanxi Province was chosen as the study area, and remotely sensed vegetation temperature condition index (VTCI) and leaf area index (LAI) which are closely related to crop growth were selected as the growth monitoring indicators. The principal component analysis (PCA) combined with Copula function were used to construct univariate (VTCI or LAI) and bivariate (VTCI and LAI) winter wheat yield estimation models at the county scale. The results showed that the liner regression model of comprehensive values of LAI and winter wheat yield constructed based on PCA-Copula had the highest accuracy ($R^2 = 0.567, P < 0.001$). However, when the liner regression model of comprehensive values of VTCI and LAI and winter wheat yield ($R^2 = 0.524, P < 0.001$) was used to estimate the yield of winter wheat in each county (district) in the study area from 2012 to 2017, the distribution of the error between the estimated yield and the actual yield was more concentrated, and the maximum error value was also smaller, which was more reliable than the results of the winter wheat yield estimation model based on a single variable. These results indicated that the bivariate yield estimation model constructed by PCA-Copula was more suitable for large-scale winter wheat yield estimation.

Key words: winter wheat; yield estimation; principal component analysis; Copula function; vegetation temperature condition index; leaf area index

收稿日期: 2020-09-23 修回日期: 2020-11-10

基金项目: 国家自然科学基金项目(41871336)

作者简介: 王鹏新(1965—),男,教授,博士,主要从事定量遥感及其在农业中的应用研究, E-mail: wangpx@cau.edu.cn

0 引言

冬小麦是我国主要粮食作物之一,对冬小麦进行大规模的长势监测和产量估测,不仅能为地方政府制定生产计划提供科学依据,也能为粮食安全提供保障。近年来,遥感技术以其快速、简便、宏观、无损及客观等优点被广泛应用于农业生产的各个环节^[1]。国内外学者通常使用可以表征作物生长信息的植被指数(Vegetation index, VI)来建立与作物产量间的经验模型,进而进行作物估产。以往的研究中常用的植被指数有归一化植被指数(Normalized difference vegetation index, NDVI)^[2]、增强型植被指数(Enhanced vegetation index, EVI)^[3]、比值植被指数(Ratio vegetation index, RVI)^[4]等。利用遥感技术获得的叶面积指数(Leaf area index, LAI)可以反映作物光合作用、呼吸作用、蒸腾作用及碳氮循环等生物物理过程^[5],常作为作物长势监测和产量估测的重要参数^[6]。然而作物单产不仅与植被指数及其相关的长势参数有关,还与土壤水分有密切联系。在NDVI和地表温度(Land surface temperature, LST)的散点图呈三角形区域分布的基础上,文献^[7]提出条件植被温度指数(Vegetation temperature condition index, VTCI)的干旱监测方法,弥补了单一遥感干旱监测指数的不足,并在干旱监测、预测和影响评估等领域得到了广泛的应用。在此基础上,综合考虑作物长势与干旱胁迫可以更加准确地描述作物生长状态,从而提高作物产量估测精度。

Copula函数是构造多元联合分布和随机变量间相关结构的重要工具。近年来,Copula函数以其在多变量分析中的便利性引起了学者们的广泛关注,它能够将变量的相关性结构与边缘分布分开处理,考虑变量间的非线性与非正态关系,没有任何限制^[8-10]。VERGNI等^[11]将Copula理论应用于农业干旱领域,利用Student Copula获得了向日葵相对发病率 and 相对严重程度的联合概率以及双变量返回期,为干旱规划和管理提供了有用的信息。张迎等^[12]综合降雨和径流两种要素,基于Archimedean Copula构建了一种能够综合表征气象干旱和水文干旱的新型综合干旱指标,并用其表征渭河流域的干旱演变特征。针对Copula函数构建多元随机变量间联合分布时参数求解困难、不同Copula函数的参数不尽相同的问题,王鹏新等^[13-14]分别利用主成分分析法(Principal component analysis, PCA)和核主成分分析法(Kernel PCA, KPCA)提取VTCI的主成分因子,基于Copula函数评估冬小麦主要生育期干旱对其产量的影响,验证了Copula函数对基于VTCI

的干旱影响评估研究具有较好的适用性。然而作物的长势和产量是多个指标共同作用的结果,仅靠单一的指标进行作物长势监测研究难以反映不同指标在作物生长与产量估测过程中的相互影响^[15]。因此,本文以陕西省关中平原为研究区域,选取VTCI和LAI为长势监测指标,将PCA应用于提取单个、两个特征变量的主成分数据,再结合无需求解参数的乘积Copula函数构建综合长势监测指标,并建立综合长势监测指标与冬小麦单产之间的线性回归模型,以期获得更准确的冬小麦单产估测方法。

1 材料与方法

1.1 研究区域概况

关中平原位于陕西省中部的渭河流域,西起宝鸡大散关,东至渭南潼关,北到陕北黄土高原,南至秦岭,坐标为106°22'~110°24'E, 33°57'~35°39'N。行政区域包括西安市、铜川市、宝鸡市、咸阳市、渭南市和杨凌国家农业高新技术产业示范区,其中杨凌国家农业高新技术产业示范区由于面积较小,故将其划分到咸阳市境内。关中平原是陕西省的农业基地,也是全国粮食主要生产基地之一^[16]。该区域年平均气温在6~13℃之间,年平均降雨量在550~700 mm之间,但分布极不均匀,11月到次年5月易发生干旱^[17]。受全球气候变暖影响,农业干旱灾害不断加剧,由干旱造成的冬小麦减产现象时有发生^[18-19]。

1.2 试验数据

1.2.1 时间序列VTCI的生成

选取关中平原2012—2017年每年3—5月的Aqua-MODIS遥感数据的日地表反射率产品(MYD09GA)和日地表温度产品(MYD11A1),获取日LST和日NDVI,通过最大值合成法分别逐像素取10 d内所包含的多日LST和NDVI的最大值作为该像素的LST和NDVI值,即获得旬LST和旬NDVI最大值合成产品。将多年某一句的LST和NDVI最大值合成产品再一次运用最大值合成技术,生成多年旬尺度LST和NDVI最大值合成产品。基于最小值合成技术将多年某一句LST的最大值合成产品逐像素取最小值,得到多年旬尺度LST最大-最小值合成产品;通过多年旬尺度NDVI和LST最大值合成产品确定研究区域VTCI的热边界,通过多年旬尺度NDVI最大值合成产品和多年旬尺度LST最大-最小值合成产品确定VTCI的冷边界。VTCI计算公式为^[20-21]

$$VTCI = \frac{LST_{\max}(NDVI_i) - LST(NDVI_i)}{LST_{\max}(NDVI_i) - LST_{\min}(NDVI_i)} \quad (1)$$

其中 $LST_{\max}(NDVI_i) = a + bNDVI_i$ (2)

$LST_{\min}(NDVI_i) = a' + b'NDVI_i$ (3)

式中 $LST(NDVI_i)$ ——研究区域内某一像素的 NDVI 值为 $NDVI_i$ 时的地表温度

$LST_{\max}(NDVI_i)$ ——研究区域内当 NDVI 值等于 $NDVI_i$ 时的所有像素地表温度的最大值, 称作热边界

$LST_{\min}(NDVI_i)$ ——研究区域内当 NDVI 值等于 $NDVI_i$ 时的所有像素地表温度的最小值, 称作冷边界

a, b, a', b' ——待定系数, 由研究区域 NDVI 和 LST 的散点图近似获得

VTCI 的取值范围为 $[0, 1]$, 其中, VTCI 值越小, 表示旱情越严重, 值越大, 则相反。根据越冬后冬小麦的生长情况, 将越冬后的生育时期划分为返青期(3 月上旬—中旬)、拔节期(3 月下旬—4 月中旬)、抽穗—灌浆期(4 月下旬—5 月上旬)和乳熟期(5 月中旬—下旬)4 个生育时期^[22-23]。取某一生育时期内多旬 VTCI 的平均值作为该生育时期的 VTCI 值, 叠加关中平原的行政矢量图, 取各县(区)所包含像素的 VTCI 的平均值作为该地区该生育时期的 VTCI 值。

1.2.2 时间序列 LAI 的生成

选取 2012—2017 年冬小麦主要生育时期的 MCD15A3H 产品, 与 MOD15A2(Terra - MODIS) 和 MYD15A2(Aqua - MODIS) 产品相比, MCD15A3H 具有较高的时间分辨率(4 d)和空间分辨率(500 m), 对于监测农作物的生长和物候特性更为有利。原始叶面积指数产品由于云和大气等因素的影响, 会导致 LAI 数据出现骤降现象, 从而降低准确性。为了解决这个问题, 应用上包络线 Savitzky - Golay(S - G) 滤波对原始时间序列 LAI 进行平滑处理^[24], 经上包络线 S - G 滤波平滑处理后的叶面积指数数据更加符合冬小麦的生长情况。

为了使 LAI 和 VTCI 具有相同的取值范围, 将 S - G 滤波后的 LAI 进行归一化处理; 逐像素取所在旬 LAI 的最大值作为该旬的 LAI, 取各生育时期所包含的多旬 LAI 的平均值作为该生育时期的 LAI 值, 通过叠加研究区域的行政矢量图, 取各县(区)所包含像素的 LAI 的平均值作为该地区该生育时期的 LAI 值。

1.2.3 冬小麦单产数据来源

《陕西统计年鉴》记录了冬小麦的产量数据。

因此, 关中平原各县(区)2012—2017 年的单产数据均来源于其所在市(西安市、咸阳市、宝鸡市、渭南市)的统计年鉴。

1.3 PCA - Copula 法

主成分分析是一种多元统计方法, 可通过降维技术将多个变量转换为少数几个主成分, 这些主成分能够反映原始变量的大部分信息, 通常表示为原始变量的线性组合。由于主成分因子是新的互相独立的变量, 因此在建立 Copula 函数时无需求解参数, 计算简便。

1.3.1 主成分分析

首先根据样本数据构建矩阵 $A_{n \times p}$, 计算协方差矩阵 $R_{p \times p}$, 进而求出矩阵的特征值 $\lambda_i (i = 1, 2, \dots, p)$ 及其对应的单位特征向量 e_i , 各主成分的贡献率 D_i 用来反映其信息量^[13], 计算公式为

$$D_i = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \times 100\% \quad (4)$$

最终选择几个主成分, 即 $P_{C1}, P_{C2}, \dots, P_{Cm}$ 中的 m 通常由主成分的累计贡献率 $G(m)$ 来确定, 计算式为

$$G(m) = \frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i} \times 100\% \quad (5)$$

根据经验, 当 $G(m)$ 大于 85% 时, 认为能够反映原来的信息, 对应的 m 就是选取的前 m 个主成分。第 r 个主成分的线性表示为

$$P_{Cr} = a_{1r}X_1 + a_{2r}X_2 + \dots + a_{pr}X_p \quad (6)$$

式中 P_{Cr} ——第 r 个主成分因子

a_{pr} ——所选指标在第 r 个主成分线性组合中的系数

p ——原变量所选指标的个数

X_p ——所选指标(如 VTCI、LAI)的值

1.3.2 Copula 函数

Copula 函数是定义在 $[0, 1]$ 上均匀分布的随机向量的联合分布函数。根据 Sklar 定理^[25], 联合分布函数与相关结构函数之间存在一一对应的关系, Copula 函数形式可表示为

$$F(x_1, x_2, \dots, x_j) = C_\theta(F_1(x_1), F_2(x_2), \dots, F_j(x_j)) = C(u_1, u_2, \dots, u_j) \quad (7)$$

其中

$$u_j = F(x_j)$$

式中 F ——随机变量分布函数

j ——样本容量

C ——Copula 函数

u_j ——随机变量 x_j 的最优边缘分布函数

由于采用 PCA 方法提取样本数据的 m 个主成分因子间互无相关性,故可以通过相应的乘积 Copula 获取综合长势监测指标,即

$$I = C(u_1, u_2, \dots, u_m) = \prod_{i=1}^m u_i \quad (8)$$

式中 I ——基于前 m 个主成分确定的综合长势监测指标

1.3.3 边缘分布函数确定与优度评价

由给定样本集合求解随机变量边缘分布函数的方法包括参数法和非参数法。其中参数法是假定总体服从某种已知的分布,利用极大似然法估计分布函数的参数,这种方法依赖于实现对总体分布的假设。常用的分布线型包括皮尔逊 III 型分布、正态分布、广义极值分布、对数分布、指数分布等^[26]。首先通过 Kolmogorov - Smirnov (K - S) 方法检验样本是否服从指定的分布,原假设为样本服从指定分布,备择假设为不服从指定分布。当输出检验值 $h = 1$ 时,在显著性水平下拒绝原假设; $h = 0$ 时,则在显著性水平下接受原假设。若该样本某一主成分的多个指定分布均通过 K - S 检验,则通过各边缘分布的理论频率与经验频率之间的均方根误差 (Root mean square error, RMSE) 及赤池信息准则值 (Akaike information criterion, AIC) 评估拟合效果, RMSE 和 AIC 越小,表明拟合效果越好^[27]。一维随机分布的经验频率^[28]计算公式为

$$Q_s = \frac{s - 0.44}{N + 0.12} \quad (9)$$

式中 N ——样本数

s ——从最小样本开始观察的顺序值

Q_s ——第 s 项的累计经验频率

表 1 冬小麦 4 个生育时期不同长势监测指标的主成分贡献率及其线性表达式

Tab.1 Principal component contribution rates and linear expressions of different growth monitoring indices at four growth stages of winter wheat

指标	贡献率/ %	累计贡献 率/%	主成分表达式
VTCI	主成分 1	64.78	$P_{CV1} = 0.982V_1 + 0.170V_2 + 0.076V_3 + 0.031V_4$
	主成分 2	25.61	$P_{CV2} = -0.412V_1 + 0.438V_2 + 0.579V_3 + 0.673V_4$
LAI	主成分 1	91.68	$P_{CL1} = 0.371L_1 + 0.643L_2 + 0.577L_3 + 0.341L_4$
	主成分 2	7.10	$P_{CL2} = 0.763L_1 + 0.169L_2 - 0.392L_3 - 0.486L_4$
VTCI 和 LAI	主成分 1	59.44	$P_{CVL1} = 0.133V_1 + 0.315V_2 + 0.345V_3 + 0.299V_4 + 0.357L_1 + 0.423L_2 + 0.432L_3 + 0.426L_4$
	主成分 2	14.41	$P_{CVL2} = 0.530V_1 - 0.029V_2 - 0.454V_3 - 0.601V_4 + 0.293L_1 + 0.211L_2 + 0.135L_3 + 0.054L_4$
	主成分 3	12.27	$P_{CVL3} = 0.716V_1 + 0.419V_2 + 0.193V_3 + 0.143V_4 - 0.358L_1 - 0.302L_2 - 0.184L_3 - 0.003L_4$

注: P_{CV1} 、 P_{CV2} 分别表示 VTCI 的前 2 个主成分因子; P_{CL1} 、 P_{CL2} 分别表示 LAI 的前 2 个主成分因子; P_{CVL1} 、 P_{CVL2} 、 P_{CVL3} 分别表示 VTCI 和 LAI 的前 3 个主成分因子; V_1 、 V_2 、 V_3 、 V_4 分别表示冬小麦返青期、拔节期、抽穗-灌浆期、乳熟期的 VTCI; L_1 、 L_2 、 L_3 、 L_4 分别表示冬小麦返青期、拔节期、抽穗-灌浆期、乳熟期的 LAI。

RMSE 计算公式为

$$V_{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Q_{sir} - Q_{tir})} \quad (10)$$

式中 Q_{sir} ——第 r 个主成分第 i 个数据的经验频率
 Q_{tir} ——第 r 个主成分第 i 个数据的理论频率
 n ——第 r 个主成分因子中数据的个数,取 n 为 144

AIC 计算公式为

$$V_{AIC} = n \ln V_{RMSE} + 2\sigma \quad (11)$$

式中 σ ——所选模型中参数的个数

当样本总体分布未知或不符合常见分布的线型时,则不易用参数法对边缘分布作出估计。非参数法则避开了线型选择的问题,经验分布与核密度估计均属于非参数检验方法,它们可以从样本数据本身出发研究数据分布的特征,估计未知的分布函数,不受限于事先对总体分布做出假设。

2 结果与分析

2.1 冬小麦 4 个生育时期各长势监测指标的主成分分析

由于 VTCI 和经归一化处理后的 LAI 的取值范围均为 $[0, 1]$, 因此直接通过 PCA 方法对原始数据进行降维处理,根据主成分相应累计贡献率达到 85% 以上的原则选取主成分 (表 1)。VTCI 第 1 主成分的贡献率为 64.78%, 第 2 主成分贡献率为 25.61%, 前 2 个主成分的累计贡献率为 90.39%; LAI 的第 1 主成分贡献率达到 91.68%, 第 2 主成分贡献率为 7.10%, 前 2 个主成分的累计贡献率为 98.78%; VTCI 和 LAI 双变量数据、共 8 项指标经 PCA 处理后的前 3 个主成分的贡献率分别为 59.44%、14.41% 和 12.27%, 其累计贡献率为 86.12%。

2.2 主成分因子边缘分布函数的确定

基于目前常用的分布线型,选取正态分布、广义极值分布和对数分布分别拟合冬小麦 4 个生育时期各长势监测指标的主成分因子,应用极大似然法获得对应的边缘分布函数参数,利用 K-S 检验对各主成分的边缘分布函数进行拟合优度评价(表 2)。可以看出,各长势监测指标主成分因子的分布拟合情况分为 2 种,指定的 3 种分布均未通过 K-S 检验与仅通过正态分布和广义极值分布的 K-S 检验。以 VTCI 的第 1、2 主成分为例,阐述边缘分布函数的选择过程。VTCI 第 1 主成分的正态分布、广义极值分布和对数分布均未通过 K-S 检验,因此选用非参数法拟合该主成分的经验分布。由图 1 可见,VTCI 第 1 主成分因子经验分布和核分布估计曲线的走势基本吻合,因此优选更加光滑的核分布估计结果作为该主成分的经验分布。VTCI 第 2 主成分的正态分布和广义极值分布均能通过 K-S 检验,依据边缘分布的理论频率与经验频率之间 RMSE 及 AIC 确定最优边缘分布(式(9))。经计算,广义极值分布的拟合结果($V_{RMSE} = 0.013$, $V_{AIC} = -616.592$)优于正态分布的拟合结果($V_{RMSE} = 0.016$, $V_{AIC} = -592.383$),因此 VTCI 的第 2 主成分优选广义极值分布。同理可以得出,LAI 的第 1 主成分优选广义极值分布,第 2 主成分选择核分布;VTCI 和 LAI 的第 1、2、3 主成分均优选广义极值分布(表 2)。

表 2 冬小麦 4 个生育时期不同长势监测指标主成分因子的 K-S 检验值(h)及其优选分布

Tab.2 K-S test value (h) and optimal distribution of principal components of different growth monitoring indices at four growth stages of winter wheat

指标	正态分布	广义极值分布	对数分布	优选分布
VTCI	主成分 1	1	1	核分布
	主成分 2	0	0	广义极值分布
LAI	主成分 1	0	0	广义极值分布
	主成分 2	1	1	核分布
VTCI 和 LAI	主成分 1	0	0	广义极值分布
	主成分 2	0	0	广义极值分布
	主成分 3	0	0	广义极值分布

2.3 综合长势监测指标与冬小麦单产间的线性回归分析

以 VTCI 为例,阐述综合长势监测指标的构建过程。从表 1 中 VTCI 第 1 主成分 P_{CV1} 的表达式看,返青期 VTCI 的系数最大,其他生育时期的系数比较小,说明返青期 VTCI 对 P_{CV1} 的影响最大。从第 2 主成分 P_{CV2} 的表达式看, V_1 的系数为负数,说明随着返青期 VTCI 的减小, P_{CV2} 增大,可能是由于返青期

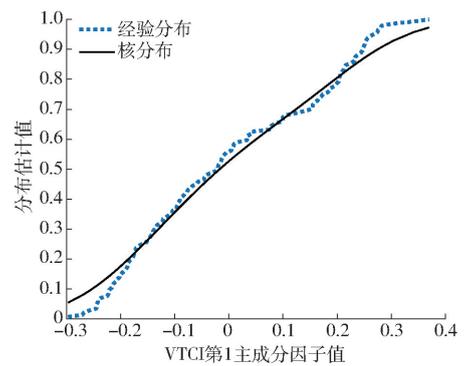


图 1 冬小麦 4 个生育时期 VTCI 第 1 主成分的经验分布和核分布估计曲线

Fig. 1 Empirical distribution and kernel distribution estimation curves of the first principal component of VTCI at four growth stages of winter wheat

的干旱的补偿效应造成的,即在冬小麦返青期发生一定程度的水分胁迫不会对作物生长造成太大的影响,反而能够提高作物在生育后期的水分利用效率,使产量增加^[29]; P_{CV2} 在拔节期 VTCI、抽穗-灌浆期 VTCI、乳熟期 VTCI 上有均相近的正系数,说明这 3 个生育时期对 P_{CV2} 的重要性都相似。基于此,分别以各主成分的贡献率在其累计贡献率中所占的比重将主成分综合,进一步获取前 2 个主成分的综合线性表达式为

$$P_{CV} = 0.717P_{CV1} + 0.283P_{CV2} \quad (12)$$

将 VTCI 前 2 个主成分的线性表达式(表 1)代入式(12)得到综合评价值为

$$P_{CV} = 0.664V_1 + 0.246V_2 + 0.218V_3 + 0.213V_4 \quad (13)$$

由式(13)可见,返青期 VTCI 的系数最大,乳熟期 VTCI 的系数最小。但拔节期是干旱对冬小麦生长影响的关键因素,其次为抽穗-灌浆期,返青期和乳熟期的影响相对较小^[22-23]。该模型的构建没有考虑前 2 个主成分各自的分布特征,因此仅采用主成分分析得到的综合评价值 P_{CV} 存在不足,而将主成分分析与 Copula 函数结合的方法不受各个单因子变量边缘分布的影响。因此,根据各主成分对应的最优边缘分布特征建立联合分布(式(8)),进而构建不同的综合长势监测指标。

由于表 1 中选择各长势监测指标的主成分个数均基于经验方法,没有理论支持该方法反映出的信息量最优,因此利用 PCA-Copula 方法计算基于 VTCI 和 LAI 的包含不同主成分个数的综合长势监测指标,进而与冬小麦单产进行线性回归分析(图 2、3、4)。结果表明,综合 VTCI 包含第 1 个主成分时(记为综合 VTCI-1)与冬小麦单产的相关性较低($R^2 = 0.066$, $P = 0.002$)(图 2);综合 VTCI 包含

前2个主成分(记为综合VTCI-2)与冬小麦的相关性达到极显著水平($R^2 = 0.246, P < 0.001$),此时综合VTCI-2中包含的主成分也符合主成分个数的选取原则。

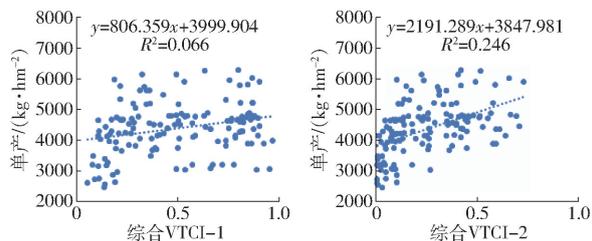


图2 PCA-Copula方法计算的冬小麦综合VTCI与单产间的线性回归模型

Fig. 2 Linear regression models of winter wheat yields and comprehensive values of VTCI calculated by PCA-Copula method

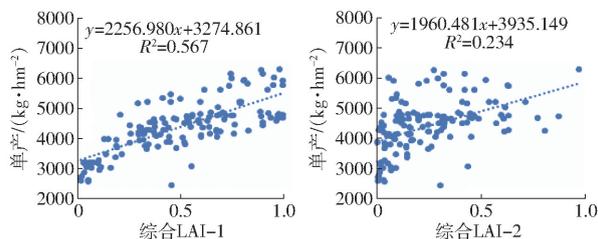


图3 PCA-Copula方法计算的冬小麦综合LAI与单产间的线性回归模型

Fig. 3 Linear regression models of winter wheat yields and comprehensive values of LAI calculated by PCA-Copula method

综合LAI包含第1个主成分(记为综合LAI-1)与冬小麦单产的相关性($R^2 = 0.567, P < 0.001$)优于综合LAI包含前2个主成分(记为综合LAI-2)与冬小麦单产的相关性($R^2 = 0.234, P < 0.001$)(图3)。这是由于建立LAI的第2主成分 P_{CV2} 表达式(表1)中抽穗-灌浆期LAI、乳熟期LAI系数均为负值,即 P_{CV2} 在抽穗-灌浆期LAI、乳熟期LAI均有中等程度的负载荷。LAI与作物的长势及产量呈正相关,因此表达式中负系数与实际不符,加入后会带来噪声,使精度下降。分析表明,利用PCA-

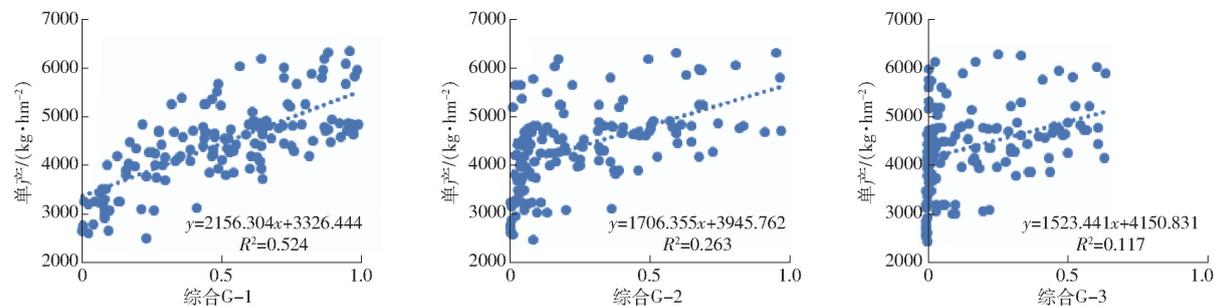


图4 PCA-Copula方法计算的冬小麦综合VTCI和LAI与单产间的线性回归模型

Fig. 4 Linear regression models of winter wheat yields and comprehensive values of VTCI and LAI calculated by PCA-Copula method

Copula方法进行作物长势监测和产量估测时,主成分的选取并非越多越好,不仅需分析所采用方法的理论基础,而且需分析建立的各主成分表达式中各长势监测指标的系数与作物长势间的关系。

综合VTCI和LAI包含第1个主成分(记为综合G-1)与冬小麦单产之间的回归结果($R^2 = 0.524, P < 0.001$)优于综合VTCI和LAI包含前2个主成分(记为综合G-2)与冬小麦单产之间的回归结果($R^2 = 0.263, P < 0.001$),以及综合VTCI和LAI包含前3个主成分(记为综合G-3)与冬小麦单产之间的回归结果($R^2 = 0.117, P < 0.001$)(图4)。随着第2、3主成分的加入,综合VTCI和LAI与冬小麦单产之间的相关性逐渐降低。这是由于建立VTCI和LAI的第2主成分 P_{CV2} 表达式(表1)中拔节期VTCI、抽穗-灌浆期VTCI、乳熟期VTCI系数均为负值,其中 P_{CV2} 在乳熟期VTCI有中等程度的负载荷;建立VTCI和LAI的第3主成分 P_{CV3} 表达式(表1)中的4个负系数均与LAI相关。VTCI已被证明能够准确地反映作物水分胁迫信息,LAI与作物的长势和产量密切相关, P_{CV2} 、 P_{CV3} 表达式中负系数指标的加入会导致综合值变小,精度下降,不符合实际情况。

由此可见,尽管没有满足主成分的累计贡献率不低于85%的要求,但包含综合长势监测指标第1个主成分的综合LAI-1、综合G-1相比于综合VTCI-2与冬小麦产量之间的关系更为紧密,以PCA-Copula方法为基础计算作物综合长势监测指标时,仅仅考虑主成分分析的理论基础不够全面,更需考虑各长势监测指标对作物长势监测结果产生的实际影响,从而建立与作物长势和产量最密切的综合长势监测指标。

2.4 冬小麦单产估测的精度评价

基于包含最优主成分个数的各综合长势监测指标分别建立估产模型,计算关中平原24个县(区)的估测单产。利用综合VTCI-2建立的冬小麦估产模型估测单产与实际单产的均方根误差为

733.64 kg/hm², 相对误差为 0.05% ~ 57.76%, 平均相对误差为 13.99%; 利用 LAI-1 建立的冬小麦估产模型估测单产与实际单产的均方根误差为 556.28 kg/hm², 相对误差为 0.18% ~ 73.21%, 平均相对误差为 10.31%; 利用综合 G-1 建立的冬小麦估产模型估测单产与实际单产的均方根误差为 582.81 kg/hm², 相对误差为 0.03% ~ 54.79%, 平均相对误差为 11.03%。相比之下, 基于综合 VTCI-2 建立的估产模型精度较低; 基于综合 LAI-1 和综合 G-1 建立的估产模型总体估产误差较小, 其中综合 LAI-1 建立的估产模型计算的估测单产存在部分年份或部分地区的冬小麦估测结果与实际偏差较大的情况。

为了进一步验证估测模型的精度, 将基于综合 LAI-1、综合 G-1 建立的冬小麦估产模型分别用于分析 2012—2017 年关中平原 24 个县(区)估测单产与实际单产的均方根误差和平均相对误差(表 3)。基于综合 LAI-1 构建的冬小麦单产回归模型各县(区)估测单产与实际单产均方根误差为 126.74(三原县) ~ 1 148.87 kg/hm²(眉县), 估测单产与实际单产的平均相对误差以三原县最低, 为 2.35%, 以蒲城县最高, 为 26.40%。基于综合 G-1 构建冬小麦单产回归模型各县(区)估测单产与实际单产均方根误差为 125.91(大荔县) ~ 1 061.63 kg/hm²(阎良区), 估测单产与实际单产的平均相对误差以大荔县最低, 为 2.56%, 以合阳县最高, 为 23.62%。上述结果表明, 基于综合 LAI-1 建立估产模型的决定系数(R^2)虽略高于基于综合 G-1 建立的估产模型, 但具体到各县(区)的估产时, 误差范围大于综合 G-1 所建立的估产模型, 说明利用单一 LAI 构建的综合 LAI 得到的冬小麦估产结果容易存在偏差。相对于单一指标构建的综合 VTCI 或综合 LAI, 基于双变量构建的综合干旱指标(综合 VTCI 和 LAI)得到的冬小麦估产结果更加全面可靠, 可用于研究区域冬小麦单产的估测。

3 讨论

已有研究表明 Copula 函数对于基于 VTCI 的干旱影响评估具有良好的适用性^[13-14], LAI 可以反映作物在不同时间阶段生长发育的动态特征和健康状况, 是表征作物长势与进行产量预报的重要参数。鉴于此, 本文基于冬小麦越冬后主要生育期 VTCI、LAI 为长势监测指标, 尝试利用 PCA-Copula 法分别建立冬小麦主要生育期单变量(综合 VTCI 或综合 LAI)、双变量(综合 VTCI 和 LAI)的估产模型。由主成分分析提取出的少数的几个主成分因子几乎

表 3 关中平原各县(区)2012—2017 年冬小麦估测单产与实际单产对比

Tab. 3 Comparison of estimated and actual winter wheat yields for each county (district) in Guanzhong Plain from 2012 to 2017

县(区)	综合 LAI-1		综合 G-1	
	均方根误差/ (kg·hm ⁻²)	平均相对 误差/%	均方根误差/ (kg·hm ⁻²)	平均相对 误差/%
阎良	966.06	15.92	1 061.63	17.23
临潼	366.32	7.40	410.32	7.19
长安	219.44	4.08	292.03	5.29
蓝田	434.63	10.37	461.19	9.86
周至	396.65	8.22	286.74	5.63
户县	163.40	2.80	273.66	5.27
三原	126.74	2.35	310.34	5.25
泾阳	312.41	6.50	364.77	6.01
乾县	183.39	3.34	414.86	7.70
礼泉	376.07	6.96	385.54	7.31
永寿	376.02	7.81	606.26	12.65
淳化	254.04	6.24	462.79	10.34
武功	639.29	13.54	567.55	11.01
兴平	186.92	3.51	325.99	5.68
凤翔	725.66	15.46	636.42	12.96
岐山	459.17	7.68	584.19	9.54
扶风	884.75	15.19	1 039.82	17.33
眉县	1 148.87	21.88	1 023.69	19.37
大荔	182.63	3.06	125.91	2.56
合阳	603.50	20.19	712.77	23.62
澄城	403.84	12.15	484.47	15.12
蒲城	988.39	26.40	767.10	20.27
白水	330.69	9.73	449.49	14.08
富平	716.17	16.65	645.60	13.53

保留了原始变量的全部信息, 在大多实际研究工作中, 选取主成分的个数有多种方法, 常使用的原则是主成分的特征值大于 1 或累计贡献率大于 85%。它的优势在于计算简单, 且对多数情况适用。但此方法没有理论支持, 属于经验性方法。基于此, 本文首先根据主成分的累计贡献率不小于 85% 的选取原则, 在此基础上尝试令结合 Copula 获得的综合长势监测指标包含不同的主成分个数, 对比它们与冬小麦单产之间的拟合效果。

综合 LAI、综合 VTCI 和 LAI 均包含第 1 个主成分时与单产间的拟合效果更好, 这是由于 LAI 的第 2 个主成分表达式、VTCI 和 LAI 的后 2 个主成分表达式中存在载荷量较大的负系数。VTCI 和 LAI 均与冬小麦的长势与产量密切相关, 两者的值越大, 作物长势越好、产量超高, 因此负系数的出现不符合实际情况, 加入后易导致精度降低。这一规律表明, 开展多指标的主成分分析时, 不仅要考虑主成分的累计贡献率, 而且更要考虑各指标的系数与作物长势

间的关系。

利用综合 LAI 构建的估产模型虽然模型的拟合精度较高,但只考虑单一指标对冬小麦产量进行估测结果存在偏差。具体到每一个县时,利用双变量指标综合 G-1 构建的冬小麦单产模型得到的估测单产误差范围比基于单变量构建的模型更集中,结果更加可靠。研究成果表明,使用 BP 神经网络和 IPSO-BP 神经网络构建双变量(VTCI 和 LAI)冬小麦估产模型的 R^2 分别为 0.310、0.342^[30]。相比之下,利用综合 G-1 构建估产模型的 R^2 更高,进一步说明基于 PCA-Copula 方法构建的双变量估产模型精度较高,可用于研究区域冬小麦单产估测。

4 结论

(1) 采用 PCA-Copula 法分别对关中平原 2012—2017 年 24 个县(区)冬小麦主要生育时期单变量(VTCI 或 LAI)、双变量(VTCI 和 LAI)进行处理,获得综合 VTCI、综合 LAI、综合 VTCI 和 LAI。结

果表明,综合 VTCI 包含前 2 个主成分时与冬小麦单产的拟合精度优于包含第 1 个主成分的情况;综合 LAI 包含第 1 个主成分时与单产的拟合精度高于综合 LAI 包含前 2 个主成分;综合 VTCI 和 LAI 包含第 1 个主成分时比包含前 2 个、3 个主成分时与单产间的拟合效果好。

(2) 综合 LAI-1 与冬小麦单产间的回归模型拟合效果最好($R^2 = 0.567, P < 0.001$),综合 G-1 与冬小麦单产间的回归模型略次之($R^2 = 0.524, P < 0.001$),综合 VTCI-2 构建的估产模型拟合效果相对较差($R^2 = 0.246, P < 0.001$)。具体到各县(区)的单产估测时,尽管综合 LAI-1 与单产间的相关性高于综合 G-1 与单产间的相关性,但基于综合 LAI-1 的估测单产相比于基于综合 G-1 的估测单产与实际单产之间的误差范围更大,说明基于单一 LAI 构建的综合长势监测指标反映的冬小麦长势与产量信息不够全面。因此,基于双变量综合 VTCI 和 LAI 构建的冬小麦估产模型更加全面可靠。

参 考 文 献

- [1] 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. 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)
- [2] 任建强, 陈仲新, 周清波, 等. MODIS 植被指数的美国玉米单产遥感估测[J]. 遥感学报, 2015, 19(4): 568-577. REN Jianqiang, CHEN Zhongxin, ZHOU Qingbo, et al. MODIS vegetation index data used for estimating corn yield in USA[J]. Journal of Remote Sensing, 2015, 19(4): 568-577. (in Chinese)
- [3] 郭锐, 朱秀芳, 李石波, 等. 山东省冬小麦单产监测与预报方法研究[J/OL]. 农业机械学报, 2020, 51(7): 156-163. GUO Rui, ZHU Xiufang, LI Shibo, et al. Monitoring and forecasting method of winter wheat yield in Shandong Province[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(7): 156-163. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20200718&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2020.07.018. (in Chinese)
- [4] SIYAL A A, DEMPEWOLF J, RESHEF I B. Rice yield estimation using Landsat ETM+ Data[J]. Journal of Applied Remote Sensing, 2015, 9(1): 095986.
- [5] 杨邦杰, 裴志远. 农作物长势的定义与遥感监测[J]. 农业工程学报, 1999, 15(3): 214-218. YANG Bangjie, PEI Zhiyuan. Definition of crop condition and crop monitoring using remote sensing[J]. Transactions of the CSAE, 1999, 15(3): 214-218. (in Chinese)
- [6] 王蕾, 王鹏新, 李俐, 等. 河北省中部平原玉米长势遥感综合监测[J]. 资源科学, 2018, 40(10): 201-211. WANG Lei, WANG Pengxin, LI Li, et al. Integrated maize growth monitoring based on gray correlation analysis and remote sense data in the central plain of Hebei Province[J]. Resources Science, 2018, 40(10): 201-211. (in Chinese)
- [7] 王鹏新, 龚健雅, 李小文, 等. 基于植被指数和土地表面温度的干旱监测模型[J]. 地球科学进展, 2003, 18(4): 527-533. WANG Pengxin, GONG Jianya, LI Xiaowen, et al. Advances in drought monitoring by using remotely sensed normalized difference vegetation index and land surface temperature products[J]. Advances in Earth Science, 2003, 18(4): 527-533. (in Chinese)
- [8] 黄健熙, 李昕璐, 刘帝佑, 等. 顺序同化不同时空分辨率 LAI 的冬小麦估产对比研究[J/OL]. 农业机械学报, 2015, 46(1): 240-248. HUANG Jianxi, LI Xinlu, LIU Diyou, et al. Comparison of winter wheat yield estimation by sequential assimilation of different spatio-temporal resolution remotely sensed LAI datasets[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(1): 240-248. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20150134&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2015.01.034. (in Chinese)
- [9] 沈婕, 梁忠民, 胡义明, 等. Copula-MCP 洪水概率预报方法研究与应用[J]. 中国农村水利水电, 2020(1): 125-129. SHEN Jie, LIANG Zhongmin, HU Yiming, et al. Research and application of Copula-MCP model in flood probabilistic forecasting[J]. China Rural Water and Hydropower, 2020(1): 125-129. (in Chinese)
- [10] DEGHANI M, ZARGAR M, MADVAR H R, et al. A novel approach for longitudinal dispersion coefficient estimation via tri-variate Archimedean Copulas[J]. Journal of Hydrology, 2020, 584(2): 124662.
- [11] VERGNI L, TODISCO F, MANNOCCHI F. Analysis of agricultural drought characteristics through a two-dimensional Copula[J]. Water Resources Management, 2015, 29(8): 2819-2835.
- [12] 张迎, 黄生志, 黄强, 等. 基于 Copula 函数的新型综合干旱指数构建与应用[J]. 水利学报, 2018, 49(6): 703-714.

- ZHANG Ying, HUANG Shengzhi, HUANG Qiang, et al. Construction and application of a new comprehensive drought index based on Copula function[J]. *Journal of Hydraulic Engineering*, 2018, 49(6): 703–714. (in Chinese)
- [13] 王鹏新, 冯明悦, 孙辉涛, 等. 基于主成分分析和 Copula 函数的干旱影响评估研究[J/OL]. *农业机械学报*, 2016, 47(9): 334–340.
WANG Pengxin, FENG Mingyue, SUN Huitao, et al. Drought impact assessment based on principal component analysis and Copula function[J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2016, 47(9): 334–340. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20160945&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2016.09.045. (in Chinese)
- [14] 王鹏新, 冯明悦, 孙辉涛, 等. 基于非线性特征的干旱影响评估研究[J/OL]. *农业机械学报*, 2016, 47(10): 325–331.
WANG Pengxin, FENG Mingyue, SUN Huitao, et al. Drought impact assessment based on nonlinear characteristics of drought [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2016, 47(10): 325–331. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20161041&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2016.10.041. (in Chinese)
- [15] INES A V, DAS N N, HANSEN J W, et al. Assimilation of remotely sensed soil moisture and vegetation with a crop simulation model for maize yield prediction[J]. *Remote Sensing of Environment*, 2013, 138: 149–164.
- [16] 李艳, 王鹏新, 刘峻明, 等. 基于条件植被温度指数的冬小麦主要生育时期干旱监测效果评价Ⅲ——干旱对冬小麦产量的影响评估[J]. *干旱地区农业研究*, 2014, 32(5): 218–222.
LI Yan, WANG Pengxin, LIU Junming, et al. Evaluation of drought monitoring effects in the main growth and development stages of winter wheat using vegetation temperature condition index Ⅲ. — Impact evaluation of drought on wheat yield[J]. *Agricultural Research in the Arid Areas*, 2014, 32(5): 218–222. (in Chinese)
- [17] 彭珂珊. 陕西农业综合发展中的干旱基本规律研究[J]. *中国农业资源与区划*, 1996(5): 24–27.
- [18] 李红梅, 范建忠, 李星敏. 近 51 a 来陕西农业干旱特征分析[J]. *陕西气象*, 2013(6): 9–12.
- [19] 王蕾, 王鹏新, 李俐, 等. 应用条件植被温度指数预测县域尺度小麦单产[J]. *武汉大学学报(信息科学版)*, 2018, 43(10): 1566–1573.
WANG Lei, WANG Pengxin, LI Li, et al. Wheat yield forecasting at county scale based on time series vegetation temperature condition index[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(10): 1566–1573. (in Chinese)
- [20] 王鹏新, 龚健雅, 李小文. 条件植被温度指数及其在干旱监测中的应用[J]. *武汉大学学报(信息科学版)*, 2001, 26(5): 412–418.
WANG Pengxin, GONG Jianya, LI Xiaowen. Vegetation temperature condition index and its application for drought monitoring [J]. *Geomatics and Information Science of Wuhan University*, 2001, 26(5): 412–418. (in Chinese)
- [21] 孙威, 王鹏新, 韩丽娟, 等. 条件植被温度指数干旱监测方法的完善[J]. *农业工程学报*, 2006, 22(2): 22–26.
SUN Wei, WANG Pengxin, HAN Lijuan, et al. Further improvement of the approach to monitoring drought using vegetation and temperature condition indexes from multi-years remotely sensed data[J]. *Transactions of the CSAE*, 2006, 22(2): 22–26. (in Chinese)
- [22] 李艳, 王鹏新, 刘峻明, 等. 基于条件植被温度指数的冬小麦主要生育时期干旱监测效果评价——I. 因子权重排序法和熵值法组合赋权[J]. *干旱地区农业研究*, 2013, 31(6): 159–163.
LI Yan, WANG Pengxin, LIU Junming, et al. Application of temperature condition index to evaluate the drought monitoring effect in main growing period of winter wheat: — I. Factor weight sorting method and entropy method[J]. *Agricultural Research in the Arid Areas*, 2013, 31(6): 159–163. (in Chinese)
- [23] 李艳, 王鹏新, 刘峻明, 等. 基于条件植被温度指数的冬小麦主要生育时期干旱监测效果评价——II. 改进层次分析法和变异系数法组合赋权[J]. *干旱地区农业研究*, 2014, 32(1): 236–239, 262.
LI Yan, WANG Pengxin, LIU Junming, et al. Application of temperature condition index to evaluate the drought monitoring effect in main growing period of winter wheat: — II. Improved analytic hierarchy process and variation coefficient method[J]. *Agricultural Research in the Arid Areas*, 2014, 32(1): 236–239, 262. (in Chinese)
- [24] 王鹏新, 苟兰, 李俐, 等. 基于时间序列叶面积指数傅里叶变换的作物种植区域提取[J]. *农业工程学报*, 2017, 33(21): 207–215.
WANG Pengxin, XUN Lan, LI Li, et al. Extraction of planting areas of main crops based on Fourier transformed characteristics of time series leaf area index products [J]. *Transactions of the CSAE*, 2017, 33(21): 207–215. (in Chinese)
- [25] NELSEN R B. An introduction to Copulas[M]. New York: Springer, 2006.
- [26] 李浩鑫, 邵东国, 尹希, 等. 基于主成分分析和 Copula 函数的灌溉用水效率评价方法[J]. *农业工程学报*, 2015, 31(11): 96–102.
LI Haoxin, SHAO Dongguo, YIN Xi, et al. Evaluation method for irrigation-water use efficiency based on principle component analysis and Copula function[J]. *Transactions of the CSAE*, 2015, 31(11): 96–102. (in Chinese)
- [27] KAO S C, GOVINDARAJU R S. A Copula-based joint deficit index for droughts[J]. *Journal of Hydrology*, 2010, 380(1–2): 121–134.
- [28] GRINGORTEN I I. A plotting rule for extreme probability paper[J]. *Journal of Geophysical Research*, 1963, 68(3): 813–814.
- [29] 岳文俊, 张富仓, 李志军, 等. 返青期水分胁迫、复水和施肥对冬小麦生长及产量的影响[J]. *西北农林科技大学学报(自然科学版)*, 2012, 40(2): 57–63, 78.
YUE Wenjun, ZHANG Fucang, LI Zhijun, et al. Effect of water stress, rewatering and fertilization on growth and yield of winter wheat during seedling establishment[J]. *Journal of Northwest A&F University (Nat. Sci. Ed.)*, 2012, 40(2): 57–63, 78. (in Chinese)
- [30] TIAN H R, WANG P X, TANSEY K, et al. An IPSO–BP neural network for estimating wheat yield using two remotely sensed variables in the Guanzhong Plain, PR China[J]. *Computers and Electronics in Agriculture*, 2020, 169: 105180.