

doi:10.6041/j.issn.1000-1298.2016.09.045

基于主成分分析和 Copula 函数的干旱影响评估研究

王鹏新¹ 冯明悦¹ 孙辉涛¹ 李俐¹ 张树誉² 景毅刚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: 干旱是关中平原主要的农业灾害之一, 准确地评估干旱的影响, 对抗旱减灾及作物稳产具有重要意义。基于关中平原 2008—2013 年冬小麦主要生育期旬尺度的条件植被温度指数 (VTCI) 干旱监测结果, 将 Copula 函数用于评估冬小麦主要生育时期干旱对其产量的影响。针对多元变量导致 Copula 函数参数求解困难的问题, 采用主成分分析法 (PCA) 提取主要生育时期的 VTCI 的主成分因子, 形成新的相互独立的指标, 进而结合 Copula 函数建立 PCA - Copula 法, 确定关中平原主要生育时期的综合 VTCI, 并构建其与冬小麦单产间的线性回归模型, 评估干旱对产量的影响。结果表明, 应用 PCA - Copula 法得到的综合 VTCI 与单产间的相关性达到极显著水平 ($P < 0.001$), 所建回归模型的拟合度与熵值法的结果相比有所提高, 决定系数由 0.39 提高到 0.49, 且对应模型的估测单产与实测单产间的均方根误差较熵值法的结果降低了 30.2 kg/hm^2 , 平均相对误差降低了 0.66%, 表明 PCA - Copula 法能较好地应用于评估冬小麦主要生育时期干旱对其产量的影响。

关键词: 主成分分析; Copula 函数; 干旱影响评估; 综合 VTCI

中图分类号: S127 **文献标识码:** A **文章编号:** 1000-1298(2016)09-0334-07

Drought Impact Assessment Based on Principal Component Analysis and Copula Function

Wang Pengxin¹ Feng Mingyue¹ Sun Huitao¹ Li Li¹ Zhang Shuyu² Jing Yigang²

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

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

Abstract: Drought is one of the most important agricultural disasters in the Guanzhong Plain, China. Assessing the influence of the droughts in the plain accurately can provide reference for drought mitigation and maintaining stable crop yields. Based on remotely sensed vegetation temperature condition index (VTCI) which was calculated at ten-day intervals for monitoring droughts in the years of 2008—2013 in the plain, the Copula function method was used to assess the effect of drought at the main growth stages of winter wheat on the yields. The mutually independent principal factors were extracted from the VTCIs at the main growth stages of winter wheat by using principal component analysis (PCA), overcoming difficulty of parameter estimation for multivariate Copula, and then incorporated into the Copula function to establish a PCA - Copula method. The comprehensive values of VTCIs at the main growth stages were determined by the PCA - Copula method, and then linear regression model between the comprehensive VTCIs and wheat yields was established to assess the effect of drought on the yields. The results showed that the linear correlation coefficient between the wheat yields and comprehensive VTCIs was at the extremely significant level ($P < 0.001$). Compared with the linear regression model based on the entropy value method, the determination coefficient of the model with the PCA - Copula method reached 0.49 from 0.39, which indicated that the fitting degree of the model was improved, and the root mean square error and average relative error between the estimated and measured yields reduced by 30.2 kg/hm^2 and

收稿日期: 2016-02-29 修回日期: 2016-03-28

基金项目: 国家自然科学基金项目(41371390)

作者简介: 王鹏新(1965—), 男, 教授, 博士生导师, 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

0.66%, respectively. These results indicated that the PCA - Copula method was a better approach for accessing the impact of droughts at the main growth stages of winter wheat on the yield.

Key words: principal component analysis; Copula function; impact assessment of drought; comprehensive vegetation temperature condition index

引言

干旱一直是制约农业发展最主要的因素之一。中国是一个农业大国,干旱对农业的影响较大,每年由各种自然灾害造成的粮食作物减产的损失中,旱灾的影响占一半以上^[1]。同时农业干旱是影响面积最广、造成农业损失最大、研究也最为复杂的一种因素。近年来,人们越来越关注如何有效地监测和评估干旱影响,从而提前采取有效的抗旱减灾策略和措施以减轻干旱灾害可能造成的损失。目前,国内外主要应用归一化植被指数(NDVI)和地表温度(LST)等参数进行农业干旱遥感监测,王鹏新等^[2-3]在 NDVI 和 LST 的散点图呈三角形区域分布的基础上,提出了基于条件植被温度指数(VTCI)的干旱监测方法,并成功应用于干旱的实时监测、预测和影响评估等研究^[4-7]。

以往的干旱影响评估研究方法主要依靠主观赋权法和客观赋权法^[6-7]。主观赋权法具有一定的主观色彩,受一定的人为因素影响;客观赋权法根据各指标的初始信息量来确定权重,其中,主成分分析(Principal component analysis, PCA)方法能够通过线性变换从多元随机变量中提取出相互独立的少数几个重要变量^[8-11]。本文对冬小麦越冬后4个生育时期的 VTCI 进行降维处理,便于与 Copula 函数相结合构建干旱影响评估模型。

Copula 函数不受单变量服从何种边缘分布的限制,可以通过边缘分布和相关性结构两部分来构造多元随机变量之间的联合分布^[12],国内外学者将 Copula 函数广泛应用于描述多种随机变量间的相关性研究^[13-15]。干旱是一种包含多种随机变量的复杂事件,传统的多变量频率分析方法在干旱研究中受到限制,Copula 函数正是构建多元随机变量联合分布的一种有效途径。SONG 等^[16]基于 Meta-elliptical、GH、AMH、Frank 和 Clayton 等连接函数研究了干旱历时、干旱烈度和干旱时间间隔的联合分布。MISHRA 等^[17]通过 Copula 函数研究了干旱的持续时间和严重程度之间的关系,进而分析地表和地下干旱在不同时间尺度的特征。建立 Copula 函数的关键是确定参数,常用的参数求解方法包括极大似然法、相关性指标法、边际函数推断法等^[18]。但是,针对评估时构建的多元随机变量间的联合分

布,参数计算一般需要建立多元偏微分方程组,求解非常困难。本文基于主成分分析方法对关中平原 2008—2013 年冬小麦主要生育时期的 VTCI 进行降维处理,形成相互独立的指标,采用 Copula 函数构建 PCA - Copula 评估模型,便于获取冬小麦主要生育时期的综合 VTCI 值,并结合此综合值评估冬小麦主要生育时期的干旱监测效果,为评价冬小麦主要生育期干旱对单产的影响和制定抗旱减灾策略提供科学依据。

1 材料与方 法

1.1 研究区域概况

陕西省中部的关中平原位于渭河流域,西起宝鸡大散关,东至渭南潼关,北到陕北黄土高原,南止秦岭,地理坐标为 106°22' ~ 110°24' E, 33°57' ~ 35°39' N。该地区土壤肥沃,地势平坦,水源丰富,机耕条件好,土地利用率高,盛产小麦,是陕西省的农业基地,也是我国重要的商品粮产区,种植模式主要为冬小麦与夏玉米轮作。关中平原属大陆性季风半湿润气候区,是气候变化的敏感区,年均温度 6 ~ 13℃,属于生态环境脆弱地带,降水量较少,年平均降水量为 500 ~ 700 mm,多集中在夏季,并存在明显的波动性。关中平原整体上气候暖干化特征显著,干旱已成为研究区域内普遍而重大的气象灾害之一^[19]。

1.2 试验数据

采用的遥感数据包括 Aqua - MODIS 的日地表温度产品 (MYD11A1) 和日地表反射率产品 (MYD09GA),所用的冬小麦单产来自陕西省统计局发布的关中平原 5 市 2008—2013 年的数据。由于铜川市位于关中平原向陕北黄土高原的过渡地带,冬小麦面积相对较小,且主要分布在其南部的渭北旱塬,因此选用关中平原 2008—2013 年其余 4 市的冬小麦单产和主要生育期的 VTCI 进行相关研究。

VTCI 是基于遥感反演的 NDVI 和 LST 特征空间呈三角形区域分布的特点提出的,主要用于监测旱情。基于日 NDVI 和日 LST,应用最大值合成技术分别生成旬 NDVI 和旬 LST 最大值合成产品,并以此计算 VTCI^[2,20],生成 2008—2013 年每年 3—5 月份以旬为单位的关中平原 4 市的 VTCI 数据。

结合关中平原冬小麦的生长情况,将冬小麦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),并将这4个生育时期称为冬小麦主要生育期^[6-7]。依据关中平原的行政边界图,取各市区内所包含像素的 VTCI 平均值作为该区域该年该旬的 VTCI 值。再根据各生育时期包含的多旬 VTCI 平均值作为该年该生育时期的 VTCI 值。

1.3 PCA - Copula 法

主成分分析又称主分量分析,是一种将多个变量通过线性变换选出几个重要变量的多元统计方法,旨在利用降维的思想,将原来数量较多的、具有一定相关性的变量,重新组合成一组新的、互不相关的综合指标,使新变量尽可能多地保留原始变量信息,且在保证主要信息的前提下,避开变量之间的线性相关^[21]。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。本文利用主成分分析方法将关中平原冬小麦4个生育时期的 VTCI 组合成一组新的互相独立的指标,再基于 Copula 函数进行干旱影响评估研究。

1.3.1 主成分分析

由于 VTCI 的取值范围为 $[0,1]$,因此直接对原始 VTCI 数据进行主成分分析。应用关中平原2008—2013年冬小麦越冬后4个生育时期的 VTCI 数据构建矩阵 $A_{n \times 4}$,获得相关系数矩阵 $R_{4 \times 4}$,并计算特征值 $\lambda_i (i = 1, 2, 3, 4)$ 及其对应的特征向量 $Q(q_i)$,进而求解各主成分贡献率 s_i 并确定主成分个数 m ,即

$$s_i = \lambda_i / \sum_{i=1}^4 \lambda_i \times 100\% \quad (1)$$

主成分的贡献率越大,对应的主成分反映的信息就越多,如果前 m 个主成分的累计贡献率 $\sum_{i=1}^m \lambda_i / \sum_{i=1}^4 \lambda_i \times 100\% > 80\%$,则可以选取 m 个因子作为主成分。

第 r 个主成分的线性表达式为

$$F_{nr} = \sum_{i=1}^k \beta_{ri} a_{ni} \quad (2)$$

式中 F_{nr} ——关中平原4市第 n 年第 r 个主成分的因子值

β_{ri} ——第 i 个生育时期的 VTCI 在第 r 个主成分线性组合中的系数

a_{ni} ——关中平原4市第 n 年第 i 个生育时期对应的 VTCI 值

k ——VTCI 的主要生育期个数,取 $k = 4$

1.3.2 Copula 函数

SKLAR 在 1959 年指出可以将一个 g 元联合分布函数分解为 g 个边缘分布函数和一个 Copula 函数,其中的 Copula 函数可以用来描述变量间的相关关系^[22]。Copula 是在 $[0,1]$ 区间上服从均匀分布的联合分布函数,可以通过边缘分布和相关性结构两部分来构造联合分布以描述变量间的相依性^[12]。Copula 函数形式为

$$F(x_1, x_2, \dots, x_j) = C_\theta(F_1(x_1), F_2(x_2), \dots, F_j(x_j)) = C(u_1, u_2, \dots, u_j) \quad (3)$$

其中

$$u_j = F_j(x_j)$$

式中 F ——随机变量分布函数

j ——样本容量 C ——Copula 函数

θ ——Copula 参数

u_j ——随机变量 X_j 的边缘分布函数

采用 PCA 方法提取 VTCI 数据 m 个主成分因子,各因子之间无相关性,即主成分因子所构建的新变量之间相互独立,则可以得到相应的乘积 Copula (又称为独立 Copula) 为

$$C(u_1, u_2, \dots, u_m) = u_1 u_2 \dots u_m \quad (4)$$

式中 u_m ——第 m 个主成分因子对应的边缘分布函数

1.3.3 边缘分布函数

目前常用的分布线型有皮尔逊 III 型分布、正态分布、Gamma 分布、指数分布、对数分布、广义极值分布等^[11]。其中,针对广义极值分布的研究源于 20 世纪 20 年代, FISHER 等^[23]在对独立同分布的极大值渐进分布进行理论研究时,提出了 3 种极值分布,第 I 型为指数型原始分布,又称 Gumbel 分布;第 II 型为柯西型原始分布,即 Frechet 分布;第 III 型为有界型分布,即 Weibull 分布。JENKINSON^[24]从理论上证明了 3 种分布模型可概括成一个通式,即具有 3 个参数的极值分布函数。后来称极值 I 型、II 型和 III 型分布为广义极值 (Generalized extreme value, GEV) 分布,其标准化分布函数为

$$F_X(x) = P(X < x) =$$

$$\begin{cases} \exp\left(-\left[1 - \varepsilon\left(\frac{x - \mu}{\sigma}\right)\right]^{\frac{1}{\varepsilon}}\right) & (\varepsilon \neq 0) \\ \exp\left(-\exp\left(\frac{x - \mu}{\sigma}\right)\right) & (\varepsilon = 0) \end{cases} \quad (5)$$

式中 ε ——形状参数 μ ——位置参数

σ ——尺度参数

通过形状参数判断极值类型, $\varepsilon = 0$ 为极值 I 型; $\varepsilon > 0$ 为极值 II 型; $\varepsilon < 0$ 为极值 III 型。

1.3.4 边缘分布的优度评价

首先通过 Kolmogorov - Smirnov (K - S) 方法检

验边缘分布函数拟合效果,再通过比较理论频率与经验频率之间的均方根误差(RMSE)以及信息准则值(Akaike information criterion, AIC)优选边缘分布函数,边缘分布优选的原则是 RMSE 值及 AIC 值越小拟合效果越好^[12]。GRINGORTEN^[25]在研究水文极值的无偏估计量时,给出的一维随机分布经验频率计算公式为

$$P_s = \frac{s - 0.44}{N + 0.12} \quad (6)$$

式中 P_s ——累计经验频率,定义为在 N 个观测值中小于第 s 个最小观测值的概率

RMSE 的计算公式为

$$V_{\text{RMSE}} = \sqrt{\frac{1}{t} \sum_{i=1}^t (P_{\text{eir}} - P_{\text{ir}})^2} \quad (7)$$

式中 P_{eir} ——第 i 年第 r 个主成分的数据对应的经验频率值

P_{ir} ——第 i 年第 r 个主成分的数据对应的理论频率值

t ——第 r 个主成分因子中数据的个数,即关中平原 4 市对应的年总数

AIC 的计算公式为

$$V_{\text{AIC}} = t \ln V_{\text{RMSE}} + 2\eta \quad (8)$$

式中 η ——所选模型参数的个数

2 结果与分析

2.1 主要生育时期 VTCI 的主成分分析

通过主成分分析方法对冬小麦主要生育时期的 VTCI 进行降维处理,根据主成分对应的特征根大于 1、且相应累计贡献率达到 80% 以上的原则,提取出主要生育时期的 VTCI 前 3 个主成分因子,并建立各主成分的线性表达式为

$$P_{C1} = 0.502V_{\text{TCI1}} + 0.324V_{\text{TCI2}} + 0.555V_{\text{TCI3}} + 0.579V_{\text{TCI4}} \quad (9)$$

$$P_{C2} = -0.055V_{\text{TCI1}} + 0.927V_{\text{TCI2}} - 0.126V_{\text{TCI3}} - 0.348V_{\text{TCI4}} \quad (10)$$

$$P_{C3} = 0.808V_{\text{TCI1}} - 0.07V_{\text{TCI2}} - 0.547V_{\text{TCI3}} - 0.109V_{\text{TCI4}} \quad (11)$$

式中 P_{C1} 、 P_{C2} 、 P_{C3} ——前 3 个主成分因子

V_{TCI1} 、 V_{TCI2} 、 V_{TCI3} 、 V_{TCI4} ——4 个生育时期的 VTCI 值

可以发现,4 个不同生育时期的 VTCI 在各主成分系数的绝对值存在差异,4 个生育时期中,拔节期 VTCI 在第 1 主成分的系数最小,而拔节期 VTCI 在第 2 主成分的系数达到最大,返青期 VTCI 与抽穗-灌浆期 VTCI 在第 3 主成分的系数绝对值较大。基于此,根据各主成分的方差贡献率获取前 3 个主成

分的综合线性表达式为

$$P_c = 0.468V_{\text{TCI1}} + 0.406V_{\text{TCI2}} + 0.328V_{\text{TCI3}} + 0.359V_{\text{TCI4}} \quad (12)$$

式中 P_c ——前 3 个主成分因子的综合值

可以看出,返青期 VTCI 的系数最大,而抽穗-灌浆期 VTCI 的系数最低,但依据相关的农学知识和专家的主观判断可知,拔节期是干旱对冬小麦生长影响的关键时期,其次为抽穗-灌浆期,返青期和乳熟期相对不重要^[6-7],因此仅采用主成分方法分析主要生育时期的 VTCI,并基于前 3 个主成分的综合线性模型存在不足,该模型的构建设没有考虑 3 个主成分因子各自的分布特征,仅是将 3 个主成分因子进行综合,且获取的各生育时期干旱对产量的影响程度与农学先验知识不相符,难以准确反映冬小麦生长各生育时期的相对重要程度。而将主成分分析与 Copula 函数相结合的方法能够根据各主成分因子对应的分布特征建立联合分布,同时不受各单因子变量边缘分布的影响,因此尝试将其用于干旱影响评估研究。

2.2 基于 PCA - Copula 法的干旱影响评估

对于 2008—2013 年研究区域的 VTCI 数据,通过主成分分析法提取了 3 个因子作为主成分因子,通过确定 3 个主成分因子的边缘分布函数,采用乘积 Copula(式(4))获取研究区每年的综合 VTCI 值。

2.2.1 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、对数分布和广义极值分布分别拟合关中平原冬小麦 4 个生育时期的 VTCI 的主成分因子值,应用极大似然法获得对应的边缘分布函数的参数,利用 K-S 检验对各主成分因子的边缘分布函数进行拟合优度评价,结果表明,3 个主成分的对数分布均不能通过 K-S 检验,广义极值分布和正态分布均能通过 K-S 检验,其中,利用广义极值分布拟合各生育时期的 VTCI 的前 3 个主成分因子时,前 2 个主成分的形状参数小于零,属于极值 III 型,即 Weibull 分布;第 3 个主成分的形状参数大于零,属于极值 II 型,即 Frechet 分布。

基于正态分布及各主成分相对应的极值分布类型,进一步获取各边缘分布对应的 RMSE 和 AIC (表 1),可以看出,基于 Weibull 分布获取的第 1 个主成分的 RMSE 和 AIC 最小;基于正态分布获取的第 2 个主成分的 RMSE 和 AIC 最小;基于 Frechet 分布获取的第 3 个主成分的 RMSE 和 AIC 最小。依据边缘分布的 RMSE 及 AIC 越小拟合效果越好的优选原则,最终确定 2008—2013 年关中平原 4 市主要生育时期的 VTCI 的 3 个主成分因子值的最优边缘

分布。结果表明,第1主成分因子值优选 Weibull 分布;第2主成分因子值优选正态分布;第3主成分因子值优选 Frechet 分布。尽管3个主成分的边缘分

布存在差异,但 Copula 函数具有不受单变量服从何种边缘分布的优越性,使其能较好地用于构建3个主成分间的联合分布。

表1 冬小麦主要生育时期的 VTCI 主成分因子的分布线型拟合效果

Tab.1 Distribution fitting results of principle components of VTCI at the main growth stages of winter wheat

主成分1			主成分2			主成分3		
边缘分布	RMSE	AIC	边缘分布	RMSE	AIC	边缘分布	RMSE	AIC
正态分布	0.07	-23.79	正态分布	0.04	-30.03	正态分布	0.09	-19.22
Weibull	0.03	-32.09	Weibull	0.05	-28.21	Frechet	0.05	-26.47

2.2.2 PCA-Copula 法与熵值法建立的综合 VTCI 与冬小麦单产间的线性回归模型比较

基于冬小麦主要生育时期的 VTCI 的3个主成分因子优选的边缘分布,结合乘积 Copula 函数建立主成分因子间的连接函数,获取综合 VTCI,建立其与单产间的一元线性回归模型,并将其作为冬小麦单产的估测模型。以往研究^[6]表明,熵值法是干旱影响评估的最佳客观赋权法,通过熵值法确定冬小麦不同生育时期的 VTCI 权重,并将其用于建立的加权 VTCI 与单产间的线性回归模型,再与 PCA-Copula 法建立的模型的结果进行对比(表2)。可以看出,尽管 PCA-Copula 法确定的综合 VTCI、熵值法确定的加权 VTCI 与单产间的相关性均达到极显著水平($P < 0.01$),但基于 PCA-Copula 法构建的回归模型的决定系数($R^2 = 0.49$)高于熵值法确定的回归模型的决定系数($R^2 = 0.39$)。基于熵值法确定回归模型的估测单产与实测单产之间的均方根误差为 358.1 kg/hm^2 ,相对误差的范围为 $0.39\% \sim 28.13\%$,平均相对误差为 7.10% ,而 PCA-Copula 法确定回归模型的估测单产与实测单产的均方根误差为 327.9 kg/hm^2 ,相对误差范围为 $0.33\% \sim 25.20\%$,平均相对误差为 6.44% ,表明 PCA-Copula 法确定的模型估测精度较高。为了进一步验证该估测模型的精度,将 PCA-Copula 法确定的估产模型用于分析 2008—2013 年 4 市估测单产与实际单产的平均相对误差,结果表明,宝鸡市的平均相对误差为 6.21% ,西安市的平均相对误差为 5.74% ,咸阳市的平均相对误差为 4.39% ,渭南市的平均相对误差为 9.44% ,表明 PCA-Copula 法可用于开展关中平原干旱对冬小麦生产及其产量的影响评估研究。

3 讨论

遥感定量反演的 VTCI 是综合地表主要参数——NDVI 和 LST 的干旱指标,其与土壤浅层水分存在较强的相关性^[26],能够较为准确地监测干

表2 PCA-Copula 法与熵值法确定的综合 VTCI 与单产之间的回归模型的对比

Tab.2 Comparison of linear regression models between wheat yields and comprehensive values of the VTCIs determined by PCA-Copula and entropy value methods

方法	回归模型	决定系数	显著性检验
PCA-Copula 法	$y = 3\ 065.9x + 3\ 960.0$	0.49	$P < 0.001$
熵值法	$y = 3\ 334.2x + 2\ 194.9$	0.39	$P = 0.001$

注:回归模型中 x 为综合 VTCI 值; y 为单产,单位为 kg/hm^2 。

旱、反映作物水分胁迫信息,作物长势和最终产量与其不同生育期发生的干旱程度密切相关。相比传统的多变量频率分析方法,Copula 函数作为一种相关性度量工具,在构建多元随机变量联合分布时不受各单因子变量边缘分布的影响。本文基于冬小麦越冬后不同生育时期 VTCI 的干旱监测结果,尝试将 Copula 函数应用于建立冬小麦主要生育时期的干旱影响评估模型,若直接建立 Copula 函数与冬小麦单产间的线性回归模型,首先需要确定4个生育时期 VTCI 变量对应的最优边缘分布,由于不同形式的 Copula 函数对变量的相关性有不同的要求,因而需要根据各变量之间的相关关系特点选择可行的四维 Copula 模型,并基于所选 Copula 函数的密度函数和各边缘分布的密度函数获取 Copula 函数的参数,对于所选 Copula 函数是否合适还需进一步的分布拟合检验,确定最优 Copula 函数,从而获取综合 VTCI。然而利用主成分分析法具有提取主成分因子的客观性,在提取的主要生育时期的 VTCI 的主成分因子中,所选的前3个主成分因子的累计贡献率高达 98% ,新变量不仅几乎保留了全部原始变量信息,而且可以直接通过用于独立变量的乘积 Copula 获取综合 VTCI,在建立联合分布过程中仅需优选3个主成分因子的边缘分布,无需求解参数,计算简便。

现有的干旱影响评估研究方法主要有主观赋权法和客观赋权法2类,其权重的确定对干旱影响评估结果均有很大影响。针对以往客观赋权法根据 VTCI 的客观数据差异来确定权重、得到不同生育时

期干旱对冬小麦生长影响程度相当的结论与农学先验知识不相符^[6-7]的问题,本文采用主成分分析方法提取出冬小麦越冬后主要生育时期的 VTCI 的前 3 个主成分因子作为新的相互独立的变量,而根据各主成分方差贡献率获取的综合线性模型反映的各生育时期相对重要程度与农学先验知识仍然不相符。因此,尝试基于获取的 3 个主成分因子,通过构造各主成分因子的边缘分布,进而采用 Copula 函数建立的 PCA-Copula 法,最终获取综合 VTCI 值,并建立其与单产之间的线性回归模型,结果表明,PCA-Copula 法能较好地用于开展关中平原干旱对冬小麦生产及其产量的影响评估研究。

4 结论

(1) 采用主成分分析法提取关中平原 2008—2013 年 4 市冬小麦主要生育时期 VTCI 的 3 个因子

作为主成分,将依据各主成分贡献率获取的前 3 个主成分的综合线性模型与 PCA-Copula 法的结果进行对比,结果表明,基于前 3 个主成分,仅通过各主成分的方差贡献率建立的综合线性模型难以准确反映冬小麦生长相对重要生育时期,而通过优选各主成分相应的边缘分布,进而通过乘积 Copula 函数建立 3 个主成分联合分布的方法能有效评估关中平原干旱对冬小麦产量的影响。

(2) 基于熵值法建立加权 VTCI 与单产之间的线性回归模型,并将其与 PCA-Copula 法构建的线性回归模型进行对比分析。结果表明,采用 PCA-Copula 法确定的综合 VTCI 与单产之间回归模型的决定系数达到 0.49,相比于熵值法的结果提高了 0.10,且 PCA-Copula 法对应模型的估测单产与实测单产之间的均方根误差较熵值法的结果降低了 30.2 kg/hm²,平均相对误差降低了 0.66%。

参 考 文 献

- 1 史舟,梁宗正,杨媛媛,等. 农业遥感研究现状与展望[J]. 农业机械学报, 2015, 46(2): 247-260.
SHI Zhou, LIANG Zongzheng, YANG Yuanyuan, et al. Status and prospect of agricultural remote sensing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2015, 46(2): 247-260. (in Chinese)
- 2 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)
- 3 王鹏新, WAN Zhengming, 龚健雅, 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, WAN Zhengming, GONG Jianya, 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)
- 4 陈阳,范建容,郭芬芬,等. 条件植被温度指数在云南干旱监测中的应用[J]. 农业工程学报, 2011, 27(5): 231-236.
CHEN Yang, FAN Jianrong, GUO Fenfen, et al. Application of the vegetation temperature condition index to drought monitoring in Yunnan Province[J]. Transactions of CSAE, 2011, 27(5): 231-236. (in Chinese)
- 5 田苗,王鹏新,韩萍,等. 基于 SARIMA 模型和条件植被温度指数的干旱预测[J]. 农业机械学报, 2013, 44(2): 109-116.
TIAN Miao, WANG Pengxin, HAN Ping, et al. Drought forecasts based on SARIMA models and vegetation temperature condition index[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(2): 109-116. (in Chinese)
- 6 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)
- 7 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)
- 8 HAO Z C, SINGH V P. Drought characterization from a multivariate perspective: A review[J]. Journal of Hydrology, 2015, 527: 668-678.
- 9 李靖华,郭耀煌. 主成分分析用于多指标评价的方法研究——主成分评价[J]. 管理工程学报, 2002, 16(1): 39-43.
LI Jinghua, GUO Yaohuang. Principal component evaluation; a multivariate evaluate method expanded from principal component analysis[J]. Journal of Industrial Engineering and Engineering Management, 2002, 16(1): 39-43. (in Chinese)

- 10 刘春景,唐敦兵,何华,等. 基于灰色关联和主成分分析的车削加工多目标优化[J]. 农业机械学报, 2013, 44(4): 293–298.
LIU Chunjing, TANG Dunbing, HE Hua, et al. Multi-objective optimization of turning based on grey relational and principal component analysis[J]. Transactions of the Chinese Society for Agricultural Machinery, 2013, 44(4): 293–298. (in Chinese)
- 11 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 CSAE, 2015, 31(11): 96–102. (in Chinese)
- 12 KAO S C, GOVINDARAJU R S. A copula-based joint deficit index for droughts[J]. Journal of Hydrology, 2010, 380(1): 121–134.
- 13 INDU J, KUMAR D N. Copula-based modeling of TMI brightness temperature with rainfall type[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(8): 4832–4845.
- 14 郭爱军,畅建霞,王义民,等. 近 50 年泾河流域降雨-径流关系变化及驱动因素定量分析[J]. 农业工程学报, 2015, 31(14): 165–171.
GUO Aijun, CHANG Jianxia, WANG Yimin, et al. Variation characteristics of rainfall-runoff relationship and driving factors analysis in Jinghe river basin in nearly 50 years[J]. Transactions of CSAE, 2015, 31(14): 165–171. (in Chinese)
- 15 刘招,田智,乔长录,等. 基于 Copula 函数的关中河流水文丰枯遭遇特征分析[J]. 干旱地区农业研究, 2013, 31(4): 245–248.
LIU Zhao, TIAN Zhi, QIAO Changlu, et al. Analysis on coincidence characteristics of high and low stream flow in central Shaanxi based on Copula function[J]. Agricultural Research in the Arid Areas, 2013, 31(4): 245–248. (in Chinese)
- 16 SONG S, SINGH V. Meta-elliptical copulas for drought frequency analysis of periodic hydrologic data [J]. Stochastic Environmental Research and Risk Assessment, 2009, 24(3): 425–444.
- 17 MISHRA A K, INES A V M, DAS N N, et al. Anatomy of a local-scale drought: application of assimilated remote sensing products, crop model, and statistical methods to an agricultural drought study[J]. Journal of Hydrology, 2015, 526: 15–29.
- 18 HAO Z C, SINGH V P. Integrating entropy and Copula theories for hydrologic modeling and analysis [J]. Entropy, 2015, 17(4): 2253–2280.
- 19 王鹏新,孙辉涛,王蕾,等. 基于 4D-VAR 和条件植被温度指数的冬小麦单产估测[J]. 农业机械学报, 2016, 47(3): 263–271.
WANG Pengxin, SUN Huitao, WANG Lei, et al. Winter wheat yield estimation based on 4D variational assimilation method and remotely sensed vegetation temperature condition index[J]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(3): 263–271. (in Chinese)
- 20 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 CSAE, 2006, 22(2): 22–26. (in Chinese)
- 21 李顺峰,张丽华,刘兴华,等. 基于主成分分析的苹果霉心病近红外漫反射光谱判别[J]. 农业机械学报, 2011, 42(10): 158–161.
LI Shunfeng, ZHANG Lihua, LIU Xinghua, et al. Discriminant analysis of apple moldy core using near infrared diffuse reflectance spectroscopy based on principal component analysis [J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(10): 158–161. (in Chinese)
- 22 NELSEN R B. An introduction to copulas[M]. New York: Springer, 2006.
- 23 FISHER R A, TIPPETT L H C. Limiting forms of the frequency distribution of the largest or smallest member of a sample[J]. Proceedings of the Cambridge Philosophical Society, 1928, 24(2): 180–190.
- 24 JENKINSON A F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements[J]. Quarterly Journal of the Royal Meteorological Society, 1955, 81(348): 158–171.
- 25 GRINGORTEN I I. A plotting rule for extreme probability paper[J]. Journal of Geophysical Research, 1963, 68(3): 813–814.
- 26 杨鹤松,王鹏新,孙威. 条件植被温度指数在华北平原干旱监测中的应用[J]. 北京师范大学学报: 自然科学版, 2007, 43(3): 314–318.
YANG Hesong, WANG Pengxin, SUN Wei. Application of the vegetation temperature condition index to drought monitoring in North China Plain[J]. Journal of Beijing Normal University: Natural Science, 2007, 43(3): 314–318. (in Chinese)