



# 利用双变量同化与PCA-Copula法的冬小麦单产估测

王鹏新<sup>1</sup> 陈弛<sup>1</sup> 张悦<sup>1</sup> 张树誉<sup>2</sup> 刘峻明<sup>3</sup>

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,陕西 西安,710014

3 中国农业大学土地科学与技术学院,北京,100193

**摘要:**为了进一步提高冬小麦产量估测的精度,基于集合卡尔曼滤波算法和粒子滤波(particle filter, PF)算法,对CERES-Wheat模型模拟的冬小麦主要生育期条件植被温度指数(vegetation temperature condition index, VTCI)、叶面积指数(leaf area index, LAI)和中分辨率成像光谱仪(moderate-resolution imaging spectroradiometer, MODIS)数据反演的VTCI、LAI进行同化,利用主成分分析与Copula函数结合的方法构建单变量和双变量的综合长势监测指标,建立冬小麦单产估测模型,并通过对比分析选择最优模型,对2017—2020年关中平原的冬小麦单产进行估测。结果表明,单点尺度的同化VTCI、同化LAI均能综合反映MODIS观测值和模型模拟值的变化特征,且PF算法具有更好的同化效果;区域尺度下利用PF算法得到的同化VTCI和LAI所构建的双变量估产模型精度最高,与未同化VTCI和LAI构建的估产模型精度相比,研究区各县(区)的冬小麦估测单产与实际单产的均方根误差降低了56.25 kg/hm<sup>2</sup>,平均相对误差降低了1.51%,表明该模型能有效提高产量估测的精度,应用该模型进行大范围的冬小麦产量估测具有较好的适用性。

**关键词:**同化;PCA-Copula;主成分分析;叶面积指数;条件植被温度指数;作物生长模型;冬小麦估产

中图分类号:P237;TP79

文献标志码:A

小麦是中国最主要的粮食作物之一,及时有效地对小麦进行大区域的长势和产量信息监测有助于优化种植结构、调整区域布局和保障国家的粮食安全<sup>[1]</sup>。作物生长模型根据作物生长发育的规律,逐日动态地模拟单点尺度的作物长势信息<sup>[2]</sup>,但应用在较大规模的尺度上时,在参数的获取以及代表空间信息的能力方面存在限制<sup>[3-4]</sup>。利用遥感技术对地观测能够反映环境因素对作物长势的综合影响,是快速在大范围内获取作物长势信息的有效手段,但遥感数据的准确性和连续性易受混合像元和大气等因素的影响。运用数据同化算法,将遥感观测数据与作物生长模型相结合可充分发挥二者的优势,在不间断时间内提高大范围作物长势监测的准确性,进而提高产量估测的精度<sup>[5-7]</sup>。当前普遍采用的同化算法可分为连续数据同化算法和顺序数据同化算法两类<sup>[8]</sup>。连续同化算法包括三维变分和四维变分算法等,通过构建代价函数的方式迭代并优化模

型的初始参数,在满足动态约束的条件下不断计算并缩小模型模拟值与遥感观测值之间的差距。顺序同化算法又称滤波算法,常用的有集合卡尔曼滤波(ensemble Kalman filter, EnKF)和粒子滤波(particle filter, PF)算法等,这类算法利用遥感观测值更新模型的状态变量,不断驱动模型向前运行,此类同化可增强模型与观测值之间误差信息的使用<sup>[9]</sup>。

叶面积指数(leaf area index, LAI)反映植被对光能的利用率和冠层结构特征,常用于遥感与作物生长模型的同化研究。文献[10]将3种不同的LAI产品与MCWLA-Wheat模型进行同化,结果显示所有的同化LAI均对产量估计结果有显著改善。由于作物生长发育受多种因素的影响,因此相比单一的同化LAI,基于多变量的数据同化策略能够获得准确度更高的作物长势监测结果及产量估测信息,已成为现今农业数据同化的研究趋势。目前基于双变量开展的数据同化研

究成果卓著,文献[11]采用EnKF将遥感观测LAI和土壤水分(soil moisture, SM)与作物模型同化以估测美国爱荷华州的玉米产量,发现在极端潮湿的条件下单独同化LAI的效果最好,在一般条件下,同时同化LAI和SM的效果最好。文献[12]利用EnKF将遥感观测的LAI和SM与WOFOST(world food studies)模型同化,以此估测冬小麦的产量,结果表明,使用同化后的变量进行估产可使估产精度显著提高,且同时同化LAI和SM具有更好的估产效果。在归一化植被指数(normalized difference vegetation index, 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[13]提出基于条件植被温度指数(vegetation temperature condition index, VTCI)的干旱监测方法。在此基础上,文献[14]使用EnKF算法将田间测量的LAI和土壤水分以及Landsat反演的LAI和土壤水分同时同化到CERES(crop environment resource synthesis)-Wheat模型中,结果表明同化更多与产量相关的变量可以有效提高冬小麦产量估测的准确性。

Copula 函数是一种构造联合分布的有效方法,结构简单、对不同类型边缘分布连结效果优良,在多变量干旱分析中具有广泛的应用性<sup>[15]</sup>。建立 Copula 评价方法进行干旱评估的重点在于 Copula 函数的选择和参数的计算,由于涉及多变量不同指标的分析,导致其参数计算较为复杂,不易求解。运用主成分分析(principal component analysis, PCA)对多变量数据进行降维处理后得到的主成分数据相互独立,结合 Copula 函数构建 PCA-Copula 法<sup>[16]</sup>能够有效解决 Copula 参数求解困难的问题。VTCI 单变量与 PCA-Copula 法的结合在干旱影响评估方面已取得较好的结果<sup>[17]</sup>,而基于双变量和同化算法与 PCA-Copula 法相结合的相关研究尚未展开。

VTCI能够反映水分胁迫对作物生长的影响, LAI可以表征作物的长势情况。考虑同化双变量数据可能会获得更准确的作物长势监测与产量估测结果, 本文以关中平原的冬小麦为研究对象, 首先应用EnKF和PF两种同化算法, 对CERES-Wheat模型模拟的VTCI、LAI和中分辨率成像光谱仪(moderate resolution imaging spectroradiometer, MODIS)数据反演的VTCI、LAI进行同化, 以获取能更合理反映冬小麦实际生长状况的VTCI和LAI变化趋势。然后, 利用PCA

与 Copula 函数结合的方法构建综合长势监测指标, 建立冬小麦估产模型, 评价估测结果精度并得到最优估产模型。

## 1 研究区域与方法

## 1.1 研究区域概况

本文研究区域关中平原位于陕西省中部的渭河流域,介于渭北北山和秦岭之间,地势西高东低,地理坐标范围为 $106^{\circ}22'E \sim 110^{\circ}24'E$ , $33^{\circ}57'N \sim 35^{\circ}39'N$ (见图1)。该地气候属于温带大陆性季风气候,四季分明,年均气温为 $6 \sim 13^{\circ}C$ ,降水集中在6月—9月,年均降水量为 $500 \sim 800 mm$ ,土质肥沃,人口密集,自古以来就是农业富庶之地,种植模式主要为冬小麦与夏玉米轮作。其中冬小麦在10月上旬播种,次年6月上旬收割,根据关中平原冬小麦的实际生长情况,将冬小麦的主要生育期划分为返青期(3月上旬—3月中旬)、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—5月下旬)4个生育时期<sup>[18]</sup>。

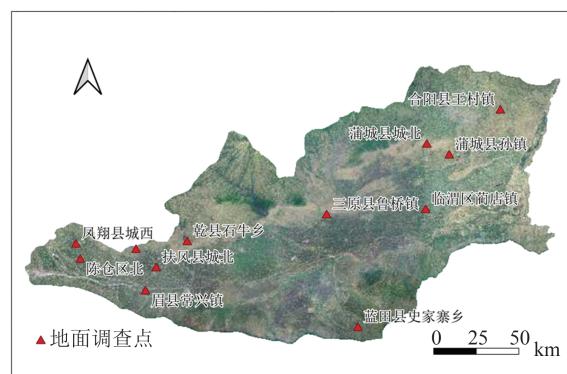


图 1 研究区位置及地面调查点分布

Fig.1 Location of Study Area and Distribution of Ground Sampling Sites

## 1.2 数据获取与处理

### 1.2.1 遥感观测 VTCI 与 LAI

首先,基于关中平原2013—2018年每年3月—5月的MYD09GA产品和MYD11A1产品,获取空间分辨率为1 000 m、时间分辨率为1 d的NDVI和LST。应用最大值合成技术分别生成旬尺度NDVI和LST的最大值合成产品;其次,基于多年某一旬的NDVI和LST最大合成值产品,利用最大值合成技术分别生成多年旬尺度NDVI和LST最大合成值产品;然后,基于每年3月—5月的旬LST最大合成值产品,运用最小值合成技术生成多年旬尺度LST最大-最小值合成产品;最

后,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,并以此计算 VTCI<sup>[13, 19]</sup>为:

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

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

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

式中,  $LST(NDVI_i)$  为某一像素的 NDVI 值为  $NDVI_i$  时的地表温度;  $LST_{\max}(NDVI_i)$ 、 $LST_{\min}(NDVI_i)$  分别表示 NDVI 值等于  $NDVI_i$  的所有像素地表温度的最大值、最小值, 分别为热边界、冷边界;  $a$ 、 $b$ 、 $a'$ 、 $b'$  为由 NDVI 和 LST 的散点图近似获得的待定系数。

取某一生育时期包含的多旬 VTCI 的平均值作为该生育时期的 VTCI 值, 本文依据关中平原各县的行政区域分布图, 取各县域内所包含像素的 VTCI 平均值作为该区域该年该生育时期的 VTCI 值。

基于 2013—2018 年冬小麦主要生育期的 MCD15A3H 产品获取空间分辨率为 500 m、时间分辨率为 4 d 的 LAI, 并利用上包络线 Savitzky-Golay(S-G) 滤波对 LAI 进行平滑处理<sup>[20]</sup>, 经 S-G 滤波处理后的 LAI 消除了原始 LAI 受云和大气影响而产生的噪声, 其 LAI 曲线变化更符合冬小麦的生长规律。逐像素取每旬包含的多时相 LAI 的最大值作为该旬的 LAI 值, 取某一生育时期包含的多旬 LAI 的平均值作为该生育时期的 LAI 值; 依据关中平原各县的行政区域分布图, 取各县域内所包含像素的 LAI 平均值作为该区域该年该生育时期的 LAI 值。由于 VTCI 和 LAI 的空间分辨率不一致, 将 VTCI 遥感影像通过最邻近法重采样至 500 m。

根据地面实验所选样点的经纬度坐标, 定位至 MODIS VTCI、LAI 影像中相应位置处的像素, 取该像素  $3 \times 3$  窗口内像素值的平均值作为单点同化实验所需的样点尺度 MODIS VTCI、LAI。

## 1.2.2 CERES-Wheat 模型模拟的 VTCI 和 LAI

CERES-Wheat 模型中输入的数据主要包括气象数据、土壤数据、田间管理数据和作物品种遗传特性参数。本文选取 2013—2018 年冬小麦主要生育期关中平原 12 个典型的冬小麦种植区为研究样点, 在冬小麦播种前、拔节期、抽穗期和

成熟期分别进行了田间实测与调查实验, 样点的位置分布如图 1 所示。气象数据由与研究样点最邻近的气象站点测得, 主要包括逐日最高和最低温度、降水量、太阳辐射量等; 土壤数据由样点实测获得, 主要包括土壤剖面各层 (0~12 cm、12~20 cm、20~50 cm、50~80 cm、80~120 cm、120~160 cm、160~200 cm) 的水分、养分, 以及其他化学和物理属性等; 田间管理数据为实地调研或实测获得的作物生长模拟所需的管理信息, 如作物品种, 种植、施肥、灌溉情况, 前茬作物种类等。为获取更加准确的模拟结果, 在驱动模型运行前采用“试错法”对作物模型进行区域校准<sup>[21]</sup>。首先根据农业技术转移决策支持系统 (decision support system for agrotechnology transfer, DSSAT) 提供的遗传参数取值赋予各遗传参数初值, 再对比模型模拟值和田间实测值的差距, 逐个调整遗传参数, 直至模型模拟值和实测值能够较好吻合, 符合条件: (1) 检验模拟的冬小麦收获日期与实地调查情况吻合; (2) 检验模拟的冬小麦 LAI 和土壤水分与实测的 LAI 和土壤水分接近<sup>[22]</sup>。

CERES-Wheat 模型无法直接获取模拟 VTCI, 利用文献[23]提出的方法依据 VTCI 与土壤浅层 (0~20 cm) 相对含水量 (relative soil moisture content, SMC) 的关系获取同化实验所需的样点旬尺度的模拟 VTCI。该模型可直接输出以天为步长的模拟 LAI, 取冬小麦生长季内每旬模拟 LAI 数据的最大值作为样点旬尺度模拟 LAI 数据。

## 1.2.3 冬小麦产量数据

本文中涉及的 24 个县(区)的冬小麦产量数据均来源于 2013—2018 年《陕西统计年鉴》, 用于建立冬小麦单产估测模型并评价模型精度。

## 1.3 数据同化方法

### 1.3.1 集合卡尔曼滤波算法

EnKF 算法包含预测和更新两个步骤<sup>[14, 24]</sup>。

1) 预测阶段: 以  $t-1$  时刻的分析值  $x_{t-1}^a$  计算第  $t$  时刻的背景值  $x_t^b$  为:

$$x_t^b = M(x_{t-1}^a) \quad (4)$$

式中,  $x$  为状态变量;  $M$  为 CERES-Wheat 模型算子。

2) 更新阶段: 利用集合变量的平均值  $\bar{x}^b$  计算集合成员的扰动量  $x_t'^b$ , 进而得到扰动集合  $X_t'^b$ , 获取背景场误差协方差  $B$ :

$$x_t'^b = x_t^b - \bar{x}^b \quad (5)$$

$$X_t'^b = (x_1'^b, x_2'^b, \dots, x_t'^b) \quad (6)$$

$$B = \frac{1}{N} X_t^{t,b} (X_t^{t,b})^T \quad (7)$$

计算第  $t$  时刻的卡尔曼增益矩阵  $K$  为:

$$K = B(B + R)^{-1} \quad (8)$$

式中,  $R$  为观测场误差协方差矩阵。

计算第  $t$  时刻的分析值  $x_t^a$  为:

$$x_t^a = x_t^b + K(x_t^m - x_t^b) \quad (9)$$

式中,  $x_t^m$  为第  $t$  时刻观测值。

在  $t+1$  时刻将  $x_t^a$  代入式(4), 进而得到  $t+1$  时刻的分析值  $x_{t+1}^a$ , 重复上述过程, 直至将所有观测数据更新完毕, 此时的分析值即为该时刻的同化值。

### 1.3.2 粒子滤波算法

PF 算法同样分为预测和更新两个阶段<sup>[25-26]</sup>。

1) 预测阶段: 以第  $t$  时刻初始状态变量  $x_t$  计算第  $t+1$  时刻的预测状态  $x_{t+1}^i (i=1, 2, \dots, N)$  为:

$$x_t^i = x_t + \theta_t^i \quad (10)$$

$$x_{t+1}^i = M(x_t^i, \eta_t) \quad (11)$$

式中,  $\theta_t^i$  为符合正态分布的模型随机噪声序列;  $x_t^i$  为系统状态;  $\eta_t$  为模型驱动参数。

将模型状态预测转换为观测预测  $y_{t+1}^i$  为:

$$y_{t+1}^i = H(x_{t+1}^i) + \gamma_{t+1} \quad (12)$$

式中,  $H$  为观测算子;  $\gamma_{t+1}$  为符合正态分布的观测随机噪声序列。

2) 更新阶段: 设初始重要性概率密度  $q(x_{t+1}^i | x_t^i, y_{t+1}) = p(x_{t+1}^i | x_t^i)$ , 其中  $p(x_{t+1}^i | x_t^i)$  表示先验概率分布函数。依据  $t+1$  时刻的  $x_{t+1}^i$ 、遥感观测值  $y_{t+1}$  和设初始重要性概率密度计算第  $t+1$  时刻  $x_{t+1}^i$  的权重。假设预测值与观测值的偏差符合正态分布, 根据正态分布的概率密度计算预测粒子的权重为:

$$w_{t+1}^i = \frac{1}{\sqrt{2\pi O_{t+1}}} \exp\left(-\frac{(x_{t+1}^i - y_{t+1})^2}{2O_{t+1}}\right) \quad (13)$$

式中,  $w_{t+1}^i$  为第  $t+1$  时刻  $x_{t+1}^i$  的权重;  $O_{t+1}$  为遥感观测值的标准差。

对  $w_{t+1}^i$  进行归一化处理, 得到归一化粒子权重  $w_{t+1}^i$ , 计算第  $t+1$  时刻的状态估计值为:

$$x_{t+1} = \sum_{i=1}^N w_{t+1}^i x_{t+1}^i \quad (14)$$

再次进入预测阶段, 重复上述过程, 直至  $t$  等于预期的时间步长为止,  $x_{t+1}$  即为该时刻的同化值。

### 1.4 综合长势监测指标与单产估测模型的构建

关中平原的冬小麦主要分布在宝鸡市中部的陈仓区、凤翔区、岐山县、扶风县, 咸阳市的永寿县、乾县、武功县、兴平市和泾阳县, 西安市的

鄠邑区、蓝田县, 渭南市中部的临渭区、富平县、蒲城县、大荔县等<sup>[27]</sup>。本文选取的 12 个典型冬小麦种植点均位于冬小麦主要分布区域内, 横跨关中平原, 在西部、中部和东部典型的冬小麦种植区内均有采样点分布, 故认为这些样点的区域特性可以代表整个关中平原。将单点尺度的同化 VTCI、同化 LAI 分别与 MODIS VTCI、LAI 建立线性回归方程。通过将单点尺度同化值与 MODIS 观测值的回归关系应用至区域尺度, 得到关中平原每年 3 月—5 月以旬为尺度的同化 VTCI、同化 LAI, 并依据冬小麦生育期的划分, 获取冬小麦各主要生育期的同化 VTCI、同化 LAI。分别将 MODIS 反演的 VTCI、LAI 和采用两种不同同化算法同化的 VTCI、LAI 运用 PCA-Copula 法<sup>[28]</sup> 建立综合长势监测指标。将基于 MODIS VTCI、EnKF VTCI 和 PF VTCI 计算得到的综合长势监测指标分别记为综合 VTCI、综合 VTCI-EnKF 和综合 VTCI-PF; 基于 MODIS LAI、EnKF LAI 和 PF LAI 计算得到的综合长势监测指标分别记为综合 LAI、综合 LAI-EnKF 和综合 LAI-PF; 基于 MODIS VTCI 和 LAI 双变量、EnKF 同化 VTCI 和 LAI 双变量、PF 同化 VTCI 和 LAI 双变量计算得到的综合长势监测指标分别记为综合 G、综合 G-EnKF、综合 G-PF。将上述所列的综合长势监测指标分别与冬小麦单产建立回归模型, 通过对比模型的精度获得最优估产模型。

## 2 实验与分析

### 2.1 CERES-Wheat 模型的标定实验和分析

基于试错法对 CERES-Wheat 模型的遗传参数进行本地化标定, 结果如表 1 所示。采用 2014 年的田间实测数据与模型模拟结果之间的误差评价模型的标定精度, 结果表明标定后的模型模拟精度较高, 所有样点 LAI 的模拟值和实测值之间的均方根误差 (root mean square error, RMSE) 为  $0.63 \text{ m}^2/\text{m}^2$ , 平均相对误差 (mean relative error, MAE) 为  $18.64\%$ , 模拟 LAI 与实测 LAI 之间的散点图如图 2 所示。SMC 的模拟值与实测值之间的 RMSE 为  $0.05 \text{ cm}^3/\text{cm}^3$ , 平均相对误差 (mean relative error, MRE) 为  $8.38\%$ ; 模拟收获日期与实际收获日期的差距均在 7 d 以内。

### 2.2 VTCI 和 LAI 同化结果与分析

将关中平原 2013—2018 年由 CERES-Wheat 模型模拟得到的冬小麦主要生育期内 SMC 的均

表 1 关中平原冬小麦的遗传参数取值

Tab.1 Genetic Coefficients of Winter Wheat in the Guanzhong Plain

| 遗传参数         | 定义               | 标定值 |
|--------------|------------------|-----|
| P1V/d        | 春化作用系数           | 46  |
| P1D/%        | 光周期系数            | 68  |
| P5/(°C·d)    | 灌浆期积温            | 500 |
| G1/g         | 开花期单位冠层重量的籽粒数    | 29  |
| G2/mg        | 最佳条件下标准籽粒质量      | 29  |
| G3/g         | 成熟期非胁迫下单株茎穗标准干质量 | 2   |
| PHINT/(°C·d) | 完成一片叶生长所需积温      | 90  |

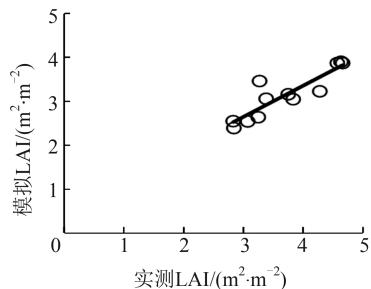


图 2 模拟 LAI 与实测 LAI 之间的散点图

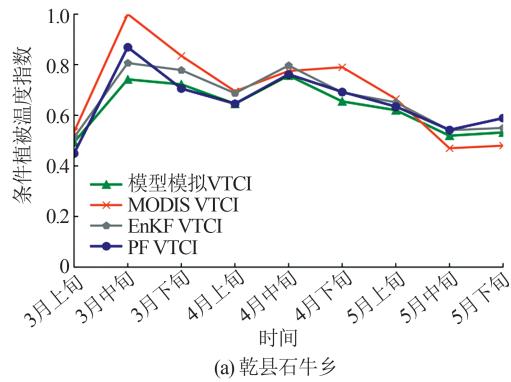
Fig. 2 Scatter Plot Between Simulated LAIs and Measured LAIs

值作为该旬的 SMC, 与观测 VTCI 进行线性回归分析。样点观测 VTCI 与样点模拟 SMC 回归方程的决定系数  $R^2$  为 0.32, 相关性达极显著水平 ( $P < 0.001$ )。同化所需的旬尺度样点模拟 VTCI 的计算如下:

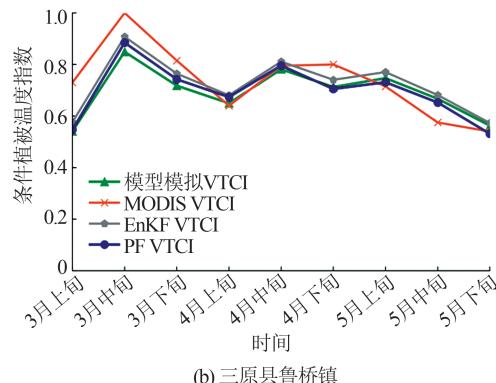
$$VTCI = 0.46SMC + 0.32 \quad (15)$$

以 2017 年的 2 个样点为例分析 VTCI 的同化效果, 其中乾县石牛乡为旱作样点, 三原县鲁桥镇为灌溉样点, 得到其 VTCI 的变化趋势如图 3 所示。

由图 3 可以看出, 乾县石牛乡 3 月中旬和 4 月下旬的 MODIS VTCI 较大, 经 EnKF 和 PF 同化后的 VTCI 受模拟值影响, 同化后的 VTCI 值有所降低。其中, 5 月中旬至 5 月下旬 VTCI 的同化值高于 MODIS 观测值和模型模拟值, 这是因为 VTCI 以旬为单位合成, 同化实验中样本的数量有限, 导致 EnKF 算法易出现滤波发散现象<sup>[29-30]</sup>; 而对于 PF 算法, 虽然利用重采样方法解决了粒子权值退化问题, 但也易造成样本贫化的现象<sup>[31-32]</sup>。三原县鲁桥镇 3 月上旬的 MODIS 观测和模拟 VTCI 均较大, EnKF 和 PF 同化后的 VTCI 值介于观测值与模拟值之间, 有效结合了二者的变化趋势。经两种同化算法同化后的 VTCI 均



(a) 乾县石牛乡



(b) 三原县鲁桥镇

图 3 2017 年不同样点 VTCI 的变化趋势

Fig.3 Changing Trend of VTCI at Different Sampling Sites in 2017

能综合反映 MODIS 观测值和模型模拟值的变化特征。

为进一步验证同化 VTCI 对降水量变化的敏感程度, 以 2017 年的两个旱作样点为例, 将 MODIS VTCI、EnKF VTCI 和 PF VTCI 分别与旬累积降水量、前 1 月的累积降水量进行对比分析, 结果如图 4 所示。

由图 4 可知, 在乾县石牛乡, MODIS VTCI 与旬累积降水量、前 1 月累积降水量之间的 Pearson 相关系数分别为 0.46 和 0.63, 同化后的 EnKF VTCI 与两种降水量之间的 Pearson 相关系数分别为 0.44 和 0.85, PF VTCI 与两种降水量之间的 Pearson 相关系数分别为 0.62 和 0.79。在岐山县蒲村镇, MODIS VTCI 与旬累积降水量、前 1 月累积降水量之间的 Pearson 相关系数分别为 0.20 和 0.60, 同化后的 EnKF VTCI 分别达到 0.23 和 0.85, PF VTCI 分别达到 0.22 和 0.93。由此可见, EnKF VTCI 和 PF VTCI 与降水量的相关性总体上均有显著提高, 相比于 MODIS VTCI, 同化后的 VTCI 总体上更符合降水量的变化趋势。

同样以 2017 年的乾县石牛乡和三原县鲁桥镇两个样点为例分析 LAI 的同化效果, 结果见图 5。由图 5 可以看出, 在乾县石牛乡, 模型模拟的

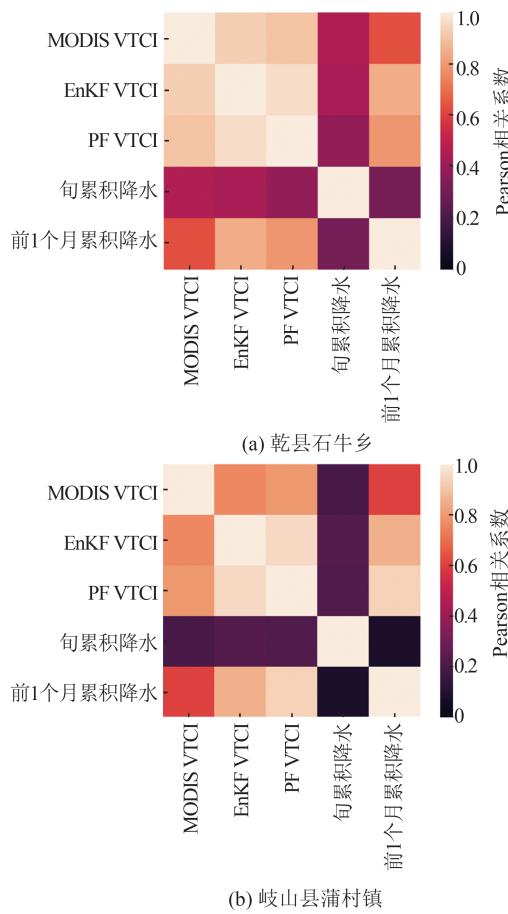


图4 2017年不同旱作样点观测VTCI、同化VTCI与降水量之间Pearson相关系数的热力图

Fig.4 Heat Map of Pearson Correlation Coefficients Between Observed and Assimilated VTCIs and Rainfall Amounts at Different Rain-fed Sampling Sites in 2017

LAI自3月上旬起上升幅度明显,到4月中旬达到最大值,然后呈现显著下降趋势;受MODIS数据分辨率及卫星运行周期等影响,MODIS LAI值总体偏低,自3月上旬起缓慢上升,5月上旬达到最大值后开始下降。经EnKF和PF算法同化后的LAI较MODIS LAI值有较大提升,更加接近模型模拟值。在三原县鲁桥镇,MODIS LAI、模型模拟LAI与经EnKF和PF算法同化后的LAI值变化趋势总体较为一致,均在4月中旬达到最大值后开始显著下降;EnKF和PF同化LAI总体上介于MODIS观测值和模拟值之间,两种同化LAI均能较好地结合遥感观测值和模型模拟值。

### 2.3 同化结果对比分析

将2013—2018年12个样点的单点EnKF VTCI和PF VTCI分别与MODIS VTCI进行线性回归分析(见表2),决定系数 $R^2$ 分别为0.32和

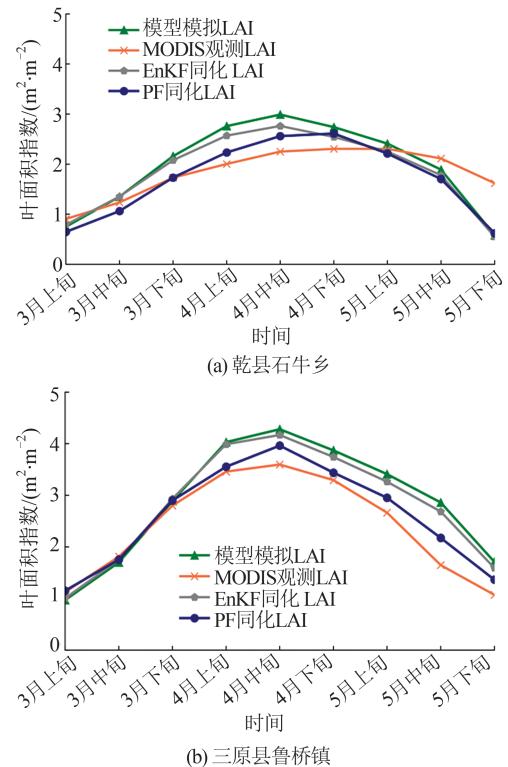


Fig.5 Changing Trend of LAI at Different Sampling Sites in 2017

0.65,均通过了5%的显著性水平检验( $P<0.001$ );单点尺度两种不同的同化VTCI与观测VTCI的RMSE分别为0.16和0.12,即PF VTCI与MODIS VTCI之间的 $R^2$ 更高, RMSE更小。同样,将2013—2018年12个样点的单点EnKF LAI和PF LAI分别与MODIS LAI进行线性回归分析, $R^2$ 分别为0.45和0.72,均能通过5%的显著性水平检验( $P<0.001$ );样点尺度两种不同的同化LAI与观测LAI的RMSE分别为 $0.97\text{ m}^2/\text{m}^2$ 和 $0.65\text{ m}^2/\text{m}^2$ ,即PF LAI与MODIS LAI之间的 $R^2$ 更高, RMSE更小。由表2可知,PF同化后的VTCI和LAI与MODIS观测值更接近。运用由此得到的线性关系计算获得研究区域每年3月—5月冬小麦VTCI、LAI的EnKF和PF同化结果。

表2 基于不同同化算法的同化VTCI和LAI精度评价  
Tab.2 Accuracy Evaluation of Assimilated VTCI and LAI Based on Different Assimilation Algorithms

| 指标        | $R^2$ | RMSE                         |
|-----------|-------|------------------------------|
| EnKF VTCI | 0.32  | 0.16                         |
| PF VTCI   | 0.65  | 0.12                         |
| EnKF LAI  | 0.45  | $0.97\text{ m}^2/\text{m}^2$ |
| PF LAI    | 0.72  | $0.65\text{ m}^2/\text{m}^2$ |

## 2.4 PCA-Copula 法确定的综合长势监测指标与单产的回归分析

### 2.4.1 综合长势监测指标的构建

利用 PCA 对 2013—2018 年冬小麦 4 个生育时期的 VTCI、EnKF VTCI、PF VTCI, 以及经归一化处理后的 LAI、EnKF LAI、PF LAI 分别进行了降维处理, 结果如表 3 所示。各长势监测指标选取的主成分个数如下: VTCI、EnKF VTCI 和 PF VTCI 均选取前 2 个主成分, 累积贡献率分别达到 89.20%、90.75% 和 89.56%; LAI、EnKF LAI 和 PF LAI 均选取前 1 个主成分, 累积贡献率分别达到 88.50%、92.26% 和 89.08%; VTCI 和 LAI、EnKF VTCI 和 LAI 以及 PF VTCI 和 LAI 均选取前 3 个主成分, 累积贡献率分别达到 90.91%、91.21% 和 90.72%。

表 3 VTCI 和 LAI 的主成分贡献率

Tab.3 Principal Component Contribution Rates of VTCI and LAI

| 指标            | 主成分   | 贡献率/% | 累积贡献率/% |
|---------------|-------|-------|---------|
| VTCI          | 第一主成分 | 59.15 | 59.15   |
|               | 第二主成分 | 30.05 | 89.20   |
| EnKF VTCI     | 第一主成分 | 59.50 | 59.50   |
|               | 第二主成分 | 31.25 | 90.75   |
| PF VTCI       | 第一主成分 | 61.88 | 61.88   |
|               | 第二主成分 | 27.68 | 89.56   |
| LAI           | 第一主成分 | 88.50 | 88.50   |
|               | 第二主成分 | 9.72  | 98.22   |
| EnKF LAI      | 第一主成分 | 92.26 | 92.26   |
|               | 第二主成分 | 5.51  | 97.77   |
| PF LAI        | 第一主成分 | 89.08 | 89.08   |
|               | 第二主成分 | 8.05  | 97.13   |
| VTCI+LAI      | 第一主成分 | 51.52 | 51.52   |
|               | 第二主成分 | 27.74 | 79.52   |
|               | 第三主成分 | 11.65 | 90.91   |
| EnKF VTCI+LAI | 第一主成分 | 59.43 | 59.43   |
|               | 第二主成分 | 19.65 | 79.07   |
|               | 第三主成分 | 12.13 | 91.21   |
| PF VTCI+LAI   | 第一主成分 | 62.59 | 62.59   |
|               | 第二主成分 | 17.15 | 80.09   |
|               | 第三主成分 | 10.98 | 90.72   |

利用 MODIS 观测 VTCI 计算综合 VTCI。经 PCA 降维处理后, VTCI 选取前 2 个主成分, 第一、二主成分均能通过正态分布和广义极值分布的 K-S(Kolmogorov-Smirnov) 检验。通过计算各边缘分布的理论频率与经验频率之间的 RMSE 和赤池信息准则值(Akaike information criterion,

AIC) 可知, 观测 VTCI 的第一主成分利用正态分布进行拟合的效果 ( $RMSE = 0.04$ ,  $AIC = -430.75$ ) 优于利用广义极值分布进行拟合的效果 ( $RMSE = 0.05$ ,  $AIC = -424.37$ ); 观测 VTCI 的第二主成分利用广义极值分布进行拟合的效果 ( $RMSE = 0.02$ ,  $AIC = -534.80$ ) 优于利用正态分布进行拟合的效果 ( $RMSE = 0.02$ ,  $AIC = -512.67$ ), 因此观测 VTCI 第一主成分优选正态分布拟合, 第二主成分优选广义极值分布拟合。基于此, 结合乘积 Copula 函数建立主成分间的联合分布, 得到综合 VTCI。利用上述同样的方法计算得到各长势监测指标的最优边缘分布函数, 分别构建综合长势监测指标。

以往进行的冬小麦估产较多采用客观赋权法对 VTCI 和 LAI 进行权重赋值<sup>[14, 22]</sup>, 根据数据的差异程度确定冬小麦不同生育时期 VTCI 或 LAI 的权重, 建立其与冬小麦单产之间的线性关系。本文通过主成分分析法对冬小麦各生育时期同化与未同化的 VTCI、LAI 进行降维处理, 不仅能有效避免因变量较多导致的信息重叠、多重共线性等问题, 而且 PCA-Copula 法构建的综合长势监测指标也能避免运用客观赋权法得到的权重结果与 VTCI 或 LAI 在冬小麦不同生育时期对产量的影响程度不符的问题<sup>[18, 33]</sup>。利用 PCA 法得到的各因子之间相互独立, 因此可以结合无需求解参数的乘积 Copula 构建综合长势监测指标, 解决了多元变量导致 Copula 函数的参数求解复杂的问题。

### 2.4.2 冬小麦估产模型的构建

考虑到主成分表达式中 4 个生育时期 VTCI 或 LAI 的系数对作物长势的影响, 在满足主成分累积贡献率的基础上, 通过对比包含不同主成分个数的综合长势监测指标与冬小麦单产的线性回归分析结果, 将包含最优主成分个数的综合长势监测指标作为估产模型的自变量<sup>[28]</sup>, 结果见表 4。由表 4 可知, 综合 VTCI 包含前 2 个主成分时与冬小麦单产的相关性最高,  $R^2$  为 0.18, 通过了 5% 的显著性检验 ( $P < 0.001$ ); 受主成分表达式中负系数数的影响, 综合 LAI 以及综合 G 均为包含前 1 个主成分时与冬小麦单产的相关性最高,  $R^2$  分别为 0.52 和 0.45, 均通过了 5% 的显著性检验 ( $P < 0.001$ )。综合 VTCI-EnKF、综合 LAI-EnKF 和综合 G-EnKF 均包含前 1 个主成分时与冬小麦单产的相关性最高; 相比于 MODIS 观测值构建估产模型的决定系数均有提升, 分别为 0.19、0.53 和

0.54,均通过了5%的显著性检验( $P<0.001$ ),其中基于综合G-EnKF构建的估产模型的决定系数提升较大。综合VTCI-PF、综合LAI-PF和综合G-PF均包含前1个主成分时与冬小麦单产的相关性最高;相比于MODIS观测与EnKF同化结果,基于PF同化构建的估产模型精度均有提高,模型的 $R^2$ 分别为0.21、0.55和0.55,均通过了5%的显著性检验( $P<0.001$ )。综上,基于同化变量进行产量估算精度均高于MODIS观测变量的估算精度,且基于PF同化变量构建估产模型的估测精度最优。PF同化结果在单点尺度和区域尺度均优于EnKF同化结果,这是由于EnKF算法假设误差基于高斯型的概率密度函数,而PF算法是对EnKF算法的改进,不受限于非线性高斯系统,通过蒙特卡洛模拟方法近似计算后验概率分布,更加符合状态变量的真实后验概率密度。

表4 冬小麦单产估测模型

Tab.4 Winter Wheat Yield Estimation Models

| 指标            | 回归模型  | 决定系数 |
|---------------|---|------|
| VTCI          | $Y(V) = 1547.05V + 3982.71$                             | 0.18 |
| LAI           | $Y(L) = 2219.71L + 3293.02$                             | 0.52 |
| VTCI+LAI      | $Y(G) = 1937.77G + 3404.26$                             | 0.45 |
| EnKF VTCI     | $Y(V_{\text{EnKF}}) = 1407.92V_{\text{EnKF}} + 3982.71$ | 0.19 |
| EnKF LAI      | $Y(L_{\text{EnKF}}) = 2161.94L_{\text{EnKF}} + 3313.03$ | 0.53 |
| EnKF VTCI+LAI | $Y(G_{\text{EnKF}}) = 2178.14G_{\text{EnKF}} + 3303.70$ | 0.54 |
| PF VTCI       | $Y(V_{\text{PF}}) = 1462.60V_{\text{PF}} + 3538.15$     | 0.21 |
| PF LAI        | $Y(L_{\text{PF}}) = 2226.60L_{\text{PF}} + 3277.80$     | 0.55 |
| PF VTCI+LAI   | $Y(G_{\text{PF}}) = 2201.01G_{\text{PF}} + 3286.14$     | 0.55 |

注:Y表示冬小麦估测单产( $\text{kg}/\text{hm}^2$ );V、 $V_{\text{EnKF}}$ 和 $V_{\text{PF}}$ 分别为综合VTCI、综合VTCI-EnKF和综合VTCI-PF;L、 $L_{\text{EnKF}}$ 和 $L_{\text{PF}}$ 分别为综合LAI、综合LAI-EnKF和综合LAI-PF;G、 $G_{\text{EnKF}}$ 和 $G_{\text{PF}}$ 分别为综合G、综合G-EnKF和综合G-PF

## 2.5 估产模型的选择与精度评价

选择表4所构建的估产模型中的较优模型,即基于综合LAI、综合LAI-EnKF、综合LAI-PF、综合G、综合G-EnKF、综合G-PF所构建的估产模型分别对2013—2018年关中平原24县(区)的冬小麦单产进行估测,并统计获取的估测单产与实际单产之间的误差。

利用综合LAI构建的冬小麦估产模型的估测单产与实际单产的RMSE为 $574.61\text{ kg}/\text{hm}^2$ ,MRE为10.72%;利用综合LAI-EnKF构建的冬小麦估产模型的估测单产与实际单产的RMSE为 $569.80\text{ kg}/\text{hm}^2$ ,MRE为10.42%;利用综合

LAI-PF构建的冬小麦估产模型的估测单产与实际单产的RMSE为 $559.22\text{ kg}/\text{hm}^2$ ,MRE为10.25%。统计结果表明同化的LAI构建的估产精度更高,且基于综合LAI-PF构建的模型估产结果更准确。

利用综合G构建的冬小麦单产估测模型的估测单产与实际单产的RMSE为 $617.91\text{ kg}/\text{hm}^2$ ,MRE为11.79%;利用综合G-EnKF构建的冬小麦单产估测模型的估测单产与实际单产的RMSE为 $564.95\text{ kg}/\text{hm}^2$ ,MRE为10.27%;利用综合G-PF构建的冬小麦单产估测模型的估测单产与实际单产的RMSE为 $561.66\text{ kg}/\text{hm}^2$ ,MRE为10.28%。由统计结果可知,基于综合G-PF构建的冬小麦单产相比基于综合G构建的估产模型RMSE降低了 $56.25\text{ kg}/\text{hm}^2$ ,MRE降低了1.51%,表明基于综合G-PF构建的单产估测模型在县域尺度估测结果较优。

为了进一步验证估测模型的精度,分别将综合LAI-PF、综合G-PF与冬小麦实际单产进行逐年线性拟合(2013—2018年),结果如图6所示,其中 $R^2(L_{\text{PF}})$ 、RMSE( $L_{\text{PF}}$ )分别表示利用综合LAI-PF构建的估产模型计算的估测单产与实际单产的 $R^2$ 和RMSE; $R^2(G_{\text{PF}})$ 、RMSE( $G_{\text{PF}}$ )分别表示利用综合G-PF构建的估产模型计算的估测单产与实际单产的 $R^2$ 和RMSE。由图6可知,除了2013年基于综合LAI-PF得到的估测单产与实际单产的 $R^2$ 高于同化双变量外,其余年份均为基于综合G-PF得到的估测单产与实际单产的 $R^2$ 较高;估测单产与实际单产RMSE的计算结果总体上PF同化双变量优于PF同化LAI的结果。以上结果进一步说明,基于PF同化VTCI和LAI构建的双变量估产模型精度更高,对于改善关中平原冬小麦产量估测精度效果显著,可用于研究区域冬小麦单产的估测。

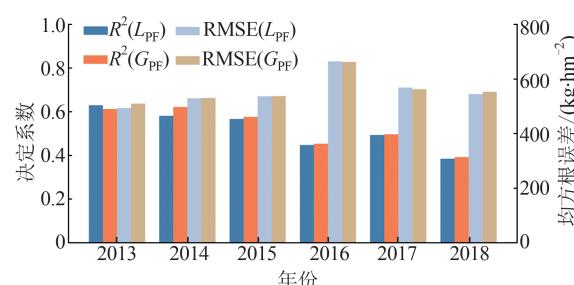


图6 估测单产与实际单产的精度评价结果  
Fig.6 Accuracy Evaluation Results of Estimated Yield and Actual Yield

运用基于综合 G-PF 构建的冬小麦估产模型对关中平原 2017—2020 年进行逐像素单产估测,结果见图 7。由图 7 可知,冬小麦在西部地区产量最高,中部地区次之,东部地区最低。其中西部和中部地区是关中平原种植作物的主要区域,冬小麦估产结果的空间分布特征与实际一致。2019 年冬小麦的产量偏低,国家统计局陕西调查总队 (<http://snzd.stats.gov.cn/>) 分析报告显示,

2018 年冬小麦播种期内,关中地区的连续降水造成了农田积水,而排水、晒田时间严重不足导致局部地区冬小麦难以下种;2019 年冬小麦生长季内的降水分布不均,且气温偏高,导致了阶段性干旱。播种面积的下降及旱情直接影响了 2019 年的冬小麦产量。2020 年,冬小麦生长季雨热适中,气候条件适宜,产量得以恢复,与估测结果相符。

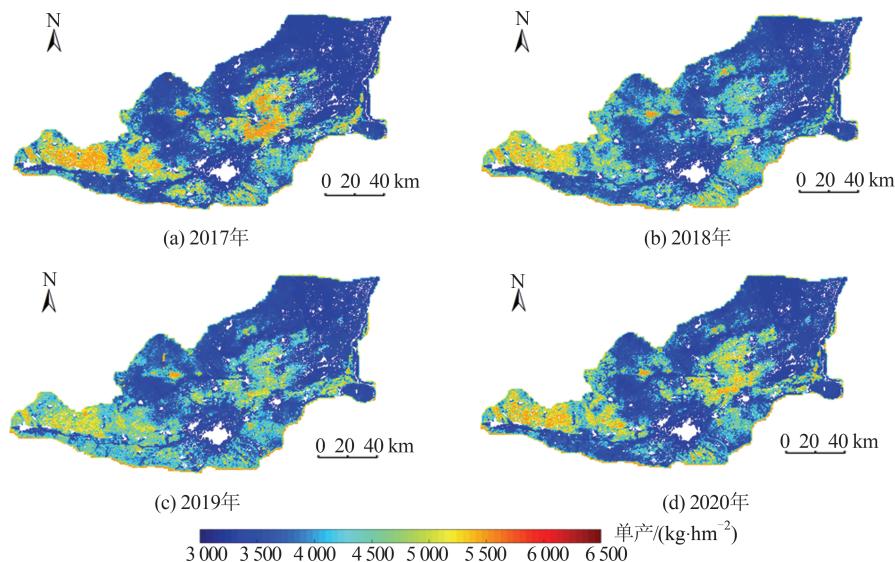


图 7 关中平原 2017—2020 年冬小麦单产估算结果

Fig. 7 Estimates of Winter Wheat Yields in the Guanzhong Plain from 2017 to 2020

### 3 结语

利用 EnKF、PF 算法同化后的 VTCI 和 LAI 均能综合反映 MODIS 观测值和 CERES-Wheat 模型模拟值的变化特征,且单点同化 VTCI 相比于 MODIS VTCI 对降水量的变化更加敏感,可以更好地反映冬小麦受水分胁迫的影响;PF 同化后的 VTCI 和 LAI 均更符合关中平原冬小麦实际的变化规律。

采用 PCA-Copula 法对 2013—2018 年冬小麦 4 个生育时期的观测 VTCI、观测 LAI 和同化 VTCI、同化 LAI 分别构建单变量、双变量的综合长势监测指标,建立冬小麦单产估测模型并进行精度评价。结果表明,经过同化后的长势监测指标所构建的估产模型均优于 MODIS 观测数据构建的估产模型,且基于 PF 同化 VTCI 和 LAI 构建的双变量估产模型 ( $R^2=0.55$ ,  $P<0.001$ ) 拟合效果最好。2013—2018 年所有县(区)估测单产与实际单产的 RMSE 为  $561.66 \text{ kg}/\text{hm}^2$ , MRE 为 10.28%。相比未同化的 MODIS 观测双变量构建的估产模型精度, RMSE 降低了  $56.25 \text{ kg}/\text{hm}^2$ ,

MRE 降低了 1.51%, 估产精度较高。研究区冬小麦的产量西部地区最高,中部地区次之,东部地区最低,估测结果与实际结果相符合,进一步说明该模型可用于研究区的冬小麦单产估测。下一步工作可以考虑使用适用于处理非线性数据的降维方法,以更大程度保留多数据集变量之间的非线性特征,提高冬小麦估产精度。

### 参 考 文 献

- [1] Lobell D B, Asner G P, Ortiz-Monasterio J I, et al. Remote Sensing of Regional Crop Production in the Yaqui Valley, Mexico: Estimates and Uncertainties [J]. *Agriculture, Ecosystems & Environment*, 2003, 94(2): 205–220
- [2] Ren Jianqiang, Chen Zhongxin, Tang Huajun, et al. Regional Crop Yield Simulation Based on Crop Growth Model and Remote Sensing Data [J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2011, 27(8): 257–264 (任建强, 陈仲新, 唐华俊, 等. 基于遥感信息与作物生长模型的区域作物单产模拟 [J]. 农业工程学报, 2011, 27(8): 257–264)

- [3] Zhao Chunjiang. Advances of Research and Application in Remote Sensing for Agriculture[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2014, 45(12): 277-293 (赵春江. 农业遥感研究与应用进展[J]. 农业机械学报, 2014, 45(12): 277-293)
- [4] Wagner M P, Slawig T, Taravat A, et al. Remote Sensing Data Assimilation in Dynamic Crop Models Using Particle Swarm Optimization [J]. *ISPRS International Journal of Geo-Information*, 2020, 9(2): 105
- [5] Bai T C, Wang S G, Meng W B, et al. Assimilation of Remotely-Sensed LAI into WOFOST Model with the SUBPLEX Algorithm for Improving the Field-Scale Jujube Yield Forecasts [J]. *Remote Sensing*, 2019, 11(16): 1945
- [6] Liu Yating, Gong Yan, Duan Bo, et al. Combining Multi-temporal NDVI and Abundance from UAV Remote Sensing Data for Oilseed Rape Growth Monitoring [J]. *Geomatics and Information Science of Wuhan University*, 2020, 45(2): 265-272 (刘雅婷, 龚龑, 段博, 等. 多时相NDVI与丰度综合分析的油菜无人机遥感长势监测[J]. 武汉大学学报·信息科学版, 2020, 45(2): 265-272)
- [7] de Wit A, Duveiller G, Defourny P. Estimating Regional Winter Wheat Yield with WOFOST Through the Assimilation of Green Area Index Retrieved from MODIS Observations [J]. *Agricultural and Forest Meteorology*, 2012, 164: 39-52
- [8] Huang J X, Gómez-Dans J L, Huang H, et al. Assimilation of Remote Sensing into Crop Growth Models: Current Status and Perspectives [J]. *Agricultural and Forest Meteorology*, 2019, 276/277: 107609
- [9] Mathieu P P, O'Neill A. Data Assimilation: From Photon Counts to Earth System Forecasts [J]. *Remote Sensing of Environment*, 2008, 112(4): 1258-1267
- [10] Chen Y, Tao F L. Improving the Practicability of Remote Sensing Data-Assimilation-Based Crop Yield Estimations over a Large Area Using a Spatial Assimilation Algorithm and Ensemble Assimilation Strategies [J]. *Agricultural and Forest Meteorology*, 2020, 291: 108082
- [11] Ines A V M, 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
- [12] Pan H Z, Chen Z X, Allard D W, et al. Joint Assimilation of Leaf Area Index and Soil Moisture from Sentinel-1 and Sentinel-2 Data into the WOFOST Model for Winter Wheat Yield Estimation [J]. *Sensors (Basel, Switzerland)*, 2019, 19(14): 3161
- [13] 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 (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)
- [14] Xie Y, Wang P X, Bai X J, et al. Assimilation of the Leaf Area Index and Vegetation Temperature Condition Index for Winter Wheat Yield Estimation Using Landsat Imagery and the CERES-Wheat Model [J]. *Agricultural and Forest Meteorology*, 2017, 246: 194-206
- [15] AghaKouchak A, Cheng L Y, Mazdiyasni O, et al. Global Warming and Changes in Risk of Concurrent Climate Extremes: Insights from the 2014 California Drought [J]. *Geophysical Research Letters*, 2014, 41(24): 8847-8852
- [16] 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 Chinese Society of Agricultural Engineering*, 2015, 31(11): 96-102 (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)
- [17] Wang Pengxin, Feng Mingyue, Sun Huitao, et al. Drought Impact Assessment Based on Principal Component Analysis and Copula Function [J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2016, 47(9): 334-340 (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]. 农业机械学报, 2016, 47(9): 334-340)
- [18] 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 (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)
- [19] Sun Wei, Wang Pengxin, Han Lijuan, et al. Fur-

- [10] Chen Y, Tao F L. Improving the Practicability of Remote Sensing Data-Assimilation-Based Crop Yield Estimations over a Large Area Using a Spatial Assimilation Algorithm and Ensemble Assimilation Strategies[J]. *Agricultural and Forest Meteorology*, 2020, 291: 108082.
- [11] Ines A V M, 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.
- [12] Pan H Z, Chen Z X, Allard D W, et al. Joint Assimilation of Leaf Area Index and Soil Moisture from Sentinel-1 and Sentinel-2 Data into the WOFOST Model for Winter Wheat Yield Estimation[J]. *Remote Sensing of Environment*, 2015, 167: 450.
- [13] 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.
- [14] Xie Yi, Wang P X, Bai X J, et al. Assimilation of the Leaf Area Index and Vegetation Temperature Condition Index for Winter Wheat Yield Estimation Using Landsat Imagery and the CERES-Wheat Model[J]. *Agricultural and Forest Meteorology*, 2017, 246: 194-206.
- [15] AghaKouchak A, Cheng L Y, Mazdiyasni O, et al. Global Warming and Changes in Risk of Concurrent Climate Extremes: Insights from the 2014 California Drought[J]. *Geophysical Research Letters*, 2014, 41(24): 8847-8852.
- [16] 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 Chinese Society of Agricultural Engineering*, 2015, 31(12): 96-102.
- [17] Wang Pengxin, Feng Mingyue, Sun Huitao, et al. Drought Impact Assessment Based on Principal Component Analysis and Copula Function[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2016, 47(9): 334-340.
- [18] Li Yan, Wang Pengxin, Liu Junming, et al. Evaluation Method for Drought Monitoring Effects in the Main Growth and Development Stages of Winter Wheat Using Vegetation Temperature Condition Index — Impact Evaluation of Drought on Wheat Yield[J]. *Agricultural Research in the Arid Areas*, 2014, 32(5): 218-222.
- [19] Sun Wei, Wang Pengxin, Han Lijuan, et al. The Improvement of the Approach to Monitoring Drought Using Vegetation and Temperature Condition Indexes from Multi-Years Remotely Sensed Data[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2013, 29(10): 255-265.
- [20] Wang Pengxin, Xun Lan, Li Li, et al. Extraction of Planting Areas of Main Crops Based on Fourier Sensing Assimilation Model[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2016, 32(22): 22-26.
- [21] Jones J W, Hoogenboom G. The DSSAT Cropping System Model[J]. *European Journal of Agronomy*, 2003, 18(3/4): 235-265.
- [22] Xie Yi, Wang Pengxin, Wang Lei, et al. Estimation of Wheat Yield Based on Crop and Remote Sensing Assimilation Model[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2006, 22(2): 22-26.
- [23] Wang Wei, Liu Xiangge, Wang Pengxin, et al. Application of 4DVAR and Ensemble Kalman Filter Assimilating Vegetation Temperature Condition Index for Winter Wheat Yield Estimation[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2011, 27(12): 184-190.
- [24] Zhang Xianfeng, Zhao Jiepeng, Wang Pengxin, et al. Application of 4DVAR and Ensemble Kalman Filter Assimilating Vegetation Temperature Condition Index for Winter Wheat Yield Estimation[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2011, 27(12): 184-190.
- [25] Bi H Y, Ma J W, Wang F, et al. An Improved Particle Filter Algorithm Based on Ensemble Kalman Filter and Markov Chain Monte Carlo Method[J]. *IEEE Transactions on Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(2): 447-459.
- [26] Xu L, Abbaszadeh P, Moradkhani H, et al. Continental Drought Monitoring Using Satellite Soil Moisture, Data Assimilation and Integrated Drought Index[J]. *Remote Sensing of Environment*, 2012, 37(7): 794-799.
- [27] Guo Xin, Wang Naijiang, Zhang Lingling, et al. Monitoring of Spatial-Temporal Change of Winter Wheat Area in Guanzhong Region Based on Google Earth Engine[J]. *Agricultural Research in the Arid Areas*, 2020, 38(3): 275-280.
- [28] Wang Pengxin, Chen Chi, Zhang Shuyi, et al. Winter Wheat Yield Estimation Based on Copula Function and Remotely Sensed LAI and VTCI[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 52(10): 255-263.
- [29] Houtekamer P L, Mitchell H L. Data Assimilation Using an Ensemble Kalman Filter Technique[J]. *Monthly Weather Review*, 1998, 126(3): 796-811.
- [30] Han Pei, Shu Hong, Xu Jianhui. Sensitivity Analysis of the Sequential Data Assimilation Methods[J]. *Journal of Geomatics Science and Technology*, 2015, 32(5): 483-488.
- [31] Djuric P M, Vemula M, Bugallo M F. Target Tracking by Particle Filtering in Binary Sensor Networks[J]. *IEEE Transactions on Signal Processing*, 2008, 56(6): 2229-2238.
- [32] Ito K, Xiong K. Gaussian Filters for Nonlinear Filtering Problems[J]. *IEEE Transactions on Automatic Control*, 2000, 45(5): 910-927.
- [33] Wang Pengxin, Feng Mingyue, Mei Shuli, et al. Analysis and Application of the Multi-Scale Characteristics of Vegetation Temperature Condition Index[J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(6): 915-921.

## Estimation of Winter Wheat Yield Using Assimilated Bi-variables and PCA-Copula Method

WANG Pengxin<sup>1</sup> CHEN Chi<sup>1</sup> ZHANG Yue<sup>1</sup> ZHANG Shuyu<sup>2</sup> LIU Junming<sup>3</sup>

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

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

3 College of Land Science and Technology, China Agricultural University, Beijing 100193, China

**Abstract:** **Objectives:** Accurate, timely and effective monitoring of the growth and yield of winter wheat over a large area can help optimize the wheat planting structure, adjust the regional layout and ensure the country's food security. Therefore, it is very important to further improve the estimation accuracy of winter wheat yield. **Methods:** Vegetation temperature condition index (VTCI) and leaf area index (LAI) at the main growth period of winter wheat, which were simulated by the CERES (crop environment resource synthesis)-Wheat model and retrieved from MODIS (moderate resolution imaging spectroradiometer) data, were assimilated by using ensemble Kalman filtering (EnKF) algorithm and particle filtering (PF) algorithm. In addition, the principal component analysis combined with the Copula function was used to develop univariate (VTCI or LAI) and bi-variate (VTCI and LAI) winter wheat yield estimation models, and the optimal model was selected to estimate winter wheat yields from 2017 to 2020. **Results:** The experimental results show that, at the sampling-sites scale, both VTCI and LAI after assimilated can comprehensively reflect the variation characteristics of MODIS observed and model simulated values, and the application of PF algorithm has a better assimilation effect. At the regional scale, the bivariate yield estimation model developed by using PF algorithm has the highest accuracy. Compared with the accuracy of the models constructed by VTCI and LAI without assimilation, the root mean square error of the optimal assimilation model is reduced by 56.25 kg/hm<sup>2</sup>, and the average relative error is reduced by 1.51%. **Conclusions:** The above results indicate that the model can effectively improve the accuracy of winter wheat yield estimation and has good applicability for large area yield estimation.

**Key words:** assimilation; PCA-Copula; principal component analysis; leaf area index; vegetation temperature condition index; crop growth model; winter wheat yield estimation

**First author:** WANG Pengxin, PhD, professor, specializes in quantitative remote sensing and its application in agriculture. E-mail: wangpx@cau.edu.cn

**Foundation support:** The National Natural Science Foundation of China (41871336, 42171332).

**引文格式:** WANG Pengxin, CHEN Chi, ZHANG Yue, et al. Estimation of Winter Wheat Yield Using Assimilated Bi-variables and PCA-Copula Method[J]. Geomatics and Information Science of Wuhan University, 2022, 47(8):1201-1212.DOI:10.13203/j.whugis20220038(王鹏新,陈弛,张悦,等.利用双变量同化与PCA-Copula法的冬小麦单产估测[J].武汉大学学报·信息科学版,2022,47(8):1201-1212.DOI:10.13203/j.whugis20220038)