doi:10.6041/j.issn.1000-1298.2020.09.016

基于 EnKF 和随机森林回归的玉米单产估测

王鹏新[1,2]　胡亚京[1,2]　李　俐[2,3]　许连香[2,3]

(1. 中国农业大学信息与电气工程学院, 北京 100083; 2. 农业农村部农业灾害遥感重点实验室, 北京 100083;
3. 中国农业大学土地科学与技术学院, 北京 100083)

摘要: 为了提高玉米的估产精度,以河北省中部平原为研究区域,采用 CERES-Maize 模型模拟 2013—2018 年 8 个典型样点玉米整个生育期的叶面积指数(LAI),将遥感反演的 LAI 与 CERES-Maize 模型模拟的 LAI 相结合,通过集合卡尔曼滤波(EnKF)同化算法实现 2013—2018 年玉米主要生育时期旬尺度 LAI 的同化,运用随机森林回归法计算同化和未同化的 LAI 权重,进而建立玉米单产估测模型,对 2015 年 53 个县(区)的玉米进行单产估测和精度评价,并分析 2013—2018 年玉米的单产时空分布特征。结果表明,采用 EnKF 算法对 8 个研究样点进行单点同化,同化 LAI 更符合玉米实际生长情况;将样点 LAI 同化值从单点尺度扩展到区域尺度,同化 LAI 图像减少了相邻像素间 LAI 陡升陡降的现象,其效果优于遥感反演的 LAI;与未同化 LAI 构建的估测模型相比,应用同化的 LAI 所建的估测模型精度明显提高,R^2 提高了 0.024 5;在 2015 年河北中部平原 53 个县(区)估产结果中,总体平均相对误差为 12.11%,RMSE 为 371 kg/hm^2,NRMSE 为 6.18%;河北中部平原玉米单产估测结果呈现个别年份波动、总体呈先减少后增加的年际变化特点,并呈现西部地区最高、北部和南部地区次之、东部地区最低的空间分布特征。

关键词: 玉米; 估产; 集合卡尔曼滤波; 叶面积指数; 随机森林回归

中图分类号: TP79　　**文献标识码**: A　　**文章编号**: 1000-1298(2020)09-0135-09　　**OSID**:

Estimation of Maize Yield Based on Ensemble Kalman Filter and Random Forest for Regression

WANG Pengxin[1,2]　HU Yajing[1,2]　LI Li[2,3]　XU Lianxiang[2,3]

(1. *College of Information and Electrical Engineering*, *China Agricultural University*, *Beijing* 100083, *China*
2. *Key Laboratory of Remote Sensing for Agri-Hazards*, *Ministry of Agriculture and Rural Affairs*, *Beijing* 100083, *China*
3. *College of Land Science and Technology*, *China Agricultural University*, *Beijing* 100083, *China*)

Abstract: In order to improve the estimation accuracy of maize, the central plain of Hebei Province was chosen as research area, and the remote sensed LAI and simulated LAI by CERES-Maize model was combined in eight typical samples from 2013 to 2018 by using the ensemble Kalman filter (EnKF) data assimilation approach. The random forest regression was used to estimate maize yield by using monitored LAI and the assimilated ones respectively. The optimal model for estimating maize yields in study area from 2013 to 2018 was selected, and the measured maize yield of the year 2015 was used to validate the accuracy of the optimal model. The results showed that the single point assimilation of eight samples using the EnKF algorithm was more consistent with the actual growth of maize. The assimilated LAIs were extended from the sampling sites to the regional scale, the phenomenon of LAIs rising and falling between adjacent pixels was reduced and the effect was better than the remote sensing inversion LAIs. Compared the yield estimation models with the monitored LAIs, the accuracy of the yield estimation models with the assimilated LAIs was improved, and the R^2 was increased by 0.024 5. The yield estimation model was applied to estimate maize yield in 53 counties (districts), in general, the average relative error of the estimated yield was 12.11%, and the root mean square error was 371 kg/hm^2, the normalized root mean square error was 6.18%. The yearly estimated yield from 2013 to 2018 in the central plain of Hebei Province was fluctuated in individual years, and the overall distribution in time was characterized by a tendency to decrease first and then increase, and the spatial distribution of maize yield was the highest in

收稿日期: 2019-12-10　修回日期: 2020-04-21

基金项目: 国家重点研发计划项目(2016YFD030060303-3)

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

the western region of the plain, following by the north and south regions, and the lowest was in the eastern region.

Key words: maize; yield estimation; ensemble Kalman filter; leaf area index; random forest regression

0 引言

玉米是我国主要粮食作物之一[1]。及时、准确地估测和预测粮食产量,对掌握作物生长状况、指导农业生产及制定粮食政策具有重要意义[2]。作物生长模型[3]可以准确模拟作物的生长发育过程和估测最终产量,但应用于大范围区域时,难以解决宏观资料的获取及参数的区域化等问题[4]。遥感技术为大范围的作物产量预测提供了一种高效的手段[5],但遥感数据不能揭示作物生长和产量形成的内在机理。目前,将遥感数据与作物生长模型结合是当前作物产量估算研究的重要内容和发展趋势之一[6]。数据同化方法[7]将遥感信息与作物长势相结合,较好地解决作物生长模型从单点研究发展到区域应用时遇到的问题。王鹏新等[8]将遥感反演的条件植被温度指数(Vegetation temperature condition index,VTCI)与 CERES-wheat 小麦生长模型模拟的土壤浅层含水率相结合,采用四维变分算法(4D-VAR)实现小麦主要生育期旬尺度的 VTCI 同化。结果表明,在单点尺度和区域尺度上,同化的 VTCI 均能更好地响应外部观测数据,更符合 VTCI 的先验知识。de WIT 等[9]采用 EnKF 算法同化 WOFOST(World food studies)模型模拟与微波遥感反演的土壤水分,预测欧洲西南部的冬小麦与玉米产量,结果表明,同化方法能够改善区域冬小麦的预测精度。

叶面积指数(Leaf area index,LAI)可表征作物的生长状态和光合作用能力,是作物长势及单产估测的重要指标[10]。作物单产是作物在整个生长阶段长势信息综合反映,不同生育时期的作物对作物单产的影响也不尽相同,可通过对不同生育时期特征变量赋予不同的权重,构建综合特征参数,进行作物单产估测,以提高估测精度[11]。王蕾等[12]采用多指标综合评价中的灰色关联分析法确定了玉米 4 个生育时期(出苗—拔节期、拔节—抽雄期、抽雄—灌浆期与灌浆—成熟期)的 LAI 作为长势监测指标的权重值,结果表明,基于灰色关联分析法的玉米长势综合指标优于单一 LAI 指标,基于该综合指标的长势监测结果与实际产量的年际变化和空间分布规律等更为相符。随机森林回归(Random forest for regression,RFR)算法是基于决策树的一种集成学习算法,对噪声和异常值有较好的容忍度,且对高维数据集具有良好的可扩展性和并行性,具有很高的预测精度[13]。随机森林算法除了可对数据集进行分类和回归外,还可进行变量重要性分析、奇异值监测、数据缺失插补、生存分析等[14]。王鹏新等[15]利用随机森林回归模型估测玉米单产,并且给出了各个变量的重要性评分,即玉米 4 个生育时期 VTCI 或 LAI 对玉米单产的影响程度。

本文以河北省中部平原夏玉米为研究对象,以 2013—2018 年玉米主要生育期(7—9 月)为研究时段,以 LAI 作为同化系统状态变量,运用集合卡尔曼滤波(Ensemble Kalman filter, EnKF)[16]算法对遥感反演的 LAI 和 CERES-Maize 模型模拟的 LAI 进行同化,利用随机森林回归法[17]分别计算同化和未同化加权 LAI,构建加权 LAI 与玉米单产的估产模型,对夏玉米单产进行估测,并应用 2015 年的单产数据对模型进行验证。

1 材料与方法

1.1 研究区域概况

研究区位于河北省中部平原,包括保定市、廊坊市、石家庄市、衡水市和沧州市的 53 个县(区),覆盖范围为 114°32′~117°36′E,36°57′~39°50′N。河北省中部平原属于黄淮海平原的一部分,是河北省重要的农耕区,盛产小麦、玉米等,是中国一年两熟作物种植区和重要的玉米生产基地[18]。该区域属于温带大陆性季风气候,降水量主要集中在夏季,年降水量在 350~700 mm 之间,占全年的 65%~70%,降水量由南向北逐渐减少。海拔在 0~100 m 之间,地势西北高,东南低,年平均气温 4~13℃,年辐射总量 4 390~5 180 MJ/m^2。根据王鹏新等[19]提出的作物分类方法提取玉米的种植区域,通过叠加河北省行政区县边界图,获得研究区 2013—2018 年各县(区)的玉米种植分布图,且在研究区选取 8 个典型的夏玉米种植区作为研究样点(图 1),依次为保定市雄县雄州镇南十里铺村、保定市定州市刘早镇北王庄村、保定市定兴县固城南太平庄村、衡水市冀州市南午镇刘瓦窑村、衡水市景县青兰乡南申庄村、衡水市饶阳县城关镇西直沃村、沧州市沧县兴济镇宋官屯村、沧州市泊头县四营乡大赵屯村。

1.2 研究方法

1.2.1 集合卡尔曼滤波算法

EnKF 算法的基本过程是对状态量假设一个集

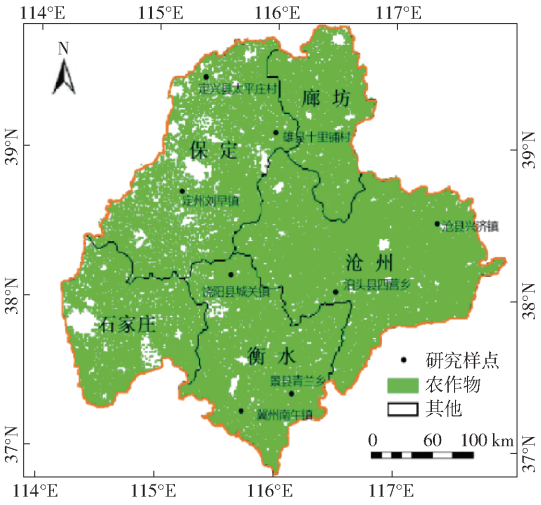


图 1　研究区域位置及玉米种植区(2017 年)

Fig. 1　Location of study area and planting area of maize (2017)

合,每个集合表示状态量的可能取值,集合的均值即为状态的最优估计。每个集合通过动态过程模拟向前积分,预测下一个时刻每个集合的状态量,对于每个有观测数据的时刻,利用观测数据对每个集合分别进行更新,此时所有集合的均值即为状态量的最优估计值,估计值的方差也可以通过 EnKF 算法进行计算[20]。

EnKF 算法包含预测和更新 2 个步骤:

(1)预测:根据 CERES - Maize 模型的模型算子 M 计算第 i 时刻的 LAI 背景值为

$$LAI_i^b = M(LAI_{i-1}^a) \tag{1}$$

式中　LAI_{i-1}^a——第 $i-1$ 时刻的 LAI 分析值

LAI_i^b——第 i 时刻的 LAI 背景值

(2)更新:计算第 i 时刻的集合卡尔曼增益矩阵为

$$\boldsymbol{K} = \boldsymbol{B}(\boldsymbol{B} + \boldsymbol{O})^{-1} \tag{2}$$

式中　$\boldsymbol{O}$——观测场误差协方差矩阵

$\boldsymbol{B}$——背景场误差标准差矩阵

计算第 i 时刻 LAI 的分析值为

$$LAI_i^a = LAI_i^b + \boldsymbol{K}(LAI_i^o - LAI_i^b) \tag{3}$$

式中　LAI_i^o——第 i 时刻 LAI 的观测值

进入第 $i+1$ 时刻,将 LAI_i^a 代入模型算子,计算更新的背景值,同时更新背景场误差标准差和观测场误差标准差,重复上述过程,直到没有外部 LAI 数据为止,此时,第 i 时刻的值即为该时刻的 LAI 同化值。

1.2.2　随机森林回归算法

随机森林回归[21](RFR)算法以决策树为基本模型,通过构建不同的训练数据集以及不同的特征空间来产生一系列有差异的决策树模型,通常使用投票或者取平均值得到最终结果。

(1)将研究区域各县(区)2013—2018 年玉米 4 个生育时期的 LAI 及玉米单产数据作为原始样本数 Q,通过 Bootstrap 重抽样得到 n 个与原始样本数据集相等的训练样本并生成 n 棵回归树$\{h(x,\theta_n), n=1,2,\cdots,N\}$,表示为

$$h(x) = \frac{1}{Q}\sum_{n=1}^{N} h(x,\theta_n) \tag{4}$$

式中　θ_n——独立同分布随机变量

n——回归树的棵数

$h(x)$——回归树

(2)在回归树生长过程中,每个分裂节点随机抽取所有变量中的 V 个变量作为当前节点的特征子集(一般选取总变量的 1/3,本文中 V 为 1)进行分裂,且分裂过程不需要剪枝。

(3)在每次进行 Bootstrap 重抽样时,未被抽中的数据接近于 1/e,约为 0.37,即袋外数据(Out of bag,OOB),用来进行内部误差估计,并基于 OOB 误差得到玉米 4 个生育时期的重要性[22],将 $x_p(p=1,2,3,4)$ 作为输入数据,则第 q 棵树上的重要性 I_q 为

$$I_q(x_p) = \frac{\sum_{q=1}^{Q_{OOB}} I(f'(x_q)) - \sum_{q=1}^{Q_{OOB}} I(f''(x_q))}{Q_{OOB}} \tag{5}$$

式中　x_p——玉米 4 个生育时期的 LAI

Q_{OOB}——袋外数据样本数

$f'(x_q)$——袋外数据中第 q 个样本值

$f''(x_q)$——随机置换变量后袋外数据第 q 个样本的估测值

$I(\cdot)$——判别函数

变量 x_p 在整个随机森林中的重要性得分为

$$I(x_p) = \sum_{n=1}^{N} I_n(x_p)/N \tag{6}$$

由于随机性的引入,模型每次得到的变量重要性评分略有差异,故将 10 次运行结果的平均值进行归一化处理,作为玉米 4 个生育时期的权重[15]。

1.2.3　加权 LAI 的确定与单产估产模型的构建

本文同化 LAI 以旬为尺度,依据河北中部平原各县行政边界图,将各县所包含像素的 LAI 的平均值作为该地区的 LAI,即得到该县旬尺度的 LAI。将夏玉米划分为出苗—拔节期(7 月上旬—中旬)、拔节—抽雄期(7 月下旬—8 月上旬)、抽雄—乳熟期(8 月中旬—9 月上旬)与乳熟—成熟期(9 月中旬—下旬)4 个主要生育时期[12]。将各主要生育时期所包含的 LAI 的平均值作为该生育时期的 LAI。由此得到河北省中部平原地区各县玉米 4 个生育期的 LAI。将玉米 4 个生育期的同化 LAI 及玉米单产数据(212 组)和未同化 LAI 及玉米单产(212 组)分别

输入随机森林回归模型,计算玉米各生育时期同化和未同化 LAI 的权重,进而计算各县(区)每年的加权 LAI。将通过随机森林回归算法得到的加权 LAI 分别与玉米统计单产进行线性回归,建立单产估测模型(本研究选用 2015 年的数据用于单产估测精度的验证,故 2015 年各县(区)的相关数据未参与模型的构建)。

1.2.4 总体技术路线

将气象数据、土壤数据、田间管理数据和作物品种遗传参数输入并驱动 CERES - Maize 模型,将模型模拟的 LAI 数据与 Terra 和 Aqua 卫星上的 MODIS 传感器获得的研究区域的 LAI 数据相结合,通过 EnKF 算法得到研究样点的同化 LAI,利用随机森林回归算法计算玉米 4 个主要生育期的权重,进而得到加权 LAI,运用玉米实际单产与加权 LAI 构建的单产估测模型估测玉米单产以及进行精度验证,同时与未同化的加权 LAI 和玉米单产数据构建的估产模型进行对比,研究技术路线如图 2 所示。

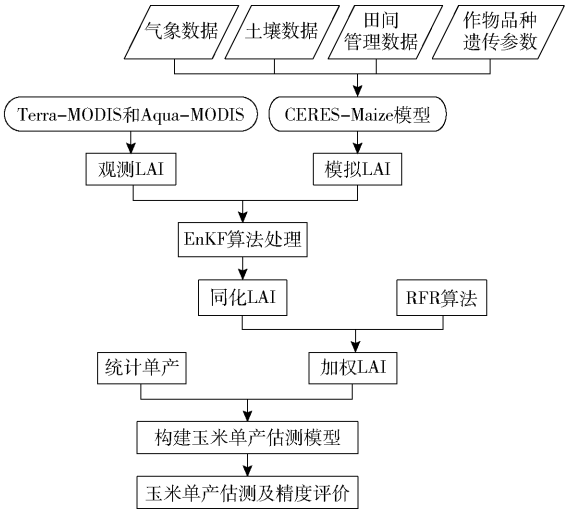


图 2 研究技术路线图

Fig. 2 Research route of technology

1.3 数据获取与处理

1.3.1 CERES - Maize 模型数据

CERES - Maize 模型[23]能够以日为时间步长模拟玉米生长发育、产量形成、氮碳水平衡过程,其输入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数据 4 方面参数。

(1)气象数据

气象数据通过 DSSAT 内嵌的 Weather Man 软件导入模型,存储在后缀为“WHT”的文件中,包括气象站点名称、站点经纬度信息、海拔以及逐日太阳总辐射量、最高气温、最低气温、降水量。2013—2018 年研究样点的逐日气象数据由中国气象科学数据共享服务网(http://cdc.cma.gov.cn/)提供,其中包括最高气温、最低气温、降水量和日照时数等,在输入气象参数前需要将日照时数转换为逐日太阳总辐射量,利用 Angstrom[24]经验计算公式

$$R_s = R_{\max}\left(a_s + b_s \frac{w}{W}\right) \tag{7}$$

式中 R_s——逐日太阳总辐射量,MJ/m²

$R_{\max}$——逐日天文辐射量,MJ/m²

a_s、b_s——经验系数,取决于地理位置和气候,河北中部平原属温带地区,本文 $a_s = 0.18$,$b_s = 0.55$

w——逐日日照时数,h

W——逐日可照时数,即最大时长,h

(2)土壤数据

CREES - Maize 模型的土壤参数包括土壤理化参数和形态属性等,主要分为表层参数和分层参数,其中表层参数包括土壤颜色、土壤类型、坡度、渗透性等;土壤参数包括土壤各层的粘粒含量和粉粒含量、田间持水量、pH 值、有机质含量、有效氮含量等,通过查阅《中国土种志》[25]获得土壤数据剖面各层的理化参数。通过 DSSAT - CERES - Maize 内置的 Sbuild 参数管理平台将土壤数据导入模型,并存储为 SOL 文件。

(3)田间管理数据

主要包括夏玉米的品种、播种日期、播种方式、播种深度、植株密度、行距、施肥日期和施肥量、灌溉日期和灌溉量等,这些参数均来自《河北夏玉米与冬小麦一体化种植》[1]和《河北玉米》[26]。通过 DSSAT - CERES - Maize 模型中的田间管理数据管理平台(XBuild)进行数据输入并存储为 WHX 文件。

(4)作物品种遗传参数

作物的遗传特性控制着其生长发育进程,直接关系到植株形态的发育与作物产量的高低[27]。CERES - Maize 模型遗传参数主要包括 P_1、P_2、P_5、G_2、G_3、P_{HINT},其中,P_1 表示出苗到幼苗期结束所需的积温(℃·d),P_2 表示光周期敏感系数,P_5 表示抽丝到收获成熟所需的积温(℃·d),G_2 为单株最大穗粒数,G_3 为最大灌浆速率参数,P_{HINT} 为完成一片叶生长所需积温(℃·d)。采用模型自带的 DSSAT - GLUE 调参程序包进行作物遗传参数标定[28],并存储在以 CUL 为后缀的文件中。

CERES - Maize 模型的输出参数保存为 OUT 文件,由于模型运行得到以日为步长的 LAI,取玉米主要生育期内每旬 LAI 数据均值作为同化所需的旬尺度模拟 LAI。

1.3.2 遥感数据

选取 2013—2018 年玉米生长季(177 ~ 265 d) MODIS 叶面积指数产品(MCD15A3H),其空间分辨率为 500 m,时间分辨率为 4 d。利用 MODIS 数据处理工具 MRT 进行拼接、重采样、投影转换和裁剪等预处理,得到研究区 LAI 产品。采用上包络线 S-G(Savitzky-Golay)滤波对原始 LAI 进行平滑处理,以消除由云、大气等因素引起的 LAI 数据骤降现象,滤波后的 LAI 曲线变化平稳,更符合玉米的生长规律[19]。将玉米每旬所包含的多时相 LAI 的最大值作为每旬的 LAI,得到旬尺度 LAI[15]。根据研究样点的经纬度坐标计算其在 MODIS 影像上的像素坐标,提取每年玉米主要生育期(7—9 月)每旬 LAI 影像中以样点为中心、3 像素 ×3 像素内所含像素的 LAI 的均值作为单点同化的外部观测数据,以涵盖研究样点的河北中部平原遥感反演的 LAI 作为区域尺度的观测数据。

1.3.3 县域产量数据

通过查阅《河北农村统计年鉴》得到研究区各县(区)2013—2016 年玉米播种面积和总产量数据,玉米单产数据由总产量和播种面积计算得到。

2 结果与分析

2.1 LAI 同化结果与分析

将 2015 年 7—9 月玉米种植区的 8 个研究样点 LAI 的遥感观测值和模型模拟值输入 EnKF 算法以获得单点同化结果(图 3),总体观察研究样点,经 EnKF 算法同化 LAI 变化趋势均能综合表达遥感观测 LAI 和模拟 LAI 变化曲线,同化 LAI 变化趋势较接近模拟 LAI 变化趋势。选择 4 个研究样点进行分析,具体表现为:保定市雄县雄州镇南十里铺村(图 3a),在 7 月上旬至 7 月下旬,遥感观测 LAI 变化趋势有一个陡升的过程,在 7 月下旬至 8 月中旬,观测 LAI 几乎没有变化,而同化 LAI 变化趋势较平稳地上升至最大值,且峰值出现在 8 月中旬左右,结果与实际夏玉米生长过程较符合,在 8 月中旬至 9 月下旬,曲线为下降阶段,同化 LAI 的变化趋势较平滑;衡水市景县青兰乡南申庄村(图 3d)在 7 月上旬至 7 月下旬,遥感观测 LAI 有一个陡升过程,在 7 月下旬至 8 月下旬,观测 LAI 缓慢上升,在曲线下降过程中,观测 LAI 先缓慢下降后快速下降,而同化 LAI 变化趋势受模拟结果的影响,曲线更为平滑,且同化 LAI 曲线的最大值出现在 8 月中旬,更符合玉米实际生长过程;同样,保定市定州市刘早镇北王庄村(图 3b)和衡水市饶阳县城关镇西直沃村(图 3c) 2 个样点,同化 LAI 变化趋势无论是上升阶段、下降阶段还是峰值出现时,均能够较好地综合模拟和遥感观测 LAI 变化趋势,且 LAI 观测值整体相对较低,同化之后,LAI 有所提升,更接近于玉米实际生长的 LAI。为了验证单点数据同化结果能较好地响应外

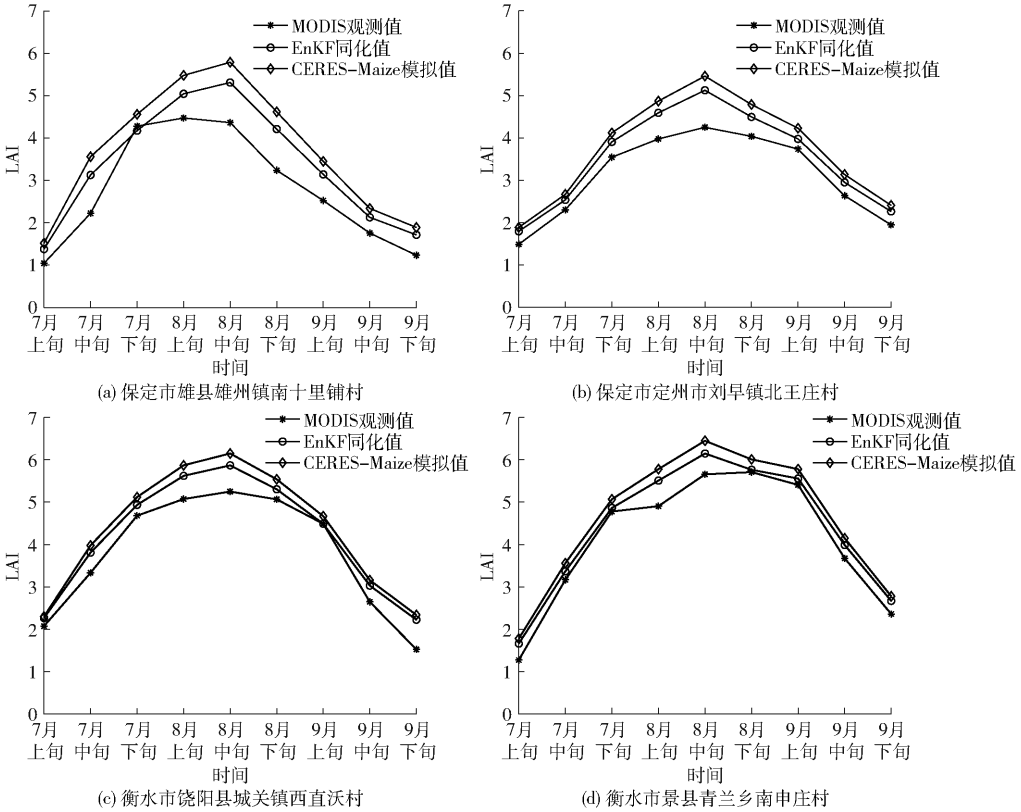


图 3　2015 年不同样点 LAI 的变化曲线

Fig. 3　Changing trends of LAI at different sampling sites in 2015

部观测数据,将 2013—2018 年 8 个样点的单点 LAI 的 EnKF 同化结果与原观测结果进行线性回归分析(图 4),可以看出,总体上 LAI 的同化结果和观测数据结果较为吻合,R^2 为 0.669 7($P<0.001$),所有样点同化 LAI 和观测 LAI 间的均方根误差(RMSE)为 0.44,归一化均方根误差(NRMSE)为 12.7%。

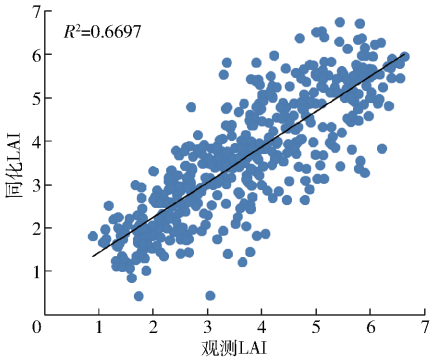
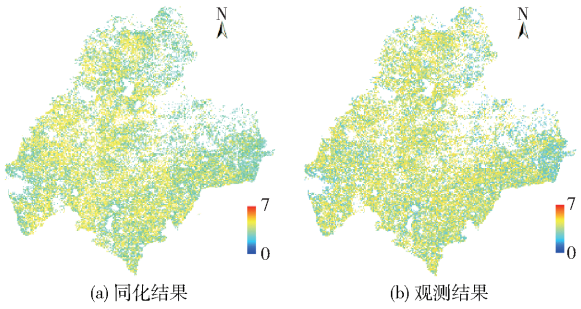


图 4 EnKF 同化 LAI 与观测 LAI 相关性分析结果

Fig. 4 Correlation analysis results between observed LAIs and assimilated LAIs using EnKF

由于所选取的 8 个样点较为均匀地分布在河北中部平原,故假设该 8 个研究样点的特性可以代表河北中部平原,将同化试验中的同化 LAI 与遥感观测 LAI 建立线性回归方程,由此可以得到每年 7~9 月以旬为单位的区域 LAI 同化结果。本文选取 2015 年 8 月上旬的区域 LAI 同化结果与观测结果(图 5)。通过对比同化和观测 LAI 的图像,发现区域尺度的同化试验中,由于结合了模型模拟值和原观测值,同化 LAI 近邻像素间 LAI 数值过渡较平稳,图像纹理更好,减少了原观测图像中数值陡升陡落



(a) 同化结果 (b) 观测结果

图 5 EnKF 同化 LAI 结果与观测 LAI 结果

Fig. 5 Oberved LAIs and assimilated LAIs using EnKF at regional scales

的情况,效果优于遥感反演结果。

2.2 区域单产的估测与验证

根据随机森林回归算法获取夏玉米各主要生育时期的权重(表 1),抽雄—乳熟期与拔节—抽雄期权重较大,乳熟—成熟期权重次之,出苗—拔节期权重最小,主要是因为抽雄—乳熟期是夏玉米从营养阶段向生殖生长阶段过渡时期,是夏玉米产量形成的关键阶段,拔节—抽雄期主要是玉米营养生长和生殖生长并进阶段,因此,LAI 对玉米单产的抽雄—乳熟期与拔节—抽雄期影响较大;乳熟—成熟期是光合产物的运输、转移时期,对夏玉米产量起着重要作用,权重略低于拔节—抽雄期;出苗—拔节期 LAI 的权重最小,表明生长前期 LAI 与玉米产量的相关性不大,依据农学知识,该生育时期为玉米的营养生长阶段,光合作用的产物用于进行根系和叶片的营养生长。综上,随机森林回归算法确定的同化与观测 LAI(表 1)的权重较为合理。

表 1 基于随机森林回归算法的玉米生育时期观测、同化 LAI 的权重和单产估测模型

Tab. 1 Weights of monitored LAI and assimilated ones at key growth stages of maize and yield estimation models

LAI	出苗—拔节期	拔节—抽雄期	抽雄—乳熟期	乳熟—成熟期	单产估测模型	R^2	P
观测值	0.180 1	0.253 1	0.369 3	0.197 5	$y=1\,340x+1\,958$	0.305 6	<0.01
同化值	0.182 0	0.253 8	0.369 4	0.194 8	$y=1\,265x+2\,306$	0.330 1	<0.01

基于玉米各主要生育期的权重,结合观测 LAI 和同化 LAI 分别计算各县(区)的加权 LAI,构建随机森林算法计算的加权 LAI 与实际单产之间的线性回归方程作为夏玉米单产模型(表 1)。结果表明,观测 LAI 和同化 LAI 与玉米单产间的相关性均达到显著水平($P<0.01$)。基于观测 LAI 和同化 LAI 的估产模型的决定系数 R^2 分别为 0.305 6 和 0.330 1。综上,同化 LAI 的单产估测模型精度更高,可作为较优的单产估测模型。

为验证估测模型的精度,应用两个单产估测模型分别对河北中部平原 2015 年各县(区)的夏玉米单产进行估测,分析 53 个县(区)估测单产与统计单产的相对误差,结果表明,基于同化 LAI 构建的估测模型中涿州市、高碑店市、雄县等 24 个市县(区)相对误差低于 10%,任丘市、河间市、故城县、吴桥县等 12 个县的相对误差均低于 15%,栾城区、孟村县等 15 个县(区)相对误差在 25% 以内,只有武邑县和阜城县相对误差超过 25%,分别为 27.20% 和 26.52%;基于同化 LAI 的玉米单产估测模型中,高碑店市统计单产为 6 905 kg/hm^2,估测单产为 6 965 kg/hm^2,相对误差为 0.87%,望都县统计单产为 7 628 kg/hm^2,估测单产为 7 732 kg/hm^2,相对误差为 1.36%,永清县统计单产为 5 640 kg/hm^2,估测单产为 5 561 kg/hm^2,相对误差为 1.4%,无极县统

计单产为 7 329 kg/hm^2，估测单产为 7 452 kg/hm^2，相对误差为 1.68%，晋州市统计单产为 7 201 kg/hm^2，估测单产为 7 070 kg/hm^2，相对误差为 1.81%，雄县统计单产为 5 785 kg/hm^2，估测单产为 5 934 kg/hm^2，相对误差为 2.58%。分析所有县（区）估测单产与统计单产相对误差、RMSE 和 NRMSE，结果表明，基于同化 LAI 单产估测模型的所有县（区）估测单产的平均相对误差为 12.11%，RMSE 为 371 kg/hm^2，NRMSE 为 6.18%，而基于观测 LAI 单产估测模型的所有县（区）估测单产的平均相对误差为 14.31%，RMSE 为 510 kg/hm^2，NRMSE 为 8.49%，表明基于同化 LAI 的夏玉米单产估测模型的估产精度较高，可以为玉米产量的早期预估提供判断依据和决策信息。

将 2013—2018 年各县（区）的夏玉米的同化加权 LAI 代入较优单产估测模型（表 1），得到 2013—2018 年河北中部平原玉米区域单产估测结果（图 6），结果表明，2013 年西部地区（包括石家庄市和保定市）玉米估测单产为 6 500 kg/hm^{2}左右，东部（包括沧州市）为 6 100 kg/hm^{2}左右，南部（包括衡水市）在 6 700 kg/hm^2 左右，北部（包括廊坊市）在 6 100 kg/hm^{2}左右，2014 年玉米单产估测值略低于 2013 年；2015、2016 年玉米单产估测值略高于 2014 年，西部地区在 6 900 kg/hm^2 左右，东部地区在 6 200 kg/hm^{2}左右，南部地区在 6 800 kg/hm^2 左右，北部地区在 6 000 kg/hm^{2}左右；2017 年西部玉米单产估测在 7 000 kg/hm^{2}左右，东部地区在 6 300 kg/hm^{2}左右，北部地区在 6 400 kg/hm^{2}左右，南部地区在 6 100 kg/hm^{2}左右；2018 年西部地区和南部地区玉米单产在 6 900 kg/hm^{2}左右，东部地区和北部地区在 6 300 kg/hm^{2}左右。以 2018 年为例，西部地区玉米单产为 6 956 kg/hm^2，东部地区为 6 253 kg/hm^2，南部地区为 6 859 kg/hm^2，北部地区为 6 124 kg/hm^2。从年际变化看，河北中部平原 2013—2018 年各县（区）的单产个别年份有波动，但总体呈先减少后增加的趋势。从单产的空间分布看，整体呈现西部地区最高、北部和南部地区次之、东部地区最低。从研究年份看，2017 年玉米单产最高，2014 年玉米单产较低，原因可能是 2017 年降水量充沛，玉米单产高于往年平均水平，2014 年玉米生育期内发生阶段性干旱且局部地区旱情较重，玉米单产下降。

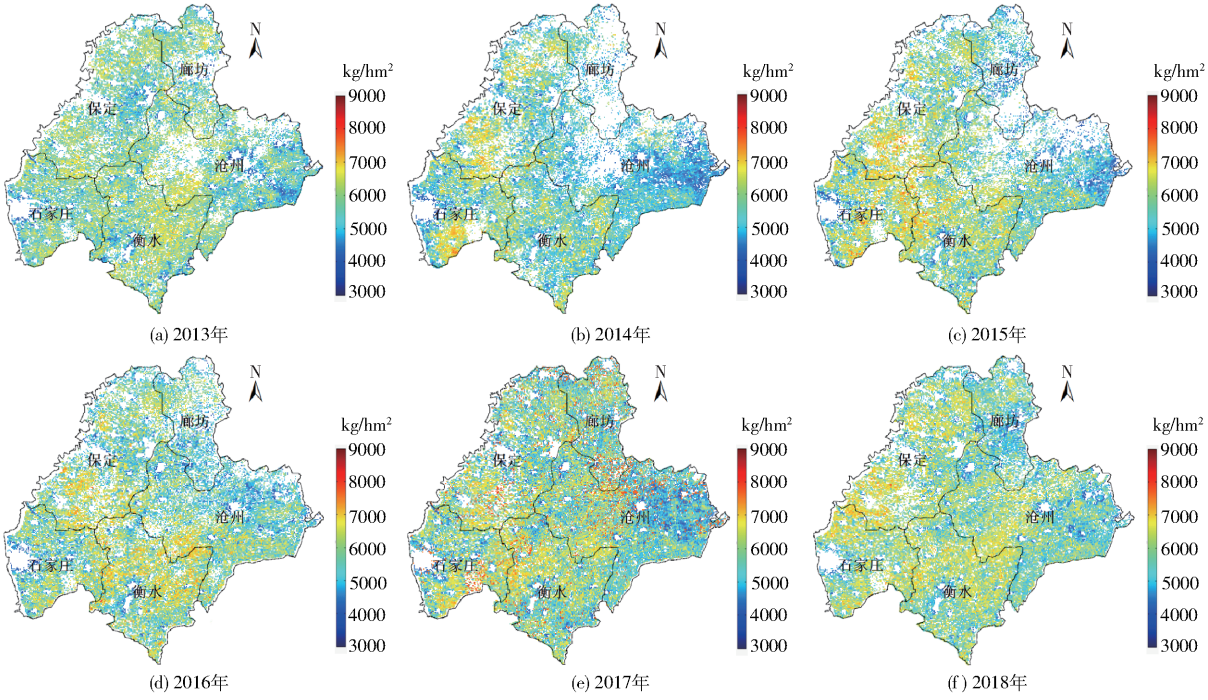


图 6　2013—2018 年各县（区）玉米单产估测结果

Fig. 6　Estimated yields of maize in different counties from 2013 to 2018

3　讨论

同化算法在估产研究方面，往往只针对某一年的遥感数据估算农作物的单产且只针对同化效果进行描述，本文选择河北中部平原 2013—2018 年为研究时段，采用集合卡尔曼滤波算法同化 LAI 并分析了玉米年内单产及年际间单产的估测结果以及误差，结果表明，同化后的玉米单产估测精度比未同化有明显提高，且有效地提高了区域粮食产量的估测精度。但本文中 LAI 是遥感与作物模型进行同化的唯一状态变量，难以完整描述作物受光照、温度、降水、气候、肥料不同要素相互作用的产量形成机理过

程,如干旱指数是影响作物产量的重要因素[29],因此,将 LAI 与其他变量结合进行多变量同化估产研究是未来研究工作的重点。

利用 CERES - Maize 模型模拟叶面积指数的误差与不确定性有多个方面,其中作物模型中的气象、土壤、作物品种与田间管理对模拟结果产生较大的影响,本文采用的天气资料均是中国气象数据中心的资料,而有些气象数据与作物田间观测样点位置上有偏离,因此天气数据无法准确地反映实际的天气条件,这是造成模拟误差原因之一。另外,作物生长过程中实际发生的病虫害等也会影响模拟效果,本文模拟试验未考虑这类实际情况。

4 结论

(1)采用 EnKF 算法对 8 个研究样点进行单点同化,同化 LAI 能较好地综合模型模拟与遥感观测值的优点,更符合玉米实际生长情况;将样点 LAI 同化值从单点尺度扩展到区域尺度,同化 LAI 图像的纹理性更好,效果优于遥感反演 LAI,更适合区域估产。

(2)基于同化 LAI,采用随机森林回归算法确定的单产估测模型的精度较高,R^2 为 0.330 1,观测 LAI 的 R^2 提高了 0.024 5。在 2015 年各县(区)估产结果中,大多数区域单产估测结果比实际结果偏高,基于同化 LAI,所有县(区)估测单产的平均相对误差为 12.11%,RMSE 为 371 kg/hm^2,NRMSE 为 6.18%,比观测 LAI 估测单产的平均相对误差减少了 2.2 个百分点,RMSE 减少了 139 kg/hm^2,NRMSE 减少了 2.31 个百分点。

(3)对 2013—2018 年研究区域估产结果的时间演变规律和空间变化特征进行分析,结果表明,河北中部平原玉米单产估测结果呈现个别年份波动、总体上先减少后增加的年际变化特点,并呈现西部地区最高、北部和南部地区次之、东部地区最低的空间分布特征。

参 考 文 献

[1] 韩文霆,彭星硕,张立元,等. 基于多时相无人机遥感植被指数的夏玉米产量估算[J/OL]. 农业机械学报,2020,51(1):148 - 155.
HAN Wenting,PENG Xingshuo,ZHANG Liyuan, et al. Summer maize yield estimation based on vegetation index derived from multi-temporal UAV remote sensing[J/OL]. Transactions of the Chinese Society for Agricultural Machinery,2020,51(1):148 - 155. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag = 1&file_no = 20200116&journal_id = jcsam. DOI:10.6041/j.issn.1000-1298.2020.01.016. (in Chinese)

[2] 程志强, 蒙继华. 作物单产估算模型研究进展与展望[J]. 中国生态农业学报, 2015, 23(4): 402 - 415.
CHENG Zhiqiang, MENG Jihua. Research progress and prospect of crop yield estimation model[J]. Chinese Journal of Eco-Agriculture, 2015,23(4): 402 - 415. (in Chinese)

[3] 谢云, JAMES R K. 国外作物生长模型发展综述[J]. 作物学报, 2002, 28(2):190 - 195.
XIE Yun, JAMES R K. A review of the development of foreign crop growth models[J]. Acta Agronomica Sinica, 2002, 28(2): 190 - 195. (in Chinese)

[4] DORIGO W A, ZURITA-MILLA R, de WIT A J W, et al. A review on reflective remote sensing and data assimilation technique for enhanced agroecosystem modeling[J]. International Journal of Applied Earth Observation and Geoinformation, 2007, 9(2): 165 - 193.

[5] MA Yuping, WANG Shili, ZHANG Li, et al. Monitoring winter wheat growth in North China by combining a crop model and remote sensing data[J]. International Journal of Applied Earth Observation and Geoinformation, 2008, 10(4): 426 - 437.

[6] 靳华安,王锦地,柏延臣,等. 基于作物生长模型和遥感数据同化的区域玉米产量估算[J]. 农业工程学报, 2012, 28(6):162 - 173.
JIN Hua'an, WANG Jindi, BO Yanchen, et al. Estimation on regional maize yield based on assimilation of remote sensing data and crop growth model[J]. Transactions of the CSAE, 2012, 28(6):162 - 173. (in Chinese)

[7] JIN X, KUMAR L, LI Z, et al. A review of data assimilation of remote sensing and ciopmodels[J]. European Journal of Agronomy, 2018, 92: 141 - 152.

[8] 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/OL]. 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 four-dimensional variational assimilation method and remotely sensed vegetation temperature condition index[J/OL]. Transactions of the Chinese Society for Agricultural Machinery,2016,47(3):263 - 271. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20190337&flag = 1&journal_id = jcsam. DOI:10.6041/j.issn.1000-1298.2019.03.037. (in Chinese)

[9] de WIT A J W, VAN DIEPEN C A. Crop model data assimilation with the ensembel Kalman filter for improving regional crop yield forecasts[J]. Agricultural and Forest Meteorology,2007,146(1/2): 38 - 56.

[10] 姚延娟,范闻捷,刘强,等. 玉米全生长期叶面积指数收获测量法的改进[J]. 农业工程学报,2010, 26(8): 189 - 194.
YAO Yanjuan, FAN Wenjie, LIU Qiang, et al. Improvement of leaf area index harvesting measurement method for maize

throughout growth period[J]. Transactions of the CSAE, 2010, 26(8):189 - 194. (in Chinese)
[11] 徐新刚,王纪华,黄文江,等. 基于权重最优组合和多时相遥感的作物估产[J]. 农业工程学报, 2009, 25(9):137 - 142.
XU Xin'gang, WANG Jihua, HUANG Wenjiang, et al. Estimation of crop yield based on weight optimization combination and multi-temporal remote sensing data[J]. Transactions of the CSAE, 2009, 25(9): 137 - 142. (in Chinese)
[12] 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):2099 - 2109.
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):2099 - 2109. (in Chinese)
[13] 钟登华,田耕,关涛,等. 基于混沌时序-随机森林回归的堆石坝料加水量预测研究[J]. 水力发电学报, 2018, 37(8): 1 - 12.
ZHONG Denghua, TIAN Geng, GUAN Tao, et al. Prediction of rockfill dam material watering volume based on chaotic time series and random forest regression[J]. Journal of Hydroelectric Engineering, 2018, 37(8):1 - 12. (in Chinese)
[14] 陈妙金,汪小钦,吴思颖. 基于随机森林算法的水土流失影响因子重要性分析[J]. 自然灾害学报, 2019, 28(4):209 - 219.
CHEN Miaojin, WANG Xiaoqin, WU Siying. Importance analysis of influence factors of soil and water loss based on random forest algorithm[J]. Journal of Natural Disasters, 2019, 28(4): 209 - 219. (in Chinese)
[15] 王鹏新,齐璇,李俐,等. 基于随机森林回归的玉米单产估测[J/OL]. 农业机械学报, 2019, 50(7):237 - 245.
WANG Pengxin, QI Xuan, LI Li, et al. Estimation of corn yield based on random forest regression[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(7): 237 - 245. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20190726&flag = 1&journal_id = jcsam. DOI: 10.6041/j.issn.1000-1298.2019.07.026. (in Chinese)
[16] NAN Tongchao, WU Jichun. Groundwater parameter estimation using the ensemble Kalman filter with localization[J]. Hydrogeology Journal, 2011, 19(3):547 - 561.
[17] 董婷,任东,邵攀,等. 基于多源遥感数据和随机森林的综合旱情指标构建[J/OL]. 农业机械学报, 2019, 50(8):200 - 212.
DONG Ting, REN Dong, SHAO Pan, et al. Construction of comprehensive drought index based on multi-source remote sensing data and random forest[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(8):200 - 212. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20190822&flag = 1&journal_id = jcsam. DOI:10.6041/j.issn.1000-1298.2019.08.022. (in Chinese)
[18] 代立芹,李春强,魏瑞江. 河北省夏玉米气候适宜度及其变化特征分析[J]. 生态环境学报, 2011, 20(6):1031 - 1036.
DAI Liqin, LI Chunqiang, WEI Ruijiang. Analysis of climate suitability and its variation characteristics of summer maize in Hebei Province [J]. Chinese Journal of Ecology, 2011, 20(6):1031 - 1036. (in Chinese)
[19] 王鹏新,荀兰,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)
[20] 马建文. 数据同化算法研发与试验[M]. 北京: 科学出版社, 2013.
[21] RASTGOU M, BAYAT H, MANSOORIZADEH M, et al. Estimating the soil water retention curve: comparison of multiple nonlinear regression approach and random forest data mining technique[J/OL]. Computers and Electronics in Agriculture, 2020,174:105502. https://doi.org/10.1016/j.compay.2020.105502.
[22] GONG H R, SUN Y R, SHU X, et al. Use of random forests regression for predicting IRI of asphalt pavements[J]. Construction and Building Materials, 2018, 189:890 - 897.
[23] ADNAN A, JIBRIN M, ALPHA Y K, et al. Using CERES - Maize model to determine the nitrogen fertilization requirements of early maturing maize in the Sudan Savanna of Nigeria[J]. Journal of Plant Nutrition,2017,40(7):1066 - 1082.
[24] LEE K H. Estimation of daily solar radiation at the missing point for water quality impact assessment in Nakdong River Watershed: comparison of modified angstrom model and transmittance interpolation model[J]. Journal of Environmental Impact Assessment,2012,21(1):219 - 227.
[25] 马同生. 中国土种志[M]. 北京:中国农业出版社, 1996.
[26] 贾银锁,谢俊良. 河北玉米[M]. 北京:中国农业科学技术出版社, 2008.
[27] JONES J W, HOOGENBOON G, PORTER C H, et al. The DSSAT cropping model[J]. European Journal of Agronomy, 2003, 18(3 - 4):235 - 265.
[28] HE J, DUKES M, JONES J, et al. Applying GLUE for estimating CERES - Maize genetic and soil parameters for sweet corn production[J]. Transactions of the ASABE, 2009, 52(6): 1907 - 1921.
[29] KUMAR M N, MURTHY C S, SESHA SAI M V R, et al. On the use of standardized precipitation index(SPI) for drought intensity assessment[J]. Meteorological Applications, 2009, 16(3): 381 - 389.