

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/327572922>

基于MCMC 方法的WOFOST 模型参数标定与不确定性分析

Article · September 2018

CITATIONS

0

READS

529

1 author:



Jianxi Huang

China Agricultural University

220 PUBLICATIONS 3,563 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Crop parameters quantitative retrieval and monitoring with remote sensing-A special issue of Remote Sensing (ISSN 2072-4292). This special issue belongs to the section "Remote Sensing in Agriculture and Vegetation". [View project](#)



A Special Issue of Remote Sensing (ISSN 2072-4292): "Earth Observations and Crop Models for Sustainable Agricultural Management" [View project](#)

基于 MCMC 方法的 WOFOST 模型参数标定与不确定性分析

黄健熙^{1,2}, 黄海¹, 马鸿元¹, 李颖^{3,4}, 侯英雨⁵, 何亮⁵, 朱德海^{1,2}

(1. 中国农业大学土地科学与技术学院, 北京 100083; 2. 农业部农业灾害遥感重点实验室, 北京 100083;

3. 中国气象局河南省农业气象保障与应用技术重点开放实验室, 郑州 450003; 4. 河南省气象科学研究所, 郑州 450003; 5. 中国气象局国家气象中心, 北京 100081)

摘要: 为探究作物生长模型参数的自动标定技术及其不确定性分析方法, 该研究以郑州农业气象试验站为试验点, 利用融入了 snooker 更新 (snooker update) 的 DE-MC (differential evolution Markov chain, 差分进化马尔科夫链) 方法实现对 WOFOST (world food studies) 作物生长模型的参数标定和不确定性定量评价。snooker 更新增加了 DE-MC 算法中候选样本的多样性, 从而实现利用更少的并行链对多维参数空间进行有效采样, 较适合于 WOFOST 模型参数众多的特性。结果表明: 相比于模型默认值, 采用 MCMC (Markov chain Monte Carlo, 马尔科夫链-蒙特卡洛) 标定后的参数, 叶面积指数 (leaf area index, LAI) 模拟精度可提高 51.40%~53.07%, 产量模拟精度提高 8.25%~8.88%。标定参数中, SPAN、SLATB070、SLATB040、AMAXTB130 和 SLATB00 的后验分布可近似为高斯分布, 其中 SPAN 的不确定性最低。带入后验参数集合进行模型, LAI 在三叶期至返青期之间以及拔节期至抽穗期之间模拟的不确定性较大; 产量模拟的不确定性随时间不断增大, 至乳熟期前后达到稳定。该方法能够实现对多参数复杂作物生长模型的参数标定和不确定性分析, 对作物模型参数估计及提高模拟精度具有重要作用。

关键词: 气象; 模型; 作物; MCMC 方法; WOFOST 模型; 参数标定; 不确定性分析

doi: 10.11975/j.issn.1002-6819.2018.16.015

中图分类号: S512.1

文献标志码: A

文章编号: 1002-6819(2018)-16-0113-07

黄健熙, 黄海, 马鸿元, 李颖, 侯英雨, 何亮, 朱德海. 基于 MCMC 方法的 WOFOST 模型参数标定与不确定性分析[J]. 农业工程学报, 2018, 34(16): 113—119. doi: 10.11975/j.issn.1002-6819.2018.16.015 <http://www.tcsae.org>
Huang Jianxi, Huang Hai, Ma Hongyuan, Li Ying, Hou Yingyu, He Liang, Zhu Dehai. Markov Chain Monte Carlo based WOFOST model parameters calibration and uncertainty analysis[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(16): 113—119. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2018.16.015 <http://www.tcsae.org>

0 引言

气候变化、人口膨胀、资源短缺等一系列问题对农业的高效持续发展提出了更高要求, 如何在宏观上准确地进行作物估产, 获取作物生长关键性参数, 以帮助确保粮食安全, 已经越来越成为研究的热点。基于机理过程模拟的作物生长模型普遍用于农作物产量预报、参数估计、灾害影响评估^[1-7], 还有学者将其成功应用于作物重金属、营养品质监测等^[8-9], 在不同区域环境、不同作物研究中均得到广泛验证。

作物模型机理性模拟的准确性依赖于众多模型参数、气象驱动数据、土壤参数的准确获取, 模型应用的区域越大, 数据的直接、准确获取越困难, 模拟的精度也越难以保证。实际应用中, 往往将一定的地理区域进行人为的均质化, 在此区域内共用全部或者部分模型参数, 由此不可避免的引起代表性误差。为了更准确反映出区域模型模拟的空间变异性, 降低以点代面导致的误

差, 近年来众多学者将遥感观测数据对作物模型进行约束, 通过数据同化在空间上对模型参数进行优化, 取得了大量的成果^[3-5, 10-12]。然而, 诸多研究中鲜有对模型参数标定的详尽描述, 也缺乏模型不确定性的定量表达, 使得模型在更大区域上的应用可靠性受到质疑。

所谓模型参数标定, 就是调整模型参数使其适用于特定的研究区域。传统方法是采用一定数量的实测值, 通过试错法标定出一套“最适”参数, 这种标定方法很难准确代表空间变异性的结果的综合。并且由于数据代表性、测量精度、标定方法等限制, 常常基于一组观测值会得到多组相异参数, 即模型的“异参同效”现象。因此, 无论是“以点代面”的区域模拟, 还是更大范围的模型同化应用, 都迫切需要更为科学精确的参数标定方法以及标定后参数不确定性的定量评价方法。

基于贝叶斯理论的 MCMC (Markov chain Monte Carlo, 马尔科夫链-蒙特卡洛) 方法能够结合观测数据和参数的先验知识得到模型参数的后验分布, 从而利用参数后验样本的均值或中值作为参数标定结果, 并且可以通过后验样本的均方根偏差定量的表达模型参数的不确定性。MCMC 方法在水文模型^[13-15]、森林管理模型^[16]、林冠蒸腾模型^[17]等模型参数标定研究中已得到较多应

收稿日期: 2018-02-06 修订日期: 2018-07-17

基金项目: 国家自然科学基金资助 (41671418, 41471342)

作者简介: 黄健熙, 男, 副教授, 博士, 博士生导师, 主要从事农业定量遥感等研究。

用, 并且显示出较强的适用性, 其在作物模型参数标定与不确定性评价中也逐渐得到应用^[18-20]。中国的作物模型应用面临地形复杂、气候多样空间异质性挑战, 因此模型参数也具有更大的不确定性。利用 MCMC 方法对中国小麦种植区作物模型参数较大范围的标定、不确定性评价以及比较目前还少有报道, 因此需要进一步研究探索。

本研究实现利用融入 snooker 更新算法的 DE-MC (differential evolution Markov chain, 差分进化马尔科夫链) 方法在 Python 环境下对 WOFOST 作物生长模型参数进行标定和不确定性分析。标准的 DE-MC 算法由 Braak 等^[21]于 2006 年提出, 并通过引入 snooker 更新算法进一步改进, 研究表明融入 snooker 更新的 DE-MC 算法能够通过至少 3 条的并行链实现对多达 50~100 维参数空间的后验估计^[22]。在河南郑州农业气象站点上, 结合多年实测 LAI (leaf area index, 叶面积指数) 和产量数据得到参数后验分布, 实现对模型参数的估计和不确定性评价。

1 数据及模型

1.1 研究数据

本文所用的试验观测数据来自河南郑州农业气象试验站 2008—2012 年的观测数据, 其中 2008—2011 年数据用于模型参数标定, 2012 年数据用于标定结果验证。研究区均位于种植区划中的黄淮冬麦区, 耕作模式主要为一年两熟, 小麦生育期降水约 120~140 mm, 播种期至成熟期大于 0 °C 积温 2 100~2 300 °C, 无霜期 180~200 d。冬小麦播种期一般为 10 月份上中旬, 成熟期一般为次年 5 月下旬到 6 月上旬。

作物数据主要包括从播种到成熟各个生育期的时间、生长状况、管理措施、土壤情况、产量及产量结构等。其中, LAI 观测值分别对应三叶期、分蘖期、越冬始期、返青期、拔节期、抽穗期和乳熟期 7 个生育时期; 模型运行所需气象数据来自中国科学院青藏高原研究所制作中国区域地面气象要素数据集^[23], 该数据是用于驱动陆面模型模拟的栅格数据集, 包括近地面气温、近地面气压、近地面空气比湿、近地面全风速、地面向下短波辐射、地面向下长波辐射与降水率共 7 个要素。

1.2 WOFOST 模型

WOFOST 模型是荷兰瓦赫宁根大学开发的一个机理性过程模型, 基于光截获和二氧化碳同化作为驱动过程来描述作物生长, 并将物候发育作为生长控制过程, 分别用 0、1、2 代表作物出苗、开花、成熟 3 个生育期 (development stage, DVS)。

本研究采用 Python 下的 PCSE 5.3.1 (python crop simulation environment, PCSE) 软件包进行 WOFOST 的运行。PCSE 提供了作物模型模拟的环境、读取数据 (如气象、土壤、农艺管理等) 的工具、模拟生物物理过程 (如物候、呼吸、蒸散发等) 的模块。相比于传统 FORTRAN 版本的作物模型, PCSE 是用纯 Python 代码编写, 使其具有更加灵活、易于修改和拓展的特性, 并可轻松连接数据库、图形用户界面、可视化工具和数值统计软件包。有关 PCSE 更详细的介绍可以参考网站: <https://pcse.readthedocs.org> 或者 <http://github.com/ajwdewit/pcse>。

1.3 标定参数选择

本研究利用 LAI 和产量观测值对模型参数进行标定,

与积温有关的作物参数直接依据作物物候发育日期与近地面气温计算得到。其余作物参数通过 Sobol 方法^[24-25]进行全局敏感性分析, 以 Sobol 全局敏感性指数大于 0.05 作为参数敏感的标准, 选取对不同生育期 LAI 以及成熟期产量敏感的参数作为待标定参数。

2 研究方法

2.1 MCMC 方法

贝叶斯理论能够结合模型参数的先验知识和模型输出对应的观测值, 实现对模型参数的后验估计。MCMC 方法即在贝叶斯理论框架下, 构造以参数后验分布为平稳分布的马尔科夫链, 从而得到参数的后验样本, 基于这些样本实现对参数数字特征的推断。贝叶斯理论用公式表达如下

$$p(\theta/y) = \frac{f(y/\theta)g(\theta)}{\int f(y/\theta)g(\theta)d\theta} \quad (1)$$

式中 θ 和 y 分别代表 WOFOST 模型的参数和模拟输出值 (如 LAI、产量等); $p(\theta/y)$ 为参数的后验概率密度函数; $f(y/\theta)$ 为观测值似然函数; $g(\theta)$ 为参数的先验分布。研究中, 将模型的驱动数据, 如气象数据、降水等, 记为 x , 则式 (1) 可以改为:

$$p(\theta/x, y) = \frac{f(y/\theta, x)g(\theta/x)}{\int f(y/\theta, x)g(\theta/x)d\theta} \propto f(y/\theta, x)g(\theta) \quad (2)$$

引入 MCMC 方法的目的是通过结合实测数据得到模型参数后验样本和并定量评价其不确定性, 研究中, 模型参数的不确定性用 MCMC 方法标定得到的参数后验样本的均方根偏差 RMSD 计算:

$$\text{RMSD}_\theta = \sqrt{\frac{\sum_{i=1}^{N-1} (\theta_i - \bar{\theta})^2}{N-1}} \quad (3)$$

式中 N 表示参数 θ 后验样本个数; θ_i 表示第 i 个后验参数; $\bar{\theta}$ 表示参数 θ 后验样本均值, 当 θ 服从高斯分布时, RMSD_θ 可以作为 θ 总体标准差的无偏估计。

为便于比较不同参数样本内部的离散程度, 对每个参数后验分别再计算相对均方根偏差 RRMSD, 定义为:

$$\text{RRMSD}_\theta = \frac{\text{RMSD}_\theta}{|\bar{\theta}|} \quad (4)$$

式中 RMSD_θ 表示参数 θ 的均方根偏差, $\bar{\theta}$ 表示参数 θ 后验样本均值。

2.2 DE-MC 算法及 snooker 更新

差分进化马尔科夫链 (differential evolution Markov chain, DE-MC) 算法^[21]通过多条平行链的运行, 以实现更好的探索参数空间。但标准 DE-MC 算法中并行链数必须大于参数空间的维数。为了提高采样效率, Braak 等^[22]在提出 DE-MC 算法的基础上, 引入 snooker 更新算法^[26-27]以部分取代其中的平行方向更新 (parallel direction update), 克服了标准 DE-MC 中并行链数必须大于维数的限制, 发展成为 DE-MC_(Z) 算法。本文基于这种思想实现算法的构建, 主要步骤为:

1) 初始化 $M_0 \times d$ 维的矩阵 Z , 即依据先验分布随机生成 M_0 组参数, 其中 $M_0 > \max(d, N)$, 然后复制矩阵 Z 中前 N 行, 记为种群 (population) X , 并设置 $M \leftarrow M_0$;

2) 设置一定的次数 K , 每逢 K 倍对 X 进行更新 (update)。

3) 将更新后的 X 添加到 Z 中, 即 $M \leftarrow M+N$;

4) 如果 X 已经达到收敛 (本研究通过比方差法 (variance ratio) [28] 进行判断), 或者总的更新次数超过了程序设定值, 则进行到第 5 步, 否则回到步骤 2;

5) 丢弃 Z 中一部分较早状态的样本 (burn in), 并进行一定比例的细化 (thin), 最终得到参数的后验样本。

其中每次对 X 的更新形成一个世代周期 (generation cycle), 并依次对链 x_1, x_2, \dots, x_N 进行更新, 首先随机生成 0 到 1 之间的随机数, 如果小于 0.1, 则进行 snooker 更新, 否则按照标准的 DE-MC 算法进行平行方向更新 [22]。

snooker 更新的步骤为:

1) 选择另外一条处于状态 z 的链;

2) 沿着 x_i-z 方向按如下步骤依概率密度采样,

① 选择另外 2 条分别处于状态 z_{R1} 和 z_{R2} 的随机链 R_1 和 R_2 ,

② 将 z_{R1} 和 z_{R2} 正射投影到 x_i-z 方向, 并分别生成 z_{P1} 和 z_{P2} ,

③ 提议 $x^* \leftarrow x_i + \gamma_s(z_{P1} - z_{P2})$, 其中 γ_s 为缩放因子, 取 1.2~2.2 之间的随机数,

④ 计算 Metropolis 系数: $r = \frac{\pi(x^*) \|x^* - z\|^{d-1}}{\pi(x_i) \|x_i - z\|^{d-1}}$,

⑤ 以概率 $\min(1, r)$ 接受提议, 否则保持 x_i 状态不变。

3 结果与分析

3.1 Sobol 敏感性分析结果

Sobol 全局敏感性分析结果表明, 对三叶期 LAI 敏感的参数包括 SLATB00 (specific leaf area at development stage of 0, 生育期为 0 时的比叶面积)、TBASEM (lower threshold temperature for emergence, 出苗的低温阈值)、TDWI (initial total crop dry weight, 作物初始干物质质量)、SLATB020 (specific leaf area at development stage of 0.2, 生育期为 0.2 时的比叶面积); 对分蘖期 LAI 敏感的参数包括 SLATB00、SLATB020、TBASEM、AMAXTB00 (maximum CO₂ assimilation rate at development stage of 0, 生育期为 0 时的单叶最大二氧化碳同化速率); 对越冬期 LAI 敏感的参数包括 SLATB00、SLATB020、TBASEM、AMAXTB00、TBASE (lower threshold temperature for aging of leaves, 叶龄的低温阈值)、TMNFTB_min (low temperature threshold when reduction factor of gross assimilation rate is 0, 总同化率降低因子为 0 时的低温阈值); 对返青期 LAI 敏感的参数包括 SLATB00、SLATB020、TMNFTB_min、TBASEM、TBASE、AMAXTB00; 对拔节期 LAI 敏感的参数包括 SLATB040 (specific leaf area at development stage of 0.4, 生育期为 0.4 时的比叶面积)、SLATB00、TBASE、SLATB020、AMAXTB00、SLATB070 (specific leaf area at development stage of 0.7, 生育期为 0.7 时的比叶面积)、TMNFTB_min、RGRLAI (maximum relative increase in LAI, 叶面积指数最大增长率); 对抽穗期 LAI 敏感的参

数包括 SLATB070、SLATB040、SLATB00、TBASE、KDIFTB100 (extinction coefficient for diffuse visible light at development stage of 1, 生育期为 1 时的散射消光系数)、AMAXTB00、SLATB020、AMAXTB100 (maximum CO₂ assimilation rate at development stage of 1, 生育期为 1 时的单叶最大二氧化碳同化速率); 对乳熟期 LAI 敏感的参数包括 SPAN (life span of leaves growing at 35 Celsius, 在 35 °C 时叶面积的生命周期)、SLATB070、KDIFTB100; 对成熟期产量敏感的参数 SPAN、CVO (efficiency of conversion into storage organs, 储存器官的同化物转换效率)、AMAXTB130 (maximum CO₂ assimilation rate at development stage of 1.3, 生育期为 1.3 时的单叶最大二氧化碳同化速率)、SLATB070、AMAXTB100、KDIFTB100、SLATB040。

上述敏感性分析是以 2008 年郑州农业气象试验站上数据运行分析得到的, 由于 TBASEM 的改变会引起生育期在时间上的“漂移”, 因而设置该参数为默认值, 不进行标定。最终综合敏感性分析结果, 选取 SLATB00、SLATB020、AMAXTB00、SLATB070、TBASE、AMAXTB100、KDIFTB100、SLATB040、TMNFTB_min、SPAN、CVO、AMAXTB130、RGRLAI、TDWI 共计 14 个参数作为待标定参数。

3.2 MCMC 标定

3.2.1 参数先验分布

对 14 个参数设置为以模型默认值为初始值、在参数范围内的均匀分布。参数初始值及范围如表 1。

表 1 待优化的 14 个参数的初始值及范围
Table 1 Initial value and range of 14 parameters for optimization

参数 Parameters	定义 Definition	初始值 Initial value	范围 Range
SLATB00	生育期为 0 时的比叶面积/(hm ² ·hm ⁻²)	0.002 12	[0.0007,0.0042]
SLATB020	生育期为 0.2 时的比叶面积/(hm ² ·hm ⁻²)	0.002 12	[0.0007,0.0042]
AMAXTB00	生育期为 0 时的单叶最大二氧化碳同化速率/(kg·hm ⁻² ·h ⁻¹)	35.83	[20,70]
SLATB070	生育期为 0.7 时的比叶面积/(hm ² ·hm ⁻²)	0.002 12	[0.0007,0.005]
TBASE	叶龄的低温阈值/°C	0	[-5,5]
AMAXTB100	生育期为 1 时的单叶最大二氧化碳同化速率/(kg·hm ⁻² ·h ⁻¹)	38.83	[20,70]
KDIFTB100	生育期为 1 时的散射消光系数	0.60	[0.44,1]
SLATB040	生育期为 0.4 时的比叶面积/(hm ² ·hm ⁻²)	0.002 12	[0.0007,0.0035]
TMNFTB_min	TMNF 为 0 时的低温阈值/°C	0	[-7,1]
SPAN	在 35 °C 时叶面积的生命周期/d	31.3	[17,50]
CVO	储存器官的同化物转换效率/(kg·kg ⁻¹)	0.709	[0.45,0.85]
AMAXTB130	生育期为 1.3 时的单叶最大二氧化碳同化速率/(kg·hm ⁻² ·h ⁻¹)	38.83	[20,70]
RGRLAI	叶面积指数最大增长率/(hm ² ·hm ⁻² ·d ⁻¹)	0.008 17	[0.007,0.02]
TDWI	作物初始干物质重/(kg·hm ⁻²)	210	[150,300]

3.2.2 似然函数的确定

本研究中认为 LAI 和产量服从以观测值为期望的高斯分布, 由于概率密度可能非常小, 为了避免位数舍入引起的误差, 所有概率密度的计算均通过取对数形式计算, 由此确立似然函数:

$$\log L_{LAI} = -0.5(x - x_{obs})^T \Sigma^{-1}(x - x_{obs}) - 0.5K \log(2\pi) - \log(\det \Sigma) \quad (5)$$

$$\log L_{yield} = -0.5 \left(\frac{Y - Y_{obs}}{\sigma} \right)^2 - 0.5 \log(2\pi) - \log \sigma \quad (6)$$

$$\log L = \log L_{LAI} + \log L_{yield} \quad (7)$$

式中 \log 表示取对数; L_{LAI} 和 L_{yield} 分别表示 LAI 和产量的似然函数; L 表示似然函数 (likelihood function) 的返回值; 矢量 x 和 x_{obs} 分别表示对应不同生育期 LAI 的模型模拟值和观测值; Σ 表示 LAI 观测值的协方差矩阵, 本研究中不同生育期 LAI 观测值相互独立, 且方差均取

0.2; K 表示空间维数, 即 LAI 观测值的个数; $\det \Sigma$ 表示 Σ 的行列式值; Y 和 Y_{obs} 分别表示成熟期产量的模型模拟值和观测值; σ 表示产量观测值的标准差, 本研究中取观测值的 10%。

3.2.3 参数的后验分布

本研究中设定 DE-MC 种群 (并行链) 数目为 3, 每 5 次采样进行一次种群进化 (链的更新), 每 5000 次迭代通过比方差法进行一次收敛性判断, 计算诊断指标 \hat{R} [28], $\hat{R} > 1$ 时, 表明链没有达到收敛, 若 $\hat{R} \approx 1$, 表明链已经处于静止状态。本研究中当连续超过 3 次诊断指标 $\hat{R} < 1.05$, 则认为马尔科夫链达到收敛。收敛后舍弃前 2000 次采样 (burn-in), 为解决自相关问题, 并每隔 5 次对链进行细化, 最终得到参数的后验样本及其分布。图 1 为 14 个参数先验及后验样本的分布图。表 2 列出了最终得到参数后验样本的均值、中值、95% 置信区间、RMSD 以及 RRMSD, 结果均保留小数点后 4 位。

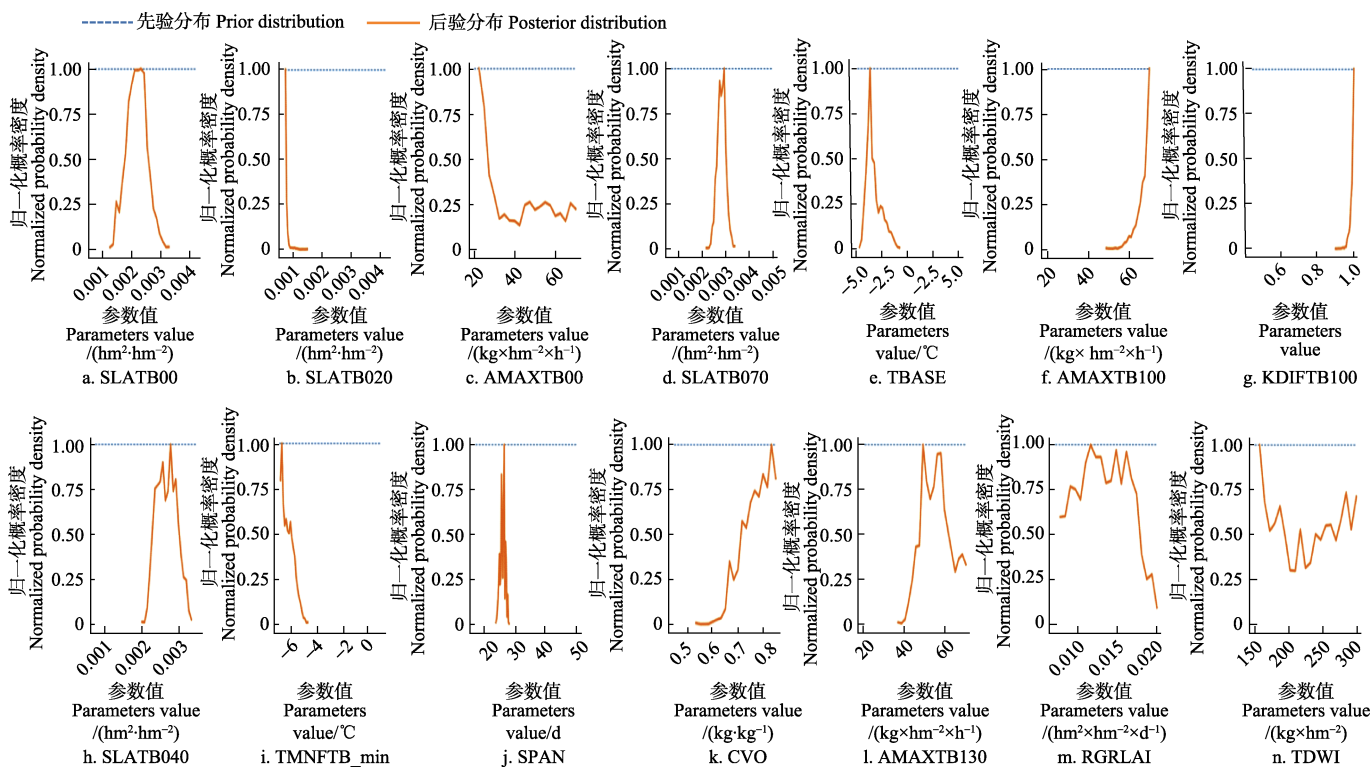


图 1 14 个参数的先验及后验分布

Fig.1 Prior and posterior distribution of 14 parameters

从图 1 可以看出, 相对于各参数的先验分布, 参数的后验分布均出现较大改变, 表明标定参数选取的有效性。由于 WOFOST 模型无法准确模拟越冬期小麦实际生长情况, 因而在标定参数包含越冬期前后 LAI 观测值的情况下, 部分参数的后验分布表现出极端变化, 表现为: SLATB020 集中分布在参数区间的左边界, 这是因为模型运行时, 比叶面积与 LAI 模拟值成正比例关系, 生育期为 0.2 较为接近越冬期, 由于 WOFOST 无法模拟低温导致的叶片损伤减少, 其 LAI “模拟期望值” 很大程度上高于实际 LAI 值, 因而在实际 LAI 观测值的标定下, 导致该时期比叶面积极端减小以使模型模拟出较小的 LAI 值。同理, 生育期为 1 时大致处于抽穗期附近, 此阶段 LAI

值基本处于整个生育期的峰值, 因而在抽穗期 LAI 峰值的影响下, AMAXTB100 和 KDIFTB100 集中分布在参数区间的右边界。

除个别极端分布的参数, 表 2 中参数的后验样本均值和中值基本相一致。后验参数的 95% 置信区间一定程度上反映了参数的合理取值区间, 与表 1 中参数的初始范围相比, 其范围缩小的越多, 也一定程度表明该参数标定后的不确定越小。RMSD 和 RRMSD 描述了参数样本内部相对于均值的离散程度, 可作为参数不确定性的定量评价。在参数符合高斯分布时, 参数样本的 RMSD 是参数总体高斯分布的标准差 σ 的无偏估计。由图 1 分布, 可近似认为 SLATB00、SLATB070、SLATB040、SPAN、

AMAXTB130 符合高斯分布，以 RRMSD 为指标，不确定性由小到大依次为 SPAN、SLATB070、SLATB040、AMAXTB130、SLATB00。

表 2 参数后验样本的均值、中值、95%置信区间、RMSD、RRMSD

Table 2 Mean, median, 95% confidence interval, RMSD and RRMSD of parameters posterior sample

参数 Parameters	平均值 Mean	中值 Median	95%置信区间 95% confidence interval	均方根 偏差 RMSD	相对均方 根偏差 RRMSD /%
SLATB00	0.002 1	0.002 1	[0.0015,0.0027]	0.000 3	15.99
SLATB020	0.000 7	0.000 7	[0.0007,0.0008]	0.000 1	8.89
AMAXTB00	39.039 2	35.154 3	[20.7706,67.2242]	16.036 8	41.08
SLATB070	0.002 8	0.002 8	[0.0025,0.0031]	0.000 2	6.17
TBASE	-3.471 4	-3.697 3	[-4.3592,-2.0130]	0.724 9	20.88
AMAXTB100	66.716 7	67.725 7	[59.9660,69.8827]	3.150 2	4.72
KDIFTB100	0.988 9	0.992 8	[0.9653,0.9995]	0.011 1	1.12
SLATB040	0.002 6	0.002 6	[0.0022,0.0031]	0.000 3	10.25
TMNFTB_min	-6.359 7	-6.425 2	[-6.9471,-5.5272]	0.457 7	7.2
SPAN	25.787 1	25.867 4	[24.3982,26.9296]	0.800 9	3.11
CVO	0.766 3	0.773 5	[0.6625,0.8421]	0.056	7.31
AMAXTB130	54.605 4	54.430 2	[43.9446,67.7776]	7.105 5	13.01
RGLRLAI	0.012 8	0.012 7	[0.0078,0.0178]	0.003 2	24.71
TDWI	223.691 2	226.067 8	[153.9404,295.5126]	47.632 2	21.29

3.3 标定运行结果

3.3.1 标定精度验证

以 2012 年生长期站点上气象数据为驱动，分别以参数初始值、后验均值和中值带入模型运行，与观测值进行比较，结果如图 2 所示。计算 LAI 模拟值和对应观测值间的均方根误差 (RMSE) 作为精度评价指标，公式为：

$$RMSE_{LAI} = \sqrt{\frac{\sum_{i=1}^n (LAI_i - LAI_i^{obs})^2}{n}}$$
 (8)

式中 LAI_i 表示第 i 个 LAI 模拟值， LAI_i^{obs} 表示第 i 个对应的观测值， n 表示 LAI 观测值个数。结果显示：参数初始值（模型默认参数）、后验均值和后验中值带入模型模拟的 LAI 值的 RMSE 分别为 1.79、0.84、0.87，相比于初始值，标定后 LAI 模拟精度分别提高了 53.07%（后验均值）和 51.40%（后验中值）；产量模拟精度从初始值的 75.15% 提高至 83.40%（后验均值）和 84.03%（后验中值），模拟精度提高 8.25%~8.88%。这表明利用 MCMC 标定得到的后验样本均值和中值在很大程度上能够实现对观测数据的拟合。此外，由图 2 可以发现，相比于初始状态，标定后并非在每个观测节点上都提高了精度，但是却在整体上提高了与观测数据的拟合精度。标定后的 LAI 曲线较好的捕捉到了 LAI 的峰值，但是在越冬期前后的 LAI 模拟值具有较大误差，主要是由于 WOFOST 模型开发之初主要是为了模拟欧洲的作物生长，并未实现中国冬小麦越冬期情况模拟^[29]。

标定后的产量模拟精度较初试状态有所提高，但相对观测值仍有一定误差，这很大程度上是因为用于标定的 2008—2011 年的产量数据仅仅是 4 个标量值，并不能

够充分实现对产量形成相关参数的准确标定，进一步研究可以引入生育期内的穗重数据进行标定，以提高产量标定精度。

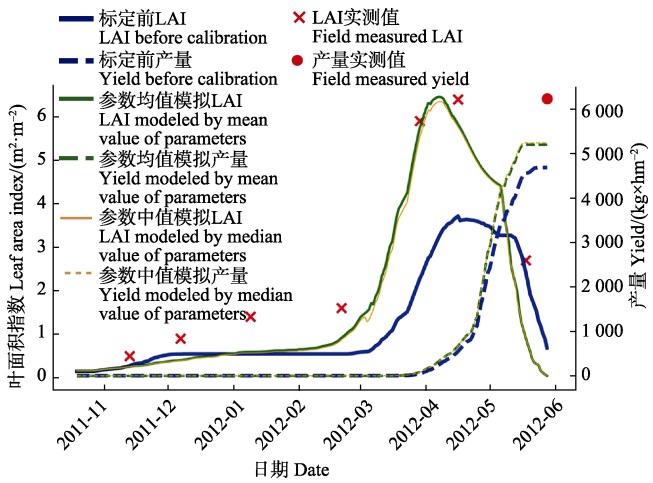


图 2 参数标定前后模型模拟结果对比
Fig.2 Comparison of model simulated results before and after parameters calibration

3.3.2 模型输出的不确定性评价

将后验样本作为参数集合，分别带入 WOFOST 模型并输出得到 LAI 和产量时间序列上的模拟集合，结果如图 3 所示。由阴影部分的走势可以直观的反映出后验参数集模拟结果所代表的不确定性，整体上 LAI 模拟的不确定性在三叶期至返青期之间以及 LAI 曲线峰值（拔节期至抽穗期之间）较大，其它时期相对稳定。产量模拟的不确定性在开始模型模拟籽粒形成开始逐渐扩大，并在乳熟期达到最大，至成熟期基本趋于稳定。

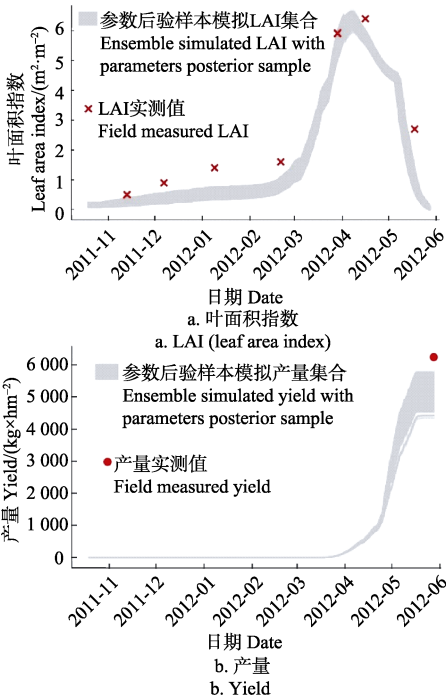


图 3 参数标定后模型模拟 LAI 和产量的不确定性
Fig.3 Uncertainty of model simulated LAI and yield with calibrated parameters dataset

4 结论

本研究以郑州农业气象试验站 2008—2011 年观测数据作为标定数据,选取 14 个敏感参数,运用融入 snooker 更新的 DE-MC 方法实现参数的标定,并用 2012 年的观测数据进行了验证,得到结论如下:

1) 以模型默认参数为初始值,通过 DE-MC 方法可以实现参数的自动标定,相比于模型默认参数模拟结果,参数标定后 LAI 模拟精度可提高 51.40%~53.07%,产量模拟精度提高 8.25%~8.88%。

2) 以生育期内 LAI 观测值和成熟期产量为标定数据,SPAN、SLATB070、SLATB040、AMAXTB130 和 SLATB00 的后验分布可近似为高斯分布,其中 SPAN 的不确定性最低。

3) 带入后验参数集合进行模型, LAI 在三叶期至返青期之间以及拔节期至抽穗期之间模拟的不确定性较大;产量模拟的不确定性随时间不断增大,至乳熟期前后达到稳定。

本研究使用的标定数据为农业气象站点上的观测数据,虽然作物品种一致,但年际间的栽培管理措施并未严格进行控制(如播种密度等),导致标定结果的精度可能受到一定影响。后续可在更多地区,采用具有定量控制的实验数据开展研究,以进一步验证本方法的有效性。

[参 考 文 献]

- [1] Wang H, Zhu Y, Li W, et al. Integrating remotely sensed leaf area index and leaf nitrogen accumulation with RiceGrow model based on particle swarm optimization algorithm for rice grain yield assessment[J]. *Journal of Applied Remote Sensing*, 2014, 8(1): 83674.
- [2] Huang J, Tian L, Liang S, et al. Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model[J]. *Agricultural & Forest Meteorology*, 2015, 204: 106—121.
- [3] He B, Li X, Quan X, et al. Estimating the aboveground dry biomass of grass by assimilation of retrieved LAI into a crop growth model[J]. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 2015, 8(2): 550—561.
- [4] 王东伟, 王锦地, 梁顺林. 作物生长模型同化 MODIS 反射率方法提取作物叶面积指数[J]. *中国科学: 地球科学*, 2010(1): 73—83.
Wang Dongwei, Wang Jindi, Liang Shunlin. Assimilating MODIS reflectance into crop growth model to extract crop leaf area index [J]. *Scientia Sinica Terrae*, 2010(1): 73—83. (in Chinese with English abstract)
- [5] Fang H, Liang S, Hoogenboom G, et al. Corn-yield estimation through assimilation of remotely sensed data into the CSM-CERES-Maize model[J]. *International Journal of Remote Sensing*, 2008, 29(10): 3011—3032.
- [6] 张建平, 赵艳霞, 王春乙, 等. 基于 WOFOST 作物生长模型的冬小麦干旱影响评估技术[J]. *生态学报*, 2013, 33(6): 1762—1769.
Zhang Jianping, Zhao Yanxia, Wang Chunyi, et al. Evaluation technology on drought disaster to yields of winter wheat based on WOFOST crop growth model[J]. *Acta Ecologica Sinica*, 2013, 33(6): 1762—1769. (in Chinese with English abstract)
- [7] 马玉平, 王石立, 李维京. 基于作物生长模型的东北玉米冷害监测预测[J]. *作物学报*, 2011, 37(10): 1868—1878.
Ma Yuping, Wang Shili, Li Weijing. Monitoring and predicting of maize chilling damage based on crop growth model in Northeast China[J]. *Acta Agronomica Sinica*, 2011, 37(10): 1868—1878. (in Chinese with English abstract)
- [8] Jin M, Liu X, Wu L, et al. An improved assimilation method with stress factors incorporated in the WOFOST model for the efficient assessment of heavy metal stress levels in rice[J]. *International Journal of Applied Earth Observations & Geoinformation*, 2015, 41: 118—129.
- [9] 李振海. 基于遥感数据和气象预报数据的 DSSAT 模型冬小麦产量和品质预报[D]. 杭州: 浙江大学, 2016.
Li Zhenhai. Predicting Winter Wheat Yield and Quality by Integrating of Remote-sensing Data and the Weather Forecast Data Into the DSSAT Model[D]. Hangzhou: Zhejiang University, 2016. (in Chinese with English abstract)
- [10] Huang J, Sedano F, Huang Y, et al. Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation[J]. *Agricultural & Forest Meteorology*, 2016, 216: 188—202.
- [11] Ma H, Huang J, Zhu D, et al. Estimating regional winter wheat yield by assimilation of time series of HJ-1 CCD NDVI into WOFOST-ACRM model with Ensemble Kalman Filter[J]. *Mathematical & Computer Modelling*, 2013, 58(3-4): 753—764.
- [12] Morel J, Todoroff P, Bégué A, et al. Toward a satellite-based system of sugarcane yield estimation and forecasting in smallholder farming conditions: A case study on Reunion Island[J]. *Remote Sensing*, 2014, 6(7): 793708.
- [13] Rwasoka D T, Madamombe C E, Gumindoga W, et al. Calibration, validation, parameter indentifiability and uncertainty analysis of a 2-parameter parsimonious monthly rainfall-runoff model in two catchments in Zimbabwe[J]. *Physics & Chemistry of the Earth Parts A/b/c*, 2014, 67-69(12): 36—46.
- [14] Hassan A E, Bekhit H M, Chapman J B. Using Markov Chain Monte Carlo to Quantify Parameter Uncertainty and Its Effect on Predictions of a Groundwater Flow Model[M]. Elsevier Science Publishers B. V., 2009.
- [15] Gallagher M, Doherty J. Parameter estimation and uncertainty analysis for a watershed model[J]. *Environmental Modelling & Software*, 2007, 22(7): 1000—1020.
- [16] Van O M, Reyser C, Bohn F J, et al. Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe[J]. *Forest Ecology & Management*, 2013, 289(2): 255—268.
- [17] Li X, Yang P, Ren S, et al. An improved canopy transpiration model and parameter uncertainty analysis by Bayesian approach[J]. *Mathematical & Computer Modelling An International Journal*, 2010, 51(11, 12): 1368—1374.
- [18] Toshichika I, Masayuki Y, Motoki N. Parameter estimation and uncertainty analysis of a large-scale crop model for paddy rice: Application of a Bayesian approach[J]. *Agricultural & Forest Meteorology*, 2009, 149(2): 333—348.
- [19] Marin F, Jones J W, Boote K J. A Stochastic Method for Crop Models: Including Uncertainty in a Sugarcane Model[J]. *Agronomy Journal*, 2017, 109(2).
- [20] 何亮, 侯英雨, 赵刚, 等. 基于全局敏感性分析和贝叶斯方法的 WOFOST 作物模型参数优化[J]. *农业工程学报*, 2016, 32(2): 169—179.
He Liang, Hou Yingyu, Zhao Gang, et al. Parameters optimization of WOFOST model by integration of global sensitivity analysis and Bayesian calibration method[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2016, 32(2): 169—179. (in Chinese with English abstract)

- [21] Braak C J F T. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces[J]. *Statistics & Computing*, 2006, 16(3): 239–249.
- [22] Braak C J T, Vrugt J A. Differential evolution Markov chain with snooker updater and fewer chains[J]. *Statistics & Computing*, 2008, 18(4): 435–446.
- [23] 何杰, 阳坤. 中国区域高时空分辨率地面气象要素驱动数据集[Z]. 寒区旱区科学数据中心, 2011.
- [24] Sobol, I. M. *Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates*[M]. Elsevier Science Publishers B. V., 2001.
- [25] Zhang X Y, Trame M N, Lesko L J, et al. Sobol sensitivity analysis: A tool to guide the development and evaluation of systems pharmacology models[J]. *Cpt Pharmacometrics & Systems Pharmacology*, 2015, 4(2): 69.
- [26] Gilks W R, Roberts G O, George E I. Adaptive direction sampling[J]. *Journal of the Royal Statistical Society*, 1994, 43(1): 179–189.
- [27] Liang F, Wong W H. Real-parameter evolutionary Monte Carlo with applications to Bayesian mixture models[J]. *Publications of the American Statistical Association*, 2001, 96(454): 653–666.
- [28] Brooks S P, Roberts G O. Convergence assessment techniques for Markov chain Monte Carlo[J]. *Statistics & Computing*, 1998, 8(4): 319–335.
- [29] 马玉平, 王石立, 张黎. 针对华北小麦越冬的 WOFOST 模型改进[J]. *中国农业气象*, 2005, 26(3): 145–149.
Ma Yuping, Wang Shili, Zhang Li. Study on improvement of WOFOST against overwinter of wheat in North China[J]. *Chinese Journal of Agrometeorology*, 2005, 26(3): 145–149. (in Chinese with English abstract)

Markov Chain Monte Carlo based WOFOST model parameters calibration and uncertainty analysis

Huang Jianxi^{1,2}, Huang Hai¹, Ma Hongyuan¹, Li Ying^{3,4}, Hou Yingyu⁵, He Liang⁵, Zhu Dehai^{1,2}

(1. *College of Land Science and Technology of China Agricultural University, Beijing 100083, China*; 2. *Key Laboratory of Remote Sensing for Agri-Hazards, Ministry of Agriculture, Beijing 100083, China*; 3. *Henan Key Laboratory of Agrometeorological Support and Applied Technique, China Meteorological Administration, Zhengzhou 450003, China*; 4. *Henan Institute of Meteorological Sciences, Zhengzhou 450003, China*; 5. *National Meteorological Center, China Meteorological Administration, Beijing 100081, China*)

Abstract: It is generally accepted that the model parameters calibration is an essential step before application. However, the traditional methods of obtaining a set of “optional parameters” based on a certain number of observations, fail to represent the uncertainties and get reliable estimates. To overcome this problem, we introduced the DE-MC (differential evolution Markov chain) algorithm with snooker update into the WOFOST (world food studies) model parameters calibration. The main objectives were: 1) to calibrate the WOFOST model by DE-MC algorithm with snooker update; 2) to evaluate the uncertainty of the model parameters after calibration; 3) to evaluate the uncertainty of the model outputs after calibration. Observational data including LAI (leaf area index) in different growth stages and the final yield of winter wheat in Zhengzhou Agrometeorological Experimental Station, were used to calibrate WOFOST model at potential mode with this algorithm. The crop parameters related to accumulated temperature were calculated directly from the crop phenological development date and near-surface temperature. The remaining crop parameters were analyzed by Sobol method. Taking Sobol global sensitivity index greater than 0.05 as the threshold for sensitive parameters, 14 parameters were then selected. The calibrated parameters were defined as the uniform distribution over their value interval. The likelihood function represented the mismatch of the model output with the measured observations, by which the parameters’ priori distribution converted to posterior distribution. The likelihood function for yield was set as a Gaussian distribution with the observational data as expected value and 10% of the observational data as standard deviation. Similarly, the likelihood function for LAI was set as a multidimensional Gaussian distribution with observational LAI as the expected vector and a diagonal matrix as the covariance matrix. At last, we found that: 1) Compared with the simulation results with model default parameters, the LAI simulation accuracy could be increased by 51.40%-53.07% after the parameter calibration, and the yield simulation accuracy is improved by 8.25%-8.88%; 2) The posterior distribution of life span of leaves growing at 35 Celsius (SPAN), specific leaf area at development stage of 0.7 (SLATB070), specific leaf area at development stage of 0.4 (SLATB040), maximum CO₂ assimilation rate at development stage of 1.3 (AMAXTB130), and specific leaf area at development stage of 0 (SLATB00) could be approximated as a Gaussian distribution with SPAN having the minimum uncertainty; 3) Running model with the posterior parameters set, the uncertainty of simulated LAI from the three-leaf stage to the re-greening stage and from the jointing stage to the heading stage is larger; the uncertainty of the simulated yield increases with time before the milky ripe stage, stayed unchanged until maturity. It was concluded that DE-MC with snooker update was an effective algorithm for parameters estimation of WOFOST model, the calibration had the potential to reduce the uncertainty of the model parameters and the calibrated model was able to model the observational data with some degree of skill. The calibration data used in this study was the observation data from the agrometeorological site. Although the crop varieties were consistent, the inter-annual cultivation management measures were not strictly controlled (eg.: seeding density), and the accuracy of the calibration results might be affected. Subsequent experiments can be carried out in more areas using experimental data with quantitative control to further verify the effectiveness of this algorithm.

Keywords: meteorology; models; Crops; MCMC method; WOFOST model; parameters calibration; uncertainty analysis