

基于贝叶斯推断的 DNDC 模型参数校正与不确定性评价研究*

秦发侶^{1,2} 赵永存^{1,2†} 史学正^{1,2} 于东升^{1,2} 徐胜祥^{1,2}

(1 土壤与农业可持续发展国家重点实验室(中国科学院南京土壤研究所), 南京 210008)

(2 中国科学院大学, 北京 100049)

摘要 实现土壤有机碳(SOC)动态模拟结果的不确定性定量评价是农田管理决策成败的关键之一。采用贝叶斯推断和马尔科夫链蒙特卡洛方法(MCMC),对 DNDC 模型模拟江苏省宜兴市一个具有 22 年稻-麦轮作历史的田块的 SOC 动态变化,进行模型参数校正和模拟结果的不确定性定量评价。结果表明, DNDC 模型适宜长期监测田块的 SOC 动态变化的模拟,但模型的模拟结果存在一定的不确定性;在 DNDC 模型输入参数数据质量不明的情况下,利用贝叶斯推断和 MCMC 方法能够有效地实现模型输入参数的自动校正和 SOC 模拟结果不确定性的定量评价,从而为实现区域或国家尺度农田 SOC 动态模拟的不确定性定量评价提供理论和方法依据。

关键词 贝叶斯推断;马尔科夫链蒙特卡洛方法;参数校正;不确定性评价;DNDC 模型

中图分类号 S11+4 **文献标识码** A

基于过程的 DNDC (Denitrification-Decomposition) 模型已经被广泛应用于中国农田土壤有机碳(SOC)的变化模拟及预测^[1-5],然而,在模拟和预测过程中不可避免地存在不确定性^[6-8]。一般而言,模型的不确定性是由于模型对复杂世界的抽象、概括而形成的结构误差和模型使用时输入参数的不确定性导致的^[7-8]。对于模型的使用者而言,模型结构是难以改变的,而由于模型输入参数的可获取性、数据质量及模型输入参数本身的变异性等因素所导致的输入参数的不确定性也会随着 DNDC 模拟 SOC 动态变化的过程加以传递,从而对 SOC 动态模拟结果产生深刻影响。因此,在模型结构确定的情况下定量 DNDC 模型输入参数的不确定性及其对 SOC 动态变化模拟结果的潜在影响,对于科学认知 SOC 动态变化规律、建立合理的推荐耕作管理措施,进而实现农田土壤肥力提升和农田土壤固碳具有重要的意义。

SOC 动态模拟结果中所包含的不确定性对农田及土壤资源管理决策具有重要影响,提供这些不确定性信息也是使用和交流模型结果的关键环

节^[6-9]。然而,目前关于 DNDC 模型模拟 SOC 动态的不确定性定量评价研究还相对较少,已有的研究多集中在模拟温室气体(如 CH₄, CO₂, N₂O 等)排放的不确定性评价方面^[6,9-10],相关的不确定性评价方法主要包括最敏感因子法(MSF)和蒙特卡洛法(MC),其中 MC 又可分为 DNDC 内置及非 DNDC 内置(用户自定义)两类。如 Li 等^[10]使用 MSF 和 DNDC 内置 MC 方法对比了中国、泰国和美国的三个农田长期定位点所属土壤图斑由于初始 SOC、土壤容重、pH 和黏粒含量这四个输入参数的不确定性所造成的 DNDC 模型模拟温室气体排放的置信区间差异;Hastings 等^[6]采用自定义的 MC 研究了位于瑞士的一个农田长期定位试验点土壤属性、作物管理措施和气候等的不确定性对 DNDC 模型模拟温室气体排放的影响。DNDC 内置的 MSF 方法假定对模型输出结果影响最大的 4 个参数分别为初始 SOC 含量、土壤容重、pH 及黏粒含量,同时假定 DNDC 输出结果与这 4 个参数之间的关系为线性正相关或负相关^[11]。MSF 方法将这 4 个参数的最小值和最大值组合成两个参数集,并将它们分别输入

* 中国科学院知识创新工程重要方向项目(KZCX2-EW-QN404)和中国科学院战略性先导科技专项(XDA05050509)资助

† 通讯作者, E-mail: yczhao@issas.ac.cn

作者简介: 秦发侶(1986—),男,云南玉溪人,博士研究生,主要从事土壤资源与遥感信息研究。E-mail: dielihuared@hotmail.com

收稿日期: 2013-05-20; 收到修改稿日期: 2013-09-19

DNDC 模型进行模拟,以输出结果的最小值和最大值来定量模拟结果的不确定性区间。如 Li 等^[10]认为初始 SOC、pH 的最大值和黏粒含量的最小值组合所得的参数集可以模拟得到 CH₄ 排放量的最大值;相反,初始 SOC、pH 的最小值和黏粒含量的最大值组合所得的参数集可以模拟得到 CH₄ 排放量的最小值。MC 方法则较 MSF 方法更灵活,可以自由选择模型输入参数进行评价,其实施步骤包括:(1)选取关注的模型输入参数;(2)设定输入参数的变化区间及概率分布;(3)依据模型输入参数的概率分布抽样,获得 N 个模型输入参数集;(4)将 N 个模型输入参数集输入模型获得 N 个模型输出结果;(5)对 N 个模型输出结果进行统计分析,用置信区间来表征模型输出结果的不确定性^[9]。DNDC 内置的 MC 方法默认输入参数均为均匀分布,即每个参数集对模型输出结果贡献的权重是相等的。用户自定义的 MC 方法则可根据模型输入参数实测数据的统计特征来建立相应的概率分布函数(比如正态分布等)。理论上讲,当模型输入参数的实测数据较为充足时,由于根据实测数据统计推断出的模型输入参数的不确定性区间较为接近输入参数的真实不确定性区间及概率分布,因此能得到较 DNDC 内置的 MC 方法更为可靠的模拟结果置信区间。然而,当模型输入参数的实测数据极为有限时,根据实测数据来统计模型输入参数的可能变化范围变得十分困难。针对此种情况,常用的方法是根据专家经验,或者采用与研究区域类似、实测数据较为齐全的其他区域的输入参数变化区间来代替此区域的输入参数的不确定性区间。此时,所得的输入参数的不确定性区间的准确度并不能得到保证,因此,若基于此输入参数变化区间使用 MSF 和 MC 方法定量 DNDC 模型的输出结果的置信区间,则该置信区间的可靠性有待进一步评估。

贝叶斯推断使用实测值来校正和改进模型参数,该方法整合了模型参数的先验概率分布和模拟结果的似然性评价,可以定量后验参数和模型输出结果的不确定性。在模型输入参数的实测数据较少的情况下,以马尔科夫链蒙特卡洛(MCMC)模拟实现的贝叶斯推断提供了一种潜在有效的不确定性评价方法。目前,该方法已被成功用于水文、生态模型的不确定性评价^[12-14],但在 DNDC 模拟 SOC 的不确定性评价方面还未见报道。本研究基于贝叶斯推断,采用 MCMC 模拟实现的贝叶斯推断对 DNDC 模型模拟江苏省宜兴市水田长期定位观测点

SOC 变化的输入参数进行校正,并定量评价了 SOC 模拟结果的不确定性置信区间,旨在为区域或国家尺度农田 SOC 动态模拟的不确定性定量评估提供理论和方法依据。

1 材料与方法

1.1 DNDC 模型简介

DNDC 模型由美国新罕布什尔州州立大学李长生教授领导的团队开发,具有 SOC 动态模拟能力,是目前被广泛接受的 SOC 动态模拟模型之一。该模型将 SOC 库分为植物残遗体、微生物、活性和惰性等 4 个库,每个库又分为两至三个具有不同分解速率的子库。由用户提供的气候、土壤、作物和耕作管理措施 4 个模块的输入参数共同驱动,完成从作物生长,植物残遗体进入土壤,碳库分配及各子碳库分解过程的模拟。在这些过程中,土壤温度、湿度、pH、Eh 和各种基质(如可溶性有机碳(DOC)等)共同决定了各子碳库的大小及其特殊的分解速率,从而决定了 SOC 的动态变化过程^[11]。关于 DNDC 模型的详细介绍,可参考文献[15-17]。

1.2 数据来源

DNDC 模型的验证数据来源于江苏省宜兴市的一个土壤类型为漂洗型水稻土的农田长期定位观测点(E 119°30'00", N 31°10'00")。该观测点记录了 1986-2007 年的 22 年间稻-麦轮作试验数据,包括 1986 年的土壤基本属性(表 1)和 22 年间逐年的 SOC 含量(图 1)、水稻小麦产量、化肥及有机肥施用量。气象数据包括 1986-2007 年日最高气温、日最低气温和日降水量均来自中国气象局国家气象信息中心构建的气象数据共享服务网(<http://cdc.cma.gov.cn/>)。本研究先将表 1 中的实测值及相应的气象数据、作物管理数据输入 DNDC 模型进行 1986-2007 年间该观测点的 SOC 动态变化模拟,然后将模拟结果与实测的逐年 SOC 数据进行对比以检验 DNDC 模拟该观测点 SOC 动态变化的可行性。

MCMC 方法的先验数据来源于《中国土种志》中记载的 55 个漂洗型水稻土剖面信息。为了检验 MCMC 方法在输入数据先验知识有限、数据质量不明条件下的 DNDC 模型输入参数校正效果,本文将各输入参数的先验概率密度分布统一设定为最简单的均匀分布,分布的上下限则根据 55 个剖面信息统计所获得的土壤属性的最大值和最小值来设定

(表 1)。以 MCMC 方法获得 SOC 的后验不确定性区间是否覆盖实测的 22 年间逐年的 SOC 含量来检验模型输出结果不确定性区间的可靠性;同时,将

MCMC 方法所获得的输入参数的不确定性区间与表 1 的 6 个输入参数实测值进行比较,检验 MCMC 方法的输入参数推断结果是否准确。

表 1 DNDC 模型 6 个输入参数的实测值、先验不确定性区间的统计特征

Table 1 Statistics of on-site-observed values and priori uncertainty intervals of the six DNDC model input parameters

		初始 SOC Initial SOC (g kg ⁻¹)	黏粒含量 Clay content (%)	容重 Bulk density (g cm ⁻³)	pH	秸秆还田量 Crop residue returned to soil (%)	灌溉指数 Irrigation index
实测值 Site-observed value		9.3	34	1.21	7.0	5	0.9
先验不确定性区间 Priori uncertainty interval	最小值 Minimum	9.1	3	0.98	5.5	5	0.7
	均值 Mean	16.7	33	1.33	6.7	23	0.85
	最大值 Maximum	24.3	62	1.67	7.8	40	1

1.3 DNDC 模型模拟 SOC 动态变化可行性检验标准

采用相关系数(r)和均方根误差(RMSE)来评价模拟结果,其计算方法分别如下:

$$r = \frac{\sum_{i=1}^n (y_i^o - \bar{y}^o)(y_i^m - \bar{y}^m)}{\sqrt{[\sum_{i=1}^n (y_i^o - \bar{y}^o)^2][\sum_{i=1}^n (y_i^m - \bar{y}^m)^2]}} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i^m - y_i^o)^2}{n}} \quad (2)$$

式(1)、式(2)中 y_i^o 为第 i 年的 SOC 实测值, y_i^m 为 SOC 模拟值, n 为 SOC 的实测和模拟年数, \bar{y}^o 和 \bar{y}^m 分别为 SOC 的 n 年实测和模拟平均值。 r 越接近 1, 模型模拟值与实测值之间的拟合度越高; RMSE 越小, 模拟结果越准确。

1.4 贝叶斯推断

贝叶斯推断过程可表示为^[12,18]:

$$p(\theta | Y) = \frac{p(Y | \theta)p(\theta)}{\int p(Y | \theta)p(\theta) d\theta} \propto p(Y | \theta)p(\theta) \quad (3)$$

式(3)中, $\theta = [\theta_{1m}, \theta_{2m}, \dots, \theta_{dm}]$ 为 DNDC 模型的输入参数集, $Y = [y_{1m}, y_{2m}, \dots, y_{sm}]$ 为输出结果集, d 为模型的输入参数维数, s 为模型的输出结果维数, m 为 MCMC 方法抽取的参数集的个数。

本研究仅针对 DNDC 众多输入参数中的 6 个关键参数进行校正,并推断由这 6 个参数的不确定性导致的 SOC 模拟的不确定性区间,因此取 $d = 6, s = 1$ 。 $p(\theta)$ 为输入参数的先验概率密度, $p(Y | \theta)$ 为数据似然性的概率密度, $p(\theta | Y)$ 为输入参数的后验概率密度。式(3)表示模型输入参数的后验概率密度依比例收敛于 (\propto) 模型输入参数的先验概率密度

与数据似然性概率密度的乘积。

数据似然性 $p(Y | \theta)$ 采用似然函数来描述,本研究中假设其服从正态分布^[12-14],即:

$$p(Y | \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(y^o - y^m)^2}{2\sigma^2}\right] \quad (4)$$

当有 n 对实测值和模拟值时,有^[12,14]:

$$p(Y | \theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(y_i^o - y_i^m)^2}{2\sigma_i^2}\right] \quad (5)$$

式(5)中, σ_i 为第 i 个 SOC 实测值的测量误差,根据国标 HJ-615-2011《土壤有机碳测定重铬酸钾氧化-分光光度法》,本研究假定 SOC 测量误差为实测值的 10%,即 $\sigma_i = 0.1y_i^o$ 。式(4)、式(5)中其他符号的含义同式(1)和式(2)。

本研究假定 DNDC 模型输入参数的先验分布为均匀分布,采用贝叶斯推断和 MCMC 方法从输入参数的先验不确定性区间中随机抽样,共建立 $m = 22\ 000$ 个 DNDC 模型输入参数集,将其输入 DNDC 模型模拟 1986 - 2007 年间的 SOC 逐年变化。根据 SOC 模拟结果与农田长期定位实测点相应年份的 SOC 实测值的似然性(相似程度)来逆向选择满足条件的输入参数集及对应的 SOC 模拟值,最终选择了 1 940 个满足条件的样点集。这些样点集被用于重建 DNDC 输入参数的后验概率分布,并统计 SOC 模拟结果的不确定性置信区间,以此实现 SOC 模拟结果的不确定性定量评价。

1.5 Metropolis-Hastings 算法

MH(Metropolis-Hastings)算法通过遍历模型输出结果的样本空间,根据输出结果与实测值的似然性来选择满足条件的样点,并通过已选择的满足条件的输出结果来逆向推断输出参数,从而重构输出

结果和输入参数后验分布的一种简单实用的 MC-MC。MH 算法中,马尔科夫链的起点可以是样本空间中的任意一点,之后按照一定的规则生成新点。在本文中,由点 θ 生成新点 θ' 的规则为 $q(\theta'|\theta) \equiv N_d(\theta, \bar{S}_\theta)$, N 为正态分布, d 为输入参数的维度, \bar{S}_θ 为输入参数的协方差矩阵^[19]。新点被接受的概率按下式来确定:

$$\alpha = \frac{p(Y|\theta')p(\theta')}{p(Y|\theta)p(\theta)} \quad (6)$$

按式(6)计算得 α 后,从(0,1)中产生一个随机数 β ,若 $\beta \leq \alpha$,则新点 θ' 被接受,设 $\theta = \theta'$,进行下一个样点的生成;若 $\beta > \alpha$,则新点 θ' 不被接受,需返回 θ 并重新生成 θ' ,直至 θ' 被接受,并设 $\theta = \theta'$,进行下一个样点的生成^[19]。如此不断地产生新样点,直至由被接受的样点组成的马尔科夫链达到收敛。Robert 和 Rosenthal^[21] 建议通过调节 \bar{S}_θ 来使得样点被接受的概率介于 10% ~ 40% 之间,以便达到高效的马尔科夫链收敛。

1.6 应用软件

MH 算法的实现是一个由起始样点集开始,不断生成并筛选新样点集,直至满足条件的样点集数量达到足以构建平稳的马尔科夫链的一个实时动态过程。每生成一个新样点集后需要将此样点集输入 DNDC 模型进行一次模拟,并提取模型模拟结果,计算模型模拟结果和 22 个逐年 SOC 实测值之间的似然性,以式(6)来决定新样点集可否被保留为后验样点集。MH 算法的实现需要用户编写程序。本研究采用 python 编程语言来实现 MH 算法。对于由 MH 算法所筛选出的 1 940 个样点集所构成的马尔科夫链的后续分析则采用具有数据统计分析功能的 R 软件来完成。R 软件中提供了专门分析马尔科夫链是否达到收敛的扩展包(boa)^[20],采用其中的 Heidleberger 和 Welch 方法。

2 结果与讨论

2.1 DNDC 模型的适用性评估

为检验 DNDC 模型模拟农田长期定位点 SOC 动态变化的可行性,将表 1 中的点位实测数据及相应的气象数据和作物管理措施数据输入 DNDC 模型进行模拟,并对比 1986 - 2007 逐年间的 SOC 模拟值与对应实测值(图 1)。可知 DNDC 模型基本能够重现出各年份 SOC 的动态变化趋势。DNDC 模型输出的 SOC 值与实测的 SOC 值之间的 r 为 0.92,

RMSE 值为 0.7 g kg^{-1} ,表明 DNDC 模型的 SOC 模拟值与实测值之间的拟合度较高,模型模拟精度可以接受。然而,从图 1 可知 DNDC 模型在重现 SOC 的剧烈波动方面还略显不足。主要表现为 DNDC 模拟的 SOC 值在部分年份仍与实测值有一定的偏差,最为明显的是在 1994 年至 2007 年间模型的模拟值未能够再现实测值所呈现出的一个较为明显的波峰和波谷。对于这些偏差,可能有模型结构和数据采集分析两方面的原因。模型结构方面可能是由于 DNDC 模型现有的模型结构并没有完全地反映复杂自然环境条件下 SOC 对某些极端条件(如干旱、暴雨等近年来常见的气象事件)的变化响应机制;数据采集分析方面,可能如徐胜祥^[22]所指出的,长期定位实测试验中 SOC 的测定存在着很多难以克服的困难,如采样位置、深度、测定方法的微小变化以及采样人员的变动等均可能导致 SOC 实测结果较为明显的年际变化。

IPCC 曾建议将过程模型用作估算由人类活动引起的 SOC 变化的指导工具之一^[23]。通过图 1 可以发现,模型模拟结果虽然能够在很大程度上反映 SOC 的总体演变趋势,但在局部年份并未能完全地再现年际间的波动。因此,简单地将模型的单一模拟值用于 SOC 变化的定量评价存在一定的风险。如图 1 中 1994 - 2007 年间, SOC 的模拟值与实测值间尚有一定的偏差。因此,当将过程模型用于 SOC 动态变化估算时,稳健的方法必须要考虑模型模拟结果所存在的不确定性。而为了准确地定量并在一定程度上减小模型模拟结果的不确定性,对于模型输入参数的校正及其不确定性评价就显得尤为重要。

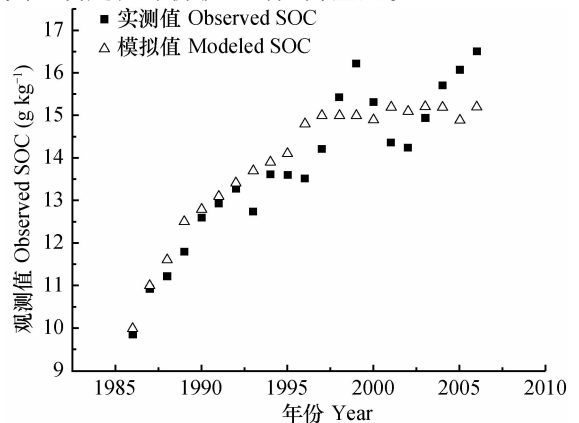


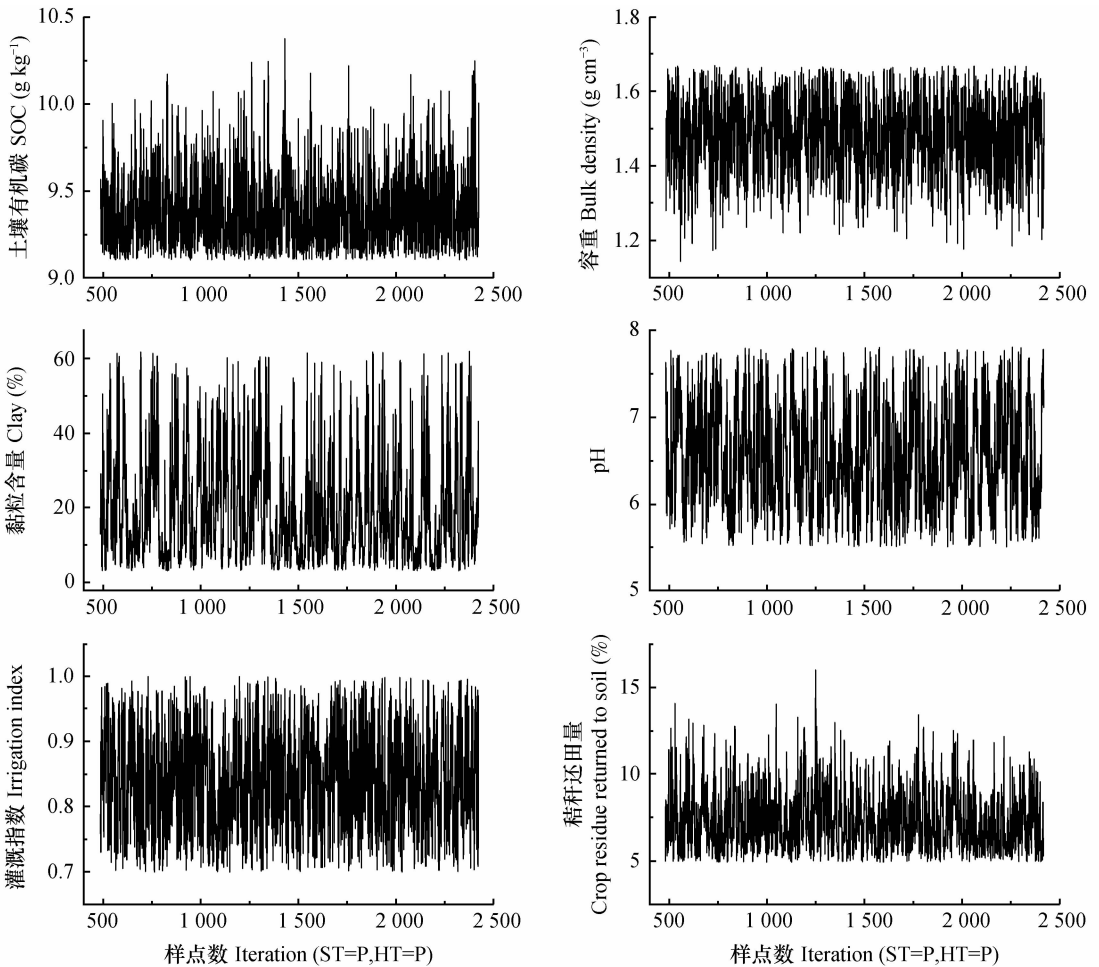
图 1 1986 - 2007 年间 SOC 的 DNDC 模型模拟值与实测值间的对比

Fig. 1 Comparison of the DNDC model simulated SOC with the on-site-observed SOC in the period between 1986 and 2007

2.2 后验马尔科夫链的平稳性和收敛性

MH 算法从 6 个输入参数的先验分布区间内共抽取了 22 000 个样点集输入至 DNDC 模型进行模拟,并筛选出了其中符合条件的 2 426 个样点集构成了 6 个输入参数的后验马尔科夫链。在此过程中,样点集的接受率约为 11%,介于 Robert 和 Rosenthal 建议的区间^[21]。为保证根据后验马尔科夫链统计推断的输入参数的后验不确定性区间和模型模拟的 SOC 的不确定性区间具有统计学意义,本研究选用 R 软件扩展包 boa^[20]中的 Heidleberger 和 Welch 方法对此 2 426 个后验样点集进行平稳性测试。结果表明 6 个输入参数的后验马尔科夫链均达到平稳和收敛状态,且

只有秸秆还田量需要剔除前 486 个样点。因此,去除 6 个输入参数的马尔科夫链的前 486 个样点,再对剩下的 1 940 个后验样点集重复平稳性测试。结果表明,各参数的马尔科夫链均达到了收敛,且样点数量也达到了统计后验参数概率分布的最小样点集的数量要求(图 2)。从各输入参数的马尔科夫链轨迹图(图 2)可知各输入参数的后验马尔科夫链并未呈现出明显的单调上升或单调下降的趋势,而是在整个可能的后验样本空间内无规律的跳动,基本达到了遍历后验不确定性空间的要求^[18]。因此,由此 1 940 个样点集所统计推断出的各输入参数的后验概率密度分布函数是可靠的。



注: ST=P表示马尔科夫链通过了用Heidleberger和Welch方法进行的平稳性测试; HT=P表示马尔科夫链的样点数足以统计稳定的输入参数的后验概率密度分布 Note: ST=P means that the Markov chain has passed the Heidleberger and Welch stationarity test, and HT=P means that the iteration is deemed to estimate the posterior marginal distribution

图 2 剔除掉 486 个不稳定样点后各输出参数的马尔科夫链轨迹图

Fig. 2 Trace plots of six Markov Chains after 486 iterations of each chain were burnt in

2.3 模型输入参数的后验分布特征

根据达到收敛和平稳状态的后验马尔科夫链可以统计各输入参数的后验分布特征。各输入参

数的实测值、先验概率密度分布、后验概率密度分布如图 3 所示。对比各输入参数的先验和后验概率密度分布,可知初始 SOC、秸秆还田量的后验不确定

性区间远远窄于先验不确定性区间,并且后验概率密度分布呈现出了明显的峰度分布,说明模型模拟的 SOC 对于这两个输入参数十分敏感;而灌溉指数和 pH 这两个输入参数的后验不确定性区间并未明显地变窄,并且后验概率密度分布在整个后验不确定性区间内呈现出略微不同于均匀分布的低峰度分布,表明模型模拟的 SOC 对于这两个输入参数的敏感程度很低,这两个参数存在的作用是约束 SOC 以外的其他模型输出结果(如 N_2O)^[6];对于容重和黏粒含量这两个输入参数,后验概率密度分布在后验不确定性区间的特定区域内均呈现出了一定的峰度,说明模型模拟的 SOC 对这两个输入参数的敏

感程度弱于初始 SOC 和秸秆还田量,但是强于灌溉指数和 pH。

对比各输入参数的实测值和其后验不确定性区间(图 3),可知根据后验马尔科夫链所统计的 6 个输入参数的后验不确定性区间均包含了各自的实测值。且在 6 个输入参数中,对于模拟 SOC 影响越大的输入参数,其后验马尔科夫链所统计的不确定性区间的峰值越接近实测值。这表明基于贝叶斯推断的 MCMC 方法完成的 DNDC 模型输入参数的不确定性区间估计是可靠的;也说明了敏感参数的数据质量对 DNDC 模型模拟 SOC 起着非常重要的作用^[10,24]。

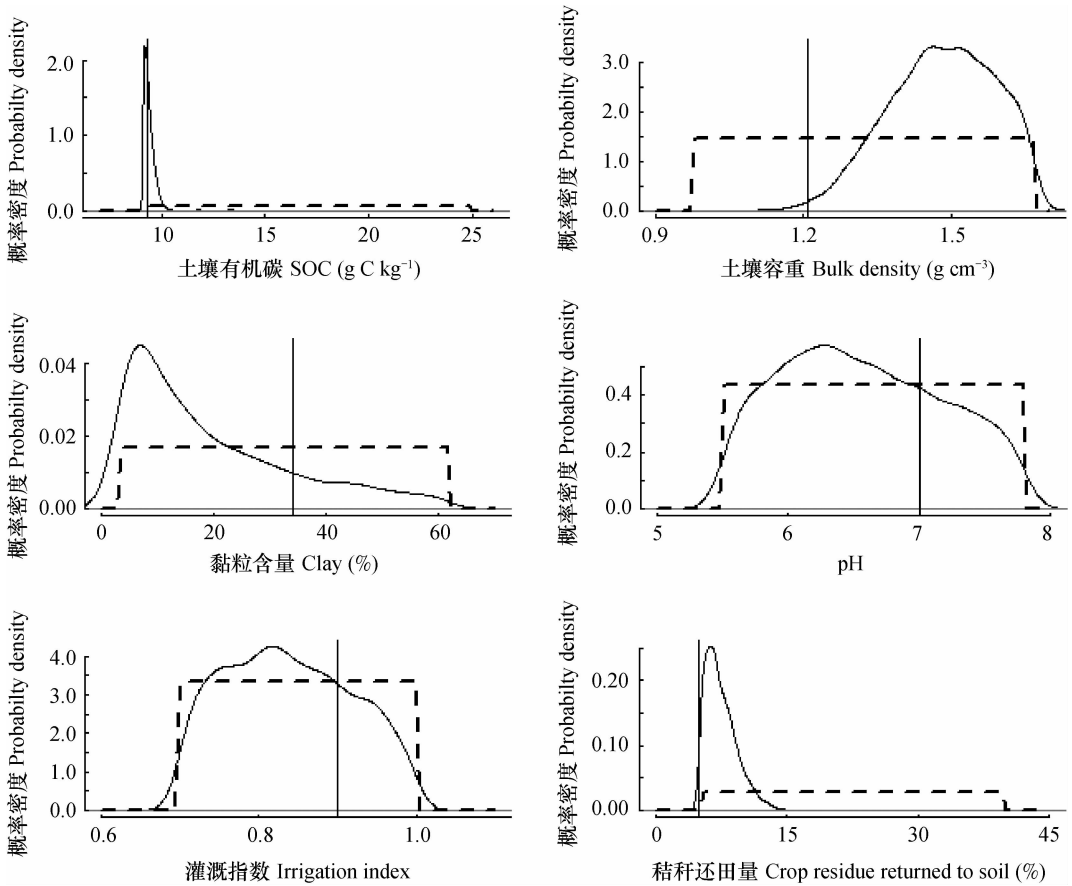


图 3 基于 MCMC 方法统计的 6 个 DNDC 模型输入参数的先验概率密度分布(虚曲线)、后验概率密度分布(实曲线)和农田长期定位点实测值(垂直的直线)

Fig. 3 Priori marginal distributions (dashed curve), posterior marginal distributions (solid curve), and on-site-observed value (vertical line) for six input parameters of the DNDC model

2.4 模型输出结果的不确定性区间

输出结果的不确定性区间可表征 DNDC 模型模拟 SOC 动态变化的可靠性。将 MH 方法筛选出的达到收敛状态的 1 940 个后验样点集经 DNDC 模型模拟,根据输出的 SOC 值构建不确定性区间及其 95% 置信区间。对比实测值(如图 4),可知 MCMC

方法估计的不确定性区间包含了所有的 SOC 实测值,但作为制定详细农田管理政策的参考或者估计 SOC 的动态变化而言,这一区间可能不够精确,实际指导意义不大。相对于整个后验不确定性区间,95% 置信区间能够覆盖大部分的(约 70%)SOC 实测值,但还有少部分的(约 30%)SOC 实测值未被包

含。究其原因,一方面可能是由于尚有其他敏感因子(如化肥、有机肥投入)存在的不确定性未纳入到不确定性分析中来;另一方面也可能是由于 DNDC 模型对于模拟此处的 SOC 动态变化尚存在一定的结构误差。针对以上问题,在今后的 SOC 模拟中需要将更多的参数纳入考虑,并尽可能提高输入参数的数据质量;同时还应该注重不确定性评价方法的选择和使用,如可通过在贝叶斯推断过程中引入高阶参数来刻画 DNDC 模型的结构不确定性^[14],从而达到提高不确定性评价精度的目的。

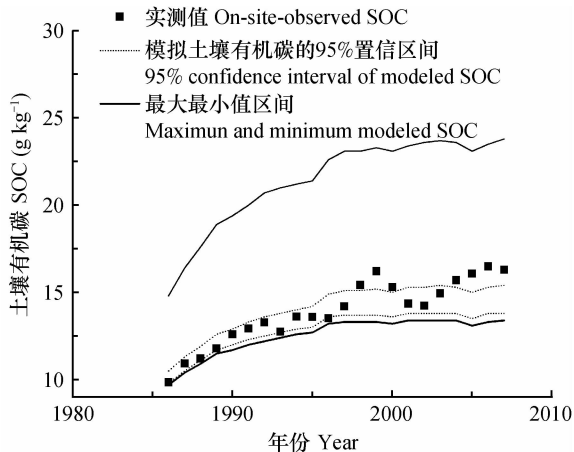


图4 基于MCMC方法统计的DNDC模型模拟的SOC值的最大值、最小值(实线)、95%置信区间(点虚线)与位于江苏省宜兴市长期农田定位观测点的SOC实测值(实点)

Fig. 4 MCMC-based statistics of the DNDC model simulated minimum and maximum value (solid lines), 95% confidence interval (dotted lines), and the site-observed values (black dots) of SOC, respectively

3 结论

DNDC模型模拟的SOC基本能够再现位于江苏省宜兴市的农田长期定位点的SOC动态变化趋势;贝叶斯推断和MCMC方法可以作为DNDC模型参数校正和输出结果不确定性区间评价的有效方法。然而,尚有一部分SOC实测值未被贝叶斯推断和MCMC方法构建的95%置信区间所覆盖。因此,在今后的研究中一方面应该将更多的SOC敏感参数纳入分析,并提高敏感参数数据收集的准确度;另一方面应改进贝叶斯推断和MCMC方法,将DNDC模型可能存在的结构误差以增加高阶参数的方式纳入到考虑中,从而达到更好地评价模型结构误差和估计模型不确定性区间的目的。

参考文献

- [1] 刘清,孙波,解宪丽,等. 县域尺度红壤丘陵区水稻土有机碳模拟. 土壤学报,2009,46(6):1059—1067. Liu Q, Sun B, Xie X L, et al. The spatial-temporal dynamic change and simulation of county-scale paddy soil organic carbon in red soil hilly region (In Chinese). *Acta Pedologica Sinica*, 2009, 46(6): 1059—1067
- [2] Xu S X, Shi X Z, Zhao Y C, et al. Carbon sequestration potential of recommended management practices for paddy soils of China, 1980—2050. *Geoderma*, 2011, 166(1): 206—213
- [3] Xu S X, Shi X Z, Zhao Y C, et al. Spatially explicit simulation of soil organic carbon dynamics in China's paddy soils. *Catena*, 2012, 92: 113—121
- [4] Xu S X, Shi X Z, Zhao Y C, et al. Modeling carbon dynamics in paddy soils in Jiangsu Province of China with soil databases differing in spatial resolution. *Pedosphere*, 2011, 21(6): 696—705
- [5] Zhang L M, Yu D S, Shi X Z, et al. Simulation soil organic carbon change in China's Tai-Lake paddy soils. *Soil and Tillage Research*, 2012, 121: 1—9
- [6] Hastings A F, Wattenbach M, Eugster W, et al. Uncertainty propagation in soil greenhouse gas emission models: An experiment using the DNDC model and at the Oensingen croplands. *Agriculture, Ecosystems & Environment*, 2010, 136: 97—110
- [7] Ogle S M, Breidt F J, Easter M, et al. Scale and uncertainty in modeled soil organic carbon stock changes for US croplands using a process-based model. *Global Change Biology*, 2010, 16(2): 810—822
- [8] Ogle S M, Breidt F J, Easter M, et al. An empirically based approach for estimating uncertainty associated with modelling carbon sequestration in soils. *Ecological Modelling*, 2007, 205(3/4): 453—463
- [9] Wang G S, Li C S. A review on parameterization and uncertainty in modeling greenhouse gas emissions from soil. *Geoderma*, 2012, 170: 206—216
- [10] Li C S, Mosier A, Wassmann R, et al. Modeling greenhouse gas emissions from rice-based production systems: Sensitivity and up-scaling. *Global Biogeochemical Cycles*, 2004, 18(1): 1—19
- [11] Li C S. User's guide for the DNDC model (version 9.3). Institute for the Study of Earth, Oceans, and Space. University of New Hampshire, Durham, NH, USA, 2009
- [12] Oijen M V, Rougier J, Smith R. Bayesian calibration of process-based forest models: Bridging the gap between models and data. *Tree Physiology*, 2005, 25: 915—927
- [13] Freni G, Mannina G, Viviani G. Urban runoff modeling uncertainty: Comparison among Bayesian and pseudo-bayesian methods. *Environmental Modeling & Software*, 2009, 24: 1100—1111
- [14] Wang G S, Chen S L. Evaluation of a soil greenhouse gas emission model based on Bayesian inference and MCMC: Model uncertainty. *Ecological Modelling*, 2013, 253: 97—106
- [15] Giltrap D L, Li C S, Saggart S. DNDC: A process-based model of greenhouse gas fluxes from agricultural soils. *Agriculture, Ecosystems & Environment*, 2010, 136(3/4): 292—300

- [16] Li C S. Quantifying greenhouse gas emissions from soils: Scientific basis and modeling approach. *Soil Science and Plant Nutrition*, 2007, 53(4):344—352
- [17] Wang L G, Qiu J J, Tang H J, et al. Modelling soil organic carbon dynamics in the major agricultural regions of China. *Geoderma*, 2008, 147(1/2):47—55
- [18] Rahn K, Butter-Bahl K, Werner C. Selection of likelihood parameters for complex models determines the effectiveness of Bayesian calibration. *Ecological Informatics*, 2011, 6:333—340
- [19] Ntzoufras I. Bayesian modeling using WinBUGS. Hoboken, New Jersey, USA: John Wiley & Sons, Inc., 2009:31—81
- [20] Smith B J. boa: An R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software*, 2007, 21(11):1—36
- [21] Robert G O, Rosenthal J S. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 2001, 16(4):351—367
- [22] 徐胜祥. 中国水稻土有机碳演变模拟与固碳潜力估算. 南京:中国科学院南京土壤研究所, 2011. Xu S X. Spatially explicit simulation of soil organic carbon dynamics and carbon sequestration potential for paddy soils of China (In Chinese). Nanjing: Institute of Soil Science, Chinese Academy of Sciences, 2011
- [23] IPCC. Guidelines for national greenhouse gas inventories. <http://www.ipcc-nggip.iges.or.jp/public/2006gl/index.html>, 2006
- [24] Qin X B, Wang H, Li Y E, et al. A long-term sensitivity analysis of the denitrification and decomposition model. *Environmental Modeling & Software*, 2013, 43:26—36

INPUT PARAMETERS CALIBRATION AND UNCERTAINTY ESTIMATION OF THE DNDC MODEL BASED ON BAYESIAN INFERENCE

Qin Falü^{1,2} Zhao Yongcun^{1,2†} Shi Xuezheng^{1,2} Yu Dongsheng^{1,2} Xu Shengxiang^{1,2}

(1 State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China)

(2 University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract It is vitally important to accurately estimate soil organic carbon (SOC) dynamics modeling uncertainty for successful decision-making in cropland management. In this study, a Bayesian inference using the Markov Chain Monte Carlo (MCMC) method was used to calibrate the parameters and estimate the output uncertainty interval for the Denitrification-Decomposition (DNDC) model in modeling SOC dynamics of the long-term monitored croplands with 22-year rice-wheat rotation history in Yixin County, Jiangsu Province of China. Results indicate that though there is some uncertainty, the DNDC model is suitable for modeling SOC dynamics of the long-term monitored croplands. In cases that input parameters are not certain in quality, Bayesian inference using the MCMC method could be an effective way for automatically calibrating input parameters and estimating uncertainty interval of the modeled SOC. The Bayesian inference using the MCMC method applied in this paper could also instruct the estimation of the SOC dynamics modeling uncertainty at region or country scale.

Key words Bayesian inference; Markov Chain Monte Carlo method; Parameter calibration; Uncertainty estimation; Denitrification-Decomposition model

(责任编辑:檀满枝)