

## 耦合敏感参数实时识别的新型数据同化算法研究 ——以湖泊藻类模拟为例<sup>\*</sup>

张 帅<sup>1,2</sup>, 彭福利<sup>3</sup>, 季雨来<sup>2</sup>, 张 京<sup>2</sup>, 张奇谋<sup>2</sup>, 李 琪<sup>2</sup>, 钱 瑞<sup>2</sup>, 齐凌艳<sup>1,4 \*\*</sup>, 黄佳聪<sup>2\*\*</sup>

(1:安徽师范大学地理与旅游学院,芜湖 241003)

(2:中国科学院南京地理与湖泊研究所, 中国科学院流域地理学重点实验室,南京 210008)

(3:中国环境监测总站,北京 100012)

(4:资源环境与地理信息工程安徽省工程技术研究中心,芜湖 241003)

**摘要:** 数据同化是提升复杂机理过程模型精度的关键技术之一,而湖泊藻类模型的敏感参数具有随时间动态变化的特征,导致数据同化过程中无法精准更新某一时段的敏感参数,影响数据同化的模型精度提升效果。针对上述问题,本研究耦合了参数敏感性分析与集合卡尔曼滤波,研发了一种能够实时识别模型敏感参数的新型数据同化算法;为验证研发算法的效率,依托巢湖的高频水质自动监测数据,测试算法对藻类动态模型的精度提升效果。测试结果表明:研发算法能够精准跟踪模型敏感参数的动态变化,并根据监测数据实时更新模型敏感参数,实现了水质高频自动监测数据与藻类动态模型的深度融合,藻类生物量模拟精度提升了 55%,即纳什系数(*NSE*)从 0.49 提升到 0.76,模拟精度提升效果也显著优于传统数据同化算法(*NSE*=0.63)。研发算法可应用于其它水生态环境模型的数据同化,为水生态环境相关要素的精准模拟预测提供关键技术支撑。

**关键词:** 巢湖;集合卡尔曼滤波;参数敏感性分析;模型

### A new data assimilation method coupled with real-time detection of sensitive parameters: An example of phytoplankton modeling in lakes<sup>\*</sup>

Zhang Shuai<sup>1,2</sup>, Peng Fulì<sup>3</sup>, Ji Yulai<sup>2</sup>, Zhang Jing<sup>2</sup>, Zhang Qimou<sup>2</sup>, Li Qi<sup>2</sup>, Qian Rui<sup>2</sup>, Qi Lingyan<sup>1,4 \*\*</sup> & Huang Jiacong<sup>2\*\*</sup>

(1: College of Geography and Tourism, Anhui Normal University, Wuhu 241003, P.R.China)

(2: Key Laboratory of Watershed Geographic Sciences, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing 210008, P.R.China)

(3: China National Environmental Monitoring Centre, Beijing 100012, P.R.China)

(4: Engineering Technology Research Center of Resources Environment and GIS, Wuhu 241003, P.R.China)

**Abstract:** Data assimilation is a critical method to improve the performance of complex process-based models. However, the sensitive parameters for lake models are generally changing over time. Therefore, it is challenging to accurately update the sensitive parameters for a specific period, which affects the performance of data assimilation. To address the problem, this study developed a new data assimilation method by coupling the methods of parameter sensitivity analysis and the Ensemble Kalman Filter. The new method aimed to identify the model's sensitive parameters in real time. To evaluate its performance, we collected the high-frequency water quality automatic monitoring data of Lake Chaohu, and investigated the performance improvement of a phytoplankton dynamic model using a new data assimilation method. Our investigation results showed that the developed method was able to identify the

\* 2021-12-18 收稿;2022-03-18 收修改稿。

中国科学院青年创新促进会项目(2019313)、安徽省自然科学基金青年项目(1908085QD151)、江苏省水利科技项目(2019025,2020042,2020032,2021036)和安徽师范大学大学生创新创业训练计划项目(202010370193)联合资助。

\*\* 通信作者; E-mail: qly1107@126.com, jchuang@niglas.ac.cn.

sensitive parameters of the model in each simulation period, and updated them based on the measured data to achieve better performance. The simulation accuracy of phytoplankton biomass increased by 55%, i.e., the Nash-Sutcliffe Efficiency ( $NSE$ ) increased from 0.49 to 0.76. This performance is better than that of the traditional data assimilation method ( $NSE=0.63$ ). The method can be applied to the data assimilation of other ecological and environmental models, technically supporting an accurate prediction of environmental and ecological factors.

**Keywords:** Lake Chaohu; Ensemble Kalman Filter; parameter sensitivity analysis; model

湖库富营养化与藻类水华是我国面临的突出水环境问题,2020年《中国环境状况公报》数据显示,在我国110个重点监测的湖库中有29%处于富营养化状态,其中三湖(太湖、滇池、巢湖)全部处于富营养化状态<sup>[1]</sup>,造成蓝藻水华频发<sup>[2]</sup>;2018年夏季,巢湖湖心区和西北岸暴发了面积约121.38 km<sup>2</sup>的蓝藻水华,严重影响周边居民日常生活和用水安全<sup>[3]</sup>. 提前预测藻类未来变化,并采取应急措施能够有效缓解蓝藻水华造成的危害<sup>[4]</sup>;近年来,我国在巢湖等210个重点湖库设置了349个监测站,积累了长时间历史系列的人工与自动监测数据,并在此基础上,构建了包含白洋淀<sup>[5]</sup>、太湖<sup>[6]</sup>、巢湖<sup>[7]</sup>等湖泊的藻类动态模型。但湖泊藻类动态变化过程十分复杂<sup>[8]</sup>,数值模拟难度大,亟需依托与日俱增的海量监测数据,发展模型—数据融合算法,提升模型模拟精度<sup>[9]</sup>.

数据同化是在模型和观测的误差估计基础上,在模型运行过程中,实时融合观测数据的方法<sup>[10]</sup>;是融合模型与数据、提升模拟精度的重要技术之一<sup>[11]</sup>. 最初应用于数值天气预报,为天气预报提供初始场的数据处理技术<sup>[12]</sup>;目前已广泛应用于大气<sup>[13]</sup>、海洋<sup>[14]</sup>等领域,在生态环境领域也有相关应用<sup>[15]</sup>,包括流域水文模型、湖泊藻类模型等,并取得了良好应用效果<sup>[16]</sup>. 集合卡尔曼滤波(Ensemble Kalman Filter, EnKF)是数据同化领域主流的算法之一,其通过蒙特卡洛方法计算状态的预报误差协方差,用集合的思想解决了实际应用中背景误差协方差矩阵的估计和预报困难的问题,能够有效降低数据同化计算量,具有程序设计相对简单、可应用于非线性系统、易于实现并行运算等优点<sup>[11,17]</sup>. 但集合卡尔曼滤波应用于湖泊藻类动态模型中,缺乏模型敏感参数动态变化的现实<sup>[18-20]</sup>,导致模型模拟精度提升效果有限,还存在可能增加模型不确定性的不足<sup>[21]</sup>. 参数敏感性分析能够有效识别复杂模型的关键参数,是实时跟踪敏感参数动态变化的有效方法<sup>[22]</sup>,有望与数据同化耦合,实现对模型敏感参数的靶向更新,精准提升模型模拟精度.

本研究通过集合卡尔曼滤波与参数敏感性分析的耦合,开发了一种能够实时识别敏感参数的新型数据同化算法,为湖泊藻类动态模型模拟精度提升提供技术支持. 新型数据同化算法将参数敏感性分析耦合到集合卡尔曼滤波,在模型运行每一阶段,根据敏感性分析定位敏感参数,采用集合卡尔曼滤波同步同化状态变量和敏感参数. 由此可精确定位并调整每阶段敏感参数,实时校正模型运行轨迹,减少模型模拟不确定性,提升模型模拟精度. 为验证新算法,以巢湖为研究对象,依据水质自动监测数据,将新型数据同化算法应用于湖泊藻类动态模型<sup>[23]</sup>,检验新算法对模型精度的提升效果,研究成果也为其它湖库水质模型数据同化提供技术参考.

## 1 材料与方法

### 1.1 研究区概况

巢湖( $31^{\circ}26' \sim 31^{\circ}42'N, 117^{\circ}17' \sim 117^{\circ}51'E$ )位于安徽省中部、合肥市内,由巢湖、肥东、肥西和庐江一市三县环抱,湖长54.5 km,最大宽度21 km,平均水深3 m,面积约787 km<sup>2</sup>,容积约20亿m<sup>3</sup>(图1),是沿岸居民生活生产的主要水源<sup>[24]</sup>. 巢湖入湖河流约33条,其中6条主要入湖河流为杭埠河、白石天河、兆河、柘皋河、南淝河和派河;出湖河流主要为裕溪河,连接长江. 近年来随着流域内人口增加,工农业迅速发展,城镇工业废水、生活污水排入,导致水体营养盐增加,蓝藻水华事件频发,水质恶化,影响周围居民饮用水安全<sup>[25]</sup>. 此外,蓝藻水华还分泌毒素,危害人类健康<sup>[26]</sup>. 因此,巢湖富营养化已成为突出的生态、环境和社会问题<sup>[27]</sup>.

### 1.2 数据

巢湖水质监测数据来自8个自动监测站的逐日监测(图1),监测指标包括水温(WT, °C)、总磷(TP, mg/L)、总氮(TN, mg/L)及叶绿素a浓度(Chl.a, μg/L)等,监测时间为2020年1—5月. 气象数据来自国家气象信息中心(<http://data.cma.cn/>)的巢湖站(图1),位于巢湖东部沿岸,代表巢湖气象条件;气象指标

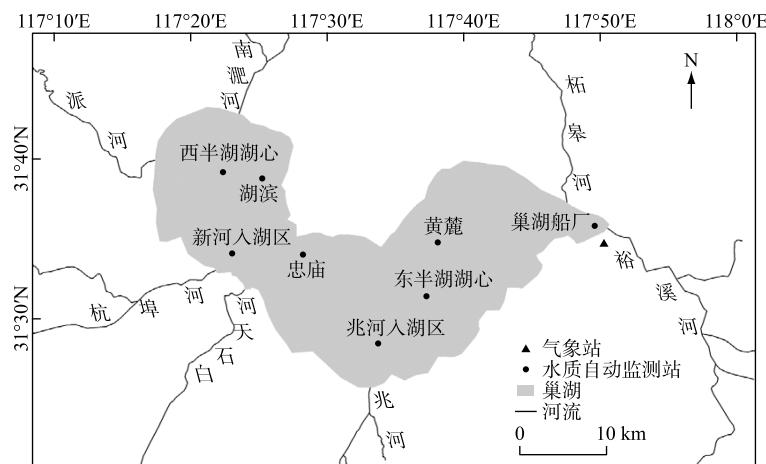


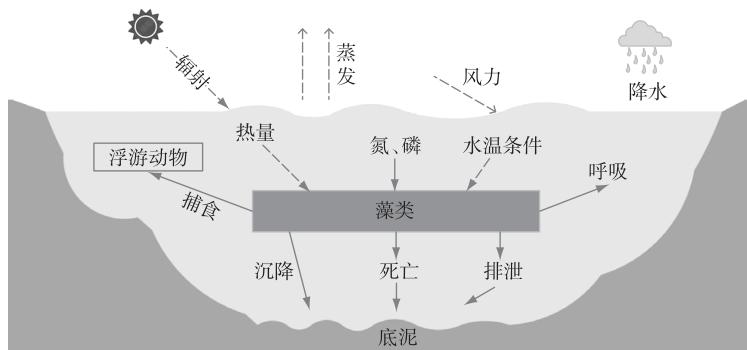
图 1 巢湖与水质自动监测站分布

Fig.1 Locations of Lake Chaohu and its automatic water quality monitoring stations

包括最低气温(℃)、最高气温(℃)、平均气温(℃)、日照时长(h)、平均风速(m/s)及降水量(mm)等。以上数据为湖泊藻类动态模型运行与校验数据。

### 1.3 巢湖藻类动态模型

目前已开发了系列湖泊藻类模拟模型,如澳大利亚西澳大学水研究中心的 ELCOM-CAEDYM 模型<sup>[28]</sup>、威廉玛丽大学海洋学院维吉尼亚海洋科学研究所的 Environmental Fluid Dynamics Code (EFDC) 模型等<sup>[29]</sup>。但此类模型多需要海量、高时空分辨率的数据输入,部分数据难以实时获取,不适用于湖泊藻类的短期(1~3 d)模拟预测<sup>[30]</sup>。为解决模型复杂数据输入问题,国内外结合研究区特征,开发了少量数据输入的湖泊藻类动态模型<sup>[31-32]</sup>。本研究采用黄佳聪等构建的湖泊藻类动态模型<sup>[23]</sup>(图 2)。

图 2 湖泊藻类动态模型概念图<sup>[23]</sup>Fig.2 Conceptual diagram of phytoplankton dynamic model for lakes<sup>[23]</sup>

该模型采用叶绿素  $a$  浓度表征藻类生物量,藻类动态变化模拟分为两个模块(表 1、表 2):①藻类生消模块:用于模拟藻类在太阳辐射、水温、营养盐等条件下的生长,以及在排泄、沉降、呼吸、死亡、被浮游动物捕食等机理过程中的消亡;②基于二维水动力的藻类运移模块:用于模拟藻类在水动力作用下的运移过程。其它机理过程的数学表达可以参考文献[23]。为了兼顾模型运行效率与算法耦合效率,本研究仅考虑了点位的藻类模拟,即模型为零维模型,仅考虑藻类生长与消亡等过程导致的藻类动态变化,不考虑水动力输移对藻类的影响。相比于其它模型,该模型输入数据易实时获取,能满足国内湖泊藻类模拟预测的需求。

表 1 湖泊藻类生消过程表达<sup>[23]</sup>  
Tab.1 Equations of phytoplankton growth  
and death in lakes<sup>[23]</sup>

| 模块   | 模拟方程  |
|------|---|
| 主控方程 | $\frac{dChl}{dt} = (U - RA - MA - SA - GA - EA) Chl$  |
| 生长过程 | $U = U_{\max} \cdot f(t) \cdot f(I) \cdot f(N, P)$<br>$f(t) = \exp(-2.3(T - T_{\text{opt}})/16)^2$<br>$f(I) = (I_z/I_{\text{opt}}) \times \exp(1 - I_z/I_{\text{opt}})$<br>$f(P) = DP/(KP + DP)$<br>$f(N) = DN/(KN + DN)$ |
| 消亡过程 | $RA = k_r \cdot \vartheta^{T-20}$<br>$MA = k_e \cdot \vartheta^{T-20}$<br>$EA = k_e \cdot \vartheta^{T-20}$<br>$SA = K/Z$<br>$GA = GR_{\max}((F - F_{\min})/(F_s + F - F_{\min}))$  |

扰动  $n$  次并计算模型输出,依次求出相邻两次扰动的敏感程度,取平均值作为目标参数的相对敏感度  $S$  (式(6)) (表 3). 该方法简单高效,既方便耦合模型又能兼顾运行效率. 根据本研究湖泊藻类动态模型的机理和结构,将模型的参数敏感程度划分为 5 个等级:不敏感(I,  $|S| < 0.25$ );弱敏感(II,  $0.25 \leq |S| < 0.5$ );一般敏感(III,  $0.5 \leq |S| < 0.75$ );比较敏感(IV,  $0.75 \leq |S| < 1.0$ );极度敏感(V,  $|S| \geq 1.0$ ). 将敏感程度  $|S| > 0.5$  作为敏感参数用于调整,敏感度阈值 0.5 是根据模型的机理结构及模拟效果综合率定的,以此为阈值调整参数能够获得的最优模拟结果.

本研究开发的新型数据同化算法将参数敏感性分析耦合到集合卡尔曼滤波过程中,使其能够精确定位并调整每阶段模型敏感参数,实时有效校正模型运行轨迹,提升模型模拟精度. 综合考虑集合卡尔曼滤波对参数的调整过程以及每一阶段参数敏感性分析的时效性,本研究提出的新型数据同化算法运行分 3 个步骤:预测、分析与更新(表 3).

①预测:与集合卡尔曼滤波数据同化的预测步相同,根据上一时刻状态变量(即叶绿素  $a$  浓度)后验集合,使用湖泊藻类动态模型模拟该时刻状态变量先验集合(式(1)),对先验集合求期望(式(5))即可得状态变量的模拟值,该模拟值在该时刻实测数据可用之前都有效.

②分析:使用扰动分析法定位该时刻敏感参数. 根据扰动分析法流程,本研究对湖泊藻类动态模型每个待分析参数的扰动范围为: $\pm 1\%$ 、 $\pm 5\%$ 、 $\pm 10\%$ 、 $\pm 15\%$ 、 $\pm 20\%$ ,共扰动 10 次. 使用(式(6))计算出该时刻每个参数的敏感度  $S$ ,将敏感度  $|S| > 0.5$  作为敏感参数输出到更新步.

③更新:当该时刻实测数据可用时即进入更新步,使用集合卡尔曼滤波更新状态变量和敏感参数集合. 先更新状态变量,根据(式(7))更新状态变量集合,状态变量先验集合被更新为后验集合,对其求期望(式(10))即可得该时刻状态变量的同化值;再更新敏感参数,由于耦合了参数敏感性分析,每一步需要更新的参数维度不同,因此将参数维度固定为一维,然后依次调整,如有多个参数处于敏感状态则依次执行调整参数步(式(11)),如没有参数处于敏感状态则不执行调整参数步. 考虑到集合卡尔曼滤波的假设是各参数误差相互独立且服从高斯分布<sup>[11]</sup>,因此涉及到多个敏感参数需要调整时,依次调整每个参数效果等同于同步调整;然后使用(式(13))对所有参数集合求期望即可得该时刻模型参数的同化值. 该时刻状态变量和参数的同化值将作为模型下一时刻预测步的初值,依此递推,运行流程如图 3 所示.

## 1.5 数据同化方案设计

本研究提出的新型数据同化算法以集合卡尔曼滤波为主体改进而成. 为测试其在传统集合卡尔曼滤波基础上的提升效果,同步设计基于传统集合卡尔曼滤波数据同化算法方案用于对比分析. 目前集合卡尔曼滤波在水文水质模型中主要有 2 种同化方案,一是仅同化状态变量<sup>[35]</sup>,二是同步同化状态变量和模型参

## 1.4 新型数据同化算法

基于集合卡尔曼滤波的数据同化在模型运行的每一阶段会对状态变量和模型参数进行调整<sup>[18]</sup>. 算法运行分两步骤:预测和更新. 预测:利用模型模拟状态变量先验值集合;更新:根据实测值统计卡尔曼增益并估计状态变量和参数,作为下一时刻的模型输入<sup>[33]</sup>. 算法递推式运行,每阶段的更新只依赖于上一刻钟集合空间,运行效率高<sup>[11]</sup>,但难以应对模型参数敏感性动态变化的问题.

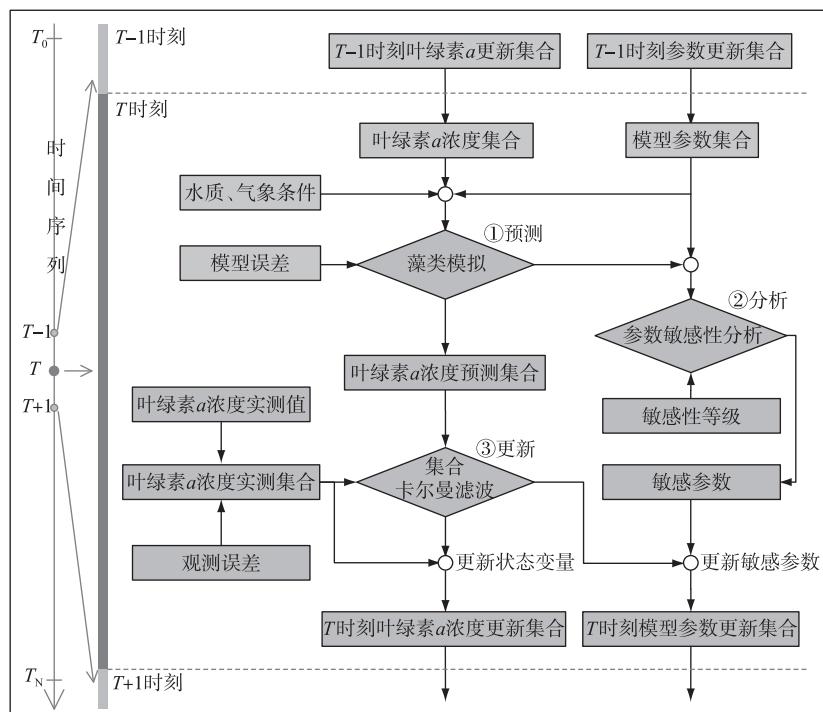
敏感性分析能够在多参数模型中有效识别关键参数,从而减少敏感参数对模型的影响,使模型模拟更加精确<sup>[22]</sup>. 本研究决定采用的扰动分析法简单实用且应用普遍<sup>[34]</sup>,采用此方法对湖泊藻类动态模型的参数进行敏感性分析. 扰动分析法基本思路如下:选定一个参数作为敏感性分析目标,其它参数都保持不变,将目标参数在一定范围内取值

表 2 藻类模型参数与变量<sup>[23]</sup>  
Tab.2 Parameters and variables of the phytoplankton model<sup>[23]</sup>

| 类型   | 简称               | 描述                   | 单位                            | 初值     |
|------|------------------|----------------------|-------------------------------|--------|
| 参数   | $U_{\max}$       | 藻类最大生长速率             | $d^{-1}$                      | 1.145  |
|      | $T_{\text{opt}}$ | 藻类生长的最佳水温            | °C                            | 27     |
|      | $\alpha$         | 水及非藻类吸收短波辐射的平均消光系数   | $m^{-1}$                      | 0.45   |
|      | $\beta$          | 藻类吸收短波辐射的平均消光系数      | $\mu\text{g Chl}/(L \cdot m)$ | 0.016  |
|      | $h$              | 水深                   | m                             | 3.0    |
|      | $I_{\text{opt}}$ | 饱和光强                 | $\text{MJ}/(m^2 \cdot d)$     | 12     |
|      | $G_{\text{sc}}$  | 太阳常数                 | $\text{MJ}/(m^2 \cdot d)$     | 1367   |
|      | $\varphi$        | 纬度                   | (°)                           | 31.5   |
|      | $KP$             | 影响藻类磷吸收的米氏常数         | $\mu\text{g/L}$               | 10     |
|      | $KN$             | 影响藻类氮吸收的米氏常数         | $\mu\text{g/L}$               | 22     |
|      | $K$              | 藻类沉降速率               | $m/d$                         | 0.0864 |
|      | $GR_{\max}$      | 浮游动物最大捕食速率           | $d^{-1}$                      | 0.09   |
|      | $F_{\min}$       | 可供捕食藻类最小浓度           | $\mu\text{g/L}$               | 100    |
|      | $F_s$            | 可供捕食藻类米氏常数           | $\mu\text{g/L}$               | 500    |
|      | $k_m$            | 藻类死亡引起叶绿素 $a$ 减少速率系数 | $d^{-1}$                      | 0.027  |
|      | $k_r$            | 藻类呼吸引起叶绿素 $a$ 减少速率系数 | $d^{-1}$                      | 0.17   |
|      | $k_e$            | 藻类排泄引起叶绿素 $a$ 减少速率系数 | $d^{-1}$                      | 0.01   |
|      | $\vartheta$      | 温度因子                 | —                             | 1.08   |
| 待解变量 | $T$              | 水体温度                 | °C                            | —      |
|      | $r$              | 光衰减系数                | m                             | —      |
|      | $Chl$            | 叶绿素 $a$ 浓度           | $\mu\text{g/L}$               | —      |
|      | $I$              | 水体表面光强               | $\text{MJ}/(m^2 \cdot d)$     | —      |
|      | $S$              | 日照时数                 | h                             | —      |
|      | $S_0$            | 昼长                   | h                             | —      |
|      | $k$              | 离心校正因子               | —                             | —      |
|      | $U$              | 藻类生长速率               | $d^{-1}$                      | —      |
|      | $RA$             | 藻类呼吸速率               | $d^{-1}$                      | —      |
|      | $SA$             | 藻类沉降速率               | $d^{-1}$                      | —      |
|      | $MA$             | 藻类死亡速率               | $d^{-1}$                      | —      |
|      | $GA$             | 藻类被捕食速率              | $d^{-1}$                      | —      |
|      | $EA$             | 藻类排泄速率               | $d^{-1}$                      | —      |
|      | $\delta$         | 太阳赤纬                 | °                             | —      |
|      | $\omega_s$       | 日落时角                 | °                             | —      |
|      | $DP$             | 水体可溶性磷               | $\mu\text{g/L}$               | —      |
|      | $DN$             | 水体可溶性氮               | $\mu\text{g/L}$               | —      |
|      | $RA_{\max}$      | 藻类最大呼吸速率             | $d^{-1}$                      | —      |
|      | $F$              | 可供捕食藻类浓度             | $\mu\text{g/L}$               | —      |

数<sup>[18]</sup>. 为详细对比差异, 分别将这两种方案应用到模型查看其对模型的提升效果. 同时, 还需一种完全凭借湖泊藻类动态模型的模拟方案作为评价上述 3 种数据同化方案效果的基准. 因此, 本研究共有了 4 种方案, 包括 1 个纯模型方案和 3 个数据同化方案(仅同化状态变量、同步同化状态变量和模型参数及新型数据同化算法方案), 下面详细介绍 4 种方案的策略及运行步骤.

1) Reference: 不使用数据同化算法, 仅依据湖泊藻类动态模型模拟叶绿素  $a$  浓度. 将 2020 年 1 月 1 日实测的水质数据和 2020 年 1 月 2 日的气象数据作为初值输入, 将该阶段模拟的叶绿素  $a$  浓度作为下一阶段模型的初值, 依此递推, 持续模拟至 2020 年 6 月 1 日. 该方案旨在了解湖泊藻类动态模型本身的模拟精度, 为

图 3 新型数据同化算法流程图<sup>[18]</sup>Fig.3 New data assimilation algorithm flow chart<sup>[18]</sup>

3 种数据同化方案的模拟值提供对比基准。

2) EnKF\_A: 使用集合卡尔曼滤波数据同化算法同化状态变量(即叶绿素 a 浓度), 不同化模型参数。输入初值时使用蒙特卡洛方法分别对状态变量和边界条件添加误差并生成集合, 模拟出下一阶段叶绿素 a 浓度集合, 对集合求期望即可获得模拟值。当获得叶绿素 a 浓度实测值时, 使用集合卡尔曼滤波同化状态变量集合, 对集合求期望获得同化值, 作为模型下一阶段状态变量输入, 依此递推。

3) EnKF\_B: 使用集合卡尔曼滤波数据同化算法同步同化状态变量和模型参数。考虑到使用的湖泊藻类动态模型参数众多, 仅同步同化一个参数, 即  $U_{\max}$  (藻类最大生长速率), 其为模型主控方程参数, 是模型初值阶段最敏感的参数。初值输入时还需要同步生成待同化参数集合, 模拟过程同上方方案相同, 当获得叶绿素 a 浓度实测值时, 使用集合卡尔曼滤波同步同化状态变量和参数集合, 分别对集合求期望获得状态变量和参数的同化值, 作为模型下一阶段状态变量和该参数输入, 依此递推。

4) EnKF\_C: 使用本研究开发的新型数据同化算法同步同化状态变量和敏感参数。考虑到湖泊藻类动态模型参数较多(18个参数), 全部用于敏感性分析影响同化效率, 因此选取其中8个总体较为敏感的参数, 用于新型数据同化算法方案中模型每一阶段的敏感性分析, 表4列出了用于敏感性分析的8个参数及其初值。其中 p4(藻类最大生长速率)即为同步同化模型参数方案(EnKF\_B)的同化参数, 该参数是模型主控方程的参数, 在初始状态是模型最为敏感的参数, 其余7个参数也是模型各模块的主要控制部分。因此, 通过分析上述8个参数的敏感性能表征模型的总体敏感程度。新算法初值输入时同步生成待分析敏感度参数集合, 模拟过程依然相同, 当获得叶绿素 a 浓度实测值时, 先使用扰动分析法定位该阶段模型的敏感参数, 然后使用集合卡尔曼滤波依次同化状态变量和敏感参数集合, 分别对集合求期望获得状态变量和敏感参数的同化值, 作为模型下一阶段状态变量和敏感参数输入, 依此递推。

以上4种方案中, Reference 为纯模型模拟方案, 其模拟值作为评价3种数据同化方案对模型精度提升效果的基准; EnKF\_A 和 EnKF\_B 为传统集合卡尔曼滤波数据同化算法2种主要应用方案; EnKF\_C 即为本

表 3 新型数据同化算法公式<sup>\*</sup>  
Tab.3 Equations of the new data assimilation algorithm

| 步骤    | 新型同化算法公式   | 编号   |
|-------|--|------|
| 集合初始化 | $\hat{x}_{t-1}^{(1)}, \hat{x}_{t-1}^{(2)}, \hat{x}_{t-1}^{(3)}, \dots, \hat{x}_{t-1}^{(N)}$                                    |      |
| ①预测   | $\hat{x}_t^{(i)} = f(\hat{x}_{t-1}^{(i)}, u_{t-1}^{(i)}, \tilde{\theta}_{t-1}^{(i)}) + \xi_t^{(i)}, i = 1, 2, \dots, N$        | (1)  |
|       | $u_{t-1}^{(i)} = u_{t-1} + \eta^{(i)}, \eta^{(i)} \sim N(0, \eta), \eta = 0.1u_{t-1}$  | (2)  |
|       | $\tilde{\theta}_{t-1}^{(i)} = \theta_0 + \sigma^{(i)}, \sigma^{(i)} \sim N(0, \sigma)$   | (3)  |
|       | $\xi_{t-1}^{(i)} \sim N(0, \xi), \xi = 0.1f(\hat{x}_{t-1}^{(i)}, u_{t-1}^{(i)}, \tilde{\theta}_{t-1}^{(i)})$                   | (4)  |
|       | $\hat{x}_t = E[\hat{x}_t^{(i)}] = \frac{1}{N} \sum_{i=1}^N \hat{x}_t^{(i)}, i = 1, 2, \dots, N$                                | (5)  |
| ②分析   | $S = \sum_{i=1}^{n-1} \frac{(R_{i+1} - R_i) / R_{\text{mean}}}{(p_{i+1} - p_i) / p_{\text{mean}}} / n - 1, i = 1, 2, \dots, N$ | (6)  |
| ③更新   | $\hat{x}_t^{(i)} = \hat{x}_t^{(i)} + K_t^x(x_t + \omega^{(i)} - \hat{x}_t^{(i)}), i = 1, 2, \dots, N$                          | (7)  |
|       | $\omega^{(i)} \sim N(0, \omega), \omega = r \times x_t$  | (8)  |
|       | $K_t^x = \sum_t (\sum_t^{\hat{x}\hat{x}} + \sum_t^{xx})^{-1}$  | (9)  |
|       | $\hat{x}_t = E[\hat{x}_t^{(i)}] = \frac{1}{N} \sum_{i=1}^N \hat{x}_t^{(i)}, i = 1, 2, \dots, N$                                | (10) |
|       | $\theta_t^{(i)} = \tilde{\theta}_{t-1}^{(i)} + K_t^\theta(x_t + \omega^{(i)} - \hat{x}_t^{(i)}), i = 1, 2, \dots, N$           | (11) |
|       | $K_t^\theta = \sum_t (\sum_t^{\theta\theta} + \sum_t^{xx})^{-1}$   | (12) |
|       | $\theta_t = E[\theta_t^{(i)}] = \frac{1}{N} \sum_{i=1}^N \theta_t^{(i)}, i = 1, 2, \dots, N$                                   | (13) |

\* 预测式中,  $f$  表示湖泊藻类动态模型,  $\hat{x}_t^{(i)}$  表示  $t$  时刻第  $i$  个集合的状态变量预测先验值;  $\hat{x}_{t-1}^{(i)}$  表示  $t-1$  时刻第  $i$  个集合的状态变量预测后验值,  $u_{t-1}$  表示  $t-1$  时刻模型的边界条件向量, 包括氮磷、日照时长、水温、风速及风向等实测数据,  $u_{t-1}^{(i)}$  为  $u_{t-1}$  加上随机向量  $\eta^{(i)}$  生成的第  $i$  个边界条件向量;  $\eta^{(i)}$  来自均值为零协方差为  $\eta$  的正态分布, 其中  $\eta$  取边界条件数值的 10%;  $\theta_0$  为模型参数的初始值;  $\tilde{\theta}_{t-1}^{(i)}$  为  $\theta_0$  加上随机向量  $\sigma^{(i)}$  生成的第  $i$  个模型参数向量,  $\sigma^{(i)}$  来自均值为零协方差为  $\sigma$  的正态分布,  $\sigma$  为参数误差协方差矩阵;  $\xi_{t-1}^{(i)}$  来自均值为零协方差为  $\xi$  的正态分布,  $\xi$  为模型误差协方差矩阵;  $\hat{x}_t$  为  $t$  时刻状态变量预测集合的期望。分析式中,  $S$  表示目标参数相对敏感度;  $p_{i+1}$  和  $p_i$  分别为第  $i+1$  和第  $i$  次目标参数的取值,  $p_{\text{mean}}$  为两者均值;  $R_{i+1}$  和  $R_i$  分别表示第  $i+1$  和第  $i$  次模型输出值,  $R_{\text{mean}}$  为两者均值。更新式中,  $\hat{x}_t^{(i)}$  表示  $t$  时刻第  $i$  个集合的状态变量后验值,  $x_t$  表示  $t$  时刻状态变量观测值;  $\omega^{(i)}$  来自均值为零协方差为  $\omega$  的正态分布, 其中  $\omega$  表示观测值误差协方差;  $K_t^x$  表示状态变量的卡尔曼增益;  $\sum_t^{\hat{x}\hat{x}}$  为状态变量预测集合的误差协方差矩阵,  $\sum_t^{xx}$  为状态变量实测值的误差协方差矩阵;  $\hat{x}_t$  为状态变量后验集合期望, 即状态变量同化值;  $\theta_t^{(i)}$  表示  $t$  时刻第  $i$  个模型参数后验值,  $K_t^\theta$  表示参数更新的卡尔曼增益;  $\sum_t^{\theta\theta}$  为模型参数集合与状态变量预测集合的误差协方差矩阵;  $\theta_t$  表示  $t$  时刻模型参数后验集合期望, 即  $t$  时刻模型参数的初始均值。

研究提出的新型数据同化算法。

### 1.6 数据同化效果评价

为评价模拟精度, 采用纳什系数(Nash-sutcliffe model efficiency, NSE)衡量模拟值和实测值的符合程度; 采用均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)和平均绝对百分比误差(mean absolute percentage error, MAPE)衡量模拟值和实测值的误差<sup>[36]</sup>。

## 2 结果与分析

由于集合卡尔曼滤波中测量误差和模型误差两参数无法显式给出<sup>[11]</sup>。为保证数据同化效果, 先使用数值

表4 敏感性分析的8个参数<sup>[29]</sup>  
Tab.4 Eight parameters for sensitivity analysis<sup>[29]</sup>

| 编号 | 参数          | 描述                 | 单位       | 初值     |
|----|-------------|--------------------|----------|--------|
| p1 | $\vartheta$ | 温度因子               | —        | 1.08   |
| p2 | $\alpha$    | 水及非藻类吸收短波辐射的平均消光系数 | $m^{-1}$ | 0.45   |
| p3 | $K$         | 藻类沉降速率             | $m/d$    | 0.0864 |
| p4 | $U_{max}$   | 藻类最大生长速率           | $d^{-1}$ | 1.145  |
| p5 | $GR_{max}$  | 浮游动物最大捕食速率         | $d^{-1}$ | 0.09   |
| p6 | $k_r$       | 藻类呼吸引起叶绿素a减少速率系数   | $d^{-1}$ | 0.17   |
| p7 | $T_{opt}$   | 最适藻类生长的水温          | ℃        | 27     |
| p8 | $k_e$       | 藻类排泄引起叶绿素a减少速率系数   | $d^{-1}$ | 0.01   |

模拟实验测试;然后依次实施4种方案并根据性能指标评价模拟精度和误差,对比分析4种方案模拟效果.

## 2.1 数据同化算法参数率定

集合卡尔曼滤波中模型误差和测量误差两参数相互影响,需同步率定<sup>[11]</sup>. 模型误差主要来源于模型结构、参数及边界条件误差等<sup>[37]</sup>,为合理评估模型误差,使用模型模拟值与实测值计算8个自动监测站MAPE,取其中最小值(10%)和最大值(50%)作为模型误差率定区间边界;测量误差主要来源于测量仪器及环境(气温、风力等)条件等,综合考虑仪器精度及测量过程,测量误差率定区间边界取定为1%~10%. 同步率定时使用RMSE评价数据同化算法模拟误差,RMSE取值越小说明同化效果越佳. 根据数值模拟结果所示(图4),当模型误差和测量误差分别取18%和3%时,数据同化效果最佳,由此即确定了集合卡尔曼滤波模型误差和测量误差取值. 研究表明集合卡尔曼滤波过程中集合个数越多提升效果越好<sup>[17]</sup>,但考虑到效率问题,本研究中集合卡尔曼滤波的集合个数取定为100,此集合数已被证明在大多数模型中都足够<sup>[38]</sup>. 集合卡尔曼滤波的初值对同化结果影响不大,因此将模型初始状态的误差设置为20%.

根据上述讨论,3种数据同化方案中集合卡尔曼滤波的参数如下:模型误差为18%,测量误差为3%,集合个数为100,初始状态误差为20%.

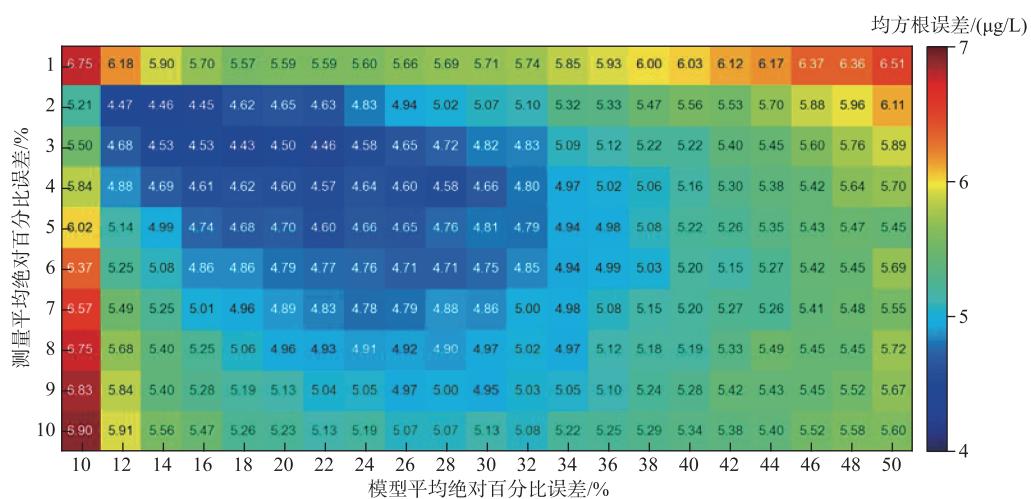


图4 不同模型误差与测量误差背景下的同化结果

Fig.4 Assimilation results in case of different models and measurement errors

## 2.2 4种方案结果对比分析

图5展示了8个自动监测站2020年1月1日—6月1日逐日叶绿素a浓度实测值和4种方案的模拟曲

线,并给出了4种方案模拟曲线的一项拟合指标(*NSE*)和3项误差指标(*RMSE*、*MAE*、*MAPE*)。从图5叶绿素*a*浓度实测值看出,时间上,多数自动监测站叶绿素*a*浓度峰值出现在4月份;空间上,巢湖西部和南部水域的叶绿素*a*浓度总体高于东部和北部水域;从叶绿素*a*浓度极大值模拟看,数据同化后的模拟值大部分略低于实测值,因多数藻类实测的极大值梯度大、回落迅速,模型往往无法及时反应,其模拟值具有滞后性且难以达到峰值高度,经过滤波后会取到一个中间值,因此滤波模拟结果多数略低于实测极大值而又高于模型模拟结果;另外,还需注意蓝藻水华暴发期藻类上浮堆积在水面可能会遮住自动监测探头,监测数据不够准确,存在出现异常峰值的可能。以叶绿素*a*浓度实测值为依据,下面详细分析4种方案模拟效果。

1) Reference: 模型模拟的叶绿素*a*浓度曲线仅能够反映实测值大致变化趋势,且在模拟预测叶绿素*a*浓度峰值上难以发挥作用,模拟结果误差较大。从性能评价指标得出模型模拟精度在8个自动监测站差距较大,新河入湖区站模拟值的*NSE*仅为0.1,而东半湖湖心站模拟值*NSE*则达到了0.69。8个自动监测站平均*NSE*为0.49,平均*RMSE*为7.31  $\mu\text{g}/\text{L}$ ,平均*MAPE*为44%,可见模型总体模拟效果并不理想。

2) EnKF\_A: 从集合卡尔曼滤波同化状态变量方案模拟曲线看出,经过同化状态变量的模型模拟曲线更加贴近实测值。从性能评价指标得出,8个自动监测站模拟值的平均*NSE*提升到了0.63,平均*RMSE*降低了2.47  $\mu\text{g}/\text{L}$ ,平均*MAPE*也降低了9%,说明同化状态变量能够提升模型模拟精度。

3) EnKF\_B: 从集合卡尔曼滤波同步同化状态变量和模型参数方案模拟曲线看出,相较于仅同化状态变量,同步同化模型参数后模型模拟精度提升有限。尽管8个自动监测站平均*NSE*提升了0.08,但注意到其中3个站(湖滨站、东半湖湖心、兆河入湖区)的*NSE*不增反减,而平均*NSE*的提升主要来自新河入湖区站,从0.26提升到0.53。这反映不考虑模型参数敏感性动态变化而对固定参数调整难以有效校正模型运行轨迹,反而会因不当调整参数使同化效果更差,存在给模型模拟带来额外不确定性的隐患。

4) EnKF\_C: 从本研究提出的耦合参数敏感性分析的新型数据同化方案模拟曲线看出,新型数据同化方案能够更加准确地模拟预测叶绿素*a*浓度,在上述2种数据同化方案上有了进一步提升。从性能评价指标得出该方案的模型模拟值平均*NSE*达到了0.76,平均*RMSE*为4.27  $\mu\text{g}/\text{L}$ ,平均*MAPE*下降到23%。无论总体上还是在每个自动监测站,考虑了参数敏感性的新型数据同化算法对传统集合卡尔曼滤波数据同化效果都有稳定的提升,对叶绿素*a*浓度峰值有更加准确的模拟。考虑到本研究设计的3种数据同化方案参数条件都相同,因此能够确定模型模拟精度的提升来源于新型数据同化算法对敏感参数的实时定位和调整。这说明耦合了参数敏感性分析的新型数据同化算法能够有效定位模型运行每阶段的敏感参数,调整模型运行轨迹,增加模型稳定性,提高模型叶绿素*a*浓度模拟精度。

### 3 讨论

案例测试结果表明:巢湖藻类模型的参数敏感性存在显著的动态变化过程,提出新型数据同化算法能够实时跟踪模型敏感参数,实现模型模拟精度的最大化提升。在新型数据同化算法方案中,根据每阶段模型模拟值,将待分析敏感性的8个参数依次输入扰动分析模块即可获得其敏感度*S*。以湖滨站1—5月参数敏感性的动态变化(图6)为例,p1(温度因子)和p4(藻类最大生长速率)在多数时段敏感度等级都在Ⅲ级( $|S| \geq 0.5$ ,一般敏感)及以上,敏感性总体较高。尤其是p4,在模型运行时间段有62 d的敏感度都在Ⅴ级( $|S| \geq 1.0$ ,极度敏感)以上,是模型最敏感的参数,但也有39 d处于不敏感状态。表明:受边界条件等要素影响,参数敏感度在模型运行各阶段不断变化,敏感参数在模型动态运行过程中存在转为不敏感的可能。其它参数虽多不敏感,但随着模型的运行也有敏感时刻。例如,p3(藻类沉降速率)在模型运行前20 d都不敏感,但在1月21日变为极度敏感参数( $|S| \geq 1.0$ ),其在当天敏感程度超过了p4( $1.0 \geq |S| \geq 0.75$ ),表明随着模型的动态运行,原本不敏感参数也可能受边界条件影响转为敏感参数。

通过对比分析可知,新型数据同化算法能够显著提升模型模拟精度的原因在于耦合了参数敏感性分析,使其能够实时定位模型敏感参数并有效调整,校正模型运行轨迹,使模型模拟更加精确。在本研究的湖泊藻类动态模型应用中,参数敏感性分析方法选择了成熟的扰动分析法,从模型中选取了8个具有代表性参数(表4)用于敏感性分析,定位到敏感参数后使用集合卡尔曼滤波更新敏感参数,实时校正模型运行轨迹。而传统集合卡尔曼滤波数据同化算法则仅能够调整固定的参数,难以应对模型运行中的敏感参数动态

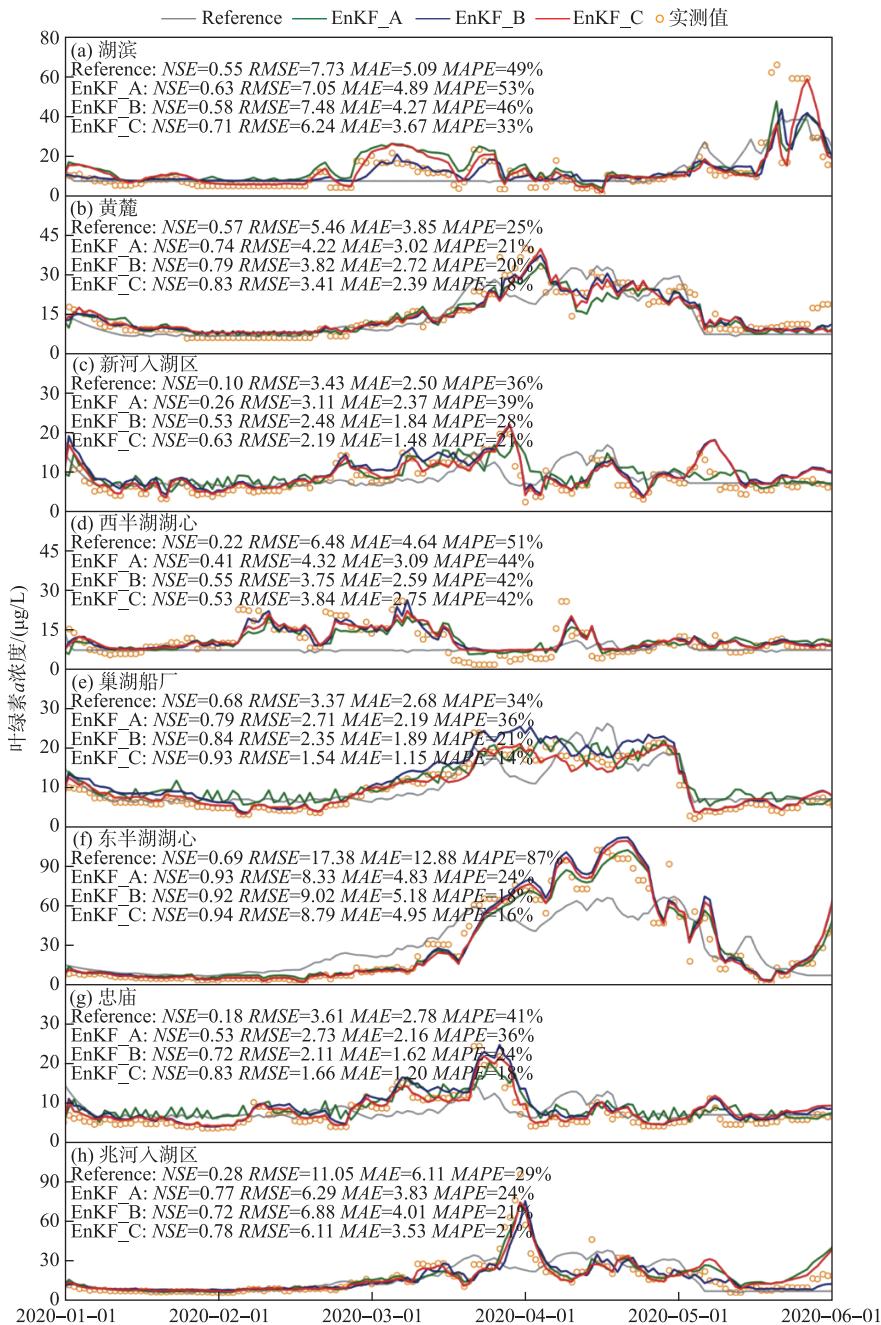


图 5 不同数据同化方案的叶绿素 *a* 浓度模拟效果比较  
Fig.5 Comparison of the simulated chlorophyll-*a* concentration performance from different data assimilation strategies

变化,无法有效校正模型运行轨迹,甚至不当的参数调整还会增加模型不确定性。通过两种模型同化方案(EnKF\_B、EnKF\_C)中参数值动态变化,能够看出新算法相较于传统算法对模型参数的优化调整过程,考虑到站点多、时序长,图7仅展示湖滨站1月份两种方案中参数值的动态变化率,选择湖滨站是为了方便从图

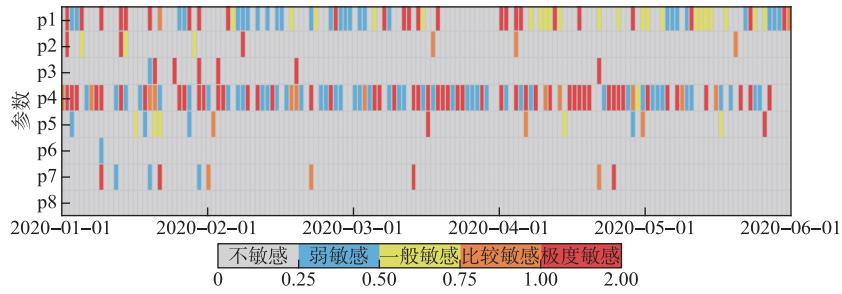


图 6 基于扰动分析法的参数敏感性动态变化过程(湖滨站,参数 p1~p8 含义见表 4)

Fig.6 The parameter sensitivity dynamic change process based on the disturbance analysis method  
(Hubin Station, the meaning of parameters p1-p8 are shown in Tab.4)

6 中同步观察各参数 1 月份的敏感度。从图 7a 看出,由于固定同化参数 p4(藻类最大生长速率),EnKF\_B 方案仅能且不断的调整该参数,即使是在该参数处于不敏感状态(如 1 月 5 日)依然调整,而其它参数即使处于敏感状态也不会进行调整。而从图 7b 看出,EnKF\_C 方案则能够根据模型敏感参数动态变化灵活改变对参数的调整,合理应对敏感参数动态变化。以 p4 为例,当其处于敏感状态时(如 1 月 2-4 日)会及时进行调整,当其处于非敏感状态时(如 1 月 6 日)则不会进行调整。由此可见,本研究提出的新型数据同化算法正是通过耦合参数敏感性分析,实时动态调整敏感参数,实现模型模拟精度的稳定提升,具有广泛应用前景。

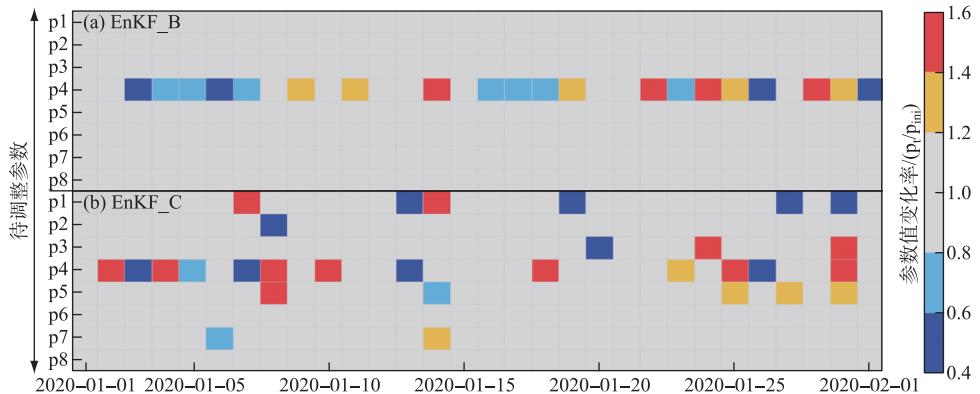
图 7 不同数据同化方案(EnKF\_B(a) 和 EnKF\_C(b))的参数动态变化  
(湖滨站,2020 年 1 月,参数 p1~p8 含义见表 4)

Fig.7 The change of parameter values for the assimilation strategies of EnKF\_B (a) and EnKF\_C (b)  
(Hubin Station, January 2020, parameters (p1-p8) descriptions can be found in Tab.4)

本研究模拟时段为 2020 年 1—5 月,主要是由于在 5—10 月的蓝藻水华暴发期间,藻类水华堆积可能影响探头准确测定,自动监测数据质量受影响;同时,根据孔繁翔等提出的湖泊蓝藻水华形成四阶段理论<sup>[39]</sup>,本研究模拟时间段(1—5 月)主要处于藻类复苏和生物量增加的初期,也是决定全年蓝藻水华情势的重要时期,开展此时段的蓝藻水华预测也有一定意义。

#### 4 结论

本研究通过耦合数据同化算法与敏感参数识别算法,实现了模型敏感参数的实时识别与同化,解决了传统数据同化算法难以适应模型敏感参数动态变化的问题;改进算法应用于巢湖藻类动态模拟,测试结果表明:改进算法实现了湖泊藻类动态模型与高频自动监测数据的实时融合,显著提升了模型模拟精度(模拟

误差减少 42% ), 对藻类峰值的模拟精度提升尤为明显; 算法与模型采用松散耦合模式, 可移植性强, 可应用于其它水生态环境模型的数据同化, 随着我国河湖多源监测数据的剧增, 算法具有广阔应用前景.

## 5 参考文献

- [ 1 ] Yang CH, Li JY, Yin HB. Phosphorus internal loading and sediment diagenesis in a large eutrophic lake (Lake Chaohu, China). *Environmental Pollution*, 2022, **292**: 118471. DOI: 10.1016/j.envpol.2021.118471.
- [ 2 ] Tong YH. Research on short-term prediction of algal bloom in reservoirs [ Dissertation ]. Hangzhou: Zhejiang University, 2012. [ 全玉华. 库区藻类水华短期预测研究 [ 学位论文 ]. 杭州: 浙江大学, 2012. ]
- [ 3 ] Zhang M, Shi XL, Yang Z et al. The variation of water quality from 2012 to 2018 in Lake Chaohu and the mitigating strategy on cyanobacterial blooms. *J Lake Sci*, 2020, **32**(1): 11-20. DOI: 10.18307/2020.0102. [ 张民, 史小丽, 阳振等. 2012—2018 年巢湖水质变化趋势分析和蓝藻防控建议. 湖泊科学, 2020, **32**(1): 11-20. ]
- [ 4 ] Yu PX, Gao R, Zhang DZ et al. Predicting coastal algal blooms with environmental factors by machine learning methods. *Ecological Indicators*, 2021, **123**: 107334. DOI: 10.1016/j.ecolind.2020.107334.
- [ 5 ] Zhu YY, Jin X, Tang WZ et al. Comprehensive analysis of nitrogen distributions and ammonia nitrogen release fluxes in the sediments of Baiyangdian Lake, China. *Journal of Environmental Sciences*, 2019, **76**: 319-328. DOI: 10.1016/j.jes.2018.05.024.
- [ 6 ] Wu ZS, Wang XL, Chen YW et al. Assessing river water quality using water quality index in Lake Taihu Basin, China. *Science of the Total Environment*, 2018, **612**: 914-922. DOI: 10.1016/j.scitotenv.2017.08.293.
- [ 7 ] Lin C, Xiong JF, Ma RH et al. Identifying the critical watershed regions creating lake nutrient enrichment (SLRs) based on a watershed-lake integrated perspective—A case study of Chaohu Lake Basin, China. *Ecological Indicators*, 2021, **121**: 107037. DOI: 10.1016/j.ecolind.2020.107037.
- [ 8 ] Zhang YY, Shao QX. Uncertainty and its propagation estimation for an integrated water system model: An experiment from water quantity to quality simulations. *Journal of Hydrology*, 2018, **565**: 623-635. DOI: 10.1016/j.jhydrol.2018.08.070.
- [ 9 ] Zhang HX, Huo SL, Wang R et al. Hydrologic and nutrient-driven regime shifts of cyanobacterial and eukaryotic algal communities in a large shallow lake: Evidence from empirical state indicator and ecological network analyses. *Science of the Total Environment*, 2021, **783**: 147059. DOI: 10.1016/j.scitotenv.2021.147059.
- [ 10 ] Narayan A, Marzouk Y, Xiu DB. Sequential data assimilation with multiple models. *Journal of Computational Physics*, 2012, **231**(19): 6401-6418. DOI: 10.1016/j.jcp.2012.06.002.
- [ 11 ] Evensen G. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 2003, **53**(4): 343-367. DOI: 10.1007/s10236-003-0036-9.
- [ 12 ] Davakis E, Andronopoulos S, Kovalets I et al. Data assimilation in meteorological pre-processors: Effects on atmospheric dispersion simulations. *Atmospheric Environment*, 2007, **41**(14): 2917-2932. DOI: 10.1016/j.atmosenv.2006.12.031.
- [ 13 ] Hu S, Shi LS, Zha YY et al. Simultaneous state-parameter estimation supports the evaluation of data assimilation performance and measurement design for soil-water-atmosphere-plant system. *Journal of Hydrology*, 2017, **555**: 812-831. DOI: 10.1016/j.jhydrol.2017.10.061.
- [ 14 ] Baduru B, Paul B, Banerjee DS et al. Ensemble based regional ocean data assimilation system for the Indian Ocean: Implementation and evaluation. *Ocean Modelling*, 2019, **143**: 101470. DOI: 10.1016/j.ocemod.2019.101470.
- [ 15 ] Chen C, Huang JC, Chen QW et al. Assimilating multi-source data into a three-dimensional hydro-ecological dynamics model using Ensemble Kalman Filter. *Environmental Modelling & Software*, 2019, **117**: 188-199. DOI: 10.1016/j.envsoft.2019.03.028.
- [ 16 ] Cho KH, Pachepsky Y, Ligaray M et al. Data assimilation in surface water quality modeling: A review. *Water Research*, 2020, **186**: 116307. DOI: 10.1016/j.watres.2020.116307.
- [ 17 ] Liu Z, Li ZJ, Hu LM et al. Ensemble Kalman filter based data assimilation in the Delft3D-BLOOM lake eutrophication model. *J Lake Sci*, 2017, **29**(5): 1070-1083. DOI: 10.18307/2017.0505. [ 刘卓, 李志杰, 胡柳明等. 基于集合卡尔曼滤波的湖泊富营养化模型 Delft3D-BLOOM 数据同化. 湖泊科学, 2017, **29**(5): 1070-1083. ]
- [ 18 ] Huang JC, Gao JF, Liu JT et al. State and parameter update of a hydrodynamic-phytoplankton model using ensemble Kalman filter. *Ecological Modelling*, 2013, **263**: 81-91. DOI: 10.1016/j.ecolmodel.2013.04.022.
- [ 19 ] Lei XH, Tian Y, Zhang Z et al. Correction of pumping station parameters in a one-dimensional hydrodynamic model using

- the Ensemble Kalman filter. *Journal of Hydrology*, 2019, **568**: 108-118. DOI: 10.1016/j.jhydrol.2018.10.062.
- [20] Li G, Chen C, He XX et al. Modify of data assimilation model for lake algae dynamic model. *Water Resources Protection*, 2021, **37**(4): 156-165. [李港, 陈诚, 何欣霞等. 湖泊藻类动态模型数据同化模式的改进. 水资源保护, 2021, **37**(4): 156-165.]
- [21] Song XD, Bryan BA, Almeida AC et al. Time-dependent sensitivity of a process-based ecological model. *Ecological Modelling*, 2013, **265**: 114-123. DOI: 10.1016/j.ecolmodel.2013.06.013.
- [22] Luo C, Li ZF, Xi Q et al. Sensitivity analysis of hydrological and water quality parameters of HSPF model. *Journal of Agro-Environment Science*, 2014, **33**(10): 1995-2002. DOI: 10.11654/jaes.2014.10.017. [罗川, 李兆富, 席庆等. HSPF 模型水文水质参数敏感性分析. 农业环境科学学报, 2014, **33**(10): 1995-2002.]
- [23] Huang JC, Gao JF, Hörmann G. Hydrodynamic-phytoplankton model for short-term forecasts of phytoplankton in Lake Taihu, China. *Limnologica*, 2012, **42**(1): 7-18. DOI: 10.1016/j.limno.2011.06.003.
- [24] Liu HQ, Ren HK, Niu XX et al. Extraction of cyanobacteria bloom in Chaohu Lake based on Sentinel-2 remote sensing images. *Ecology and Environmental Sciences*, 2021, **30**(1): 146-155. [刘海秋, 任恒奎, 牛鑫鑫等. 基于 Sentinel-2 遥感影像的巢湖蓝藻水华提取方法研究. 生态环境学报, 2021, **30**(1): 146-155.]
- [25] Wu ZS, Lai XJ, Li KY. Water quality assessment of rivers in Lake Chaohu Basin (China) using water quality index. *Ecological Indicators*, 2021, **121**: 107021. DOI: 10.1016/j.ecolind.2020.107021.
- [26] Li SM, Liu JP, Song KS et al. Analysis on spatial and temporal character of algae bloom in Lake Chaohu and its driving factors based on landsat imagery. *Resources and Environment in the Yangtze Basin*, 2019, **28**(5): 1205-1213. DOI: 10.11870/cjlyzyyhj201905020. [李晟铭, 刘吉平, 宋开山等. 基于 Landsat 影像巢湖蓝藻水华暴发时空变化特征及其驱动因素分析. 长江流域资源与环境, 2019, **28**(5): 1205-1213.]
- [27] Kibuye FA, Zamyadi A, Wert EC. A critical review on operation and performance of source water control strategies for cyanobacterial blooms: Part I-chemical control methods. *Harmful Algae*, 2021, **109**: 102099. DOI: 10.1016/j.hal.2021.102099.
- [28] Ułańczyk R, Kliś C, Łozowski B et al. Phytoplankton production in relation to simulated hydro- and thermodynamics during a hydrological wet year—Goczałkowice reservoir (Poland) case study. *Ecological Indicators*, 2021, **121**: 106991. DOI: 10.1016/j.ecolind.2020.106991.
- [29] Kim J, Lee T, Seo D. Algal bloom prediction of the lower Han River, Korea using the EFDC hydrodynamic and water quality model. *Ecological Modelling*, 2017, **366**: 27-36. DOI: 10.1016/j.ecolmodel.2017.10.015.
- [30] Lek S. Uncertainty in ecological models. *Ecological Modelling*, 2007, **207**(1): 1-2. DOI: 10.1016/j.ecolmodel.2007.03.015.
- [31] Chen Q, Zhao J, Gao Q et al. Trophic state footprint index model and its application to Dianchi Lake, China. *Ecological Indicators*, 2021, **132**: 108317. DOI: 10.1016/j.ecolind.2021.108317.
- [32] Ruley JE, Rusch KA. Development of a simplified phosphorus management model for a shallow, subtropical, urban hypereutrophic lake. *Ecological Engineering*, 2004, **22**(2): 77-98. DOI: 10.1016/j.ecoleng.2004.02.005.
- [33] Burgers G, van Leeuwen PJ, Evensen G. Analysis scheme in the Ensemble Kalman Filter. *Monthly Weather Review*, 1998, **126**(6): 1719-1724. DOI: 10.1175/1520-0493(1998)126:1719:asitek>2.0.co;2.
- [34] Choi JY, Choi CH. Sensitivity analysis of multilayer perceptron with differentiable activation functions. *IEEE Transactions on Neural Networks*, 1992, **3**(1): 101-107. DOI: 10.1109/72.105422.
- [35] Patil A, Ramsankaran R. Improved streamflow simulations by coupling soil moisture analytical relationship in EnKF based hydrological data assimilation framework. *Advances in Water Resources*, 2018, **121**: 173-188. DOI: 10.1016/j.advwatres.2018.08.010.
- [36] Bennett ND, Croke BFW, Guariso G et al. Characterising performance of environmental models. *Environmental Modelling & Software*, 2013, **40**: 1-20. DOI: 10.1016/j.envsoft.2012.09.011.
- [37] Lai RX. Data assimilation and parameters inversion of water-sediment numerical model [Dissertation]. Beijing: Tsinghua University, 2015. [赖瑞勋. 水流泥沙数学模型的数据同化与参数反演[学位论文]. 北京: 清华大学, 2015.]
- [38] Evensen G. Data assimilation. The Ensemble Kalman filter. Berlin: Springer, 2006.
- [39] Kong FX, Gao G. Hypothesis on cyanobacteria bloom-forming mechanism in large shallow eutrophic lakes. *Chinese Journal of Plant Ecology*, 2005, **25**(3): 589-595. [孔繁翔, 高光. 大型浅水富营养化湖泊中蓝藻水华形成机理的思考. 生态学报, 2005, **25**(3): 589-595.]