

DOI:10.7522/j.issn.1000-0240.2022.0004

CHEN Chong, ZHANG Wei, XING Qinghui, et al. Uncertainty analysis of groundwater model based on data assimilation[J]. Journal of Glaciology and Geocryology, 2022, 44(6):1912-1924. [陈冲, 张伟, 邢庆辉, 等. 基于数据同化的地下水模型不确定性分析[J]. 冰川冻土, 2022, 44(6):1912-1924.]

# 基于数据同化的地下水模型不确定性分析

陈 冲<sup>1</sup>, 张 伟<sup>2</sup>, 邢庆辉<sup>1</sup>, 豆沂宣<sup>1</sup>

(1. 中国石油大学(北京)信息科学与工程学院, 北京 102249; 2. 中国科学院 西北生态环境资源研究院  
冰冻圈科学国家重点实验室/可可托海站, 甘肃 兰州 730000)

**摘 要:** 黑河流域中下游地下水系统受上游冰冻圈融水和降雨的补给, 由气候变暖导致的冰冻圈萎缩致使中下游地下水系统的稳定性面临更多的风险。地下水模型是地下水系统稳定性评估的有效手段, 但是地下水模型参数往往存在较大的不确定性。为此, 本文提出了基于数据同化算法的不确定性分析方法, 通过包含观测资料信息减小模型不确定性。采用所提方法分析了(基于MODFLOW构建)黑河流域中游地下水模型中13个参数的不确定性, 讨论了算法超参数的影响及其最优取值, 分析了地下水模型参数的不确定性。实验结果证明数据同化算法可有效减小地下水模型参数的不确定性, 观测资料的种类与数量对参数不确定性的减小起到重要作用; 不同地下水模型参数的不确定性不同, 地表水与地下水相互作用频繁的区域参数不确定性较大; 含水层渗透系数、含水层给水度以及灌溉回流系数对模型输出的地下水位输出影响显著, 河床水力传导系数对模型输出的河流流量影响较大。本研究将为地下水研究提供更加可靠的模型方法, 为西北内流区地下水哺育的绿洲生态系统稳定可持续研究提供重要支撑。

**关键词:** 地下水模型; 不确定性分析; 数据同化; MODFLOW

**中图分类号:** P641 **文献标志码:** A **文章编号:** 1000-0240(2022)06-1912-13

## 0 引言

地下水系统模拟已经成为地球科学的基本研究方法, 在地球系统科学领域被誉为第二次哥白尼革命<sup>[1]</sup>, 被证明是最有价值和最实用的工具之一<sup>[2-3]</sup>。地下水模型能够刻画水文过程的整体和局部行为; 多次模拟, 选择最优的设计; 通过情景分析进行情景预测, 提出应对策略。围绕黑河中游地表水、泉水和地下水, 大量学者利用模型进行了有益探索: 陈冲等<sup>[4]</sup>建立了黑河中游地区地下水模型, 并初步分析了上游径流与耕地面积对地下水资源的影响; 王旭升等<sup>[5]</sup>总结了20年来黑河流域的地下水模型研究, 并指出地下水模型应加强与关联过程(地表水、土壤水、水利工程)的集成; 程国栋等<sup>[6]</sup>阐述了黑河流域生态—水文过程集成研究的进展及展望。目前的地下水模型一般基于确定性的参数、

边界条件, 采用确定性机制来构建模型。然而, 由于实际过程的复杂性及非线性因素, 导致模型往往存在一定不确定性。正如George所言, 所有关于真实系统的模型都是“错误”的<sup>[7]</sup>, 即模型永远不能完美表达实际系统。地下水模型的不确定性来源一般可以分为: 模型数据的不确定性、模型结构的不确定性以及模型参数的不确定性。模型数据的不确定性一般指由于监测误差或者数据缺失导致的不确定性, 只能通过提高监测技术和数据收集频率来降低。而模型结构和参数不确定性则来源于建模过程、模型参数标定及验证过程。这些不确定性将对进一步的科学研究以及决策提出等带来诸多麻烦甚至风险。因此在使用地下水模型时, 对其进行不确定性分析是十分必要的。

Oberkampf等<sup>[8]</sup>介绍了用来估计、分析模型中所

收稿日期: 2020-11-30; 修订日期: 2021-09-13

基金项目: 国家自然科学基金项目(62006247); 国家重点研发计划项目(2019YFC1510501)资助

作者简介: 陈冲, 副教授, 主要从事地下水模型研究. E-mail: chenchong@cup.edu.cn

有类型不确定性的多种方法。Beven<sup>[9]</sup>于1989年在对物理模型的适用性进行讨论的过程中,首次提出了对水文模型进行不确定性分析的概念。地下水模型不确定性分析方法主要分为蒙特卡罗(Monte Carlo)法、矩方程法与贝叶斯法三种<sup>[10]</sup>。虽然MC法是一种被广泛采用的不确定性分析方法<sup>[11-15]</sup>,但是MC法的缺点也不容忽视,其需对模型的参数进行大量的采样才能保证算法的正确性和准确性,无法适用于计算消耗大的模型。矩方程法通过随机偏微分方程直接求解模拟结果的各阶统计矩,已在地下水模型不确定性研究中得到初步应用<sup>[16-18]</sup>。贝叶斯法利用观测资料修正水文地质参数分布,其既可以用于模型参数识别反演<sup>[19]</sup>,也能够用于对参数进行不确定性分析,其优点是通过更新参数分布,以更加准确的评估模型不确定性。利用贝叶斯法进行地下水模型不确定性分析的研究相对较少<sup>[20-21]</sup>。数据同化(DA, Data Assimilation)是基于贝叶斯理论的参数更新方法。卡尔曼滤波算法(KF, Kalman Filter)是由Kalman提出的顺序数据同化算法<sup>[22]</sup>,其利用观测资料自回归地对模型的状态变量进行更新,并在更新的整个过程中保证状态变量估计值的误差最小。自提出以来,KF算法已被广泛应用于模型参数与状态的估计中。黄春林等<sup>[23]</sup>基于集合卡尔曼滤波(EnKF, Ensemble Kalman Filter)利用土壤水分观测数据同化了简单生物圈模型(SiB2, Simple Biosphere Model 2),探讨了简单生物圈模型的单点土壤水分同化方案;褚楠等<sup>[24]</sup>进一步采用双EnKF方法同时估计SiB2模型中土壤水分与土壤属性参数,提高了模型对土壤水分的估计精度。Li等<sup>[25]</sup>实现了采用EnKF同时估计地下水模型的参数与状态变量。Sly等<sup>[26]</sup>基于EnKF算法采用SWOT(Surface Water and Ocean Topography)模型的输出对GHMs(Global Hydrological Models)进行了同化,提升了全球尺度水文模拟的精度。然而,仅有少量研究基于数据同化方法进行模型不确定性分析。例如,Li等<sup>[25]</sup>采用了ESMDA(Ensemble Smoother with Multiple Data Assimilation)分析了焉耆盆地水文模型参数的不确定性。

本文基于前期工作中在黑河流域中游构建的地下水模型<sup>[4]</sup>,拟结合ESMDA与EnKF实现地下水模型参数的不确定性分析,探讨ESMDA与EnKF算法超参数对模型不确定性分析效果的影响,利用算法定量分析地下水模型参数的不确定性、对不同观

测数据的敏感性以及不同参数的不确定性范围。

## 1 研究区域

黑河流域是我国第二大内陆河流域,位于我国西北部干旱区,面积约为 $14\times 10^4\text{ km}^2$ 。流域上游的祁连山区是我国重要的冰冻圈分布区,气温较低、蒸发弱,年平均降水量大于300 mm,丰富的冰冻圈融水与降水是整个流域的主要水源。流域下游为额济纳旗盆地,为典型的大陆性气候,具有降水少(年均降水量为43.7 mm)、蒸发强(年均蒸发量为2 248.8 mm)、日照时间长、昼夜温差大等特性<sup>[27]</sup>。流域内支流共35个,由于过度开采及蒸发,绝大部分支流在山前消失,中游主要河流只有黑河干流与梨园河(图1)。黑河发源于祁连山北麓,由东南流向西北,途径青海、甘肃及内蒙古三个省区,最终注入居延海,全长约821 km,水量主要由降水、冰雪融水以及地下水出流组成。莺落峡水文站为黑河出山口径流量观测站,多年(1945—2012年)平均年径流量约为 $15.9\times 10^8\text{ m}^3$ ,是中游张掖市、临泽县、高台县及下游金塔东部和额济纳旗绿洲等地城市工业、生活用水的主要水源<sup>[28]</sup>。黑河流域中游( $38^{\circ}38'\sim 39^{\circ}53'\text{ N}$ ,  $98^{\circ}53'\sim 100^{\circ}44'\text{ E}$ ; 图1)为典型的干旱气候,年降水量稀少(90~160 mm),潜在蒸发量大(2 000~2 500 mm),总面积约9 016  $\text{km}^2$ <sup>[29-30]</sup>。河流流出祁连山后,一部分被引入灌溉渠系和供水系统,其余则沿河床下泄,并于沿途渗入地下,补给地下水,黑河在山前冲积扇区损失33%的径流量<sup>[31]</sup>。长期以来,黑河中游农业生产与下游湿地、尾间湖等生态环境之间存在着用水矛盾,且由于社会、经

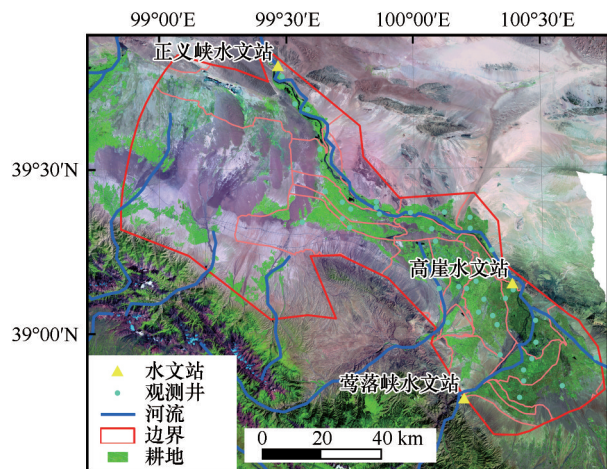


图1 研究区地理位置

Fig. 1 Study area and observation sites

济发展和人口的增长,用水矛盾日益突出。

在前期工作中,我们构建了黑河流域中游的地下水模型,采用了SRTM(Shuttle Radar Topography Mission)的DEM(Digital Elevation Model)数据<sup>[32]</sup>、土地利用<sup>[33-34]</sup>、抽水<sup>[35]</sup>、灌溉<sup>[36]</sup>、地下水位<sup>[37]</sup>以及河流径流量数据。将SRTM 90 m×90 m的数据处理为1 km×1 km的网格数据对研究区含水层系统进行空间离散化。时间上,将所获取数据(1986年1月—2010年12月)离散为月时间间隔,采用以每个自然月为一个应力期的非稳定流(暂态流, Transient state)模拟。地下水位数据包括42口观测井处的观测水位,河流径流量包括莺落峡、高崖及正义峡处观测的径流数据。经过校正(1986年1月—2008年12月)与验证(2009年1月—2010年12月),模型对研究区内地下水位及径流量的拟合程度较好,所构建的地下水模型比较真实地反映了研究区地下水系统的情况<sup>[4]</sup>。

## 2 研究方法

本文采用ESMDA与EnKF结合的方法ESMDA-EnKF对研究区地下水模型的参数进行不确定性分析。图2给出了使用ESMDA算法进行不确定性分析的程序流程。在进行不确定性分析之前,首先要对算法进行初始化,并对模型的参数进行设置,根据给定的概率分布(一般假定为正态分布或者对数正态分布)生成模型参数集合。将模型参数预测值分别输入地下水模型中,计算得到模型输出

值;综合地下水模型输出、观测值、观测误差协方差以及参数预测值集合输入ESMDA-EnKF算法中,从模型输出值和观测值集合计算得到增量场,根据集合中样本的误差统计计算出卡尔曼增益,由增量场与卡尔曼增益计算得到预测结果的更新量,将更新量叠加到初始场得到分析值集合;根据预先算法执行次数判定是否达到程序运行结束条件。

### 2.1 集合卡尔曼滤波(Ensemble Kalman Filter)

EnKF采用蒙特卡罗方法随机产生参数集合,对状态变量进行预测,并根据获取的观测信息对状态变量进行更新,已经有许多文献对其理论和具体算法做了详细的论述<sup>[38-39]</sup>。此处仅简要回顾一下EnKF的主要的工作原理。

在EnKF中,定义第 $t$ 个时刻的参数和状态向量集合为:

$$X_t = \begin{pmatrix} A \\ B \end{pmatrix}_t \quad (1)$$

式中: $X$ 为 $N_x$ 维状态变量; $A$ 为 $N_a$ 维参数向量; $B$ 为 $N_b$ 维系统状态向量;且

$$N_x = N_a + N_b \quad (2)$$

则状态变量在第 $t+1$ 时刻的预报值 $X_{t+1}^f$ 为

$$X_{t+1}^f = M_t(X_t^a) + w_k, \quad w_k \sim N(0, W_k) \quad (3)$$

式中: $f$ 代表状态变量的预测值(forecast); $a$ 代表状态变量的分析值(analysis); $X_{t+1}^f$ 为第 $i$ 个状态变量在 $t+1$ 时刻的预测值; $M_t(\cdot)$ 为非线性模型算子; $X_{i,t}^a$ 为 $t$ 时刻的分析值; $w_k$ 为期望为0,方差为 $W_k$ 的高斯白噪声。

在分析步骤(analysis step)中,采用公式(4)对观测数据进行扰动

$$Y = Y_{t+1}^0 + \varepsilon, \quad \varepsilon \sim N(0, R_e) \quad (4)$$

式中: $Y_{t+1}^0$ 为 $t+1$ 时刻的观测值; $\varepsilon$ 为期望为0,方差为 $R_e$ 的高斯白噪声,则第 $t+1$ 时刻状态变量的分析值采用以下公式进行更新:

$$X_{t+1}^a = X_{t+1}^f + \frac{P_{t+1}^f H^T}{H P_{t+1}^f H^T + R_e} (Y - H X_{t+1}^f) \quad (5)$$

$$P_{t+1}^a = (I - K_{t+1} H) P_{t+1}^f \quad (6)$$

其中,

$$P_{t+1}^f = \frac{1}{N-1} \sum_{i=1}^N (X_{i,t+1}^f - \bar{X}_{t+1}^f) (X_{i,t+1}^f - \bar{X}_{t+1}^f)^T \quad (7)$$

$$\bar{X}_{t+1}^f = \frac{1}{N} \sum_{i=1}^N X_{i,t+1}^f \quad (8)$$

并令,

$$K_{t+1} = \frac{P_{t+1}^f H^T}{H P_{t+1}^f H^T + R_e} \quad (9)$$

式中: $X_{i,t+1}^a$ 为第 $t+1$ 时刻状态变量的分析值; $N$ 为集

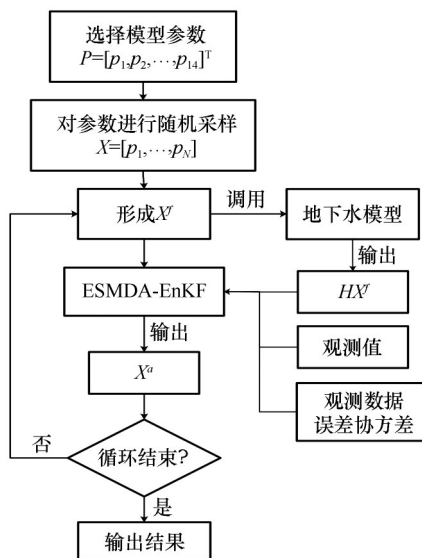


图2 ESMDA-EnKF算法不确定性分析流程图

Fig. 2 Diagram of uncertainty analysis using ESMDA-EnKF



合大小;  $I$  为单位矩阵;  $H(\cdot)$  为观测算子, 用以将状态变量从状态空间映射到观测空间;  $K_{t+1}$  为  $t+1$  时刻的卡尔曼增益;  $P_{t+1}^f$  为  $t+1$  时刻状态变量预报值的误差协方差矩阵;  $P_{t+1}^a$  为  $t+1$  时刻状态变量分析值的误差协方差矩阵;  $\overline{X_{t+1}^f}$  为  $t+1$  时刻状态变量预报值的均值。

## 2.2 多重数据同化集合平滑器 (Ensemble Smoother with Multiple Data Assimilation)

对 ESMDA 算法主要分为 3 个方面进行介绍, 首先介绍集合平滑器 (ES, Ensemble Smoother); 接下来在 ES 算法中引入膨胀系数构成 ESMDA 算法框架; 最后, 介绍 ESMDA 与 EnKF 算法的结合, 并给出伪代码的实现。

### 2.2.1 集合平滑器 (Ensemble Smoother, ES)

EnKF 属于顺序数据同化算法, 其根据  $t$  时刻状态变量值初始化模型, 预测  $t+1$  时刻模型的状态变量, 在  $t+1$  时刻利用观测资料对状态变量的预测值进行加权更新, 得到当前时刻状态变量的最优估计值, 多用于对状态变量进行实时标定, 以实现实时模型。然而, ES 多用于参数反演以及参数的不确定性分析中, 采用所有可用观测资料对参数进行更新。公式如下:

$$X^a = X^f + \frac{P^f H^T}{HP^f H^T + R_e} (Y_{\text{all}} - HX^f) \quad (10)$$

$$P^a = (I - KH) P^f \quad (11)$$

其中,

$$P^f = \frac{1}{N-1} \sum_{i=1}^N (X_i^f - \overline{X^f}) (X_i^f - \overline{X^f})^T \quad (12)$$

$$\overline{X^f} = \frac{1}{N} \sum_{i=1}^N X_i^f \quad (13)$$

并令,

$$K = \frac{P^f H^T}{HP^f H^T + R_e} \quad (14)$$

式中: 各个符号定义与 EnKF 中类似。

### 2.2.2 ESMDA

ES 的本质是将所有时刻的观测资料输入同化算法中进行一次参数更新<sup>[40]</sup>。然而, 正如算法名字所示, ESMDA 在 ES 的基础上设定一个算法执行次数, 进行多次数据同化, 在每个循环中, ESMDA 并不是简单重复了 ES 过程, 而是在每个循环中对观测误差添加了一个膨胀系数  $\alpha_i$ , 并令

$$\sum_{i=1}^{N_e} \frac{1}{\alpha_i} = 1 \quad (15)$$

式中:  $N_e$  为循环次数。显然地,  $\alpha_i$  有许多种取值方式, 然而文献<sup>[41]</sup>指出, 随着循环次数的增加而逐

渐减小的膨胀系数对于同化效果并无明显提升。因此, 在每次循环中, 采用相同的膨胀系数。

### 2.2.3 ESMDA-EnKF

ESMDA-EnKF 在每次循环中采用 EnKF 对参数进行更新, 其更新公式如下:

第  $i$  次循环:

$$X^a = X^f + K (Y_{\text{all}} - HX^f) \quad (16)$$

$$P^a = (I - KH) P^f \quad (17)$$

$$K = \frac{P^f H^T}{HP^f H^T + \alpha_i R_e} \quad (18)$$

其中,  $Y_{\text{all}}$  为经过  $\alpha_i R_e$  扰动的观测资料。

ESMDA-EnKF 的具体实现如下伪代码所示:

输入:  $N_a$ , 观测资料  $O_{\text{all}}$ ,  $X^f$ ,  $R_e$

输出:  $X^a$ ,  $K$ ,  $P^a$

read( $O_{\text{all}}$ )

for  $i = 1$  to  $N_a$

$X^{f'} = X^f - \text{mean}(X^f)$

$Y_{\text{all}} = O_{\text{all}} + \text{sqrt}(\alpha_i * R_e)$

$HX^{f'} = HX^f - \text{mean}(HX^f)$

$P^{f'} H^T = (X^{f'} * HX^{f'}) / (N - 1)$

$HP^{f'} H^T = (HX^{f'} * \text{transpose}(HX^{f'})) / (N - 1)$

$K = P^{f'} H^T / (HP^{f'} H^T + \alpha_i * R_e)$

$X^a = X^f + K * (Y_{\text{all}} - HX^f)$

$P^a = P^f - K * P^{f'} H^T$

update( $X^f$ ,  $P^f$ );

return( $K$ ,  $X^a$ ,  $P^a$ )

### 2.3 地下水模型

黑河流域中游地下水模型基于 MODFLOW (MODular three-dimension finite-difference groundwater FLOW model) 构建, MODFLOW 采用有限差分法将时间与空间离散化以解决地下水在三维空间中的流动问题。研究区平面面积约为 9 016 km<sup>2</sup>, 采用 1 km×1 km 的正方形网格将研究区含水层系统在水平方向上进行空间离散化, 离散化之后, 研究区在平面上剖分成为 132 行×165 列网格 (如图 3)。研究区内定义为活动单元, 研究区外定义为非活动单元。时间上, 以每月为一个应力期, 每天为一个时间步。研究区边界条件参考文献<sup>[42]</sup>以及自然边界确定 (图 3)。由于地下水分水岭的存在, A~E 设置为无水流边界; E 处为正义峡水文站, 黑河从此处流出研究区; 据调查资料分析, 北山地区地下水含水介质与南部祁连山相似, 但由于降水稀少, 无常年地表径流, 地下水含量无法与祁连山区相比较<sup>[27]</sup>, 因此, 将 D~E 边界假设为无水流边界; 由

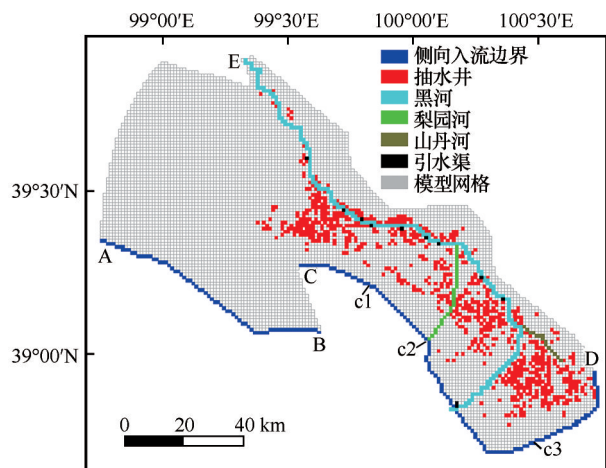


图3 黑河中游模型概化及边界设置

Fig. 3 The conceptualization and boundary conditions of the groundwater model

于多种侧向流从南部祁连山流入研究区,导致A~D边界最为复杂,因此,将此边界分为3段进行定义:如图3所示,A~B与C~D定义为定流量边界条件(第二类边界条件或Neumam条件),B~C之间定义为无流量边界。垂直方向上,模型顶部边界为空气—土壤接触面边界,模型底板定义为不透水边界。黑河干支流采用MODFLOW内的河流(STR, SStreamflow-Routing)程序包<sup>[43]</sup>进行模拟。农业灌溉通过采用地下水补给程序包(RCH, ReCHarge)<sup>[44]</sup>在网格上添加补给率实现。蒸散发采用了MODFLOW内置的蒸散发程序包(EVT, EVapo-Transpiration)进行模拟。机井采用机井程序包(Well)进行模拟。

## 2.4 模型评价

本文采用均方根误差(RMSE, Root Mean Square Error)对模拟结果进行量化评价,方程如下:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (M_i - O_i)^2} \quad (19)$$

式中: $N$ 为观测值总数(观测井个数×应力期个数); $M_i$ 和 $O_i$ 分别是模型模拟值和观测值。

## 3 结果与讨论

### 3.1 算法参数分析

在使用ESMDA-EnKF算法对地下水模型进行不确定性分析的过程中,算法存在一些超参数直接影响到同化算法对模型参数的更新效果(例如:算法执行次数(即膨胀系数) $N_e$ 、观测资料数 $n$ 、参数集合大小 $N$ 等),因此,本文首先为ESMDA-EnKF确定最优超参数,之后采用算法分析模型参数的不确定

性以及模型参数对观测数据的敏感性。

#### 3.1.1 ESMDA执行次数对不确定性分析的影响

ESMDA采用EnKF算法对地下水模型参数进行更新,因此,EnKF算法的执行次数将直接影响模型参数的更新效果。然而,目前尚没有关于执行次数的理论研究,而在文献[41]给出的例子中,分别执行了2次、4次算法以对比同化效果。考虑到计算消耗问题,本节评价了执行1次、2次、3次、4次EnKF算法后的同化效果(图4)。

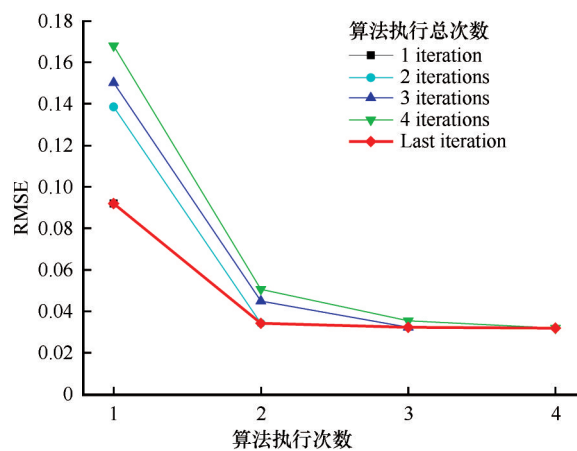


图4 不同执行次数对ESMDA-EnKF算法效果的影响

Fig. 4 Results of ESMDA-EnKF for different iterations

图4中展示了EnKF算法执行不同次数后的同化效果,Last iteration代表采用不同的执行次数时,最后一次执行结束之后的同化效果。由图中可以看出随着EnKF算法对模型参数的更新,地下水位的RMSE值逐渐降低,表明地下水位模拟值逐渐接近地下水位的观测值。经过随机采样之后的第一次、第二次参数更新对RMSE影响较大;第三次及第四次参数更新之后,与第二次参数更新之后的RMSE值相比,变化不大。这表明,使用观测数据对地下水模型参数进行的前2次更新最为有效,可能是因为:(1)地下水模型的参数采用随机初始化,因此,ESMDA在随机模型参数的基础上进行的第一次更新力度相对较大;(2)ESMDA采用了所有时刻的观测数据进行更新,最大程度上利用了观测数据中的信息,因此,能够更加有效且大幅度的更新参数。

#### 3.1.2 观测资料数对不确定性分析的影响

数据同化的目的即是最大限度的融合不同来源、不同时间、空间分辨率的直接、间接观测数据以用于提高模型的估计精度,以获得更加准确的状态

变量的时空分布<sup>[45-46]</sup>。然而由于成本原因,对系统内相关因素进行无限制观测是不现实的。因此,研究观测资料对同化效果的影响是必要的。选择42个水位观测井处的观测水位以及高崖和正义峡水文观测站处的黑河流量观测数据,分别研究观测数为7、14、21、28、35、42以及44时对同化效果的影响。图5显示了经过2次参数更新过程之后,不同观测资料数的RMSE值。由于涉及到不同类型的观测值、输出值(地下水位、流量),所以,图中对地下水位、流量的观测值及输出值进行了归一化处理,将观测值、输出值限定于[0, 1]之间。由图5的总体趋势上可以看出,随着观测资料的增加,参数逐渐接近真实值,地下水位的RMSE值逐渐降低。然而,观测数从7个增加至14个对同化效果影响不明显;观测井数增加至21个之后,同化效果基本保持不变;增加流量观测(42至44)对同化效果影响较为显著。由此可见,一种类型的观测数据确实会提升同化效果,但是当一种类型的观测数据增加到一定程度时(本实验中为21个),观测数据对同化效果的影响有限;然而,继续增加不同类型的观测数据(河流流量观测数据)将进一步提升同化效果。

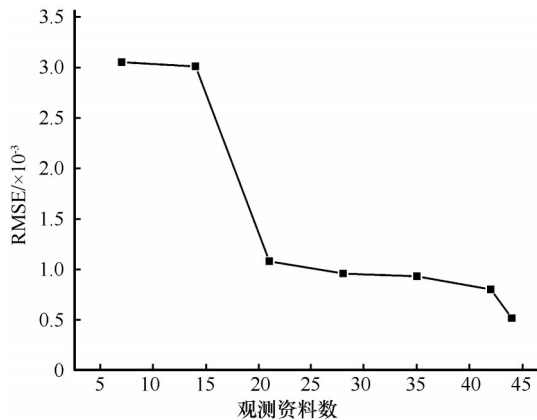


图5 不同的观测数对ESMDA-EnKF算法效果的影响  
Fig. 5 Results of ESMDA-EnKF for different number of observations

### 3.1.3 集合大小对不确定性分析的影响

EnKF采用样本集合的方式表示模型状态变量的先验概率分布,估计误差协方差<sup>[47]</sup>。Hamill等<sup>[48]</sup>研究发现,EnKF中的背景误差协方差估计和滤波函数的最优相关尺度等均与样本集合大小相关;当样本集合比较小时,协方差的估计噪音较大、最优相关尺度较小,出现滤波发散现象,且集合大小直

接影响算法运行时间。因此,采用不同的集合大小运行算法,以探索集合大小对算法同化效果的影响。图6显示了模型参数集合大小分别为30、60、90、120、150情况下,经过两次参数更新之后的RMSE值。由图中可以看出,当参数集合大小为30时,并没有出现滤波发散现象,且增加采样对同化效果的影响并不明显。因此,本文将集合大小设置为30以平衡算法效果与计算消耗。

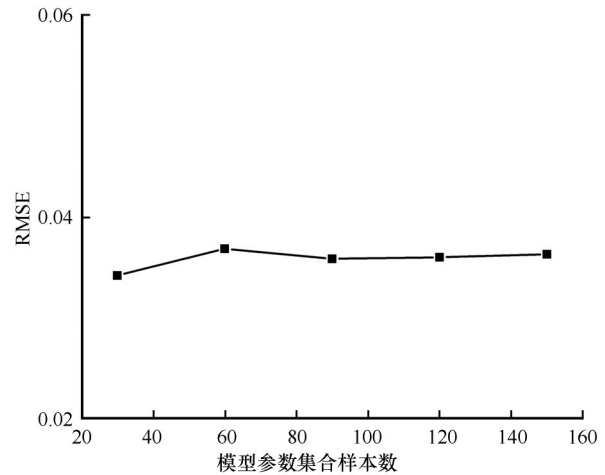


图6 不同集合大小对ESMDA算法效果的影响  
Fig. 6 Results of ESMDA for different ensemble size

### 3.2 地下水模型参数不确定性分析

基于以上分析确定了算法参数之后,本文利用ESMDA-EnKF对地下水模型参数进行了不确定性分析,模型参数分布见图7,相关设置见表1。从模型参数先验概率分布中进行30次随机采样,使用42个水位观测井、2个径流观测站处的观测数据对模型参数进行2次( $N_a=2$ )更新。模型参数随ESMDA-EnKF运行所得的不确定性如图8(注:由于参数与观测点数较多,图8中只显示了部分参数与观测点处的更新情况)。由图8可以看出,无论参数的初始分布如何,每次执行ESMDA-EnKF算法都会对模型参数进行更新,使参数值更接近其标定值。第一次执行算法对模型参数的更新效果显著,大多模型参数在进行第一次更新之后,参数值已经较为接近参数标定值,不确定性明显减小;在进行第二次更新之后,模型参数均分布于标定值附近,模型参数的不确定性进一步减小。图8第三列显示了各个参数在ESMDA-EnKF算法执行完成之后的分布情况,由此分布情况可以看出,各个分区的渗透系数(P1~P8)的不确定性显著减小,基本收敛于最优值附近。对比各个分区的渗透系数可以看



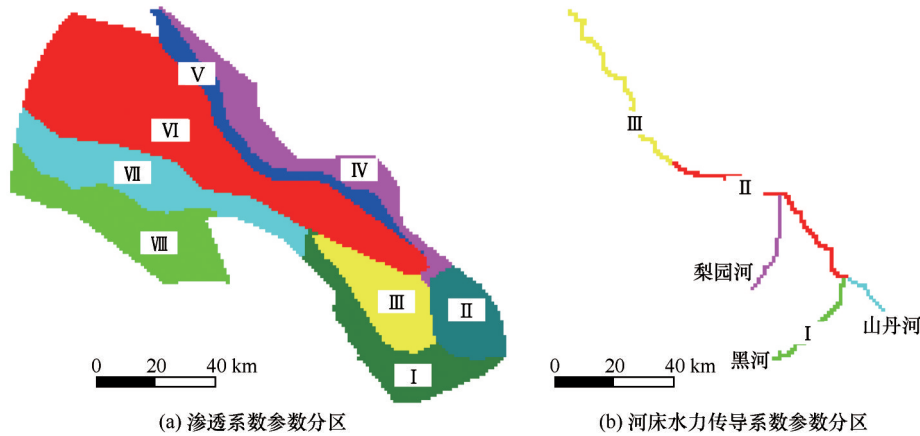


图 7 地下水模型参数分区  
Fig. 7 Zonation of the parameters: hydraulic conductivities of the aquifer (a); hydraulic conductance of the streambed (b)

表 1 进行不确定性分析的参数及其设置  
Table 1 Parameters involved in uncertainty analysis

参数名称	描述	初始分布	均值	方差
P1	含水层子分区 I 内的渗透系数	对数正态分布	23	0.30
P2	含水层子分区 II 内的渗透系数	对数正态分布	10	0.30
P3	含水层子分区 III 内的渗透系数	对数正态分布	90	0.30
P4	含水层子分区 IV 内的渗透系数	对数正态分布	3	0.30
P5	含水层子分区 V 内的渗透系数	对数正态分布	20	0.30
P6	含水层子分区 VI 内的渗透系数	对数正态分布	20	0.30
P7	含水层子分区 VII 内的渗透系数	对数正态分布	50	0.30
P8	含水层子分区 VIII 内的渗透系数	对数正态分布	50	0.30
P9	黑河子分区 I 河床水力传导系数	对数正态分布	48 447	1 500.00
P10	黑河子分区 II 河床水力传导系数	对数正态分布	22 832	1 500.00
P11	黑河子分区 III 河床水力传导系数	对数正态分布	10 000	1 500.00
P12	含水层给水度	均匀分布	0.1	0.10
P13	灌溉回流系数	均匀分布	0.3	0.30

出,经过 ESMDA-EnKF 算法更新之后,第Ⅷ个分区的渗透系数(P8)的分布最分散,不确定性最大,这可能是由于在此分区内观测数据较少,导致算法无法对参数进行有效更新。在算法执行完成之后,黑河子分区 I 与子分区 II 的河床水力传导系数(P9 与 P10)虽然也收敛到了最优值附近,但是其分布比黑河子分区 III 的河床水力传导系数(P11)分散(即其标准差比较大)。这可能是由于在黑河子分区 I 与子分区 II 河段,黑河与地下水相互作用频繁(子分区 I 河段河流补给地下水,子分区 II 河段地下水出流转化为河流),导致河床水力传导系数不确定性较大。由图 8 可见,ESMDA-EnKF 算法能够对模型参数进行有效更新,从而减小模型参数的不确定性。

地下水位随 ESMDA-EnKF 运行所得的不确定

性如图 9 所示。由图 9 第一列可以看出,受模型参数不确定性的影响,研究区内地下水位[图 9(a)、(b)]以及河流径流量[图 9(c)、(d)]均存在较大的不确定性,且随着时间变化,不确定性区间有变大的趋势,这可能是受预报前期模型运行结果不确定性的累积影响,模型运行后期地下水位与河流径流量运行结果产生了更大的不确定性。图 9 第二、三列分别为 ESMDA-EnKF 算法运行 1、2 次之后的模型输出值。可以看出,ESMDA-EnKF 算法运行一次之后,模型输出的地下水位与河流径流量的不确定性范围明显缩小,无论是数据起伏较小的地下水位还是数据起伏较大的河流径流量,都收敛到了最优值附近。第二次执行 ESMDA-EnKF 算法则基本消除了模型输出结果的不确定性(对应于图 9 的第三列)。

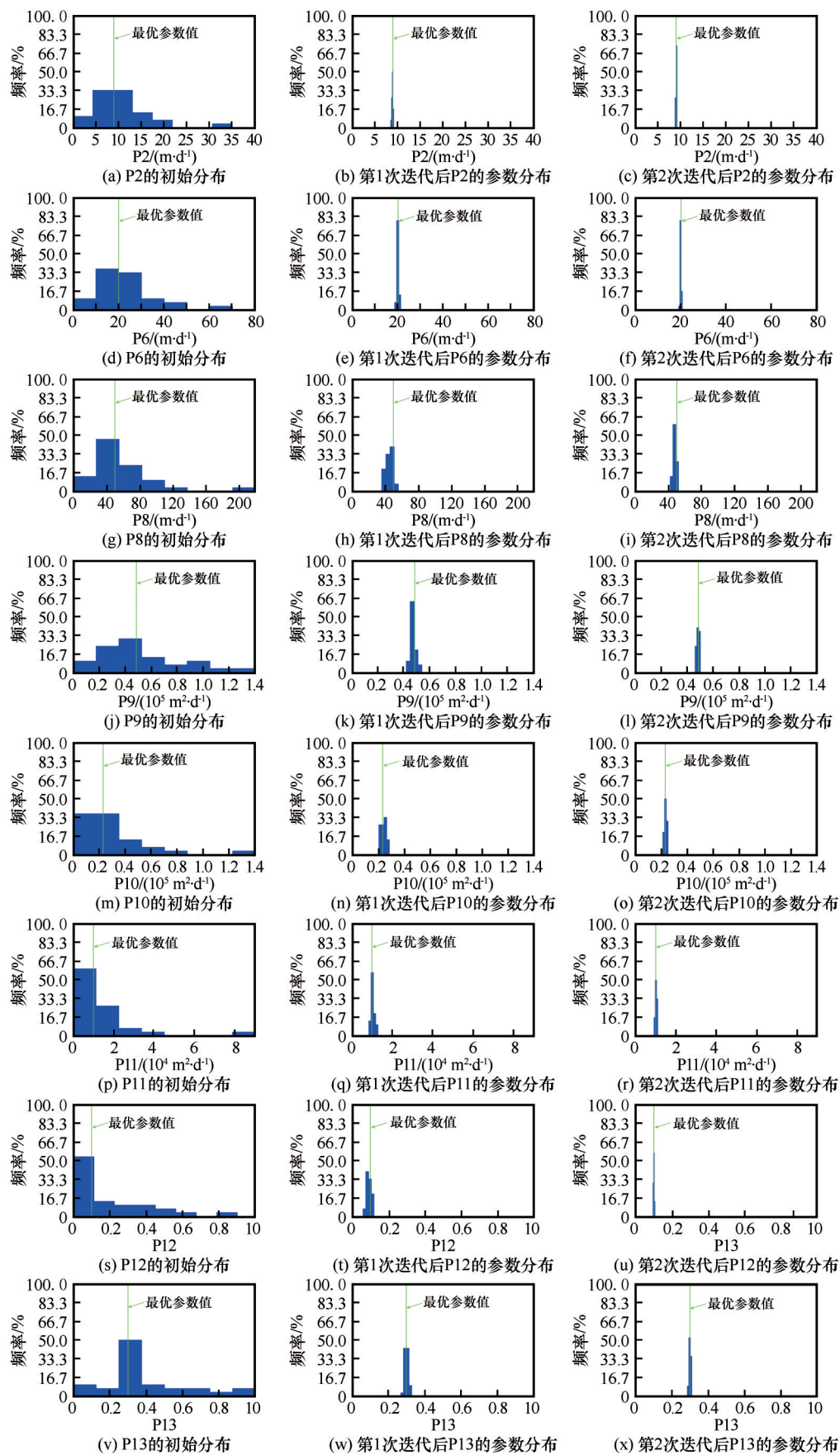


图8 经过ESMDA-EnKF更新之后的模型参数分布(参数编号解释见表1)

Fig. 8 Parameter distributions after the updating by ESMDA-EnKF (The numbers of parameters are shown in Table 1)



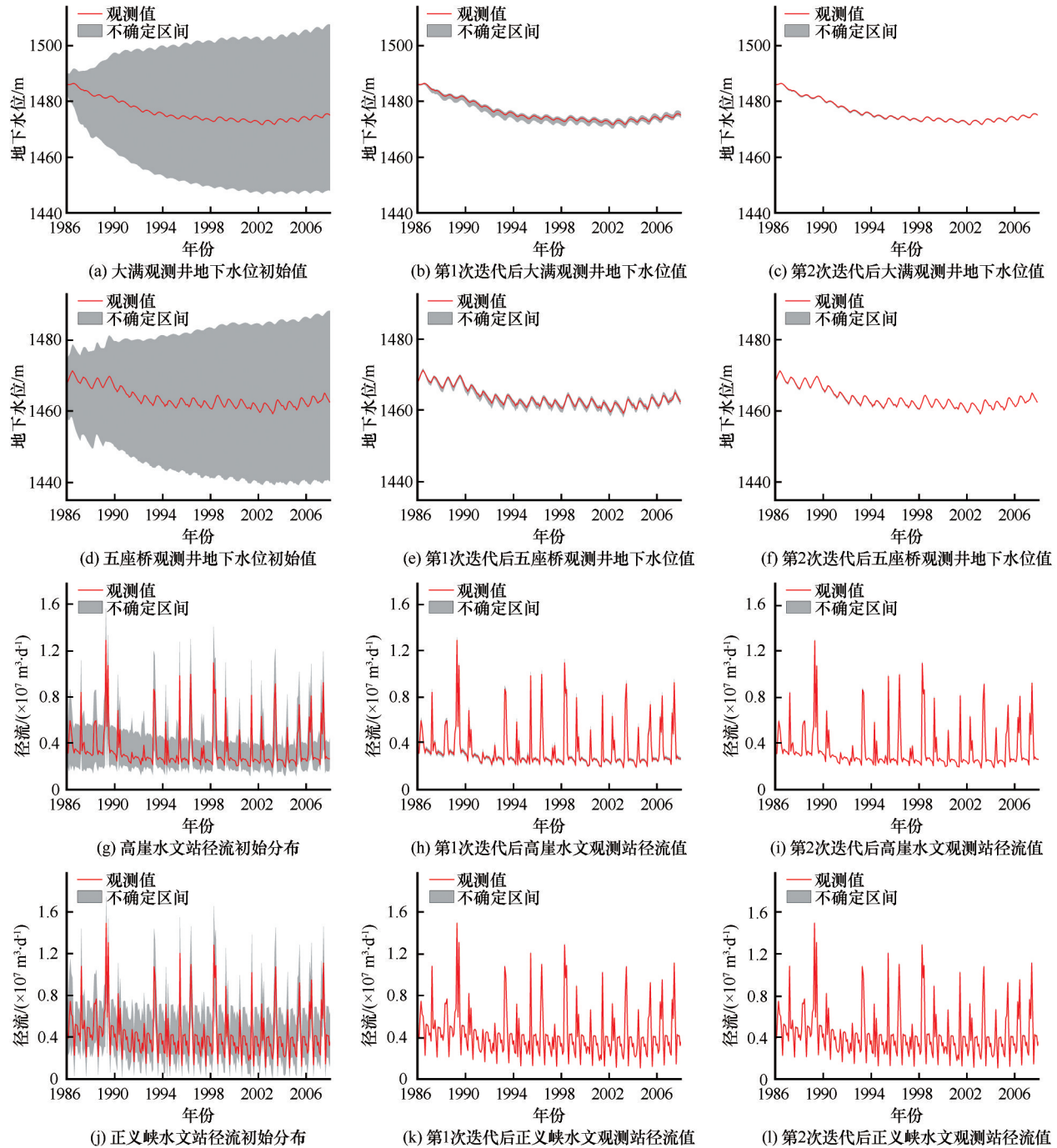


图9 使用ESMDA-EnKF对模型参数进行更新之后的模型输出值

Fig. 9 Model outputs after the update of model parameters by ESMDA-EnKF

对比图8~9可以看出,随着ESMDA-EnKF算法的执行,图8中的模型参数逐渐收敛于参数最优值,相应地,通过模型参数计算得到的模型输出也逐渐接近观测值(图9)。由算法对参数的更新可以看出,对观测数据最为敏感的模型参数为第二个参数分区以及第六个参数分区的渗透系数P2和P6。最不敏感的参数为第八个参数分区的渗透系数,这可能是因为这个区域紧邻不透水边界,且区域内没有河

流,其与其他区域的水力联系较弱,而在此分区内缺乏有效的观测数据,无法对模型参数进行有效更新。图9显示了在不同算法运行阶段的模型输出值变化,由灰色逐渐加深为黑色,分别为不同迭代次数的模型输出值,由图中可以看出含水层渗透系数(P1~P8)、含水层给水度(P12)以及灌溉回流系数(P13)对模型地下水位影响显著。由图8~9中河流流量与河床水力传导系数的同化效果可以看出,水

力传导系数对河流流量观测较为敏感。河床传导系数会影响河流与地下水之间的水量交换,然而,河流流量还受模型模拟区域内整体地下水水位以及上游来水量的影响,所以,仅使用高崖和正义峡水文站处流量观测值对河床水力传导系数进行更新,无法完全消除其不确定性。

为了定量的分析地下水模型参数的不确定性,同时进一步说明 ESMDA-EnKF 算法对地下水模型参数及输出不确定性的更新效果,本文运用 Origin 2018 软件对 ESMDA-EnKF 第 2 次迭代后的地下水模型参数以及第 2 次迭代后地下水模型最后时刻(2008 年 12 月)的输出结果(地下水位、河流流量)进行了统计分析。地下水模型参数的统计指标见表 2,地下水模型输出结果的统计指标见表 3。由表 2 中可以看出,经过 ESMDA-EnKF 算法更新后,地下水模型参数的标准差比更新之前明显变小,且参数均值与表 1 中参数均值基本相同。虽然更新之后的参数均值与参数的标定值相似,但是仍有细微差别,这可能是由于算法中仅按照相应的分布随机采样了 30 个样本,样本数较小导致抽样误差(sampling error)较大,但是根据中心极限定理(central limit theorem)可知,采样之后的均值服从正态分布,

表 2 ESMDA-EnKF 算法第 2 次迭代后地下水模型参数统计值

Table 2 Statistical values of groundwater model parameters after two iterations of ESMDA-EnKF

参数名称	最大值	最小值	均值	标准差
P2	9.06	8.84	8.99	0.06
P6	20.31	19.48	19.89	0.17
P8	51.62	42.98	47.40	0.23
P9	49 823.06	46 089.18	48 288.49	989.06
P10	24 565.18	21 157.51	23 086.16	900.69
P11	10 921.72	9 390.32	10 238.85	379.16
P12	0.10	0.09	0.10	0.0023
P13	0.30	0.28	0.29	0.0047

表 3 ESMDA-EnKF 算法第 2 次迭代后地下水模型最后时刻的输出结果

Table 3 Statistical values of groundwater model outputs in the last time step after two iterations of ESMDA-EnKF

观测井/水文站	最大值	最小值	均值	标准差
大满	1 476.45	1 476.24	1 476.25	0.10
五座桥	1 465.36	1 465.05	1 465.22	0.07
高崖水文站	2 950 870	2 848 690	2 908 010	25 756
正义峡水文站	4 606 250	4 157 380	4 311 790	92 865

因此,均值误差在可接受的范围之内。

4 结论

本文提出了基于数据同化的不确定性分析方法,通过包含观测资料信息减小模型参数的不确定性。基于已构建的黑河流域中游地下水模型,利用提出的算法对模型参数进行不确定性分析,探索了不同超参数对算法效果的影响,分析了含水层渗透系数、河床水力传导系数、含水层给水度、灌溉回流系数的不确定性,评估了不同系数对观测信息的敏感程度,主要结论如下:

(1)基于贝叶斯理论的数据同化方法是不确定性分析的有效手段,ESMDA-EnKF 算法通过包含观测资料更新参数,能够有效的减小模型参数的不确定性;对不确定性分析结果的统计分析结果表明,经过 ESMDA-EnKF 算法更新后的参数均值收敛到最优均值附近,不确定性范围明显减小。

(2)不同超参数对算法效果影响不一,观测资料的增加将提升算法对模型参数的更新程度,减小参数的不确定性,且加入新类型的观测资料会进一步减小参数的不确定性。

(3)不同模型参数的不确定性不同,地表水与地下水相互作用频繁的区域参数不确定性较大;含水层渗透系数、含水层给水度以及灌溉回流系数对模型输出的地下水位影响显著,水力传导系数对模型输出的河流流量影响较大。

参考文献(References):

[1] Schellnhuber H J. ‘Earth system’ analysis and the second Copernican revolution[J]. Nature, 1999, 402: C19.

[2] Prickett T A. Ground-water computer models—state of the Arta [J]. Groundwater, 1979, 17(2): 167-173.

[3] Bair E S. Applied groundwater modeling: simulation of flow and advective transport [J]. Groundwater, 2016, 54 (6) : 756-757.

[4] Chen Chong, He Wei, Zhou Han, et al. A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin, northwestern China[J]. Scientific Reports, 2020, 10: 3904.

[5] Wang Xusheng, Zhou Jian. Advances in numerical modeling of groundwater flow in Heihe River basin[J]. Geotechnical Investigation & Surveying, 2009, 37(9) : 35-38. [王旭升, 周剑. 黑河流域地下水流数值模拟的研究进展[J]. 工程勘察, 2009, 37(9): 35-38.]

[6] Cheng Guodong, Xiao Honglang, Fu Bojie, et al. Advances in synthetic research on the eco-hydrological process of the Heihe River basin[J]. Advances in Earth Science, 2014, 29(4): 431-437. [程国栋, 肖洪浪, 傅伯杰, 等. 黑河流域生态-水文过程集成研究进展[J]. 地球科学进展, 2014, 29 (4) : 431-437.]

- [7] Shoesmith E, Box G E P, Draper N R. Empirical model-building and response surfaces [J]. The Statistician, 1988, 37 (1): 82.
- [8] Oberkampf W L, Diegert K V, Alvin K F, et al. Variability, uncertainty and error in computational simulation [C]//ASME Proceedings of the 7th AIAA/ASME Joint Thermophysics and Heat Transfer Conference. 1998, 357-2: 259-272.
- [9] Beven K. Changing ideas in hydrology: the case of physically-based models [J]. Journal of Hydrology, 1989, 105 (1/2): 157-172.
- [10] Wu Jichun, Lu Le. Uncertainty analysis for groundwater modeling [J]. Journal of Nanjing University (Natural Sciences), 2011, 47(3): 227-234. [吴吉春, 陆乐. 地下水模拟不确定性分析[J]. 南京大学学报(自然科学版), 2011, 47(3): 227-234.]
- [11] Vrugt J A, Braak C J F, Gupta H V, et al. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? [J]. Stochastic Environmental Research and Risk Assessment, 2009, 23(7): 1011-1026.
- [12] Jin Xiaoli, Xu Chongyu, Zhang Qi, et al. Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model [J]. Journal of Hydrology, 2010, 383(3/4): 147-155.
- [13] Helton J C, Davis F J. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems [J]. Reliability Engineering & System Safety, 2003, 81(1): 23-69.
- [14] Williams S K, Acker T, Goldberg M, et al. Estimating the economic benefits of wind energy projects using Monte Carlo simulation with economic input/output analysis [J]. Wind Energy, 2008, 11(4): 397-414.
- [15] Shu Longcang, Lu Chengpeng, Li Wei. Calculation method of the exchange volume between surface water and groundwater based on uncertainty of parameters [J]. Hydrogeology & Engineering Geology, 2008, 35(5): 68-71. [束龙仓, 鲁程鹏, 李伟. 考虑参数不确定性的地表水与地下水交换量的计算方法[J]. 水文地质工程地质, 2008, 35(5): 68-71.]
- [16] Li Sen, Chen Jiajun, Ye Huihai, et al. Analysis on sensitivity of stochastic factors in numerical simulation of groundwater flow [J]. Journal of Hydraulic Engineering, 2006, 37(8): 977-984. [李森, 陈家军, 叶慧海, 等. 地下水流数值模拟中随机因素的灵敏度分析[J]. 水利学报, 2006, 37(8): 977-984.]
- [17] Shi Liangsheng, Yang Jinzhong, Li Shaolong, et al. Stochastic analysis of subsurface flow based on KL-Galerkin method [J]. Journal of Sichuan University (Engineering Science Edition), 2005, 37(5): 31-35. [史良胜, 杨金忠, 李少龙, 等. 基于KL-Galerkin解法的地下水流动随机分析[J]. 四川大学学报(工程科学版), 2005, 37(5): 31-35.]
- [18] Yao Leihua. Perturbation-coefficient awaiting determination stochastic finite element method for groundwater flow models [J]. Journal of Hydraulic Engineering, 1999, 30(7): 60-65. [姚磊华. 地下水水流模型的摄动待定系数随机有限元法[J]. 水利学报, 1999, 30(7): 60-65.]
- [19] Lu Le, Wu Jichun, Chen Jingya. Identification of hydrogeological parameters based on the Bayesian method [J]. Hydrogeology & Engineering Geology, 2008, 35(5): 58-63. [陆乐, 吴吉春, 陈景雅. 基于贝叶斯方法的水文地质参数识别[J]. 水文地质工程地质, 2008, 35(5): 58-63.]
- [20] Tang Tian, Wu Jichun, Yang Yun. Analysis on influence of aquifer heterogeneous simplification to groundwater numerical modeling [J]. Geotechnical Investigation & Surveying, 2011, 39(4): 34-42. [唐甜, 吴吉春, 杨运. 含水介质非均质概化对地下水数值模拟的影响分析[J]. 工程勘察, 2011, 39(4): 34-42.]
- [21] Mustafa S M T, Nossent J, Ghysels G, et al. Estimation and impact assessment of input and parameter uncertainty in predicting groundwater flow with a fully distributed model [J]. Water Resources Research, 2018, 54(9): 6585-6608.
- [22] Kalman R E. A new approach to linear filtering and prediction problems [J]. Journal of Basic Engineering, 1960, 82 (1): 35-45.
- [23] Huang Chunlin, Li Xin. Sensitivity analysis on land data assimilation scheme of soil moisture [J]. Advances in Water Science, 2006, 17(4): 457-465. [黄春林, 李新. 土壤水分同化系统的敏感性试验研究[J]. 水科学进展, 2006, 17(4): 457-465.]
- [24] Chu Nan, Huang Chunlin, Du Peijun. State and parameter estimation in soil moisture data assimilation based on dual ensemble Kalman filter [J]. Remote Sensing Technology and Application, 2016, 31(2): 214-220. [褚楠, 黄春林, 杜培军. 基于双EnKF的土壤水分与土壤属性参数同时估计[J]. 遥感技术与应用, 2016, 31(2): 214-220.]
- [25] Li Ning, McLaughlin D, Kinzelbach W, et al. Using an ensemble smoother to evaluate parameter uncertainty of an integrated hydrological model of Yanqi Basin [J]. Journal of Hydrology, 2015, 529: 146-158.
- [26] Wongchuig-Correa S, de Paiva R C D, Biancamaria S, et al. Assimilation of future SWOT-based river elevations, surface extent observations and discharge estimations into uncertain global hydrological models [J]. Journal of Hydrology, 2020, 590: 125473.
- [27] Li Wenpeng, Kang Weidong, Hao Aibing, et al. Regulation and optimal utilization mode of water resources in typical inland basins in Northwest China: a case study of the Heihe River basin [M]. Beijing: Geology Press, 2010. [李文鹏, 康卫东, 郝爱兵, 等. 西北典型内流盆地水资源调控与优化利用模式——以黑河流域为例[M]. 北京: 地质出版社, 2010.]
- [28] Mi Lina, Xiao Honglang, Zhu Wenjing, et al. Dynamic variation of the groundwater level in the middle reaches of the Heihe River during 1985—2013 [J]. Journal of Glaciology and Geocryology, 2015, 37(2): 461-469. [米丽娜, 肖洪浪, 朱文婧, 等. 1985—2013年黑河中游流域地下水位动态变化特征[J]. 冰川冻土, 2015, 37(2): 461-469.]
- [29] Yao Yingying, Zheng Chunmiao, Tian Yong, et al. Numerical modeling of regional groundwater flow in the Heihe River Basin, China: Advances and new insights [J]. Science China Earth Sciences, 2015, 58(1): 3-15.
- [30] Wang Penglong, Song Xiaoyu, Xu Bingxin, et al. Evaluation and advancement of water resources carrying capacity of Zhangye Prefecture in Heihe River basin [J]. Journal of Glaciology and Geocryology, 2020, 42(3): 1057-1066. [王鹏龙, 宋晓谕, 徐冰鑫, 等. 黑河流域张掖段水资源承载力评价及提升对策研究[J]. 冰川冻土, 2020, 42(3): 1057-1066.]
- [31] Ding Hongwei, Zhang Ju, Lü Zhi, et al. Characteristics and cycle conversion of water resources in the Hexi Corridor [J]. Arid Zone Research, 2006, 23(2): 241-248. [丁宏伟, 张举, 吕智, 等. 河西走廊水资源特征及其循环转化规律[J]. 干旱区研究, 2006, 23(2): 241-248.]
- [32] Jarvis A, Reuter H I, Nelson A, et al. Hole-filled seamless SRTM data V4 [DS]. International Centre for Tropical Agriculture (CIAT), 2008.
- [33] Wang Jianhua, Liu Jiyuan. Landuse/landcover data of the Hei-



- he river basin (2000)[DS]. National Tibetan Plateau Data Center, 2013. DOI: 10.3972/heihe.020.2013.db. CSTR: 18406.11. heihe.020.2013.db. [王建华, 刘纪远. 黑河流域土地利用/土地覆盖数据集(2000)[DS]. 国家青藏高原科学数据中心, 2013. DOI: 10.3972/heihe.020.2013.db. CSTR: 18406.11. heihe.020.2013.db.]
- [34] Liu Jiyan, Wang Jianhua. Landuse/landcover dataset of the Heihe river basin (1980s)[DS]. National Tibetan Plateau Data Center, 2014. DOI: 10.3972/heihe.021.2013.db. CSTR: 18406.11. heihe.021.2013.db. [刘纪远, 王建华. 黑河流域土地利用/土地覆盖数据集(80年代末)[DS]. 国家青藏高原科学数据中心, 2014. DOI: 10.3972/heihe.021.2013.db. CSTR: 18406.11. heihe.021.2013.db.]
- [35] Hu Xiaoli, Wang Jianhua, Li Xin, Landuse/landcover data of Zhangye City (2007)[DS]. National Tibetan Plateau Data Center, 2015. DOI: 10.3972/heihe.018.2013.db. CSTR: 18406.11. heihe.018.2013.db. [胡晓利, 王建华, 李新. 张掖市土地利用/土地覆盖数据集(2007)[DS]. 国家青藏高原科学数据中心, 2015. DOI: 10.3972/heihe.018.2013.db. CSTR: 18406.11. heihe.018.2013.db.]
- [36] Editor board of Series of books of First National Census for Water, Series of Books of First National Census for Water: Census for Groundwater pumping wells[M]. Beijing: China Water & Power Press, 2017. [《第一次全国水利普查成果丛书》编委会. 第一次全国水利普查成果丛书: 地下水取水井基本情况普查报告[M]. 北京: 中国水利水电出版社, 2017.]
- [37] Hu Litang. The monitoring data of the groundwater level in the middle of Heihe River basin (2005—2007)[DS]. National Tibetan Plateau Data Center, 2016. DOI: 10.3972/heihe.120.2014.db. CSTR: 18406.11. heihe.120.2014.db. [胡立堂. 黑河干流中游地区地下水位监测数据(2005—2007)[DS]. 国家青藏高原科学数据中心, 2016. DOI: 10.3972/heihe.120.2014.db. CSTR: 18406.11. heihe.120.2014.db.]
- [38] Evensen G. The ensemble Kalman Filter: theoretical formulation and practical implementation[J]. Ocean Dynamics, 2003, 53(4): 343-367.
- [39] Evensen G. Data assimilation: the ensemble Kalman filter[M]. Berlin: Springer, 2009.
- [40] Li Gaoming, Reynolds A C. Iterative ensemble Kalman filters for data assimilation[J]. SPE Journal, 2009, 14(3): 496-505.
- [41] Emerick A A, Reynolds A C. Ensemble smoother with multiple data assimilation[J]. Computers & Geosciences, 2013, 55: 3-15.
- [42] Hu L T, Chen C X, Jiao J J, et al. Simulated groundwater interaction with rivers and springs in the Heihe river basin[J]. Hydrological Processes, 2007, 21(20): 2794-2806.
- [43] Prudic D E. Documentation of a computer program to simulate stream-aquifer relations using a modular, finite-difference, ground-water flow model[R]. Carson City, Nevada: US Geological Survey Open-File Report 88-729, 1989.
- [44] Harbaugh A W. MODFLOW-2005, the US Geological Survey modular ground-water model: the ground-water flow process[M]. Reston, VA, USA: US Department of the Interior, US Geological Survey, 2005.
- [45] Zehe E, Becker R, Bárdossy A, et al. Uncertainty of simulated catchment runoff response in the presence of threshold processes: role of initial soil moisture and precipitation[J]. Journal of Hydrology, 2005, 315(1/2/3/4): 183-202.
- [46] Li Xin, Huang Chunlin, Che Tao, et al. Progress and prospect of research on land surface data assimilation system in China[J]. Progress in Natural Science, 2007, 17(2): 163-173. [李新, 黄春林, 车涛, 等. 中国陆面数据同化系统研究的进展与前瞻[J]. 自然科学进展, 2007, 17(2): 163-173.]
- [47] Evensen G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics[J]. Journal of Geophysical Research: Oceans 1994, 99(C5): 10143-10162.
- [48] Hamill T M, Whitaker J S, Snyder C. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter[J]. Monthly Weather Review, 2001, 129(11): 2776-2790.

## Uncertainty analysis of groundwater model based on data assimilation

CHEN Chong<sup>1</sup>, ZHANG Wei<sup>2</sup>, XING Qinghui<sup>1</sup>, DOU Yixuan<sup>1</sup>

(1. College of Information Science and Engineering, China University of Petroleum-Beijing, Beijing 102249, China;

2. Kокtokay Snow Station, State Key Laboratory of Cryospheric Science, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China)

**Abstract:** The groundwater system in the middle and lower reaches of Heihe River basin is replenished by the upper cryosphere meltwater and precipitation. The stability of the groundwater system in the middle and lower reaches of Heihe River basin faces more risks due to the cryosphere shrinkage caused by climate warming. Groundwater models are efficient tools for researchers in accessing the stability of groundwater system. However, the uncertainty of groundwater model parameters is an important issue. In this paper, an uncertainty analysis method based on data assimilation algorithm which reduces parameters uncertainty by including observations is proposed. The proposed method is then used to analyze the uncertainty of 13 parameters in a groundwater model of the middle reaches of Heihe River Basin (which is constructed based on MODFLOW). The influence and optimal values of hyper-parameters are discussed. The uncertainty of groundwater model parameters is analyzed using the proposed algorithm. The results show that the data assimilation algorithm can effectively reduce the uncertainty of groundwater model parameters. The diversity and quantity of observations play an important role in reducing the uncertainties. The parameters in subzones with frequent interaction between surface water and groundwater are more uncertain. The hydraulic conductivity, specific yield and irrigation backflow coefficient have significant influence on the groundwater level, while streambed hydraulic conductivity shows great influence on the streamflow. This study will provide a more reliable modelling method for groundwater research and an important support for researching the stability of groundwater system in the northwestern area.

**Key words:** groundwater model; uncertainty analysis; data assimilation; MODFLOW

(责任编辑: 丁永建)