

DOI: 10.11766/trxb202203040090

满俊, 张江江, 郑强, 尧一骏, 曾令藻. 基于数据同化的土壤水力参数反演方法: 研究进展与展望[J]. 土壤学报, 2023, 60(6): 1543–1554.
MAN Jun, ZHANG Jiangjiang, ZHENG Qiang, YAO Yijun, ZENG Lingzao. Data Assimilation for Soil Hydraulic Parameter Estimation: Progress and Perspectives[J]. Acta Pedologica Sinica, 2023, 60 (6): 1543–1554.

基于数据同化的土壤水力参数反演方法: 研究进展与展望*

满俊^{1†}, 张江江², 郑强³, 尧一骏¹, 曾令藻⁴

(1. 中国科学院土壤环境与污染修复重点实验室(南京土壤研究所), 南京 210008; 2. 长江保护与绿色发展研究院(河海大学), 南京 210024;
3. 数学与理论部(鹏城实验室), 深圳 518055; 4. 浙江省农业资源与环境重点实验室(浙江大学), 杭州 310058)

摘要: 土壤水力参数及其非均质性刻画关乎到诸多土壤与地下水等领域的量化模拟研究问题。受限于时间和采样成本, 传统的直接测定方法并不能很好地解决这个问题。随着物联网技术的发展, 与土壤水运动有关的一些状态表征量(如含水量和水头)已经能够通过传感器实时获得。如何充分融合这些观测数据信息, 反演出土壤水力参数是当前的一个研究热点。数据同化方法能够通过融合观测数据与模型预测值信息, 实现对模型参数的反演估计。本文系统分析了土壤水力参数不确定性的来源及测定方法, 阐述了常用数据同化方法的基础理论及其在土壤水力参数反演方面的应用, 并从计算效率和反演精度两方面着重论述了数据同化方法的最新前沿进展, 最后探讨了数据同化方法未来的发展方向。研究表明: 数据同化方法能够突破传统测定方法的限制, 用于土壤水力参数及其非均质性刻画。尽管如此, 由于土壤非饱和流模型的强非线性以及原位观测数据的相对稀缺性等问题的存在, 当前数据同化方法的计算效率和反演精度还有待进一步提升。未来可从发展监督式降维方法、多源多尺度数据融合以及耦合物理机制的机器学习等方面深化土壤水力参数反演方法研究, 这有利于农业土水管理、污染防治和修复等工作的合理开展。

关键词: 数值模拟; 土壤水力参数; 参数反演; 数据同化; 优化试验设计; 机器学习

中图分类号: S152 文献标志码: A

Data Assimilation for Soil Hydraulic Parameter Estimation: Progress and Perspectives

MAN Jun^{1†}, ZHANG Jiangjiang², ZHENG Qiang³, YAO Yijun¹, ZENG Lingzao⁴

(1. Key Laboratory of Soil Environment and Pollution Remediation, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China; 2. Yangtze Institute for Conservation and Development, Hohai University, Nanjing 210024, China; 3. Department of Mathematics and Theory, Peng Cheng Laboratory, Shenzhen 518055, China; 4. Zhejiang Provincial Key Laboratory of Agricultural Resources and Environment, Zhejiang University, Hangzhou 310058, China)

* 国家重点研发计划项目(2021YFC1808904)、国家自然科学基金项目(42107066)和江苏省自然科学基金项目(BK20201105)资助
Supported by the National Key Research and Development Program of China (No. 2021YFC1808904), the National Natural Science Foundation of China (No. 42107066), and the Natural Science Foundation of Jiangsu Province (No. BK20201105)

† 通讯作者 Corresponding author, E-mail: manjun91@issas.ac.cn

作者简介: 满俊(1991—), 男, 湖北鄂州人, 博士, 副研究员, 主要从事土壤污染空间识别、归趋模拟与暴露分析方面的研究。

收稿日期: 2022-03-04; 收到修改稿日期: 2022-05-30; 网络首发日期(www.cnki.net): 2022-07-10

Abstract: The characterization of soil hydraulic parameters and their heterogeneity is related to many scientific problems in soil and groundwater fields. Due to the limitation of time and sampling cost, the traditional experimental approaches cannot address this issue adequately. With the development of Internet of Things technology, the state variables related to soil water movement (such as water content and pressure head) can be acquired in real time through sensors. This has sparked some debates about how to estimate the soil hydraulic parameters using these measurements. Data assimilation methods can estimate the soil hydraulic parameters by integrating the measurements into numerical models. This paper systematically analyzes the uncertainty sources and measurement approaches of soil hydraulic parameters, expounds on the basic principles of several common data assimilation methods and their applications in soil hydraulic parameter inversion, and discusses the latest advances in data assimilation methods from aspects of computational efficiency and accuracy. Finally, the development direction of data assimilation methods is provided. The results show that the data assimilation methods can break through the limitation of the traditional experimental approach, and thus are suitable for the characterization of soil hydraulic parameters and their heterogeneity. However, limitations such as the strong nonlinearity of the unsaturated flow model, spatial heterogeneity of soil and sparsity of in-situ measurements do exist. It is, therefore, essential for us to unfold in-depth research on soil hydraulic parameter inversion from the aspects of supervised dimension reduction method, multi-source and multi-scale data fusion, and coupling of machine learning with physical mechanisms, thereby assisting agricultural soil and water management as well as the prevention, control, and remediation of pollution in agroecosystems.

Key words: Numerical modeling; Soil hydraulic parameter; Parameter inversion; Data assimilation; Optimal experimental design; Machine learning

土壤水力参数（如土壤水分特征曲线和土壤导水率^[1-2]）的测定对于土壤和地下水等研究领域均具有重要意义。比如，土壤-植物-大气连续体水热动态模拟^[3-4]、土壤水盐优化调控^[5-6]、场地土壤和地下水风险评估^[7-8]等。获取土壤水力参数的传统方法是通过直接测定。该方法成本高、耗时长，而且还会破坏土壤的天然结构。另一方面，由于土壤在空间上具有非均质性，少量监测点上获得的结果也不一定能代表整个区域土壤的性质^[9]。随着物联网技术的发展，通过信息传感和实时监控手段，能够实现对土壤数据（如含水量和水头）的自动采集。如何充分融合这些观测数据信息，反演土壤水力参数是当前的一个研究热点。

数据同化方法通过将观测数据信息与基于物理机制的模型预测结果相融合，从而实现对模型参数的反演估计。近些年来，利用数据同化方法反演土壤水力参数的研究在国内外有了一定的发展。本文旨在总结和分析这些研究进展，为土壤水力参数的获取提供参考，这对于农业水土资源管理以及污染防控具有重要的理论指导意义和实际价值。

1 土壤水力参数的不确定性

土壤非均质性是指在成土过程中受沉积环境、

成岩作用和地质构造等因素的影响，在空间分布及内部属性上均存在不均匀的变化，是造成土壤水力参数不确定性的主要原因^[10]。一般而言，自然土壤特性在水平和垂直方向上均存在非均质性，按照变异系数（Coefficient of Variation, CV）大小可将土壤性质分为强变异（ $CV > 100\%$ ）、中等变异（ $10\% \leq CV \leq 100\%$ ）和弱变异（ $CV < 10\%$ ）三类^[11]。例如，饱和导水率、土壤水扩散率等属于强变异参数，土壤颗粒（包括砂粒、粉粒和黏粒）含量属于中等变异，而土壤容重和饱和含水量则属于弱变异参数^[12]。基于大量土壤样品测定结果，Carsel 和 Parrish^[13]计算出了 12 种典型质地土壤水力参数的统计性质。雷志栋等^[14]研究发现，不同质地土壤的水力参数有很大差异，试验地块测得的土壤饱和导水率，波动幅度可达 3 个数量级；即使是相同质地的土壤，不同空间位置上的土壤水力参数也可能存在较大差异。此外，参数的变异强度还会随着研究尺度的变化而发生改变^[15-16]。

土壤非均质性刻画一直以来均是研究的热点问题^[17-18]。由于测量技术与成本的限制，不可能对空间各点上的土壤水力参数均进行测定。对于大尺度区域而言，通常所能获得的观测数据十分有限，并且在空间上呈现出不连续的特点。在未布设监测点

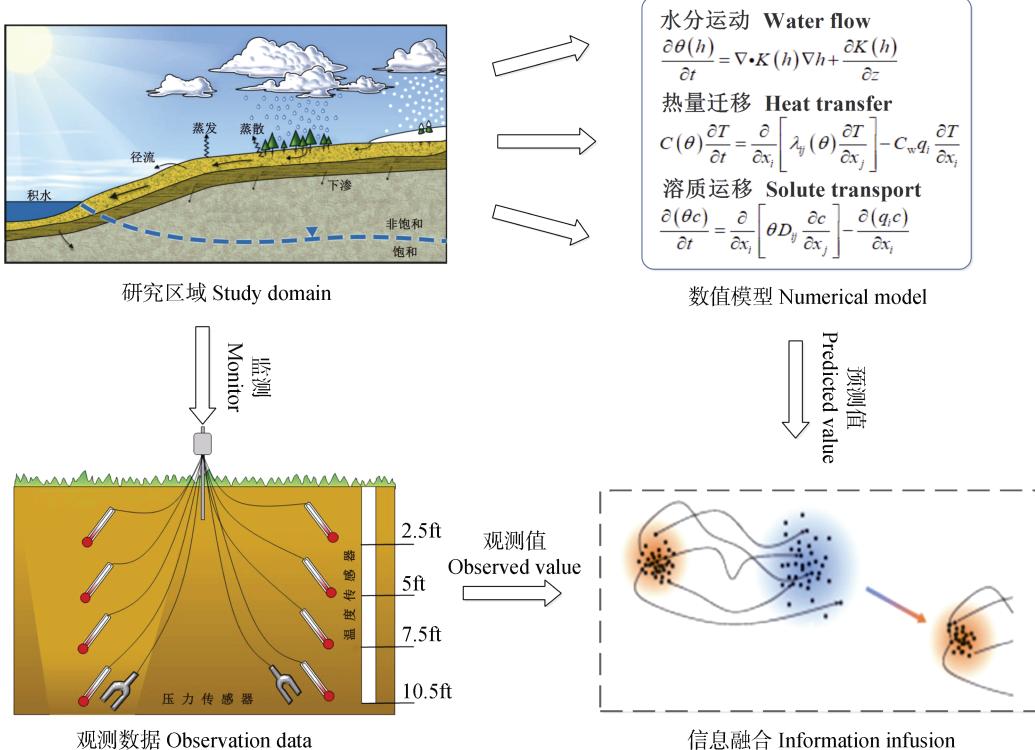
的区域, 一般以已有的观测数据为基础, 通过空间插值获得土壤水力参数, 但该方法会对土壤非均质性的刻画带来不确定性。如何准确、全面地获取空间非均质的土壤水力参数是当前面临的一个挑战。

2 土壤水力参数的测定及反演

当前, 直接测定土壤水力参数的方法主要分为实验室方法和田间方法^[19], 大多需要施加严格的初始和边界条件, 以便能用解析或半解析方法来确定。由于需要满足模型参数显式计算的条件, 直接测定法的实验分析过程往往较为耗时。举例而言, 在某些问题中, 对于不同的边界条件, 需要反复实现稳态或水力平衡状态。并且, 可能还需要对控制方程进行线性化或近似处理, 以便进行(半)解析求解。另外一种限制是要求施加相对简单的初始和边界条

件, 这对于需要精确控制边界条件的田间实验而言尤为困难。此外, 由于尺度效应的存在, 直接测定法在处理土壤非均质性方面也受到很大限制^[20]。

随着物联网技术和数值计算方法的发展, 对原位获取的观测数据进行同化, 进而识别土壤水力参数的研究得以广泛开展。为准确估计土壤水力参数, 可以对与土壤水分运动有关的状态量(如含水量和水头)进行监测, 然后将获得的观测数据信息融入土壤非饱和流模型中, 通过不断地更新参数和状态使得模型预测值与观测数据相匹配, 这一过程称为数据同化(Data assimilation)^[21-22]。如图1所示, 数据同化过程主要由数值模型、观测数据和同化方法构成。其中, 同化方法起着连接观测数据与模型预测值的关键作用, 是数据同化的核心。接下来, 本文将着重介绍近些年来有关数据同化方法方面取得的研究进展。



然而, 土壤和地下水领域的大多数参数反演问题均是不稳定的, 存在解的非唯一性。这表明, 模型参数的不确定性不能被彻底消除, 而通常只能通过融入观测数据信息来降低。在这种情况下, 优化算法提供的单一解无法准确量化参数内在的不确定性, 并且给出的结果也可能是局部而非全局最优^[25]。相较而言, 随机方法倾向于将所有可能的解均保留下来, 利用概率分布评估参数的不确定性。

近些年来, 利用数据同化方法反演土壤水力参数(如土壤水分特征曲线和土壤导水率)的研究逐渐引起重视, 其原理是通过将含水量、水头等数据信息融入系统模型以反推土壤水力参数。数值模型是模拟土壤中水、热和溶质运移的有力工具, 在水土资源管理、污染风险评估等方面具有广泛的应用。经过几十年的发展, 目前已经建立了多种土壤非饱和流数值模型, 代表性的有 HYDRUS^[26]、FEMFLOW^[27]、FEHM^[28]等, 能够用于求解控制非饱和流动过程的偏微分方程。为便于说明, 本文将土壤非饱和流模型简化成以下形式:

$$\mathbf{d} = G(\mathbf{m}) + \boldsymbol{\varepsilon} \quad (1)$$

式中, $G(\cdot)$ 表示系统模型; $\boldsymbol{\varepsilon}$ 为系统误差; \mathbf{m} 为待定的模型参数(如土壤水分特征曲线和土壤导水率); \mathbf{d} 为模型预测值(如土壤含水量、水头等)。土壤水分特征曲线定义了水头与土壤含水量之间的函数关系, 它和土壤导水率共同决定了水分和溶质的迁移速度, 其可靠性直接影响着模型的预测结果。由于模型结构和参数等存在着不确定性, 使得模型预测值也带有不确定性, 并且与观测数据之间存在着差异。模型结构的不确定性主要是由于边界条件和源汇项的可变性, 以及对初始和边界条件等缺乏了解造成的, 而模型参数的不确定性主要源于土壤非均质性。数据同化方法利用观测数据 \mathbf{d}^* 对模型预测值进行不断校准, 估计未知模型参数 \mathbf{m} , 从而降低其不确定性。目前应用较为广泛的数据同化方法包括马尔可夫链蒙特卡罗(Markov Chain Monte Carlo, MCMC)、集合卡尔曼滤波(Ensemble Kalman Filter, EnKF)及其变体^[29-31]。以下, 对这两类常用数据同化方法的基础理论及其在土壤水力参数反演方面的应用作一个简单的介绍。

3.1 马尔可夫链蒙特卡罗

MCMC是一种基于贝叶斯原理的随机方法, 通

过生成满足模型参数后验概率分布的样本表征其不确定性。根据贝叶斯原理, 后验分布可以由预先指定的先验分布和似然函数进行计算:

$$P(\mathbf{m}|\mathbf{d}^*) \propto P(\mathbf{m})P(\mathbf{d}^*|\mathbf{m}) \quad (2)$$

式中, $P(\mathbf{m})$ 为参数的先验分布, 代表获得观测数据之前对参数的初始认识, 可通过查阅文献资料和地质调查等途径获得。Freeze^[32]通过统计分析发现, 土壤饱和导水率可以用对数正态分布来描述, 而孔隙度参数则服从于正态分布。对于统计信息不明的参数, 可假定其服从于均匀分布; $P(\mathbf{d}^*|\mathbf{m})$ 称为似然函数, 表示参数取 \mathbf{m} 时观测数据 \mathbf{d}^* 的分布函数, 用于度量模型输出 $G(\mathbf{m})$ 与观测数据 \mathbf{d}^* 之间的匹配程度; $P(\mathbf{m}|\mathbf{d}^*)$ 为参数的后验分布, 反映了融合观测数据信息后对参数产生新的认识, 是贝叶斯反演的核心目标。MCMC 通过构造合适的马尔科夫链进行参数抽样, 在链条运行足够长的时间后收敛到平稳分布, 即为后验分布。有关参数后验的统计性质, 可以通过分析这些样本得到。MCMC 也可以采用多条链并行, 以提高探索参数后验区间的能力, 代表性的算法有 DREAM^[33]。

在土壤参数反演方面, 已有不少关于 MCMC 的相关研究^[34-37]。例如: 柯凤乔等^[38]基于 MCMC 和一维分层土柱实验反演土壤水力参数; 石晓蕾等^[39]应用 MCMC 推求土壤水力参数, 并量化其不确定性; Steenpass 等^[40]利用 MCMC 融合土壤表面温度和含水量数据信息, 实现了对土壤水力参数的反演估计; Younes 等^[41]研究了不同类型观测数据和溶质注入时间对土壤水力参数反演的影响。MCMC 通常需要大量的模型调用才能达到收敛, 而且所需模型调用次数随着参数个数的增加而显著上升, 对于高维问题这将造成巨大的计算量。

3.2 集合卡尔曼滤波

在 EnKF 中, 模型参数和状态的不确定性是通过集合样本来表示和传递的, 最早由 Evensen^[42-43]提出。在这里, 本文将模型参数 \mathbf{m} (即土壤水力参数)和状态 \mathbf{s} (即土壤含水量或水头)组合成一个联合状态向量 $\mathbf{x}=[\mathbf{m} \, \mathbf{s}]^T$, 包含 N 个样本的集合可以表示成 $\mathbf{X}=[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ 。EnKF 包含预测和更新两个步骤, 通常以递归的方式对观测数据进行实时同化。在预测步骤, t 时刻的状态可以通过上一时刻的状态进行预测,

$$\mathbf{x}_{t,i}^f = G(\mathbf{x}_{t-1,i}^a), \quad i=1, 2, \dots, N \quad (3)$$

式中, f 和 a 分别表示预测和更新。在获得当前时刻观测数据 \mathbf{d}_t^* (如土壤含水量或水头), 可以对状态进行如下更新,

$$\mathbf{x}_{t,i}^a = \mathbf{x}_{t,i}^f + \mathbf{K} (\mathbf{d}_t^* - \mathbf{H} \mathbf{x}_{t,i}^f) \quad (4)$$

式中, \mathbf{H} 为观测算子, 将状态向量 $\mathbf{x}_{t,i}^f$ 转化为与观测数据相对应的模型预测值; \mathbf{K} 为卡尔曼增益, 可以根据模型参数与输出之间的协方差以及模型输出的自协方差计算得到。随着更新的持续进行, 状态向量 $\mathbf{x}_{t,i}^a$ 中参数的不确定性将有效降低, 通过统计分析这些样本就可以得到土壤水力参数的反演结果。

对于多孔介质中水分运动问题, 自 Geir 等^[44]的工作开始, 研究者采用 EnKF 同时估计水头和渗透系数。随后, Hendricks 和 Kinzelbach^[45]利用 EnKF 反演空间非均质的渗透系数, Li 和 Ren^[46]则通过同化水头数据估计出 12 种不同质地土壤的水力参数。需要注意的是, 当参数分布非高斯或系统模型呈现

强非线性时, EnKF 的反演效果有时欠佳^[47]。作为 EnKF 的一种变体, 集合平滑器(Ensemble Smoother, ES) 一次性融合所有观测数据信息, 可采用迭代的方式进行更新^[48]。Li 等^[49]比较了 EnKF 和迭代 ES 方法在土壤水力参数反演方面的效果, 认为迭代 ES 效果要优于 EnKF。

4 计算效率和反演精度提升

虽然 MCMC、EnKF 及其变体在土壤水力参数反演方面得到了较为广泛的应用, 但各自也存在着不足, 即: MCMC 不适用于高维问题, EnKF 及其变体受到高斯假设的约束。由于土壤具有非均质性, 导致未知参数过多, 并且土壤非饱和流模型呈现较强的非线性, 而原位观测数据又相对稀缺, 需要发展更加高效的数据同化方法。其中, 4.1 节从不确定量化角度提升计算效率, 4.2 节和 4.3 节分别从考虑模型结构误差和优化试验设计方面提升反演精度, 4.4 节则是将近年迅猛发展的机器学习方法融入数据同化的框架, 部分前沿进展总结如表 1 所示。

表 1 数据同化方法部分前沿进展

Table 1 Advances in data assimilation methods

参考文献	算法	算法改进	应用区域	应用模型	数据类型	性能小结
Reference	Algorithm	Algorithm improvement	Application field	Application model	Data type	Performance summary
[53]	EnKF	利用多项式混沌展开 构造替代模型	数值实验	HYDRUS2D	水头	计算效率提升 4 倍以上, 但对于非均质性较强的 土壤提升有限
[59]	ES	执行多保真度 高斯过程模拟	箱体实验	OpenGeoSys	渗滤液流量	计算效率提升 20 倍以上
[62]	EnKF	考虑模型结构误差	土柱实验和 田间实验	非饱和流模型	土壤含水量	显著提高均匀流模型的 预测能力
[63]	MCMC	考虑模型结构误差	区域尺度	MODFLOW	水头	提供更准确的模型预测及 置信区间
[70]	EnKF	执行多目标优化	室内示踪实验	ParFlow 和 SLIM-FAST	浓度	节省 37.7% 的采样成本, 降低预测偏差
[71]	MCMC	执行贝叶斯试验设计	数值实验	MODFLOW 和 MT3DMS	浓度	相较于随机设计, 参数预测的均方根 误差降低 20%~50%

续表

参考文献	算法	算法改进	应用区域	应用模型	数据类型	性能小结
Reference	Algorithm	Algorithm improvement	Application field	Application model	Data type	Performance summary
[72]	MCMC	执行全局优化设计, 找到最优监测方案	数值实验	HYDRUS2D	土壤温度和 含水量	预测精度提升 50%; 土壤温度较含水量具有 更高的信息价值
[75]	EnKF	执行序贯优化设计, 找到最优监测方案	数值实验	HYDRUS2D	水头	相较于 EnKF, 饱和导水率 场预测精度提升 20%
[76]	EnKF	构造替代模型, 执行序贯优化设计	数值实验	HYDRUS2D	浓度	相较于 EnKF, 饱和导水率 场预测精度提升 25%, 并且优化设计得到的监测 方案具有唯一性
[86]	ES	利用深度神经网络构造 替代模型	数值实验	MODFLOW 和 MT3DMS	水头和浓度	计算效率提升约 10 倍
[87]	ES	利用深度神经网络构造 非线性更新公式	数值实验	MODFLOW	水头	相较于 ES, 饱和导水率场 预测精度提升 10%

4.1 不确定性量化

为了提高计算效率,可以在同化过程中使用低保真度模型。与系统模型相比,低保真度模型的模拟精度较低,但所需计算时间相对较短。一般而言,低保真度模型包括:(1)具有较低求解精度的数值模型;(2)系统模型在其低维子空间上的投影;(3)基于插值或拟合得到的替代模型等^[50-51]。Zeng 等^[52]利用多项式混沌展开构造替代模型,联合 EnKF 提出一种高效的数据同化方法,并将之用于多相流的反演模拟。Man 等^[53]进一步改进该方法,并将其应用于土壤非饱和水流的数据同化问题。多保真度模拟在近年来逐渐引起重视,它的核心思想是使用一种或多种低保真度模型来近似系统模型^[54-55]。由于多保真度模拟可以兼顾效率和精度,常被用于量化模拟过程的不确定性^[56-57],以及提高数据同化的计算效率^[58-59]。此外,降维方法(如 Karhunen-Loève 展开^[60]和分段逆回归法^[61])也可以通过减少未知参数个数,提升反演效率。

4.2 模型结构误差

在同化过程中,除了需要量化参数不确定性外,有时还需要考虑模型结构的不确定性。Erdal 等^[62]研究表明,当土壤非饱和流模型存在较大结构误差时,传统的 EnKF 无法得到合理的预测结果,故而提出同时更新模型参数、状态和结构误差。此外,

Xu 等^[63]采用高斯过程对模型结构误差进行拟合,并用 MCMC 来同时反演未知模型参数和超参数(即与模型结构误差相关的参数),以获得更加准确的预测结果。当存在多个模型时,研究者还可以采用贝叶斯模型平均量化模型结构的不确定性^[64],即随着同化过程的进行对不同模型的权重分别进行更新。Wöhling 等^[65]研究表明,在模型不确定性分析中采用贝叶斯模型平均较采用单一模型往往更加有效。

4.3 优化试验设计

一般而言,不同空间位置和时间上所采集的数据在参数反演中的效用(即信息含量)是不等的^[66]。在一定预算条件下,通常只能获得有限的观测数据。理论上,通过试验设计找到最优的监测方案,从而获得信息含量最高的观测数据,有助于提高参数反演的精度^[67]。在贝叶斯框架下,Lindley^[68]为优化试验设计提供了统一的框架,即定义一个反映试验设计目标的效用函数,然后通过最大化效用函数的期望值确定最优设计方案。

优化试验设计方法大致分为两类:全局设计(批量设计)和序贯设计。全局设计,是在所有空间位置或时间上进行设计,只设计一次。在全局设计方面,Herrera 和 Pinder^[69]利用卡尔曼滤波对地下水监测网进行时空上的优化。Kollat 等^[70]使用 EnKF 和多目标优化算法提高地下水的长期监测。Zhang 等^[71]

提出了一种高效的贝叶斯试验设计方法, 并应用于地下水污染源识别。Man 等^[72]则通过分析土壤温度和含水量的数据价值找到最优监测方案, 从而提高土壤水力参数的反演精度。当观测数据只能实时获取时, 全局设计便不容易实施。不同于全局设计, 序贯设计是在上一次反馈结果的基础上进行的, 涉及多次设计。Prakash 和 Datta^[73]在监测网中采用模拟退火优化算法进行序贯优化设计, 实现对污染源的精确描述。为提高污染物的预测精度, Gharamti 等^[74]提出了一种贪婪试验设计方法来优化选择监测井, 该方法是通过在给定时间内最大化预期信息增益实现的。Man 等^[75-76]则通过开发基于集合样本和概率配点的序贯优化设计方法, 进一步提高土壤水力参数反演的精度。Geiges 等^[77]比较了非线性参数估计问题的序贯设计和全局设计, 研究结果表明序贯优化的效果要优于全局优化。后续研究者可以根据数据的可获取性以及获取方式, 结合研究需求自主选择优化设计方法。

4.4 机器学习应用

机器学习是一种数据驱动型的建模方法, 能够通过分析数据发掘其隐含规律, 并据此对未知状态进行预测。在过去的十年中, 机器学习方法, 特别是深度学习, 已经被广泛地用于解决土壤与地下水等领域的难题^[78-79]。比如, Araya 和 Ghezzehei^[80]利用机器学习预测土壤饱和导水率及其对土壤结构扰动的敏感性; Man 等^[81]应用深度学习解析污染场地重金属的空间分布特征; Padmanabha 和 Zabarasi^[82]构建了一种条件可逆神经网络, 利用少量观测数据反演对数渗透系数场。除此之外, 机器学习还被用于校正物理模型的结构误差^[83-84]。机器学习的成功在于它能够自动提取复杂特征, 并有效地从训练数据中发掘隐含信息^[85]。

近年来, 很多学者倾向于将机器学习嵌入到数据同化的框架中。比如, Mo 等^[86]开发了一种基于深度学习的数据同化方法, 反演非高斯分布的渗透系数场; Zhang 等^[87]利用深度学习建立起了观测数据与模型参数之间的非线性映射关系, 用于刻画非高斯分布的地下介质特征。除了直接估计高维的非高斯参数场, 也有学者利用深度学习对其进行非线性降维, 然后使用数据同化方法估计降维后的特征参数。比如, Laloy 等^[88]利用生成对抗网络对非高斯场进行降维, 然后利用 MCMC 对降维后的随机空

间进行识别。作为一种数据驱动的方法, 机器学习模型的预测精度非常依赖于数据的数量和质量, 并且在训练数据未覆盖到的区域可能存在预测可靠性不足的问题。因此, 纯数据驱动的模型在实际应用当中, 很有可能得到没有物理意义的预测结果。为此, Raissi 等^[89]创造性地提出了内嵌物理机制的神经网络(Physics-Informed Neural Networks, PINNs), 该方法通过自动微分技术将物理方程以正则项的形式加入神经网络的损失函数中, 使得神经网络不仅能吸收来自数据的信息, 还能融合来自具有全局约束的方程的信息, 进而提升模型的预测鲁棒性。PINNs 的出现促进了科学机器学习领域的发展, 也催生了一些 PINNs 在土壤水文模型数据同化方面的应用研究, 比如: Tartakovsky 等^[90]将 PINNs 用于非均质渗透系数场的估计, 并能准确估计非饱和场景下的渗透系数与水头间的本构关系; Zheng 等^[91]受 PINNs 启发, 构建了内嵌物理机制的图像修复框架, 能基于少量点位观测同时估计高维参数和状态场; Wang 等^[92-93]指出工程控制条件、专家经验等亦可作为正则约束来提升模型的预测能力, 在 PINNs 的基础上进一步提出理论指导的神经网络, 将其应用于渗透系数场的高效和精准估计。虽然科学机器学习的发展方兴未艾, 但它在土壤水力参数反演中的应用还相对较少, 尤其是结合实际场地数据的研究更是稀少。

5 结论与展望

针对土壤的非均质性、土壤非饱和流模型的强非线性以及原位观测数据的相对稀缺性, 前人发展了不同的数据同化方法用于土壤水力参数反演。然而, 这些数据同化方法的应用大多停留在数值实验的层面, 在实际田间乃至区域尺度上的适用性还有待验证。考虑到实际应用中土壤的非均质特性更加复杂, 综合上述研究进展, 本文认为未来可以考虑从发展监督式降维方法、多源多尺度数据融合以及耦合物理机制的机器学习等方面进一步深化土壤水力参数反演方法研究:

(1) 发展监督式降维方法。传统的降维方法如主成分分析, 通常是基于预先设定的参数统计性质来实现的, 忽略了系统模型内在的动力学特征, 属于无监督的降维方法。监督式降维方法则通过考虑

各个参数对系统模型输出的敏感性，挖掘高维参数与模型输出之间的隐含关系，从而提高降维效率。活跃子空间^[94]是当前较为热门的监督式降维方法，利用模型输出关于各个参数的梯度信息进行降维。Yan 等^[95]通过将该方法与高斯过程回归相结合，自适应构造出地下水系统模型的替代模型。在土壤水力参数反演问题中，对参数进行监督式降维，或能用于解决实际应用中土壤空间非均质性带来的大量未知参数问题，从而有助于提高数据同化方法的计算效率。

(2) 多源多尺度观测数据信息的融合。当前大多数研究采用的是点位上的观测数据，精度较高但数量少。而区域尺度上的遥感数据虽然精度不高、测量深度也较浅，但数据量大且覆盖范围广。多尺度观测数据可能包含不同时空水平的地表-地下水文系统的有用信息，需要发展能够同化多尺度数据并获取其价值的方法。Montzka 等^[96]指出，多尺度数据同化可以通过两种方式进行：使用观测算子，或者在同化之前将观测数据重新调整到模型尺度。Zhu 等^[97]利用多尺度土壤含水量数据对分布式地下水水流模型进行同化和分析。尽管已经存在用于同化多尺度数据的方法，但在土壤水力参数反演方面的应用十分有限。一个重要的原因是观测尺度与模拟尺度不匹配，并且不同尺度土壤水分具有时空变异性。在多尺度数据同化问题中，应根据实际情况综合选择信息融合方法。另一方面，各种测量技术（如近地传感器和探地雷达等）的出现也为融合不同来源数据信息提供了可能性。例如，姚荣江等^[98]将土壤表观电导率数据融入水盐运移模型，用于提高土壤盐分的模拟精度。因此，发展有效的数据同化方法融合多源多尺度数据信息能够为土壤水力参数反演提供新思路。

(3) 机器学习与物理机制的融合。对于复杂的实际问题，现有物理模型也许不足以充分概化其过过程机制，导致基于物理模型的数据同化结果存在系统偏差。基于数据驱动的机器学习通过从数据中直接学习规律，有助于摆脱对物理模型的依赖。为了使机器学习能适应观测数据较为稀少的场景，同时增加模型预测的鲁棒性，需要将物理模型机制与机器学习方法进行有机整合，使得观测数据和物理机制信息均能得到充分利用。发展内嵌物理机制的机器学习模型用于数据同化，是突破传统数据同化方法局限性的一大机会，未来需要在复杂的实际问题中继续检验和完善这一方法框架。

参考文献 (References)

- [1] Liu J L, Xu S H, Liu H. A review of development in estimating soil water retention characteristics from soil data[J]. Journal of Hydraulic Engineering, 2004, 35 (2): 68—76. [刘建立, 徐绍辉, 刘慧. 估计土壤水分特征曲线的间接方法研究进展[J]. 水利学报, 2004, 35 (2): 68—76.]
- [2] Lv D Q, Shao M A. Models and determining methods of unsaturated soil hydraulic parameters[J]. Chinese Journal of Applied Ecology, 2004, 15 (1): 163—166. [吕殿青, 邵明安. 非饱和土壤水力参数的模型及确定方法[J]. 应用生态学报, 2004, 15 (1): 163—166.]
- [3] Kang S Z. Simulation on water and heat dynamics in soil-plant-atmosphere continuum[J]. Acta Ecologica Sinica, 1991, 11 (3): 256—261. [康绍忠. 土壤-植物-大气连续体水热动态模拟的研究[J]. 生态学报, 1991, 11 (3): 256—261.]
- [4] Li Y, Shao M A, Wang W Y, et al. Dynamic variations of soil water and heat under evaporation conditions varying with aperture ratio of plastic mulch[J]. Acta Pedologica Sinica, 2004, 41 (3): 387—393. [李毅, 邵明安, 王文焰, 等. 覆膜不同开孔程度蒸发条件下土壤水热变化动态研究[J]. 土壤学报, 2004, 41 (3): 387—393.]
- [5] Yang J S. Development and prospect of the research on salt-affected soils in China[J]. Acta Pedologica Sinica, 2008, 45 (5): 837—845. [杨劲松. 中国盐渍土研究的发展历程与展望[J]. 土壤学报, 2008, 45 (5): 837—845.]
- [6] Yang J S, Yao R J, Wang X P, et al. Research on salt-affected soils in China: History, status quo and prospect[J]. Acta Pedologica Sinica, 2022, 59 (1): 10—27. [杨劲松, 姚荣江, 王相平, 等. 中国盐渍土研究: 历程、现状与展望[J]. 土壤学报, 2022, 59 (1): 10—27.]
- [7] Li C P, Wu J, Luo F, et al. Risk assessment of soil and groundwater for an organic chemical contaminated site[J]. Soils, 2013, 45 (5): 933—939. [李春平, 吴骏, 罗飞, 等. 某有机化工污染场地土壤与地下水风险评估[J]. 土壤, 2013, 45 (5): 933—939.]
- [8] Teng Y, Luo Y M, Shen R F, et al. Research progress and perspective of the multi-medium interface process and regulation principle of pollutants in site soil-groundwater[J]. Acta Pedologica Sinica, 2020, 57 (6): 1333—1340. [滕应, 骆永明, 沈仁芳, 等. 场地土壤-地下水污染物多介质界面过程与调控研究进展与展望[J]. 土壤学报, 2020, 57 (6): 1333—1340.]
- [9] Lei Z D, Hu H P, Yang S X. A review of soil water research[J]. Advances in Water Science, 1999, 10 (3): 311—318. [雷志栋, 胡和平, 杨诗秀. 土壤水研究进展与评述[J]. 水科学进展, 1999, 10 (3): 311—318.]
- [10] Zhang Y G. Inverse method of heterogeneous hydraulic

- parameters for vadose zone flow[D]. Beijing: China University of Geosciences, 2014. [张永根. 非均质性包气带水力参数反演方法研究[D]. 北京: 中国地质大学(北京), 2014.]
- [11] Shao M A, Wang Q J, Huang M B. Soil physics[M]. Beijing: Higher Education Press, 2006. [邵明安, 王全九, 黄明斌. 土壤物理学[M]. 北京: 高等教育出版社, 2006.]
- [12] Jury W A, Horton R. Soil physics[M]. John Wiley & Sons, 2004.
- [13] Carsel R F, Parrish R S. Developing joint probability distributions of soil water retention characteristics[J]. Water Resources Research, 1988, 24 (5): 755—769.
- [14] Lei Z D, Yang S X, Xu Z R, et al. Preliminary investigation of the spatial variability of soil properties[J]. Journal of Hydraulic Engineering, 1985, 16 (9): 10—21. [雷志栋, 杨诗秀, 许志荣, 等. 土壤特性空间变异性初步研究[J]. 水利学报, 1985, 16 (9): 10—21.]
- [15] Hu W, Shao M A, Wang Q J. Scale-dependency of spatial variability of soil moisture on a degraded slope-land on the Loess Plateau[J]. Transactions of the Chinese Society of Agricultural Engineering, 2005, 21 (8): 11—16. [胡伟, 邵明安, 王全九. 黄土高原退耕坡地土壤水分空间变异的尺度性研究[J]. 农业工程学报, 2005, 21 (8): 11—16.]
- [16] Qu M M, Wang J W. Review of research on soil spatial variability[J]. Soils, 2003, 35 (1): 30—33. [区美美, 王建武. 土壤空间变异研究进展[J]. 土壤, 2003, 35 (1): 30—33.]
- [17] Peng X H, Wang Y Q, Jia X X, et al. Some key research fields of Chinese soil physics in the new era: Progresses and perspectives[J]. Acta Pedologica Sinica, 2020, 57 (5): 1071—1087. [彭新华, 王云强, 贾小旭, 等. 新时代中国土壤物理学主要领域进展与展望[J]. 土壤学报, 2020, 57 (5): 1071—1087.]
- [18] Zhang G L, Shi Z, Zhu A X, et al. Progress and perspective of studies on soils in space and time[J]. Acta Pedologica Sinica, 2020, 57 (5): 1060—1070. [张甘霖, 史舟, 朱阿兴, 等. 土壤时空变化研究的进展与未来[J]. 土壤学报, 2020, 57 (5): 1060—1070.]
- [19] Xu S H, Liu J L. Advances in approaches for determining unsaturated soil hydraulic properties[J]. Advances in Water Science, 2003, 14 (4): 394—401. [徐绍辉, 刘建立. 土壤水力性质确定方法研究进展[J]. 水科学进展, 2003, 14 (4): 394—401.]
- [20] Hopmans J. Review of inverse estimation of soil hydraulic properties. Proc. Int. Workshop , Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media, Riverside, USA: Univ. California, 1999: 643—658.
- [21] Ma J W, Qin S X. Recent advances and development of data assimilation algorithms[J]. Advances in Earth Science, 2012, 27 (7): 747—757. [马建文, 秦思娴. 数据同化算法研究现状综述[J]. 地球科学进展, 2012, 27 (7): 747—757.]
- [22] Wang Y S. Data assimilation its cause, its meaning and main procedures[J]. Marine Forecasts, 1999, 16 (1): 11—20. [王跃山. 数据同化——它的缘起、含义和主要方法[J]. 海洋预报, 1999, 16 (1): 11—20.]
- [23] Chong E, Zak S. An introduction to optimization[M]. John Wiley & Sons, 2004.
- [24] Yates S, Van Genuchten M, Warrick A, et al. Analysis of measured, predicted, and estimated hydraulic conductivity using the RETC computer program[J]. Soil Science Society of America Journal, 1992, 56 (2): 347—354.
- [25] Zhang T, Huang C L, Shen H F. Sensitivity and parameters optimization method of soil parameters to soil moisture in common land model[J]. Advances in Earth Science, 2012, 27 (6): 678—685. [张添, 黄春林, 沈焕锋. 土壤水分对土壤参数的敏感性及其参数优化方法研究[J]. 地球科学进展, 2012, 27 (6): 678—685.]
- [26] Šimůnek J, Van Genuchten M T, Šejna M. The HYDRUS software package for simulating two-and three-dimensional movement of water, heat, and multiple solutes in variably-saturated media[J]. Technical manual, version, 2006, 1: 241.
- [27] Trefry M G, Muffels C. FEFLOW: A finite-element ground water flow and transport modeling tool[J]. Groundwater, 2007, 45 (5): 525—528.
- [28] Zyvoloski G A, Robinson B A, Dash Z V, et al. Summary of the models and methods for the FEHM application—a finite-element heat-and mass-transfer code. Los Alamos National Lab. (LANL), Los Alamos, NM (United States), 1997.
- [29] Carrassini A, Bocquet M, Bertino L, et al. Data assimilation in the geosciences: An overview of methods, issues, and perspectives[J]. Wiley Interdisciplinary Reviews: Climate Change, 2018, 9 (5): e535.
- [30] Vrugt J, Stauffer P, Wöhling T, et al. Inverse modeling of subsurface flow and transport properties: A review with new developments[J]. Vadose Zone Journal, 2008, 7 (2): 843—864.
- [31] Zhou H, Gómez-Hernández J, Li L. Inverse methods in hydrogeology: Evolution and recent trends[J]. Advances in Water Resources, 2014, 63: 22—37.
- [32] Freeze A. A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media[J]. Water Resources Research, 1975, 11 (5): 725—741.
- [33] Laloy E, Vrugt J. High-dimensional posterior exploration of hydrologic models using multiple-tray DREAM (ZS) and high-performance computing[J]. Water Resources Research, 2012, 48 (1): W01526.
- [34] Scharnagl B, Vrugt J, Vereecken H, et al. Inverse modelling of in situ soil water dynamics: Investigating

- the effect of different prior distributions of the soil hydraulic parameters[J]. *Hydrology and Earth System Sciences*, 2011, 15 (10): 3043—3059.
- [35] Shi X, Ye M, Curtis G P, et al. Assessment of parametric uncertainty for groundwater reactive transport modeling[J]. *Water Resources Research*, 2014, 50 (5): 4416—4439.
- [36] Zha Y, Zhu P, Zhang Q, et al. Investigation of data assimilation methods for soil parameter estimation with different types of data[J]. *Vadose Zone Journal*, 2019, 18 (1): 190013.
- [37] Qin F L, Zhao Y C, Shi X Z, et al. Input parameters calibration and uncertainty estimation of the dn/dc model based on Bayesian inference[J]. *Acta Pedologica Sinica*, 2014, 51 (2): 247—254. [秦发倡, 赵永存, 史学正, 等. 基于贝叶斯推断的 DNDC 模型参数校正与不确定性评价研究[J]. 土壤学报, 2014, 51 (2): 247—254.]
- [38] Ke F Q, Man J, Zeng L Z, et al. Comparative study on inversion of the unsaturated hydraulic parameters using optimization and Bayesian estimation methods[J]. *Journal of Zhejiang University: Agriculture and Life Sciences*, 2016, 42 (5): 598—606. [柯凤乔, 满俊, 曾令藻, 等. 优化法与贝叶斯估计法在非饱和水力参数反演中的比较[J]. 浙江大学学报(农业与生命科学版), 2016, 42 (5): 598—606.]
- [39] Shi X L, Xu S H, Liao K H. AM-MCMC approach to estimate van genuchten model parameters[J]. *Soils*, 2012, 44 (2): 345—350. [石晓蕾, 徐绍辉, 廖凯华. 求 van Genuchten 模型参数的 AM-MCMC 方法[J]. 土壤, 2012, 44 (2): 345—350.]
- [40] Steenpass C, Vanderborght J, Herbst M, et al. Estimating soil hydraulic properties from infrared measurements of soil surface temperatures and TDR data[J]. *Vadose Zone Journal*, 2010, 9 (4): 910—924.
- [41] Younes A, Mara T, Fahs M, et al. Hydraulic and transport parameter assessment using column infiltration experiments[J]. *Hydrology and Earth System Sciences*, 2017, 21 (5): 2263—2275.
- [42] Evensen G. Data assimilation: the ensemble Kalman filter[M]. Springer Science & Business Media, 2009.
- [43] Evensen G. The ensemble Kalman filter: Theoretical formulation and practical implementation[J]. *Ocean Dynamics*, 2003, 53 (4): 343—367.
- [44] Geir N, Johnsen L M, Aanonsen S I, et al. Reservoir monitoring and continuous model updating using ensemble Kalman filter[C]. SPE Annual Technical Conference and Exhibition: OnePetro, 2003.
- [45] Hendricks Franssen H, Kinzelbach W. Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem[J]. *Water Resources Research*, 2008, 44 (9): W09408.
- [46] Li C, Ren L. Estimation of unsaturated soil hydraulic parameters using the ensemble Kalman filter[J]. *Vadose Zone Journal*, 2011, 10 (4): 1205—1227.
- [47] Wang J R, Hu L T. Advances in mathematical methods of groundwater pollution source identification[J]. *Advances in Water Science*, 2017, 28 (6): 943—952. [王景瑞, 胡立堂. 地下水污染源识别的数学方法研究进展[J]. 水科学进展, 2017, 28 (6): 943—952.]
- [48] Emerick A, Reynolds A. Ensemble smoother with multiple data assimilation[J]. *Computers & Geosciences*, 2013, 55: 3—15.
- [49] Li P, Zha Y, Tso C, et al. Data assimilation of uncalibrated soil moisture measurements from frequency-domain reflectometry[J]. *Geoderma*, 2020, 374: 114432.
- [50] Asher M, Croke B, Jakeman A, et al. A review of surrogate models and their application to groundwater modeling[J]. *Water Resources Research*, 2015, 51 (8): 5957—5973.
- [51] Razavi S, Tolson B, Burn D. Review of surrogate modeling in water resources[J]. *Water Resources Research*, 2012, 48 (7): W07401.
- [52] Zeng L, Chang H, Zhang D. A probabilistic collocation-based Kalman filter for history matching[J]. *SPE Journal*, 2011, 16 (2): 294—306.
- [53] Man J, Li W, Zeng L, et al. Data assimilation for unsaturated flow models with restart adaptive probabilistic collocation based Kalman filter[J]. *Advances in Water Resources*, 2016, 92: 258—270.
- [54] Kennedy M, O'Hagan A. Predicting the output from a complex computer code when fast approximations are available[J]. *Biometrika*, 2000, 87 (1): 1—13.
- [55] Ng L, Eldred M. Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation. 53rd AIAA/ASME/ASCE/AHS/ASC Structures2012: 1852.
- [56] Man J, Lin G, Yao Y, et al. A generalized multi-fidelity simulation method using sparse polynomial chaos expansion[J]. *Journal of Computational and Applied Mathematics*, 2021, 397: 113613.
- [57] Man J, Zhang J, Wu L, et al. ANOVA-based multi-fidelity probabilistic collocation method for uncertainty quantification[J]. *Advances in Water Resources*, 2018, 122: 176—186.
- [58] Man J, Zheng Q, Wu L, et al. Adaptive multi-fidelity probabilistic collocation-based Kalman filter for subsurface flow data assimilation: numerical modeling and real-world experiment[J]. *Stochastic Environmental Research and Risk Assessment*, 2020, 34 (8): 1135—1146.
- [59] Zheng Q, Zhang J, Xu W, et al. Adaptive multifidelity data assimilation for nonlinear subsurface flow problems[J]. *Water Resources Research*, 2019, 55 (1):

- 203—217.
- [60] Zhang D, Lu Z. An efficient, high-order perturbation approach for flow in random porous media via Karhunen–Loève and polynomial expansions[J]. *Journal of Computational Physics*, 2004, 194 (2): 773—794.
- [61] Li W, Lin G, Li B. Inverse regression-based uncertainty quantification algorithms for high-dimensional models: Theory and practice[J]. *Journal of Computational Physics*, 2016, 321: 259—278.
- [62] Erdal D, Neuweiler I, Wollschläger U. Using a bias aware EnKF to account for unresolved structure in an unsaturated zone model[J]. *Water Resources Research*, 2014, 50 (1): 132—147.
- [63] Xu T, Valocchi A, Ye M, et al. Quantifying model structural error: Efficient Bayesian calibration of a regional groundwater flow model using surrogates and a data-driven error model[J]. *Water Resources Research*, 2017, 53 (5): 4084—4105.
- [64] Gupta H, Clark M, Vrugt J, et al. Towards a comprehensive assessment of model structural adequacy[J]. *Water Resources Research*, 2012, 48 (8): W08301.
- [65] Wöhling T, Schöniger A, Gayler S, et al. Bayesian model averaging to explore the worth of data for soil-plant model selection and prediction[J]. *Water Resources Research*, 2015, 51 (4): 2825—2846.
- [66] Shi L, Song X, Tong J, et al. Impacts of different types of measurements on estimating unsaturated flow parameters[J]. *Journal of Hydrology*, 2015, 524: 549—561.
- [67] Neuman S, Xue L, Ye M, et al. Bayesian analysis of data-worth considering model and parameter uncertainties[J]. *Advances in Water Resources*, 2012, 36: 75—85.
- [68] Lindley D. Bayesian statistics: A review[M]. Society for Industrial and Applied Mathematics, 1972.
- [69] Herrera G S, Pinder G F. Space-time optimization of groundwater quality sampling networks[J]. *Water Resources Research*, 2005, 41 (12): W12407.
- [70] Kollat J, Reed P, Maxwell R. Many-objective groundwater monitoring network design using bias-aware ensemble Kalman filtering, evolutionary optimization, and visual analytics[J]. *Water Resources Research*, 2011, 47 (2): W02529.
- [71] Zhang J, Zeng L, Chen C, et al. Efficient Bayesian experimental design for contaminant source identification[J]. *Water Resources Research*, 2015, 51 (1): 576—598.
- [72] Man J, Liao Q, Zeng L, et al. ANOVA-based transformed probabilistic collocation method for Bayesian data-worth analysis[J]. *Advances in Water Resources*, 2017, 110: 203—214.
- [73] Prakash O, Datta B. Optimal characterization of pollutant sources in contaminated aquifers by integrating sequential-monitoring-network design and source identification: methodology and an application in Australia[J]. *Hydrogeology Journal*, 2015, 23 (6): 1089—1107.
- [74] Gharamit M, Marzouk Y, Huan X, et al. A greedy approach for placement of subsurface aquifer wells in an ensemble filtering framework. *International Conference on Dynamic Data-Driven Environmental Systems Science*: Springer, 2014: 301—309.
- [75] Man J, Zhang J, Li W, et al. Sequential ensemble-based optimal design for parameter estimation[J]. *Water Resources Research*, 2016, 52 (10): 7577—7592.
- [76] Man J, Zheng Q, Wu L, et al. Improving parameter estimation with an efficient sequential probabilistic collocation-based optimal design method[J]. *Journal of Hydrology*, 2019, 569: 1—11.
- [77] Geiges A, Rubin Y, Nowak W. Interactive design of experiments: A priori global versus sequential optimization, revised under changing states of knowledge[J]. *Water Resources Research*, 2015, 51 (10): 7915—7936.
- [78] Kamilaris A, Pernafeta-Boldú F. Deep learning in agriculture: A survey[J]. *Computers and Electronics in Agriculture*, 2018, 147: 70—90.
- [79] Tahmasebi P, Sahimi M. Special issue on machine learning for water resources and subsurface systems[J]. *Advances in Water Resources*, 2021, 149: 103851.
- [80] Araya S, Ghezzehei T. Using machine learning for prediction of saturated hydraulic conductivity and its sensitivity to soil structural perturbations[J]. *Water Resources Research*, 2019, 55 (7): 5715—5737.
- [81] Man J, Zeng L, Luo J, et al. Application of the deep learning algorithm to identify the spatial distribution of heavy metals at contaminated sites[J]. *ACS ES&T Engineering*, 2021, 2 (2): 158—168.
- [82] Padmanabha G, Zabaras N. Solving inverse problems using conditional invertible neural networks[J]. *Journal of Computational Physics*, 2021, 433: 110194.
- [83] Farchi A, Laloyaux P, Bonavita M, et al. Using machine learning to correct model error in data assimilation and forecast applications[J]. *Quarterly Journal of the Royal Meteorological Society*, 2021, 147 (739): 3067—3084.
- [84] Zhang Q, Shi L, Holzman M, et al. A dynamic data-driven method for dealing with model structural error in soil moisture data assimilation[J]. *Advances in Water Resources*, 2019, 132: 103407.
- [85] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. *Nature*, 2015, 521 (7553): 436—444.
- [86] Mo S, Zabaras N, Shi X, et al. Integration of adversarial autoencoders with residual dense convolutional networks for estimation of non-Gaussian hydraulic conductivities[J].

- Water Resources Research, 2020, 56(2): e2019WR026082.
- [87] Zhang J, Zheng Q, Wu L, et al. Using deep learning to improve ensemble smoother: Applications to subsurface characterization[J]. Water Resources Research, 2020, 56 (12): e2020WR027399.
- [88] Laloy E, Héault R, Jacques D, et al. Training-image based geostatistical inversion using a spatial generative adversarial neural network[J]. Water Resources Research, 2018, 54 (1): 381—406.
- [89] Raissi M, Perdikaris P, Karniadakis G E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations[J]. Journal of Computational Physics, 2019, 378: 686—707.
- [90] Tartakovsky A, Marrero C, Perdikaris P, et al. Physics-informed deep neural networks for learning parameters and constitutive relationships in subsurface flow problems[J]. Water Resources Research, 2020, 56 (5): e2019WR026731.
- [91] Zheng Q, Zeng L, Karniadakis G E. Physics-informed semantic inpainting : Application to geostatistical modeling[J]. Journal of Computational Physics, 2020, 419: 109676.
- [92] Wang N, Chang H, Zhang D. Efficient uncertainty quantification and data assimilation via theory-guided convolutional neural network[J]. SPE Journal, 2021, 26 (6): 4128—4156.
- [93] Wang N, Zhang D, Chang H, et al. Deep learning of subsurface flow via theory-guided neural network[J]. Journal of Hydrology, 2020, 584: 124700.
- [94] Constantine P G. Active subspaces: Emerging ideas for dimension reduction in parameter studies[M]. Society for Industrial and Applied Mathematics, 2015.
- [95] Yan H, Hao C, Zhang J, et al. Accelerating groundwater data assimilation with a gradient-free active subspace method[J]. Water Resources Research, 2021, 57 (12): e2021WR029610.
- [96] Montzka C, Pauwels V, Franssen H-J H, et al. Multivariate and multiscale data assimilation in terrestrial systems : A review[J]. Sensors, 2012, 12 (12): 16291—16333.
- [97] Zhu P, Shi L, Zhu Y, et al. Data assimilation of soil water flow via ensemble Kalman filter: Infusing soil moisture data at different scales[J]. Journal of Hydrology, 2017, 555: 912—925.
- [98] Yao R J, Yang J S, Zheng F L, et al. Estimation of soil salinity by assimilating apparent electrical conductivity data into HYDRUS model[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35 (13): 91—101. [姚荣江, 杨劲松, 郑复乐, 等. 基于表观电导率和 Hydrus 模型同化的土壤盐分估算[J]. 农业工程学报, 2019, 35 (13): 91—101.]

(责任编辑: 檀满枝)