

8-15-2024

Dynamic Data Driven Simulation: An Overview

Xu Xie

College of Systems Engineering, National University of Defense Technology, Changsha 410073, China

Xiaogang Qiu

College of Systems Engineering, National University of Defense Technology, Changsha 410073, China

Yizheng Bao

College of Systems Engineering, National University of Defense Technology, Changsha 410073, China

Kai Xu

College of Systems Engineering, National University of Defense Technology, Changsha 410073, China

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation. For more information, please contact xtfzxb@126.com.

Dynamic Data Driven Simulation: An Overview

Abstract

Abstract: Dynamic data driven simulation is a simulation paradigm which integrates simulation and data together. This paradigm continuously feeds real-time data into the simulation, enabling the simulation be dynamically adjusted by the data, which thus improves the simulation-based estimation and prediction capability. Due to this integration, the dynamic data driven simulation can estimate system states and predict future state evolution more accurately. This paper reviews the origins and basic concept of dynamic data driven simulation, and introduces several simulation paradigms originated from the idea of "integrating models with data", and identifies the linkages and differences among them. The particle filterbased data assimilation method and the identical-twin simulation experiment are introduced. The current research status of dynamic data driven simulation is summarized from four perspectives, i.e., application scenarios, models and data, data assimilation algorithms, and integration with new technologies. Finally, the future research directions are outlooked from five aspects, which are simulation models, measurement data, data assimilation, algorithm performance, and application areas.

Keywords

modeling and simulation, dynamic data driven simulation, data assimilation, particle filters, identical-twin simulation experiment

Recommended Citation

Xie Xu, Qiu Xiaogang, Bao Yizheng, et al. Dynamic Data Driven Simulation: An Overview[J]. Journal of System Simulation, 2024, 36(8): 1832-1842.

动态数据驱动仿真综述

谢旭, 邱晓刚, 包亦正, 许凯

(国防科技大学 系统工程学院, 湖南 长沙 410073)

摘要: 动态数据驱动仿真是一种“模型和数据相结合”的仿真范式, 它将真实系统的观测(数据)持续注入仿真(模型), 让数据动态地校正仿真(状态、参数), 以此来提高基于仿真的估计和预测能力。由于动态数据驱动仿真融合了模型预测和实时观测两方面的信息, 因此它能更准确地估计系统状态并预测状态的未来演化。梳理了动态数据驱动仿真的思想起源和基本概念, 延伸介绍了“模型和数据相结合”的思想孕育的一系列仿真范式, 并辨析了它们之间的联系和区别; 介绍了基于粒子滤波的数据同化方法和 identical-twin 仿真实验方法; 从应用场景、模型和数据、数据同化算法、与新技术的融合等4个维度综述了动态数据驱动仿真的研究现状; 从仿真模型、观测数据、数据同化、运行效率、应用领域等5个方面对动态数据驱动仿真未来研究方向进行了展望。

关键词: 建模与仿真; 动态数据驱动仿真; 数据同化; 粒子滤波; identical-twin 仿真实验

中图分类号: TP391.9 文献标志码: A 文章编号: 1004-731X(2024)08-1832-11

DOI: 10.16182/j.issn1004731x.joss.24-0127

引用格式: 谢旭, 邱晓刚, 包亦正, 等. 动态数据驱动仿真综述[J]. 系统仿真学报, 2024, 36(8): 1832-1842.

Reference format: Xie Xu, Qiu Xiaogang, Bao Yizheng, et al. Dynamic Data Driven Simulation: An Overview[J]. Journal of System Simulation, 2024, 36(8): 1832-1842.

Dynamic Data Driven Simulation: An Overview

Xie Xu, Qiu Xiaogang, Bao Yizheng, Xu Kai

(College of Systems Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: Dynamic data driven simulation is a simulation paradigm which integrates simulation and data together. This paradigm continuously feeds real-time data into the simulation, enabling the simulation be dynamically adjusted by the data, which thus improves the simulation-based estimation and prediction capability. Due to this integration, the dynamic data driven simulation can estimate system states and predict future state evolution more accurately. This paper reviews the origins and basic concept of dynamic data driven simulation, and introduces several simulation paradigms originated from the idea of "integrating models with data", and identifies the linkages and differences among them. The particle filter-based data assimilation method and the identical-twin simulation experiment are introduced. The current research status of dynamic data driven simulation is summarized from four perspectives, i.e., application scenarios, models and data, data assimilation algorithms, and integration with new technologies. Finally, the future research directions are outlooked from five aspects, which are simulation models, measurement data, data assimilation, algorithm performance, and application areas.

Keywords: modeling and simulation; dynamic data driven simulation; data assimilation; particle filters; identical-twin simulation experiment

收稿日期: 2024-02-02 修回日期: 2024-03-19

基金项目: 国家自然科学基金(62103428)

第一作者: 谢旭(1988-), 男, 副教授, 博士, 研究方向为建模与仿真、动态数据驱动仿真。

0 引言

时空动态系统(spatial-temporal dynamic system)是一类需要同时描述系统空间特征和时间特征演化规律的系统^[1-2]。随着技术的进步和社会的发展,许多时空动态系统都呈现出规模巨大、要素众多、交互复杂等特点,这类时空动态系统统称为复杂时空动态系统,例如,战争系统、交通系统。复杂时空动态系统的管理、控制、决策等需要首先对系统状态进行准确的估计和预测。一般而言,会有关于同一个复杂时空动态系统多个视角的信息^[3]:一是计算视角,即可以建立关于系统的宏观、中观或者微观层次的模型,例如,交通系统中不同层次的交通流模型^[4],并基于模型(理论分析或仿真推演)去预测系统的状态及其演化过程;二是数据视角,即通过某些观测手段获取真实系统的观测数据,基于这些观测数据也能(部分)推断真实系统的状态。基于模型的方法和基于数据的方法各有优劣。基于模型的方法计算效率高、经济成本低,但是状态估计的准确度取决于模型的质量,并且由于模型是基于历史数据建立的,因此不能有效应对真实系统中的(建模时未考虑的)突发情况;基于数据的方法能准确反映真实系统实时运行情况,但是对于没有观测到的状态变量,则不能给出准确估计。例如,在交通系统中,可以根据道路上浮动车数据(floating car data, FCD)^[5]估计路段的平均速度,并据此对道路拥堵情况进行评级,但是无法给出车流流量和车流密度等变量的估计。

因此,一个直观的想法就是将计算视角和数据视角相结合,综合利用模型和数据两方面的信息估计真实系统的状态。文献[6]通过一个室内温度估计的例子证明了综合利用模型和数据两方面信息估计系统状态,比单独使用其中任一种信息得到的估计都更为准确。较早提出“模型和数据相结合”思想的是美国空军科学研究办公室主任 Frederica Darema。2000年左右,她在主持美国自

然科学基金的一个研讨会时提出了这个理念,并在2004、2005年的ICCS会议上正式提出动态数据驱动应用系统(dynamic data driven application systems, DDDAS)的概念^[7-8]。在DDDAS中,“应用系统”可以是仿真系统,也可以是普通的软件系统。美国乔治亚州立大学的Xiaolin Hu教授把DDDAS的概念聚焦到仿真领域,在2008年申请立项的一个基金项目(CNS-0841170)中,提出了“数据驱动仿真”(data driven simulation)的概念,并在2011年正式提出了动态数据驱动仿真范式(dynamic data driven simulation, DDDS)^[9]。

动态数据驱动仿真提出以后,在很多领域得到了广泛应用,并逐步发展成为建模仿真理论和技术体系的重要分支,在历年的ACM SIGSIM PADS、WSC等重要学术会议上,还会不定期组织分会议对动态数据驱动仿真的最新进展进行专题研讨。在建模仿真领域,能体现“模型和数据相结合”思想的仿真范式有多个,它们有区别也有联系,这些概念如何辨析?当前动态数据驱动仿真的应用现状如何?如何实现模型和数据的动态融合?如何开展动态数据驱动仿真实验研究?这些问题是大部分开展动态数据驱动仿真研究的科研人员渴望快速解决的问题。为此,本文从概念、组成、方法、应用等维度对动态数据驱动仿真范式进行综述,希望就该问题为相关领域的科研人员提供一个较为全面的认识,促进动态数据驱动仿真在应用领域的进一步发展。

1 动态数据驱动仿真

仿真是有目的地利用模型进行的实验^[10]。由于在建模过程中会不可避免地引入各类误差,例如历史数据不足引起的模型误差^[11]、概念抽象引起的模型误差^[3],因此,不管多么精细的模型都不可能完全准确地描述真实系统,使用仿真来预测真实系统的状态总会有偏差,而且随着仿真时间的推进,这种偏差会累积并逐渐放大。随着传感

器、数据存储、远程数据访问等相关技术的进步，可以很容易地获取真实系统的观测数据，这些数据直接反映了真实系统当前的状态^[7-8]。一个自然的想法就是将真实系统的仿真(模型)和真实系统的观测(数据)结合起来，让数据动态地校正仿真，以此来提高基于仿真的状态估计和预测能力。这种“实时数据持续注入仿真，并动态校正仿真运行”的仿真范式就是动态数据驱动仿真^[9]。图1直观展示了动态数据驱动仿真中观测数据动态校正仿真运行的这种思想。

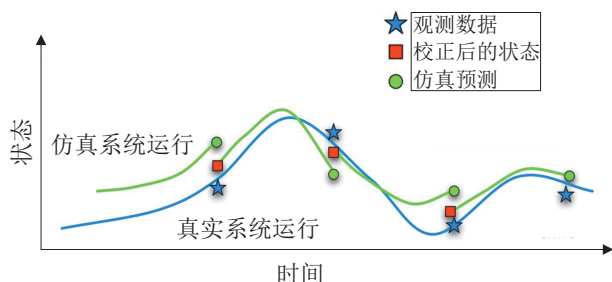


图1 动态数据驱动仿真示意图
Fig. 1 An illustration of DDDS idea

动态数据驱动仿真的一般结构如图2所示，其核心要素分别是仿真模型(也称为系统模型)、观测模型、数据同化(data assimilation)算法^[12]。系统模型描述系统状态随时间的演化；观测模型描述观测数据和系统状态之间的关系；系统模型给出了当前时刻对真实系统状态的先验估计(如图1绿色圆点所示)，同时我们从真实系统采集到一系列观测数据(如图1蓝色五角星所示)，数据同化算法则结合这两方面的信息给出真实系统状态的后验估计(如图1红色方块所示)。数据同化算法是动态数据驱动仿真的核心，将在第3节进行详细介绍。

在动态数据驱动仿真中，由于真实系统的数据被持续不断地同化到仿真中，使得仿真能够动态地调整自身的状态以更接近真实系统的状态^[9]；如果将模型参数增广到模型状态中，模型的参数也可以随着数据同化的进行和状态一同被估计出来^[13]。通过同化真实系统的数据，可以得到一个仿真状态动态逼近真实系统的状态、模型参数动

态调优的仿真，因此，动态数据驱动仿真比传统仿真具有更强的状态估计和预测能力^[12]。

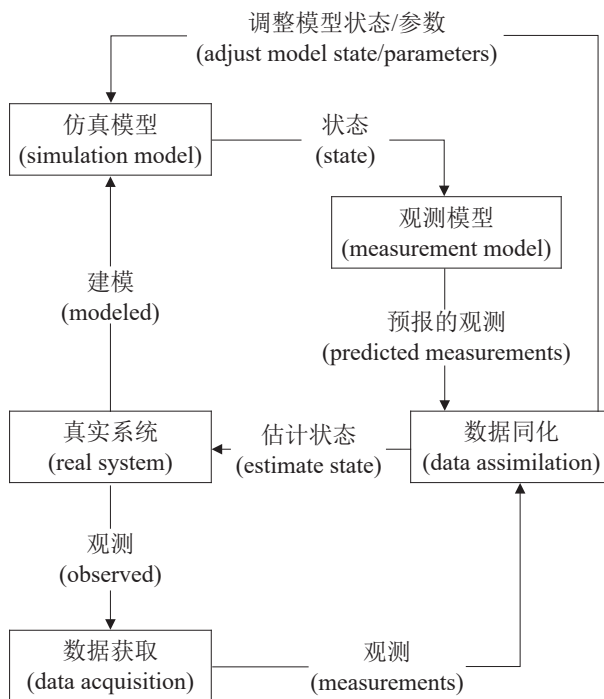


图2 动态数据驱动仿真的结构
Fig. 2 Architecture of a dynamic data driven simulation

相关文献中经常出现另一个相近的词，即“数据驱动...”(data driven ...)^[14-16]，它和“动态数据驱动...”(dynamic data driven ...)有什么区别呢?“动态数据驱动...”强调实时数据注入，它发生在仿真运行时，而“数据驱动...”一般使用的是静态的或者存档的历史数据，主要用于建模阶段的模型校正(model calibration)。

2 相关仿真范式辨析

2.1 在线仿真、共生仿真、动态数据驱动应用系统、平行仿真

建模与仿真领域还存在几个和动态数据驱动仿真紧密相关的概念，分别是共生仿真(symbiotic simulation)^[17]、动态数据驱动应用系统^[7-8]、在线仿真(on-line simulation)^[18]、平行仿真^[19-21]：

(1) 共生仿真强调真实系统和它的仿真系统之间存在一种互利共生的关系——仿真系统可以利

用真实系统的实时观测数据初始化并驱动仿真系统的运行; 反之, 真实系统可以在它的共生仿真系统中快速试验各种决策方案(what-if analysis), 从而为它的决策提供辅助支持。

(2) 动态数据驱动应用系统将应用系统或仿真系统和真实系统的观测过程有机统一成一个反馈控制系统, 使应用/仿真系统可以在运行时动态接收真实系统的数据并做出响应, 而应用/仿真系统也可以动态引导和控制真实系统的观测过程。

(3) 在线仿真是指由真实系统的数据驱动的仿真。

(4) 平行仿真是将仿真系统作为人工系统, 与真实系统协同运行、虚实互动、共同演化发展, 及相互控制的一种仿真技术应用方法。

动态数据驱动应用系统、共生仿真都强调仿真系统和真实系统之间的互利关系——真实系统的数据可以驱动仿真系统的运行, 而仿真系统反过来也可以影响真实系统。但是, 共生仿真强调仿真系统可以控制整个真实系统的运行, 而在动态数据驱动应用系统中, 仿真系统对真实系统的影响则主要体现在对观测过程的控制。在线仿真和动态数据驱动仿真的概念是一致的, 它们只强调真实系统的数据可以驱动仿真系统的运行, 但不关注仿真系统对真实系统的影响。平行仿真的思想和动态数据驱动应用系统的思想非常相似, 但是平行仿真更强调仿真系统作为一个试验台(testbed), 通过动态注入真实系统的实时数据, 实现仿真系统和真实系统的协同演化逼近, 因此仿真系统能够替代真实系统进行实验^[19]; 通过在试验台上超实时地开展计算实验, 并将实验结果作用到真实系统, 提前对真实系统进行干预, 从而达到影响真实系统演化的目的^[20]。

2.2 数字孪生

数字孪生的概念可以追溯到密西根大学 Michael Grieves 教授于 2002 年针对产品全生命周期管理(product lifecycle management, PLM)提出的

镜像空间模型(mirrored space model), 后来 NASA 的 John Vickers 将其命名为数字孪生, 并沿用至今^[22]。数字孪生本质上是物理对象的一个数字化模型, 它可以通过接收来自物理对象的数据而实时演化, 从而与物理对象在全生命周期保持一致^[23-25]。基于数字孪生可以进行仿真, 仿真结果又可以反馈给物理对象, 从而帮助物理对象进行优化和决策。目前, 数字孪生主要应用在制造领域, 可以为产品的设计、制造、运维等提供优化建议和辅助决策支持^[26]。近年来, 数字孪生的服务和功能呈现多元化趋势, 逐渐向智慧城市、健康医疗等领域拓展。

2.3 区别与联系

动态数据驱动应用系统、在线仿真、共生仿真、平行仿真、动态数据驱动仿真等范式都体现了“真实系统的数据和它的仿真相结合”的思想, 它们的基本理念、核心技术、应用场景都很相似, 只是在真实系统和仿真系统之间的相互作用关系上有少许差别, 例如, 有的强调双向的互利关系(共生仿真、动态数据驱动应用系统、平行仿真), 而有的只强调真实系统的数据可以驱动仿真系统的运行(在线仿真、动态数据驱动仿真)。数字孪生是物理对象的一个数字化模型, 它可以接收物理对象的数据进行实时演化, 从而与物理对象在全生命周期保持一致, 但是数字孪生和上面的仿真模式不在一个维度, 因为数字孪生是模型, 而另外的则是仿真。

3 基于粒子滤波的数据同化方法

动态数据驱动仿真的核心优势是它能够把真实系统的观测数据持续注入仿真, 动态估计真实系统的状态, 并据此校正仿真系统的状态, 使仿真系统更接近真实系统^[9]。这种能力是由数据同化来实现的。数据同化是一类将真实系统的观测数据和仿真预测结合起来估计真实系统状态的方法^[6,27]。

数据同化问题可以建模成一个数学问题^[27-28]。

假设系统状态随时间的演化过程可被描述为一个离散时间模型(系统模型):

$$s_k = f(s_{k-1}) + v_{k-1}, k = 1, 2, \dots,$$

其中, f 为一个关于状态 s_{k-1} 的非线性函数; v_{k-1} 为系统噪声。观测数据与系统状态之间的关系可被描述为一个观测模型:

$$m_k = g(s_k) + \varepsilon_k, k = 1, 2, \dots,$$

其中, g 为非线性函数, 描述状态 s_k 和观测 m_k 之间的映射关系; ε_k 为观测噪声。数据同化算法的目标就是根据当前采集到的所有观测数据 $m_{1:k} = \{m_i | i = 1, 2, \dots, k\}$, 给出 k 时刻状态轨迹 $s_{0:k} = \{s_i | i = 0, 1, \dots, k\}$ 的估计, 即后验概率分布 $p(s_{0:k} | m_{1:k})$ 。

常用的数据同化算法包括变分同化法(三维变分法 3D-VAR^[29]和四维变分法 4D-VAR^[30])、卡尔曼滤波^[31]、扩展卡尔曼滤波^[32]、集合卡尔曼滤波^[33]、粒子滤波^[31, 34-35]等。大多数数据同化算法都依赖某些假设, 例如, 假设系统/观测模型是线性模型, 或者假设系统/观测误差是高斯误差, 只有粒子滤波算法对模型和数据没有特殊要求^[36]。由于动态数据驱动仿真中的“仿真”一般是指微观仿真, 它的状态演化过程具有高度非线性/非高斯特征, 因此, 动态数据驱动仿真中最常用的数据同化方法是粒子滤波算法^[12]。

粒子滤波算法使用一组蒙特卡罗采样(粒子)和对应的权重 $\{s_{0:k}^i, w_k^i\}_{i=1}^{N_p}$ 来近似后验概率分布:

$$p(s_{0:k} | m_{1:k}) \approx \sum_{i=1}^{N_p} w_k^i \delta(s_{0:k} - s_{0:k}^i)$$

其中, $\delta(\cdot)$ 为狄拉克函数。粒子权重是通过重要性采样(importance sampling)来计算的, 也就是说, 如果样本 $\{s_{0:k}^i\}_{i=1}^{N_p}$ 采样于重要性分布(importance density) $q(s_{0:k} | m_{1:k})$, 那么权重 w_k^i 为

$$w_k^i \propto \frac{p(s_{0:k}^i | m_{1:k})}{q(s_{0:k}^i | m_{1:k})}$$

根据贝叶斯定理, $p(s_{0:k} | m_{1:k})$ 可以表示为

$$p(s_{0:k} | m_{1:k}) = \frac{p(s_{0:k})p(m_{1:k} | s_{0:k})}{p(m_{1:k})}$$

类似地, 有:

$$p(s_{0:k-1} | m_{1:k-1}) = \frac{p(s_{0:k-1})p(m_{1:k-1} | s_{0:k-1})}{p(m_{1:k-1})}$$

二式相除并整理, 可以得到如下递推公式:

$$p(s_{0:k} | m_{1:k}) \propto p(m_k | s_k)p(s_k | s_{k-1})p(s_{0:k-1} | m_{1:k-1})$$

假设重要性分布 $q(s_{0:k} | m_{1:k})$ 可被分解为

$$q(s_{0:k} | m_{1:k}) = q(s_k | s_{0:k-1}, m_{1:k})q(s_{0:k-1} | m_{1:k-1})$$

那么, 当新的观测 m_k 到达时, 上一时刻的样本和权重 $\{s_{0:k-1}^i, w_{k-1}^i\}_{i=1}^{N_p}$ 可以按照如下步骤更新至 $\{s_{0:k}^i, w_k^i\}_{i=1}^{N_p}$:

(1) 采样。根据分布 $q(s_k | s_{0:k-1}^i, m_{1:k})$ 产生状态 s_k^i , 并推广到状态轨迹 $s_{0:k-1}^i \sim q(s_{0:k-1} | m_{1:k-1})$, 生成 k 时刻的状态轨迹 $s_{0:k}^i \sim q(s_{0:k} | m_{1:k})$;

(2) 更新权重。根据观测数据更新粒子 $s_{0:k}^i$ 的权重:

$$w_k^i \propto \frac{p(s_{0:k}^i | m_{1:k})}{q(s_{0:k}^i | m_{1:k})} = \frac{p(m_k | s_k^i)p(s_k^i | s_{k-1}^i)}{q(s_k^i | s_{0:k-1}^i, m_{1:k})} w_{k-1}^i$$

每个粒子的权重更新完毕之后, 再进行归一化, 确保所有粒子的权重之和等于 1。

上述粒子滤波过程称为序贯重要性采样(sequential importance sampling)^[31]。在实际应用中, 一般选择系统状态转移概率分布(system transition density)作为重要性分布^[37], 也就是说 $q(s_k | s_{0:k-1}, m_{1:k}) = p(s_k | s_{k-1})$, 此时, 权重计算可简化为 $w_k^i \propto p(m_k | s_k^i)w_{k-1}^i$ 。理论证明, 当粒子数目趋于无穷时, 基于粒子和权重 $\{s_{0:k}^i, w_k^i\}_{i=1}^{N_p}$ 的蒙特卡罗近似等价于 $p(s_{0:k} | m_{1:k})$ 的解析表示^[31, 34-35]。

基于序贯重要性采样的粒子滤波算法会存在粒子退化(degeneracy)问题^[31], 也就是说一组粒子中除了极少数粒子外, 其他粒子的权重都是 0。一个有效解决粒子退化问题的方法就是在粒子权重更新之后增加一个重采样步骤, 即根据粒子的权重随机地决定该粒子是被复制还是被删除。通过重采样, 权重大的粒子被保留下来, 而权重小的粒子也不至于被完全抛弃而丧失被选优的可能。

增加了重采样步骤的序贯重要性采样粒子滤波算法的流程如图 3 所示。这个流程是当前动态数据驱动仿真应用研究普遍采用的数据同化流程^[38]。

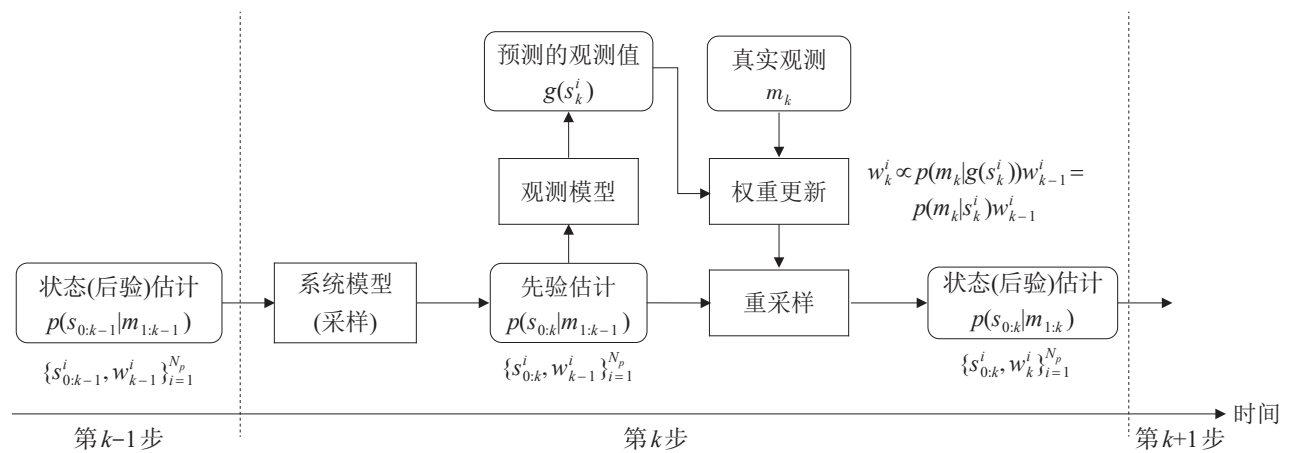


图3 基于粒子滤波的数据同化流程

Fig. 3 Data assimilation procedure based on particle filters

4 Identical-twin 仿真实验方法

在动态数据驱动仿真的相关研究中, 一般不直接把仿真系统和真实系统耦合起来, 从真实系统获取观测数据, 驱动仿真系统运行, 而是采用一种叫 identical-twin 的仿真实验方法^[37,39]。

identical-twin 仿真实验方法的基本步骤如下^[37,39]: 第一步, 运行一次仿真模型并记录相关数据, 将这次仿真看作是“真实系统”的运行; 第二步, 为记录的数据加噪声, 以此来模拟观测噪声, 并得到有噪声的观测数据; 第三步, 在第一步的仿真模型基础上加扰动, 来模拟对“真实系统”的建模, 得到“真实系统”的有误差的仿真模型。最后, 将第二步的观测数据同化到第三步的仿真模型, 得到状态估计结果, 和第一步的真实数据作对比, 从而评估动态数据驱动仿真的有效性。

基于 identical-twin 仿真实验方法开展动态数据驱动仿真研究具有多方面的优势^[40]: ①数据质量容易控制。可以在仿真数据中灵活设置不同的误差模型、噪声水平、数据采集频率等, 生成不同质量的观测数据。②在真实系统中很难获得的小概率事件或者极端条件下的观测数据, 在仿真中很容易产生。③容易开展对比研究。由于无法准确知道真实系统的状态及其演化过程, 因此,

如果在动态数据驱动仿真中直接融合真实系统, 就很难精确对比动态数据驱动仿真得到的结果与真实系统的差距。在 identical-twin 仿真实验中, “真实系统”本身也是仿真, 因此, 可以准确知道它的状态及其演化, 很容易量化评估动态数据驱动仿真的有效性。

5 动态数据驱动仿真研究现状

本节从应用场景、模型和数据、数据同化算法、与新技术的融合等4个维度综述动态数据驱动仿真的研究现状。

5.1 应用场景维度

乔治亚州立大学的Hu教授提出动态数据驱动仿真范式后, 成功将其应用于森林火灾蔓延的预测^[9,37,41]。火的蔓延过程被建模成一个基于网络的离散事件模型 DEVS-FIRE^[42-43]。火的蔓延受到火场风速、风向等随机因素的影响, 但是为了降低建模的复杂程度, DEVS-FIRE 假设仿真过程中风速和风向是恒定的。因此, 单独使用 DEVS-FIRE 预测火的蔓延显然是不准确的。另一方面, 人们可以在火场中部署温度传感器来探测火场中不同位置的温度; 传感器离火越近, 温度越高, 因此, 温度数据可以反映当前火势的蔓延情况。为此, 研究组提出了一个基于粒子滤波算法的数据同化

框架，利用实时温度数据动态校正基于 DEVS-FIRE 的仿真预测。实验结果证明这种动态数据驱动仿真范式能准确地估计火灾的蔓延。

动态数据驱动仿真范式成功应用于森林火灾蔓延预测之后，该研究组把类似的研究思路拓展到不同的应用场景，相继开展了智能建筑中(小规模)人群分布估计^[44]、交通系统中事故发生地定位^[45]等应用研究。这些应用研究的基本步骤可以归纳为 3 步：①根据应用场景建立系统模型和观测模型；②设计基于粒子滤波的数据同化算法(这一步的核心工作是根据系统状态和观测数据，推导权重计算公式)；③开展 identical-twin 仿真实验研究，验证方法的有效性。研究人员可以按上述步骤开展基于动态数据驱动仿真的应用研究。

5.2 模型和数据维度

模型(状态构成、状态演化规律等)和数据的特点直接影响动态数据驱动仿真中数据同化方法的设计。因此，一些研究重点关注模型和数据的特殊性，对数据同化方法开展了一系列有针对性的研究。

Xie 等针对离散事件系统中状态组成、状态更新的特殊性，研究了通用离散事件仿真中的数据同化问题^[12,36,46]。考虑到离散事件仿真中状态转移过程的高度非线性特征，作者提出了一个通用的基于粒子滤波算法的数据同化框架。该框架使用离散事件系统规范 DEVS 进行形式化描述，并解决了离散事件仿真数据同化中面临的非数值型状态、状态更新时间不固定、事件序列观测数据、仿真实体数量可变等难题。这个通用的数据同化框架已在工矿业系统中实体到达时间估计^[36]、城市交通系统中车辆轨迹重构^[46]、车辆密度估计^[47]等典型离散事件系统中做了验证，取得了良好效果。

传统的动态数据驱动仿真同化的数据一般是传感器采集的定量数据，例如，温度、车速，文献[1]将这一类数据称为 hard data。与之相对地，

还有一类由人报告的观测数据，例如“东南方向火势很大”，这一类数据一般是定性的、模糊的、带有报告人主观判断的，因此称之为 soft data。针对这样的定性数据，Long 等采用模糊集理论，实现定性数据的定量转换，提出了一种能够综合处理定性/定量数据的数据同化框架^[1]，并在森林火灾蔓延预测中进行了仿真实验验证，结果证实，soft data 可以作为 hard data 的有效补充，显著提升数据同化的效果。

5.3 数据同化算法维度

还有部分研究关注基于粒子滤波的数据同化算法的准确性和运行效率问题，开展了一系列优化工作。

在大多数实际应用中，人们一般把系统的状态转移概率分布作为重要性分布，以简化粒子权重的计算。但是当系统模型不够精确或者不能反映真实系统实际情况时，以系统的状态转移概率分布作为重要性分布采样产生的粒子就不能有效代表真实的状态概率分布，这将导致数据同化的结果不准确。针对这样的问题，Xue 等提出一种两阶段的重要性分布采样方法^[48]：第一阶段，根据系统状态转移概率分布采样生成一个状态；第二阶段，根据实际的观测数据再生成一个状态；最后把两个阶段生成的状态综合，作为最终采样结果。相应地，作者采用核密度估计方法计算粒子的权重。这种采样方法把观测数据引入采样过程，可以有效提升粒子对真实系统实际情况的表示能力。

由于动态数据驱动仿真中的“仿真”一般是微观仿真，具有状态维度高、状态演化复杂等特点，因此，基于粒子滤波的数据同化方法计算量很大。为此，有大量研究设计了分布式粒子滤波算法^[49]，重点关注如何在多个处理单元上并行执行采样，Bai 等则在前人研究成果基础上，重点改进了重采样过程，提出了一系列粒子迁移策略^[50]，解决重采样后各处理单元上的粒子数目差异较大

导致的负载不均衡问题。

5.4 与新技术融合维度

随着5G、物联网(internet of things, IoT)^[51]、边缘计算(edge computing)^[52-53]等信息基础设施的普及,网络、存储、计算等能力被拓展至离终端最近的边缘侧^[54],因此动态数据驱动仿真走向分布式是未来的必然趋势。Xie等提出了一种分布式动态数据驱动仿真方法^[55]——以复杂时空动态系统中的边缘基础设施为计算平台,在这些地理上分散的计算平台上运行一组相互作用的仿真进程,每个仿真进程自主选择平台附近其感兴趣的区域进行建模;每个仿真进程综合利用计算平台附近终端采集的数据、其他仿真进程的仿真结果动态地校正仿真,从而准确估计和预测系统的局部状态;这些局部状态再通过合适的方式整合起来,形成对系统全局状态的估计和预测。分布式动态数据驱动仿真具有延迟小、网络带宽要求低、数据利用更充分等优点。作为实验验证, Xie等设计了一个交通系统中车流密度估计的场景^[55],并使用3个仿真组成的分布式动态数据驱动仿真的可行性。

6 结论与展望

“模型和数据相结合”的思想催生了很多仿真范式,例如,动态数据驱动应用系统、共生仿真、在线仿真、平行仿真、动态数据驱动仿真。这些仿真范式的基本理念、核心技术、应用场景都很相似,但是在真实系统和仿真系统之间的相互作用关系上有少许差别。本文系统介绍了动态数据驱动仿真范式,并明确辨析了这些相似概念之间的联系和区别。在此基础上,详细介绍了基于粒子滤波的数据同化方法和 identical-twin 仿真实验方法。最后,从应用场景、模型和数据、数据同化算法、与新技术的融合等4个维度综述了动态数据驱动仿真的研究现状。

从动态数据驱动仿真的基本组成和推广应用两个维度分析,未来的研究可以从以下5个方面展开^[38,56]:

(1) 仿真模型。在动态数据驱动仿真中,仿真模型扮演着非常重要的角色。首先,仿真模型一般被用作系统状态转移模型,支撑粒子滤波采样过程的执行。此外,当数据同化给出真实系统状态的估计后,可以用这个估计的状态初始化仿真模型,并基于模型开展仿真实验,对真实系统的行为进行预测和分析。在动态数据驱动仿真中,仿真模型的抽象层次、详略程度必须和研究目标相匹配,否则它不能有效支撑数据同化和基于模型的预测与分析。因此,未来可以尝试研究不同抽象层次的仿真模型的数据同化问题(文献[57]对这个问题进行了初步探索),并提出仿真模型和系统噪声应当如何选择的指南。

(2) 观测数据。当前大多数应用案例主要处理数值型观测数据和高斯噪声。除此之外,还存在许多其他类型的观测数据,例如,离散事件系统中常见的事件序列数据、5.2节介绍的 soft data。如何同化这些非常规类型的数据是未来研究需要关注的问题。在一些复杂的应用中,观测数据可能来自多个不同的数据源,并具备不同的数据特征。在这类应用中,需要进行多源数据融合,以产生一致、准确和有意义的信息。如何将数据融合与数据同化相结合以支持动态数据驱动仿真也是未来一个重要的研究课题。观测数据的质量对数据同化有直接的影响。观测数据质量除了和噪声、数据采集频率等因素有关外,还受传感器部署方式的影响,因为传感器部署决定了观测数据中所蕴含的“有用”信息量。因此,在传感器资源有限的情况下,如何部署传感器以便为数据同化收集更有用的数据也是未来一个值得研究的问题。

(3) 数据同化。由于动态数据驱动仿真中系统状态具有维数高、演化过程高度非线性/非高斯等特征,因此,基于粒子滤波的数据同化算法很难

收敛。如何结合应用特点设计更高效的采样、重采样方法,以支持更准确、鲁棒性更强的系统状态估计,是基于粒子滤波的数据同化方法未来需要关注的研究问题。数据同化起源于气象和海洋学领域,这些领域的系统模型主要是微分方程模型。虽然这些领域的模型和仿真中的模型有显著差别,但是也应当密切关注这些领域数据同化方法的最新进展,研究其他领域的新方法如何应用于仿真领域的数据同化问题。基于粒子滤波的数据同化方法的性能受一系列参数的影响,例如,模型的精度、观测数据的质量、数据同化步长、粒子数量。未来需要进行系统性的灵敏度分析^[58],量化评估这些参数对数据同化结果的影响。

(4) 运行效率。动态数据驱动仿真需要注入真实系统的实时数据,因此,它要求数据同化过程中涉及到的所有计算(采样、重采样、状态估计)都必须在下一次观测数据到来之前完成。为了满足实时性的要求,需要研究如何利用并行与分布式计算来提升基于粒子滤波的数据同化方法的运行效率。虽然已有部分工作进行了这方面的探索(详见5.3节),但是还需要开展更深入的研究,尤其是当复杂系统仿真也需要进行并行或分布执行时,如何把并行与分布式粒子滤波和并行与分布式仿真有机结合起来,是未来亟需解决的问题。

(5) 应用领域。应用需求是科学研究的重要牵引。未来应当寻找更多的应用案例,解决更多实际问题。案例不同,数据同化的3个基本要素也会千差万别,会引出很多新的、需要研究的问题。

参考文献:

- [1] Long Yuan, Hu Xiaolin. Dynamic Data Driven Simulation with Soft Data[C]//Proceedings of the Symposium on Theory of Modeling & Simulation-DEVS Integrative. San Diego: Society for Computer Simulation International, 2014: 16:1-16.
- [2] Azeem Sarwar. Spatiotemporal Systems: Gradual Variations, Identification, Adaptation and Robustness[D]. Champaign: University of Illinois Urbana-Champaign, 2009.
- [3] William A Lahoz, Philipp Schneider. Data Assimilation: Making Sense of Earth Observation[J]. *Frontiers in Environmental Science*, 2014, 2: 1-16.
- [4] Martin Treiber, Arne Kesting. *Traffic Flow Dynamics: Data, Models and Simulation*[M]. Berlin: Springer Berlin Heidelberg, 2013.
- [5] Leduc G. Road Traffic Data: Collection Methods and Applications: JRC47967[R]. [S.l.]: [s.n.], 2008: 1-52.
- [6] Bouttier F, Courtier P. Data Assimilation Concepts and Methods[R]. [S.l.]: [s.n.], 1999: 1-59.
- [7] Darema F. Dynamic Data Driven Applications Systems: A New Paradigm for Application Simulations and Measurements[C]//Computational Science-ICCS 2004. Berlin: Springer Berlin Heidelberg, 2004: 662-669.
- [8] Darema F. Dynamic Data Driven Applications Systems: New Capabilities for Application Simulations and Measurements[C]//Computational Science-ICCS 2005. Berlin: Springer Berlin Heidelberg, 2005: 610-615.
- [9] Hu X. Dynamic Data Driven Simulation[J]. *SCS M & S Magazine II*, 2011, 1: 16-22.
- [10] 黄柯棣, 邱晓刚, 查亚兵, 等. 建模与仿真技术[M]. 长沙: 国防科技大学出版社, 2010.
Huang Kedi, Qiu Xiaogang, Zha Yabing, et al. *Modeling and Simulation Technology*[M]. Changsha: National University of Defense Technology Press, 2010.
- [11] Zeigler B, Praehofer H, Kim T G. *Theory of Modeling and Simulation: Integrating Discrete Event and Continuous Complex Dynamic Systems*[M]. 2nd ed. [S.l.]: Academic Press, 2000.
- [12] Xie X. *Data Assimilation in Discrete Event Simulations* [D]. Delft: Delft University of Technology, 2018.
- [13] Bai Fan, Guo Song, Hu Xiaolin. Towards Parameter Estimation in Wildfire Spread Simulation Based on Sequential Monte Carlo Methods[C]//Proceedings of the 44th Annual Simulation Symposium. San Diego: Society for Computer Simulation International, 2011: 159-166.
- [14] Zhang Bo, Zhong Jinghui, Cai Wentong. A Data-driven Approach for Pedestrian Intention Prediction in Large Public Places[C]//Proceedings of the 2022 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation. New York: Association for Computing Machinery, 2022: 33-36.
- [15] Huang Yilin, Seck M D, Alexander Verbraeck. Component-based Light-rail Modeling in Discrete Event Systems Specification DEVS[J]. *Simulation*, 2015, 91 (12): 1027-1051.
- [16] Wang Junfeng, Chang Qing, Xiao Guoxian, et al. Data Driven Production Modeling and Simulation of Complex Automobile General Assembly Plant[J]. *Computers in*

- Industry, 2011, 62(7): 765-775.
- [17] Aydt H, Turner S J, Cai Wentong, et al. Symbiotic Simulation Systems: An Extended Definition Motivated by Symbiosis in Biology[C]//2008 22nd Workshop on Principles of Advanced and Distributed Simulation. Piscataway: IEEE, 2008: 109-116.
- [18] Farzad Kamrani, Rassul Ayani. Using On-line Simulation for Adaptive Path Planning of UAVs[C]//11th IEEE International Symposium on Distributed Simulation and Real-Time Applications (DS-RT'07). Piscataway: IEEE, 2007: 167-174.
- [19] 段伟. 平行仿真的内涵、发展与应用[J]. 指挥与控制学报, 2019, 5(2): 82-86.
Duan Wei. Parallel Simulation: Motivation, Concept and Application[J]. Journal of Command and Control, 2019, 5(2): 82-86.
- [20] 王会霞. 平行仿真技术研究[J]. 航天控制, 2016, 34(6): 64-67.
Wang Huixia. Research on Parallel Simulation Technology[J]. Aerospace Control, 2016, 34(6): 64-67.
- [21] 窦林涛, 初阳, 周玉芳, 等. 平行仿真技术在指控系统中的应用构想[J]. 指挥控制与仿真, 2017, 39(1): 62-69.
Dou Lintao, Chu Yang, Zhou Yufang, et al. Conception of the Application of Parallel Simulation Technology in Command and Control System[J]. Command Control & Simulation, 2017, 39(1): 62-69.
- [22] Grieves M, Vickers J. Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems[M]//Franz-Josef Kahlen, Flumerfelt S, Anabela Alves. Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches. Cham: Springer International Publishing, 2017: 85-113.
- [23] 张霖. 关于数字孪生的冷思考及其背后的建模和仿真技术[J]. 系统仿真学报, 2020, 32(4): 1-10.
Zhang Lin. Cold Thinking about the Digital Twin and the Modeling and Simulation Techniques Behind It[J]. Journal of System Simulation, 2020, 32(4): 1-10.
- [24] Liu Mengnan, Fang Shuiliang, Dong Huiyue, et al. Review of Digital Twin About Concepts, Technologies, and Industrial Applications[J]. Journal of Manufacturing Systems, 2021, 58, Part B: 346-361.
- [25] 杨林瑶, 陈思远, 王晓, 等. 数字孪生与平行系统: 发展现状、对比及展望[J]. 自动化学报, 2019, 45(11): 2001-2031.
Yang Linyao, Chen Siyuan, Wang Xiao, et al. Digital Twins and Parallel Systems: State of the Art, Comparisons and Prospect[J]. Acta Automatica Sinica, 2019, 45(11): 2001-2031.
- [26] 陶飞, 张贺, 戚庆林, 等. 数字孪生十问: 分析与思考[J]. 计算机集成制造系统, 2020, 26(1): 1-17.
Tao Fei, Zhang He, Qi Qinglin, et al. Ten Questions Towards Digital Twin: Analysis and Thinking[J]. Computer Integrated Manufacturing Systems, 2020, 26(1): 1-17.
- [27] Nichols N K. Data Assimilation: Aims and Basic Concepts[C]//Data Assimilation for the Earth System. Dordrecht: Springer Netherlands, 2003: 9-20.
- [28] Ide K, Courtier P, Ghil M, et al. Unified Notation for Data Assimilation : Operational, Sequential and Variational[J]. Journal of the Meteorological Society of Japan. Ser. II, 1997, 75(1B): 181-189.
- [29] Wu Wanshu, Purser R J, Parrish D F. Three-dimensional Variational Analysis with Spatially Inhomogeneous Covariances[J]. Monthly Weather Review, 2022, 130(12): 2905-2916.
- [30] Lorenc A C, Rawlins F. Why does 4D-var beat 3D-Var?[J]. Quarterly Journal of the Royal Meteorological Society, 2005, 131(613): 3247-3257.
- [31] Arulampalam M S, Maskell S, Gordon N, et al. A Tutorial on Particle Filters for Online Nonlinear/non-gaussian Bayesian Tracking[J]. IEEE Transactions on Signal Processing, 2002, 50(2): 174-188.
- [32] Gillijns S, Mendoza O B, Chandrasekar J, et al. What is the Ensemble Kalman Filter and How Well Does It Work?[C]//Proceedings of the 2006 American Control Conference. Piscataway: IEEE, 2006: 4448-4453.
- [33] Geir Evensen. The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation[J]. Ocean Dynamics, 2003, 53(4): 343-367.
- [34] Djuric P M, Kotecha J H, Zhang J, et al. Particle Filtering[J]. IEEE Signal Processing Magazine, 2003, 20(5): 19-38.
- [35] Peter Jan van Leeuwen. Particle Filtering in Geophysical Systems[J]. Monthly Weather Review, 2009, 137(12): 4089-4114.
- [36] Xie X, Alexander Verbraeck. A Particle Filter-based Data Assimilation Framework for Discrete Event Simulations [J]. Simulation, 2019, 95(11): 1027-1053.
- [37] Xue Haidong, Gu Feng, Hu Xiaolin. Data Assimilation Using Sequential Monte Carlo Methods in Wildfire Spread Simulation[J]. ACM Transactions on Modeling and Computer Simulation, 2012, 22(4): 23.
- [38] Hu Xiaolin. Dynamic Data-driven Simulation: Real-time Data for Dynamic System Analysis and Prediction[M]. Singapore: World Scientific, 2023.
- [39] Hu Xiaolin, Wu Peisheng. A Data Assimilation Framework for Discrete Event Simulations[J]. ACM Transactions on Modeling and Computer Simulation,

- 2019, 29(3): 17.
- [40] Ciuffo B, Punzo V, Montanino M. The Calibration of Traffic Simulation Models. Report on the Assessment of Different Goodness of Fit Measures and Optimization Algorithms. MULTITUDE Project-COST Action TU0903: EUR25188[R]. [S.l.]: European Commission-Joint Research Centre, 2012: 1-84.
- [41] Gu Feng. Dynamic Data Driven Application System for Wildfire Spread Simulation[D]. Atlanta: Georgia State University, 2010.
- [42] Ntaimo L, Hu Xiaolin, Sun Yi. DEVS-FIRE: Towards an Integrated Simulation Environment for Surface Wildfire Spread and Containment[J]. Simulation, 2008, 84(4): 137-155.
- [43] Hu Xiaolin, Sun Yi, Ntaimo L. DEVS-FIRE: Design and Application of Formal Discrete Event Wildfire Spread and Suppression Models[J]. Simulation, 2012, 88(3): 259-279.
- [44] Wang Minghao, Hu Xiaolin. Data Assimilation in Agent Based Simulation of Smart Environments Using Particle Filters[J]. Simulation Modelling Practice and Theory, 2015, 56: 36-54.
- [45] Yan Xuefeng, Gu Feng, Hu Xiaolin, et al. Dynamic Data Driven Event Reconstruction for Traffic Simulation Using Sequential Monte Carlo Methods[C]//2013 Winter Simulations Conference (WSC). Piscataway: IEEE, 2013: 2042-2053.
- [46] Xie X, Hans van Lint, Alexander Verbraeck. A Generic Data Assimilation Framework for Vehicle Trajectory Reconstruction on Signalized Urban Arterials Using Particle Filters[J]. Transportation Research Part C: Emerging Technologies, 2018, 92: 364-391.
- [47] Wang Song, Xie Xu, Ju Rusheng. A Mesoscopic Traffic Data Assimilation Framework for Vehicle Density Estimation on Urban Traffic Networks Based on Particle Filters[J]. Entropy, 2019, 21(4): 358.
- [48] Xue Haidong, Hu Xiaolin. An Effective Proposal Distribution for Sequential Monte Carlo Methods-based Wildfire Data Assimilation[C]//2013 Winter Simulations Conference (WSC). Piscataway: IEEE, 2013: 1938-1949.
- [49] Bolic M, Djuric P M, Hong Sangjin. Resampling Algorithms and Architectures for Distributed Particle Filters[J]. IEEE Transactions on Signal Processing, 2005, 53(7): 2442-2450.
- [50] Bai Fan, Gu Feng, Hu Xiaolin, et al. Particle Routing in Distributed Particle Filters for Large-scale Spatial Temporal Systems[J]. IEEE Transactions on Parallel and Distributed Systems, 2016, 27(2): 481-493.
- [51] Al-Fuqaha A, Mohsen Guizani, Mohammadi M, et al. Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications[J]. IEEE Communications Surveys & Tutorials, 2015, 17(4): 2347-2376.
- [52] Shi Weisong, Cao Jie, Zhang Quan, et al. Edge Computing: Vision and Challenges[J]. IEEE Internet of Things Journal, 2016, 3(5): 637-646.
- [53] Shi Weisong, Schahram Dustdar. The Promise of Edge Computing[J]. Computer, 2016, 49(5): 78-81.
- [54] Mao Yuyi, You Changsheng, Zhang Jun, et al. A Survey on Mobile Edge Computing: The Communication Perspective[J]. IEEE Communications Surveys & Tutorials, 2017, 19(4): 2322-2358.
- [55] Xie X, Xu K. Distributed Dynamic Data Driven Simulations: Basic Idea and an Illustration Example[C]//2023 IEEE/ACM 27th International Symposium on Distributed Simulation and Real Time Applications (DS-RT). Piscataway: IEEE, 2023: 105-108.
- [56] Blasch E, Ravela S, Aved A. Handbook of Dynamic Data Driven Applications Systems[M]. Cham: Springer International Publishing, 2018.
- [57] Ma Yuqing, Xie Xu, Chen Hailiang. MaMiH: A New Data Assimilation Framework Based on Macro-micro Hierarchical Simulation Model[C]//Third International Conference on Advanced Algorithms and Neural Networks (AANN 2023). Bellingham: SPIE, 2023: 127912L.
- [58] Huang Yilin, Xie Xu, Cho Y, et al. Particle Filter-based Data Assimilation in Dynamic Data-driven Simulation: Sensitivity Analysis of Three Critical Experimental Conditions[J]. Simulation, 2023, 99(4): 403-415.