

# 集合模拟可视化进展\*

舒清雅<sup>1,2</sup>, 刘日晨<sup>1,2</sup>, 洪帆<sup>1,2</sup>, 张江<sup>1,2</sup>, 袁晓如<sup>1,2,3</sup>



<sup>1</sup>(机器感知与智能教育部重点实验室(北京大学),北京 100871)

<sup>2</sup>(北京大学 信息科学技术学院,北京 100871)

<sup>3</sup>(北京市虚拟仿真与可视化工程技术研究中心(北京大学),北京 100871)

通讯作者: 袁晓如, E-mail: xiaoru.yuan@pku.edu.cn

**摘要:** 近年来,集合模拟被频繁地运用于气候、数学、物理等领域。集合模拟数据通常具有多值、多变量、时变的属性,再加上其庞大的数据量,对这类数据的分析充满着挑战。集合模拟数据可视化是通过视觉和人机交互的手段,向领域专家揭示集合模拟数据中的成员差异和整体概况,从而帮助专家探索、总结和验证科学发现。从比较个体成员和概括整体成员这两个不同的分析任务,以及基于位置和基于特征这两种分析策略的角度,系统地分析了具有代表性的集合模拟可视化工作,收集并整理了各类方法的可视化形式、交互技术、应用案例。通过总结近几年的集合模拟可视化方法来讨论现有研究的趋势,并对未来研究做进一步的展望。

**关键词:** 可视化;可视分析;集合模拟可视化;大规模流场

中图法分类号: TP391

中文引用格式: 舒清雅,刘日晨,洪帆,张江,袁晓如.集合模拟可视化进展.软件学报,2018,29(2):506–523. <http://www.jos.org.cn/1000-9825/5393.htm>

英文引用格式: Shu QY, Liu RC, Hong F, Zhang J, Yuan XR. State-of-the-Art of ensemble visualization. *Ruan Jian Xue Bao/Journal of Software*, 2018, 29(2):506–523 (in Chinese). <http://www.jos.org.cn/1000-9825/5393.htm>

## State-of-the-Art of Ensemble Visualization

SHU Qing-Ya<sup>1,2</sup>, LIU Ri-Chen<sup>1,2</sup>, HONG Fan<sup>1,2</sup>, ZHANG Jiang<sup>1,2</sup>, YUAN Xiao-Ru<sup>1,2,3</sup>

<sup>1</sup>(Key Laboratory of Machine Perception (Peking University), Ministry of Education, Beijing 100871, China)

<sup>2</sup>(School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China)

<sup>3</sup>(Beijing Engineering Technology Research Center of Virtual Simulation and Visualization (Peking University), Beijing 100871, China)

**Abstract:** Ensemble simulation is increasingly popular in scientific domain such as climate research, weather report, mathematics and physics. Ensemble simulation data sets are usually multi-valued, multi-variate, time-variant and large in scale. Thus, analyzing such data sets is challenging. Ensemble visualization helps scientists to compare ensemble members and give overall summary to the whole data sets by utilizing visual encoding and human interaction. It thus helps scientists to explore, conclude and validate their findings. This article describes analytical tasks and strategies for organizing existing works on visualization and visual analysis on ensemble simulation data sets. The analytical tasks for ensemble simulation data sets include comparing individual members and summarizing whole ensemble, whereas the analytical strategies consist of location-based method and feature-based method. This article reviews major works in ensemble visualization. It gives explanation to their visual design, interaction approaches and application scenarios, along with a discussion of recent trends and future research directions.

\* 基金项目: 国家自然科学基金(61672055, 61702271); 国家重点基础研究发展计划(973)(2015CB352503); 国家重点研发计划(2016QY02D0304)

Foundation item: National Natural Science Foundation of China (61672055, 61702271); National Program on Key Basic Research Project of China (973) (2015CB352503); National Key Research and Development Program of China (2016QY02D0304)

收稿时间: 2016-10-15; 修改时间: 2017-01-22, 2017-05-11, 2017-07-14; 采用时间: 2017-07-18; jos 在线出版时间: 2017-10-09

CNKI 网络优先出版: 2017-10-09 16:20:57, <http://kns.cnki.net/kcms/detail/11.2560.TP.20171009.1620.008.html>

**Key words:** visualization; visual analysis; ensemble visualization; large-scale flow field

近年来,随着计算能力和存储技术的提高,集合模拟方法在气象、海洋、环境、物理等科学领域中的应用越来越普遍。集合模拟是指通过多种不同的参数和输入来重复模拟<sup>[1]</sup>,主要是针对具体的物理过程或自然现象进行模拟。其目的包括评估和改善模型稳定性(例如气候模型、气象模型的不确定性<sup>[2,3]</sup>)、研究模型参数影响(例如研究污染排放对全球气候影响<sup>[4]</sup>、环境参数对洪水灾难模拟的影响<sup>[5]</sup>)等。一个典型的例子是气候领域科学家用集合模拟来研究臭氧浓度的来源<sup>[6]</sup>。全球不同区域的人为污染排放不同程度地影响了全球臭氧浓度,科学家通过多组对照模拟实验来比较影响的相互大小,实验过程中,每组模拟关闭了某个地域的人为污染排放。对比每组模拟结果生成的臭氧浓度的差异,就可以帮助科学家总结地域污染对全球臭氧分布的影响。在这个例子中,模拟过程在不同参数和初始条件下重复多次( $N$  次),每次的模拟结果输出都是一个时序、三维、多变量的大规模气候数据(大多情况下,包括温度、压力、风向等变量),称为一个模拟成员。因此,整个集合模拟数据集包含了  $N$  个模拟成员,同一个时空点( $x, t$ )会包含同一种变量在不同模拟下的数值( $N$  个模拟成员对应  $N$  个数值)。总结来说,集合模拟数据同时具有多值、多变量、时变的特征且体量庞大。而这种复杂的特征和庞大的数据量为科学家的分析带来前所未有的挑战。

## 1 集合模拟数据分析与可视化

科学家主要从两种不同角度分析集合模拟数据:单个模拟成员的角度和整体模拟数据的角度。首先第一种,从单个模拟成员出发的角度,目的是比较模拟成员之间的相似与差异,由此挖掘出模拟结果与参数之间的潜在关系。针对这类分析目的的比较手段包括:比较不同成员在相同时刻的空间分布差异、比较不同成员在相同空间内的时序变化差异、比较在相同的时空范围内各成员的数值差异。在可视化上,常用的对比方法包括:将各模拟成员的可视化结果并列在一起比较、将它们重叠在一起比较等。从单个模拟成员出发的角度对数据做分析,其特点在于分析的对象明确,即集合模拟中单个或者多个模拟成员。然而,这个方法也存在局限。人的注意力和理解能力有限,并且屏幕空间不能同时容纳过多信息。因此,就成员数目而言,针对单个成员出发的分析方法不能同时支持超大数量的对比。针对超大规模集合模拟数据的分析,往往需要丰富的可视化交互手段作为辅助。

对集合模拟数据的第 2 种分析角度是从整体模拟出发,通过概括式地描述整个集合模拟数据来评估模型的稳定性。从整体模拟出发,会忽略具体单个模拟成员的个体差异,而重点关注整个模拟结果数据。在该模拟数据中,每个空间点上包含多个输出值。科学家分析多个数值的分布,描述模拟数据的不确定性信息,可以得出感兴趣的结论,诸如集合模拟中哪些地点、时间上模型输出表现得更稳定等。不确定性信息的描述方法可以是使用统计量,也可以是面向特定应用的定义方法,如基于拉格朗日方法描述流场的不确定性、用几何形状来描述模拟成员总体特征信息等。这些从整体的角度衡量集合模拟的稳定性,更贴合科学家的特定分析任务。从整体模拟出发的概括式分析,优点是能够面向成百上千的超大规模集合模拟数据:用统计、概括等方法解决了数据规模带来的难题,同时也保留了重要的信息。然而,这类方法的缺点是牺牲了单个成员的信息,科学家只能分析参数对集合模拟整体分布的影响,却不能进一步探索具体模拟成员的特征。

以上两种集合模拟数据的分析角度成为集合模拟数据可视化的动力。科学可视化用视觉和交互的手段,使科学领域研究的成果更直观地展现给领域专家,帮助专家探索、验证、总结重要发现。最初在集合模拟可视化概念未出现时,为了分析气候领域、流体力学领域中的观测数据和模拟数据,科学家们用地图、体渲染、折线图的方法探究空间、时间的分布特征<sup>[1]</sup>,而自从集合模拟愈发成为重要的实验手段后,学者们发现已有的可视化方法已逐渐无法满足这种特殊数据类型的需求<sup>[7]</sup>,它在同一空间点上有多组输出,科学家们有时对模拟输出的整体分布感兴趣,有时需要比较单个模拟结果之间的具体差别,有时又要分析模拟参数对结果的影响。于是,这种独特的数据类型和多样的分析需求推动了集合模拟可视化的诞生与发展。

集合模拟可视化研究可以追溯至上世纪 90 年代,最初脱胎于不确定性可视化<sup>[8]</sup>,之后又逐渐融入比较可视化、高维空间可视化等技术,再加上丰富的交互手段等。如今,随着计算机存储与运算能力的飞速发展,集合模拟

越来越成为科学领域中炙手可热的研究工具,集合模拟数据可视化也已成为可视化领域当下的研究热点之一。图 1 总结了近年来集合模拟可视化相关文章的数量变化趋势。我们统计了以 IEEE TVCG 等期刊为主的论文数量。从图中可以看出:这一方向的研究工作从 2000 年开始出现,自 2009 年起数量越来越多,意味着集合模拟可视化的研究近几年蓬勃发展之势。

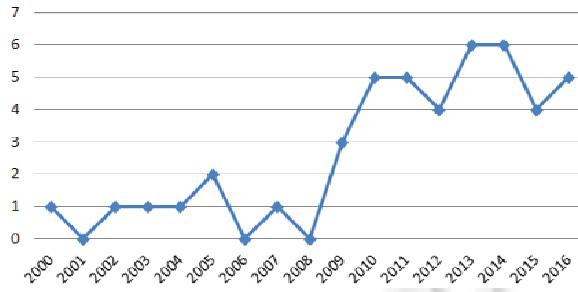


Fig.1 Number of papers in ensemble visualization recent years

图 1 近年来集合模拟可视化相关文章数量趋势

## 2 集合模拟可视化流程与分类

集合模拟可视化流程包括两步:第 1 步预处理,第 2 步可视化。第 1 步预处理,先对集合模拟数据做预处理,往往是按照不同角度对数据做聚集和抽象,比如从时间、空间、模拟成员的维度聚集数据,或者抽取各个成员的特征。第 2 步,按照不同分析任务,把比较可视化或不确定性可视化运用到集合模拟数据上。其中,比较可视化用于分析单个模拟成员,不确定性可视化用来概括整体模拟数据。两者的区别主要在于分析的关注点不同:比较可视化关注于模拟成员的差异(比如不同的污染排放条件下各个模拟之间的区别<sup>[4]</sup>)、不确定性可视化则突出集合模拟作为一个整体所具有的属性(比如土壤污染模拟中哪些区域的预测数据有更高的准确度<sup>[9,10]</sup>)。由于目标不同,比较可视化与不确定性可视化有各自侧重的技术和分析手段,我们会在文章中详细介绍。经过这两步后,科学家可以得到关于集合模拟数据的不确定性、差异分布等观察和结论。

集合模拟可视化流程中关于数据预处理和可视化的方法有多种,它们由分析的目的来决定。科学家面对如此复杂的数据,需要选择一个角度切入以作分析:是针对固定时空范围内成员之间的差异?还是针对每个成员特征属性的差异?从这两种切入角度出发,我们可以把集合模拟数据的分析策略分为两种:按照时空位置的角度、按照特征的角度。采用不同策略将决定数据预处理方法与可视化方法:第 1 种策略按每一个数据点进行聚集并以聚集结果作为可视化对象,第 2 种策略从单独的集合成员中抽取特征并以特征作为可视化对象。

按照集合模拟可视化的分析角度与分析策略,我们对相关研究分类进行综述。近年来,针对集合模拟数据可视化的研究越来越多,但整理这类研究的系统且完整的综述却尚未出现。不久前,Obermaier 等人撰文总结了集合模拟可视化研究的难点与发展方向<sup>[11]</sup>,他们提出按照“基于位置的”和“基于特征的”两种角度来划分集合模拟数据可视化,同时用代表性的 5 篇工作作为实例,以说明这一分类的适用性。“基于位置的”策略关注空间点上集合模拟成员的不同,“基于特征的”策略关注每个集合模拟成员特征之间的不同。本文在 Obermaier 等人提出的两种分析策略的基础上,引入了按照不同分析任务为出发点的另一划分维度,亦即本文第 1 节所述的“针对独立成员的比较可视化”与“概括所有成员的不确定性可视化”。本文综合分析的任务和分析的策略,从这两个维度来综述集合模拟可视化的相关研究工作。表 1 是分类结构示意图,本文按照此结构系统地整理了大多数具有代表性的集合模拟可视化工作,包括各类方法的可视化形式、交互技术以及应用案例。

**Table 1** An illustration of ensemble visualization classification**表 1** 集合模拟可视化分类结构示意图

		分析任务	
		比较成员个体	概括全体模拟
分析策略	基于位置	基于位置的比较可视化, 针对大量模拟成员的方法 <sup>[12-16]</sup> , 分析成员差异位置和大小的方法 <sup>[17-20]</sup>	基于位置的不确定性可视化, 基于单一参数 <sup>[9,10,21-24]</sup> , 多个参数 <sup>[9,10,23,25-27]</sup> , 分布函数 <sup>[12,28-30]</sup> 的不确定性描述, 矢量场集合模拟可视化 <sup>[8,14,31,33]</sup>
	基于特征	基于特征的比较可视化 <sup>[1,18,34,35]</sup>	基于特征的不确定性可视化 <sup>[36-39]</sup>

### 3 分析模拟成员个体——集合模拟数据的比较可视化

科学家通过比较不同模拟成员来发现模拟结果之间的相似性与差异性,并由此挖掘模拟输出与参数之间的潜在关系。集合模拟数据中,每一个模拟成员都是一个完整的标量场或矢量场,且大多带有时间属性。为了对多个复杂数据场进行比较,我们首先需要根据科学家的分析需求定义差异。这些差异包括:相同位置上数值差异的空间分布、相同位置上数据时序变化的差异、感兴趣区域的统计量差异、按照同一条件提取不同成员中特征之间的形状差异等等。由于不同分析需求所侧重研究的差异对象不同,所以对这些差异的比较手段也是丰富多样的。比如,为了比较多个时间序列的差别,可以显式地用曲线画出时间序列的差值变化;为了比较多个等值面,可以将它们重叠置放于同一空间中,允许用户交互式地拖动两个曲面以观察形状差别。因此,针对集合模拟数据的比较可视化有不同的比较对象和比较手段。

在梳理具体相关工作前,有必要先浏览一下常用的比较可视化方法。比较可视化根据对象的不同组织方式可分为并列(juxtaposition)、重叠(super position)和显式编码(explicit encoding)这3种手段<sup>[40]</sup>。其中,并列方法把所有比较对象依次排开;重叠方法把它们叠放在同一位置便于比对;显式编码方法先计算差异,再在可视化结果中展示差异。这3种方法分别利用人脑记忆、空间位置和计算来把用于比较的对象联系在一起。3种方法各有优点和缺陷,相互之间也可以用结合的方式取长补短。例如,在使用并列方法的同时,用显式编码方法把相同或差别的部分高亮出来,让人更容易在多个并列的对象中找到差异部分。此外,一些特殊的展现方式和辅助交互手段可以作为辅助,比如用纹理映射、序列动画的方法、交互挪动比对手段等。

按照比较对象的不同,集合模拟数据的比较可视化可以分为基于位置的和基于特征的比较可视化<sup>[11]</sup>,前者关注空间点上集合模拟成员的数值差异,后者关注集合模拟成员特征之间的差异。两者的区别可以理解成前者是对直接属性的比较;后者把特征作为成员的衍生属性来比较,加入了更多特征提取和计算的预处理步骤。

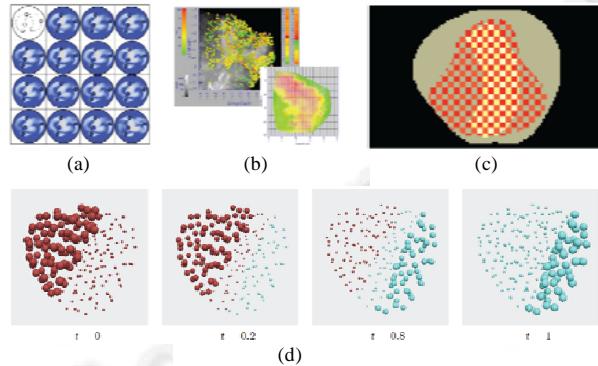
#### 3.1 基于位置的集合模拟比较可视化

基于位置的比较是指研究不同成员在相同空间位置上的差异。最直接的做法就是并列地展示所有成员,这可以让科学家以最快、最熟悉的方式看到成员之间的区别(如图2(a)所示)。但是这个方法有3个弊端:其一,仅仅依靠并列比对的方式不容易定位差异的位置和大小;其二,并列的方法一次只能同时比对两个结果,不适合成员数很多的情形;其三,若加入时间维度,则包含的信息量将变得更加庞大,使对比的难度增大。现有的工作提出了一系列的方法来解决或部分地解决这些问题。在这些方法中,有些是对可视化的组件进行新的设计来避免信息过剩的问题,有些通过牺牲一部分信息来强调另一部分信息,还有一些则花费更多的计算资源对信息先做凝练再可视化。下面我们进行具体介绍。

- 关于差异的定位

其目的是让用户更容易定位到两个成员之间差异的位置和具体大小。集合模拟数据中,模拟成员在大部分区域的大多数数据值都是非常接近的,只在局部的区域有细微的差别,而并列的方法又需要人们依靠短时的记忆和来回切换关注点来找到不同之处,所以精确地定义差异是非常困难的。一种解决方法是用重叠代替并列方法,例如把两个地图用百叶窗叠加(screen door tinting)的形式镶嵌在一起<sup>[19]</sup>(如图2(c)所示),或者用镶嵌在一起的

染色的小正方形,其中上半部分颜色表示所有成员的综合值,下半部分是当前成员对应的颜色<sup>[18]</sup>(如图 2(b)左上所示).对于大于两个成员的情形,则可以用更丰富的形状比如六角形来表示<sup>[17]</sup>.Zhang 等人<sup>[20]</sup>设计了一组三维图标,用棋盘格的方式把两个图标镶嵌在一起,表现出张量场在形状、方向和大小上的区别.重叠的地图,让用户不需要记住所有成员就可以在同一个位置比较多个成员.另一种思路是用动画(sequential animation)和闪烁(flickering)呈现多个模拟成员的可视化结果,以达到比较的目的.例如,Phadke 等人的同一篇工作中,用序列动画呈现模拟成员的空间分布<sup>[19]</sup>(如图 2(d)所示);在 Hao 等人<sup>[9]</sup>针对时序模拟数据的工作中,用空间中闪烁图标的方法依次展示成员或成员集合在三维空间中的数据分布.还有一种解决方法是直接计算成员差异,用颜色和透明度映射的方法绘制差异的地图<sup>[18]</sup>(如图 2(b)右下图所示)和体渲染结果<sup>[31]</sup>.



(a) 并列展示<sup>[11]</sup> (b) 基于颜色、透明度的差异地图<sup>[18]</sup> (c) 百叶窗叠加形式镶嵌<sup>[19]</sup> (d) 序列动画<sup>[19]</sup>

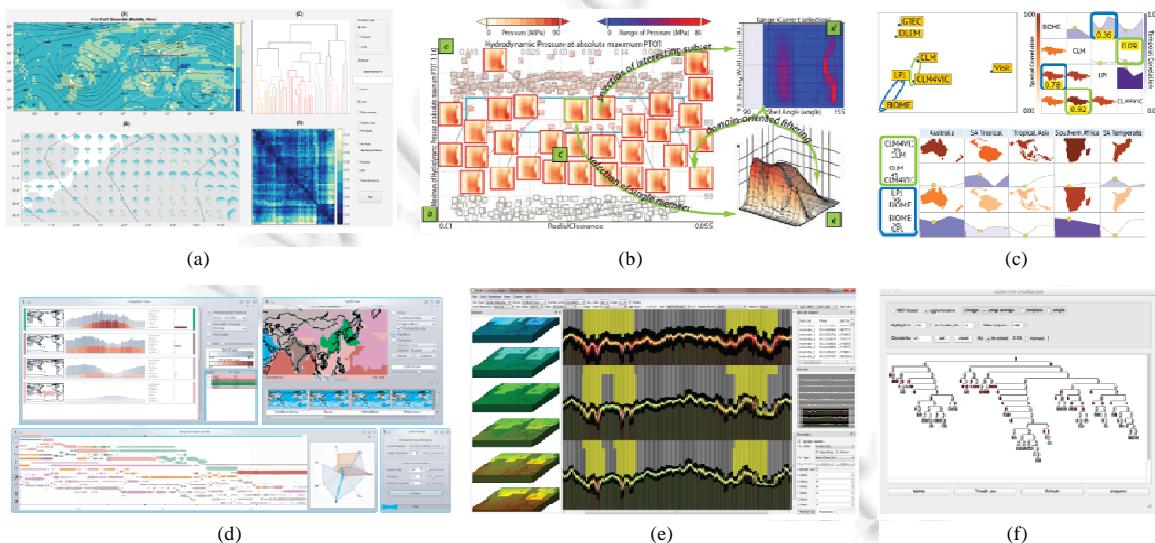
Fig.2 Location-Based comparative ensemble visualization helps locate difference among ensemble members  
图 2 基于位置的集合模拟数据的比较可视化,用于帮助发现成员之间差异位置与大小

- 关于成员个数的弊端

重叠和显式展示差别的方法虽然可以帮助用户更快地发现差异所在,但是又会遇到成员个数增多带来的问题.也就是说,用户只能一次比较有限个数的几个成员,当面对几十甚至几百个成员时,比较还是很困难的.一种解决方法是提取成员之间的相似性关系,将这些信息编码到整体的可视化布局中.比较典型的例子是按照每个模拟成员的特征属性来布置成员投影在二维空间上的位置,可以让用户很快从数量庞大的二维函数模拟结果中分析他们之间的差别<sup>[15]</sup>(如图 3(b)所示).此外,矩阵形式可以把两两模拟成员之间的相似性全部罗列出来<sup>[16]</sup>(如图 3(c)右上所示),投影的方法可以表示所有成员的相似性<sup>[16]</sup>(如图 3(c)左上图所示),树形图(dendrogram)的方式更适合展现成员之间相似性的层次结构<sup>[13]</sup>(如图 3(f)所示).Shu 等人<sup>[4]</sup>结合了信息可视化的技术,利用图可视化方法来概括地展示高相似区域在时空中的分布.用户在界面的图可视化结果中首先获取整个集合模拟数据的高相似性区域的分布,之后,通过多视图的交互手段以及比较可视化的办法来对感兴趣的子区域做具体分析(如图 3(d)所示).他们的方法被应用于气候模拟干扰实验数据上,能够帮助科学家发现不同污染情形对气候结果的影响.Jarema 等人<sup>[14]</sup>结合了矩阵和树形图两种可视化方法,用于集合模拟数据中多个模拟成员之间的相似性大小和层次结构.针对每个空间点上成员的分布,他们用高斯混合模型(Gaussian mixutre model,简称 GMM)拟合数据分布,再用染色地图(颜色对应分布模式的个数)和表示分布模式的图标展示拟合结果的具体空间分布(如图 3(a)所示).而类似于平行坐标与直方图结合的可视化方式,则可以画出每个成员在每个空间范围内的数值来比较局部的相似性<sup>[12]</sup>(如图 3(e)所示).具体来讲,每一列代表一个空间区域,上面可以画出这个区域内所有成员的分布,每一根折线都对应着一个成员,折线在每一列上的位置代表成员的数值.另一种使用平行坐标的应用<sup>[41]</sup>与这种方法相反,它用轴代表模拟成员,折线代表区域,更侧重于比较不同区域在同一成员上的差别关系.

- 关于对时序数据的支持

上述方法虽然增强了对大量成员个数的支持,但大多不支持时序的集合模拟数据.Poco 等人<sup>[16]</sup>的工作可以支持时序数据的情形.他们利用矩阵可视化在每一个矩阵元素中用折线图来表示时序变化,但也仅限于对较少数量的时间片的集合模拟数据的支持.一些针对时序的集合模拟数据的研究直接把时间信息彻底压缩成整个集合模拟过程的累加值.例如,Hao 等人在针对时序的原子模拟可视化<sup>[13]</sup>中,将时间因素引入到模拟成员之间距离的定义中.该定义有两种方式:第 1 种方法是对于两个成员,用所有时间片上的距离之和表示整个时间范围内的差别;第 2 种方法为了减少计算量,先对每一个成员做相邻时间的聚集,再用动态时间规整(dynamic time warping)的方法计算两两成员之间的距离,用该距离作为相似性的依据进行聚类,再以此计算成员的层次性结构.需要注意的是,如果不做时间维度方向的聚集,而是同时从成员和时间两个维度罗列出所有的地图,成员个数和时间片数的增长都会给用户增加比较的难度.我们可以用最直观的时间序列(time series)方法罗列出所有模拟成员在某个地点的数值变化趋势,但这需要牺牲一部分空间信息,将每个单独成员在空间聚集成一个数值或一种分布,然后再进行展示.相关的工作包括针对气象集合模拟的可视化框架 Noodles<sup>[23]</sup>,EnsembleVis<sup>[1]</sup>和 SimilarityExplorer<sup>[16]</sup>.此外,也可以同时从时间和空间两个角度做聚集的工作,避免在某个维度上过多的信息损失.例如,Bruckner 等人设计的可视化系统<sup>[42]</sup>把模拟动画序列结果按照相似性在时空上聚集.在该系统中,时间上相邻、空间上结果相似的序列可以合并,从而聚集成一个层次的结构,称为聚集的时间线(clustering timeline).用户若对某一组动画效果感兴趣,可以交互地展开分析动画的细节和具体的参数设计.



(a) 结合地图、图标、树形图、矩阵分析模拟成员相似性<sup>[14]</sup> (b) 按照特征属性投影在二维空间上展示二维函数模拟成员相似性<sup>[15]</sup>  
 (c) 利用矩阵、投影展示两两模拟成员之间的相似性<sup>[16]</sup> (d) 利用图可视化方法总结高相似区域在时空中的分布<sup>[4]</sup>  
 (e) 结合了平行坐标与直方图形式的多重图表<sup>[12]</sup> (f) 利用树形图展现成员之间相似性的层次结构<sup>[13]</sup>

Fig.3 Location-Based comparative ensemble visualization

图 3 基于位置的集合模拟数据比较可视化

总的来说,基于位置的模拟成员比较从最初的并列所有成员可视化结果开始,慢慢地加入了重叠和差异绘制的方法,之后又融入了更多丰富的手段,比如投影、矩阵、平行坐标,再到最后越来越多的计算处理方法,比如相似度聚类等,形形色色的技术加入进来,其最终目的都是为了展现不同成员在相同空间位置上的差异.

### 3.2 基于特征的集合模拟比较可视化

基于特征的方法比较各成员的特征,主要关注于如何呈现几何形状的不同,比如各个模拟成员的等值线、等值面的差别.最直接的做法是把所有成员的等高线或者流线同时在一个地图上画出,用不同色彩表示不同成

员,这样,用户可以比较每一个成员的等高线的区别.这种可视化方法称为意大利面图(spaghetti plot),被广泛地运用在集合模拟可视化框架中,尤其是针对气候集合模拟的数据<sup>[1]</sup>(如图 4(a)所示).Dasilva 等人<sup>[35]</sup>研究流场差异可视化,将其应用于磁共振扩散张量成像(diffusion tensor magnetic resonance imaging,简称 DI-MRI)数据上.他们的方法可以展示同一个流场数据在不同可视化算法下的流管(stream tube)可视化结果的差异.他们把两组可视化结果放在同一个平面中,用带颜色的流管表示表示两组结果的差异大小(如图 4(d)所示).渲染流管时,用颜色表示两组数据之间的差异正负与大小.这样的形式可以让用户更快注意到哪些位置具有更大的差异,哪些位置上两组结果非常相似.Nocke 等人<sup>[18]</sup>的研究允许用户交互式地拖动单个模拟成员的等值面,使其与另一个等值面重合,以此观察它们之间的形状差别(如图 4(b)所示).等值面的形状差别定义成点对点上的高度差,用不同颜色分别表示正负的差别,白色表示没有差别,并且给用户指示差异的具体分布位置和大小.Alabi 等人<sup>[34]</sup>在比较多个天气集合模拟数据时,将所有成员的等值曲面用切片的方式依次间隔摆放在一个三维空间中,可以方便比较不同成员等值面的差别(如图 4(c)所示).Liu 等人提出了一种新颖的基于手势(sketch-based)的特征提取方法<sup>[43]</sup>,允许用户在集合模拟数据中自行定义感兴趣的特征,比如集合流场数据中每一个成员的涡线.自定义的成员特征会被提取出来并被高亮在可视化界面中,然后再用信息可视化的方法(比如星形图、直方图等)展示各自的时间分布、形状差别和属性值.

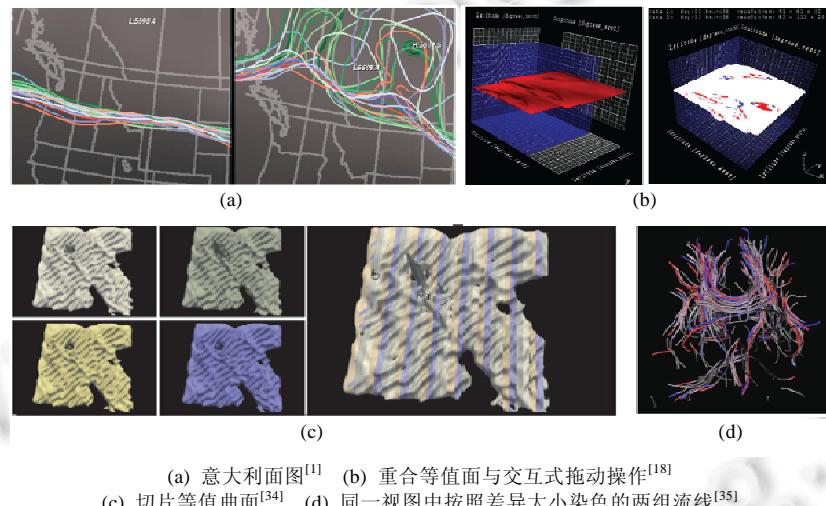


Fig.4 Feature-Based comparative ensemble visualization

图 4 基于特征的集合模拟数据比较可视化

以上都是针对集合模拟数据中的成员特征的比较手法,其特点是都基于成员各自的属性(大多是等值面、等高线).此外,它们都是从同时展示全部的特征开始,到重叠、镶嵌、显示地进行颜色编码,共同目的是展示成员各自特征的差异.然而,这类出于比较目的的可视化方法大多并不支持庞大数量成员的情况,这是由于屏幕的有限空间只能容纳少数的图形.并且,想要同时衡量数量众多的几何图形的差异不容易.因此,若是和基于位置的方法一样去计算数值差异再套用现成的可视化方法,并不能适用于基于特征的类型.由此可见,今后在针对集合模拟可视化的研究中,针对大规模成员的特征可视化研究仍然具有很大的研究空间.

基于位置和基于特征的分类分别只是对集合模拟的比较可视化分类方法中的一种,每一种分类角度都有其优点和不足.另一个值得一提的分类角度是基于比较成员个数的方法.Nocke 等人<sup>[18]</sup>对基于气候的集合模拟数据的概述和分类给我们带来了这方面的思考.他们将分析任务按照比较成员的个数划分成 3 类:针对一个、两个以及多个集合模拟成员之间的比较.针对这些不同类型的比较任务,他们总结了不同的视觉表示方法.

- 第 1 类比较任务,亦即针对一个集合模拟成员的可视化,属于时序数据可视化与多变量数据可视化的一

范畴,比如用并列的切片表示时间演化、用基于图标的方法表示多变量分布,等等,这里暂且略过不表;

- 第 1 类比较任务,亦即比较两个成员的任务.不同的数据有不同的方法:对于标量值数据,可以用基于图标或染色地图的方法;对于形状数据,可以用颜色来表示两个曲面的形状差异,或者让用户交互式地拖动两个等值面,通过观察它们的重合与缝隙来找到差异.
- 第 3 类比较任务,即针对多个模拟成员的比较,将涉及到庞大的数据量.为此,Nocke 等人提倡遵循 Shneiderman 提出的“Overview first, zoom and filter, then details on demand”<sup>[44]</sup>这一可视化策略.他们首先提供所有模拟成员的概览,比如用图形表格的方法总结各个变量的分布情况;之后,尽量利用交互手段来重点高亮出模拟数据的某一子集来展示感兴趣的细节.例如,一旦限定了感兴趣的参数和变量,涉及到的维度个数就会降低,于是,就可以采用平行坐标的方法来展示数据分布,再辅助一定的高亮交互,让用户做细节数据探索.

虽然只是简单地按照成员个数来划分,但这种思路可以给我们很多启示:首先,集合模拟数据的比较可视化是单个成员可视化的拓展,所以将单个成员数据的可视化拓展到多个成员也是可行的方法;其次,集合模拟可视化的关键仍然在于如何比较,所以重点研究如何灵活、有效地对比两个成员尤其重要.此外,另一方面的启示是 Nocke 等人考虑到的大规模集合模拟数据可视分析成为集合模拟研究的必然性.在比较的数目上,针对单个成员的比较方法使得这一类分析任务不能同时支持超大数量成员之间的对比.因为人的注意力和理解能力有限,而屏幕空间也不能同时容纳太多信息,所以针对超大规模集合模拟数据的分析往往需要更加丰富的可视化交互手段作为辅助,这将成为集合模拟数据可视化研究的未来趋势之一.

#### 4 概括全体模拟成员——集合模拟数据的不确定性可视化

除了分析单独成员的差异以外,科学家的另一个重要分析目的是概括全体成员的整体分布.从整体上概括集合模拟,可以帮助科学家评估参数稳定性和模型的敏感度.在集合模拟数据中,每一个时空点上都包含同一个变量在不同的模拟条件下产生的多组输出,如不同模拟条件下生成相应个数的风速场、温度场等.科学家这样做的目的是想知道这些输出在不同区域的分布特征,例如,天气预报输出在哪些区域更加可靠、哪些区域更不确定(更模糊)等.相对于分析单独成员差异,分析整体集合模拟的特点是从全体模拟成员出发,做整体概括性描述.然而,这一类方法往往忽略单独成员的先后顺序,这将导致忽略个体成员的独特信息.

为了描述全体模拟成员的整体分布,需要用到不确定性可视化的技术.不确定性可视化的定义、计算和视觉编码的方式展现了集合模拟数据中的不确定性大小.这里,我们分基于位置的方法和基于特征的方法两类依次加以介绍.基于位置的不确定性可视化面向每个空间点上多个数值的分布,而基于特征的方法针对单独成员内部的特征所构成的集合的分析.相对于基于位置,基于特征的方法需对衍生的特征属性做概括,比如几何形状、流体力学范畴下的时变特性等.

##### 4.1 基于位置的集合模拟不确定性可视化

在集合模拟数据可视化领域中,基于空间位置的不确定性可视化研究最受关注.因为集合模拟数据可视化的最初核心任务就是展示空间点上多个数据输出的具体分布.集合模拟可视化发展的历程不到 20 年,在早期集合模拟手段尚未普及的年代,同一个空间点上出现多组数值的数据类型是非常新颖的.虽然在多变量科学数据的可视化研究中,研究的数据也在同一空间点上具有多个数值,但多变量数据与集合模拟数据存在着本质的不同:首先,多变量数据的变量个数较少,但集合模拟数据的成员个数往往很大;其次,多变量数据可视化关注于变量之间的相关性,而集合模拟数据可视化关注成员的相似性与分布.此外,更有一些集合模拟数据自身也是多变量的.正因为这些不同,以往单变量、多变量科学数据可视化技术并不能完全适用于这种新颖的数据类型.于是,可视化领域渐渐开始研究集合模拟这种特殊的数据类型.2003 年,Luo 等人<sup>[7]</sup>针对这类数据的可视化研究做了总结.他们用“空间多值数据(spatial multivalue data)”来命名这一类型的数据.他们认为:这种新型数据相对于普通的科学数据,在三维、时序的维度上又增加了多值这一特性,对该数据的研究难点在于如何将空间中的多值信息有效地展现在二维平面中.他们的研究是集合模拟可视化领域首次对集合模拟数据以及相关可视化研究

的总结和展望。

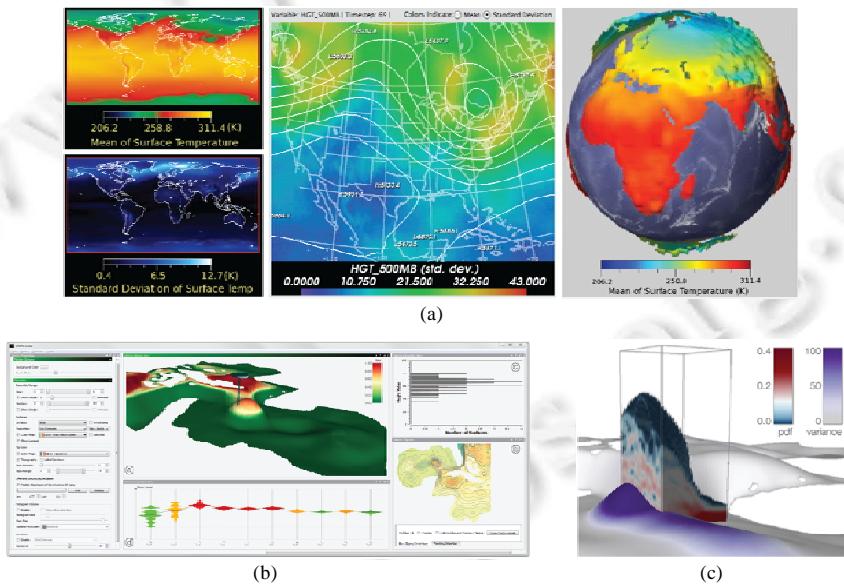
集合模拟数据的空间上不确定性分析的描述方法可以分为两种方式:基于参数的描述方法和基于分布函数的描述方法。基于参数的方法先把多值数据映射成 1 个或多个数值(可以是均值、方差、四分位数、分布函数的峰个数等统计量),然后再可视化。这种方法能够使已有的标量场可视化适用于集合模拟标量场数据。其中,单个参数的描述方法可以直接套用现成的标量场可视化方法。而对于多个参数的描述方法,将多个可视化叠加或者重新设计特定的可视化方法可以解决这个问题。用分布函数的描述方法是将分布视作一种新的数据类型,用特殊的数据存储方法和可视化方法来处理、展示数据的特征。

#### 4.1.1 基于参数的不确定性描述

在集合模拟数据可视化中,早期的常用方法是将集合模拟数据每个时空点上多组数值的分布用 1 个或多个参数来描述。在本节中,我们首先介绍该方法如何应用于标量场的集合模拟数据,分为单一参数和多个参数的情形;然后再针对矢量场的集合模拟数据情形做介绍。

##### (1) 单一参数

很多集合模拟可视化用简单的统计量来描述数据的不确定性信息,这种方法的好处是,可以使以往的科学可视化方法适用于集合模拟数据。例如,最基础的可视化方法包括着色地图(color map)。着色地图用颜色对应集合模拟数据中的均值、标准差大小,这些是科学家对集合模拟数据中最感兴趣的统计特征。Sanyal 等人著名的集合模拟数据分析系统 Noodles<sup>[23]</sup>和同时期 Potter 等人的另一个同样著名的框架 Ensemble-Vis<sup>[1]</sup>,其原型 ViSUS/CDAT<sup>[22]</sup>都用到了着色地图(如图 5(a)所示)。着色地图适用于二维的情形,三维集合模拟数据可以用着色等值面(color iso-surface)、着色曲面(color surface)等进行展示。在 Höllt 等人提出的针对海洋预报分析近海结构(off-shore structure)的系统 Ovis<sup>[9,10]</sup>中,用着色曲面表示当前时刻所有模拟成员的海面高度平均值,曲面上的颜色表示该点上数值的方差或标准差(如图 5(b)所示)。



(a) 着色地图<sup>[22]</sup> (b) 着色曲面,高度是平均值,颜色是标准差<sup>[9,10]</sup> (c) 互动式探索区域内数据分布<sup>[9,10]</sup>

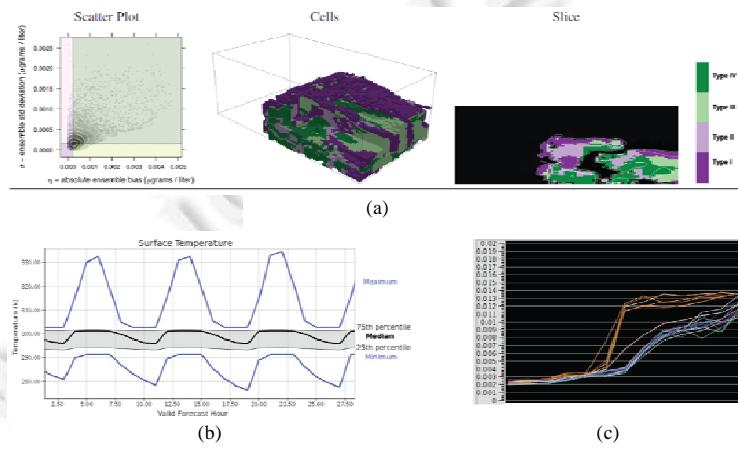
Fig.5 Location-Based uncertainty visualization for ensemble simulation data,  
uncertainty information described by parameter (I)

图 5 基于位置的集合模拟数据不确定性可视化,用参数描述不确定性信息(I)

除了利用统计参数来描述不确定性信息以外,也有一些特殊的不确定性测度满足特殊的应用需求。例如,

Gosink 等人<sup>[21]</sup>结合模拟数据和真实观测数据,定义了预测不确定性(predictive uncertainty)指标,以此表示模拟成员能否一致且准确地预测真实数据的能力(如图 6(a)所示).他们的具体方法是:用贝叶斯模型平均(Bayesian model averaging)分析方法评估了三维集合模拟数据中每个顶点上数值的不确定性,并照此把模拟空间中的点分类:一致且准确、一致但不准确、不一致但准确和不一致且不准确.他们将这种方法应用于地下水四氯化碳浓度减退模拟、土壤污染治理模拟数据上,然后,按照这个分类将集合模拟空间染成 4 种颜色,能够将不同性质、不确定性的区域高亮出来以供分析.此外,还有 Ovis 系统<sup>[9,10]</sup>,用于分析海平面高度模拟数据,并计算出科学家感兴趣的危险预计值(risk estimation),用它对曲面染色.同时,在曲面周围的小范围区域内,直接用体渲染技术(volume rendering)展现该处概率分布函数的大小,并允许用户互动式地探索某一块小区域内的分布(如图 5(c)所示),这样可以针对关键的地区分析,同时避免了三维可视化过程中过多的遮挡问题.

相对于用地图或体渲染方式展示不确定性在空间的分布,要展示不确定性随时间演化的可视化方式则相对受到屏幕空间的限制.一个较为直接的解决方案是 Ensemble-Vis 系统中的方法,并列或者逐帧展示每个时刻的不确定性地图<sup>[1]</sup>,或者通过限定空间范围,用四分位图(quartile)<sup>[23]</sup>(如图 6(b)所示)或者折线图(line chart)<sup>[1,24]</sup>方法(如图 6(c)所示)表示某个区域内所有模拟成员的平均值和方差.



(a) 预测不确定性<sup>[21]</sup> (b),(c) 四分位图<sup>[23]</sup>、折线图<sup>[1,24]</sup>表示不确定性随时间变化

Fig.6 Location-Based uncertainty visualization for ensemble simulation data, uncertainty information described by parameter (II)

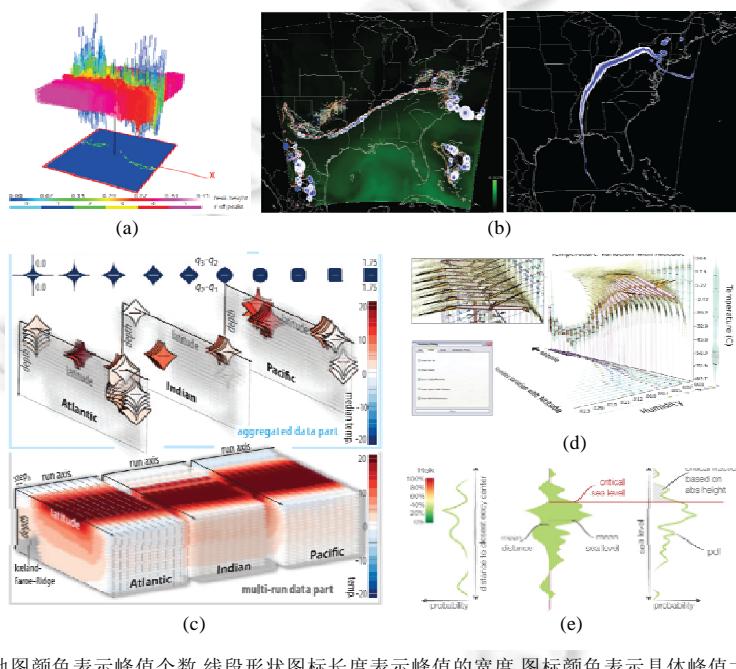
图 6 基于位置的集合模拟数据不确定性可视化,用参数描述不确定性信息(II)

## (2) 多个参数

使用单一参数来描述不确定性信息的方法可以直接套用现有的标量场可视化方法,而使用多个参数的方法则需要对视觉编码方式做一些特殊的设计.一种方法是在同一个可视化窗口中集成多种可视化形式.例如, Kao 等人<sup>[25,45]</sup>将集合模拟成员的数据值分布转化成多个参数,包括分布函数的峰的个数、峰的宽度等,然后,用基于图标(glyph-based)的可视化结合着色地图来展现(如图 7(a)所示).在这个融合的可视化形式中,地图的颜色表示峰值个数,上方线段形状图标的长度表示峰值的宽度,图标颜色表示具体峰值大小.另一个集成多种可视化方法的例子是通过嵌入精心设计的视觉小元素来传达不确定性信息.例如,气象模拟分析系统 Noodles<sup>[23]</sup>对空间点上的多组扰动压力和扰动位温的输出结果,设计了不确定图标(uncertainty glyph)(如图 7(b)左图所示)和不确定条带(uncertainty ribbon)(如图 7(b)右图所示)方法.在该设计中,同一个空间上所有模拟成员输出按照数值大小降序排列,依次在同一地点叠加对应大小且灰度递降的同心圆或条带.在该系统中,首先对所有模拟成员的输出生成一个集合平均(ensemble mean),再以集合平均生成的数据场中某个等值线为基准,于周围叠加宽度、颜色不一的同心圆或条带.这些叠加的视觉小元素可以帮助用户发现某地点或某轨迹处的数值范围、差异大小以

及是否存在异常值等信息,从而达到分析与改善模型参数的目的。此外,Keherer 等人<sup>[26]</sup>在他们的工作中也使用了基于图标的方法来表示同一个空间点上多个模拟成员的数值分布。他们把按照一定设计规则绘制的星状图标嵌入在并列显示的多个地图上(如图 7(c)上图所示)。其中,星状图的上半部分和下半部分分别表示几种四分位数之间的差别。然后,所有的模拟成员地图按照输入参数值的大小顺序堆叠(如图 7(c)下图所示),由此展示输出变量和输入参数之间的关系。

除了可以集成多种可视化方法来表示不确定性以外,另一种方法是对多个参数分别进行视觉编码。例如,可以对已有技术盒图可视化(boxplot)进行拓展。盒图可视化经常用于展示概率分布。Potter 等人设计的概括图(summary plot)<sup>[27]</sup>扩展了这种可视化方法。他们把数据分布的偏度(skew)、峰度(kurtosis)、直方图(histogram)等一并集成到盒图可视化中,以便于直观地观察和比较不同的分布特征(如图 7(d)所示)。这种方式大幅度地降低了数据的复杂程度,同时也保留了数据中的特征。Ovis 系统<sup>[9,10]</sup>中也用到了盒图可视化的扩展,称为小提琴图(violin plot)(如图 7(e)所示)。该图用外轮廓直接展示海洋模拟数据中每一个时空点上所有数值的概率分布函数,中间的横线表示中位数,颜色表示有多少比例的模拟成员满足一定的数值范围(文中例子使用的是某个表示危险程度海平面的高度)。在该系统中,每一个空间点上、每一个时刻都用一个小提琴图来展示数据分布。同一地点所有时刻的小提琴图一字排开,可以展示集合模拟数据随时间的变化。



(a) 地图颜色表示峰值个数,线段形状图标长度表示峰值的宽度,图标颜色表示具体峰值大小<sup>[25]</sup>  
 (b) 不确定图标和不确定条带<sup>[23]</sup>  
 (c) 基于图标方法,星状图标被嵌入在并列显示的地图中表示四分位数大小<sup>[26]</sup>  
 (d) 概括图<sup>[27]</sup>  
 (e) 小提琴图扩展版本,外轮廓展示概率分布函数,横线表示中位数<sup>[9,10]</sup>

Fig.7 Location-Based uncertainty visualization for ensemble simulation data,  
uncertainty information described by parameter (III)

图 7 基于位置的集合模拟数据不确定性可视化,用参数描述不确定性信息(III)

### (3) 矢量场数据

以上是针对标量场集合模拟数据的情形,相对于标量场,矢量场的集合模拟数据包含更大的数据复杂度,因此,相应的不确定性表示方法会更复杂。由于矢量场自身包含 1 个以上的数据维度,大多数矢量场的集合模拟可

视化都用 1 个以上参数来表示不确定性信息(除了 Liu 等人<sup>[46]</sup>和 Guo 等人<sup>[31]</sup>的工作采用差异场的方式以外).例如,Liu 等人<sup>[46]</sup>定义了一种集合模拟向量场的距离度量.该方法对于每个空间点出发的所有成员的迹线进行序列编码,再使用最长公共子序列(longest common subsequence,简称 LCSS)来计算集合迹线之间的相似性与差异性.图 8(f)是使用该方法计算得到的差异场.他们提出的度量定义可以很好地适应噪声和异常值的情况,更稳定、可信地描述集合模拟流场中流线的不确定性,他们的用例测试在计算与存储效率上有很好的结果.Lodha 等人的研究例举了将矢量场中不确定性信息映射到视觉元素的各种方法(如图 8(a)所示),包括不确定性图标、流线包络线(flow envelope)、动画(animation)、旋转接力棒(twirling baton)等<sup>[8]</sup>.Wittenbrink 等人<sup>[33]</sup>用不确定性图标方法把每一个空间点上矢量场的速度大小、角度方向的不确定性编码成颜色、箭头大小、方向等视觉元素(如图 8(b)所示),帮助用户更方便地比较矢量场数据中的不确定性信息的位置和大小.这种直观的可视化方法也被 Jarema 等人采纳于集合模拟矢量场数据分析系统中<sup>[14]</sup>(如图 8(d)所示),并且他们做了一定的改进,对同一位位置上矢量场的分布做混合高斯模型拟合并,用图标表示多个模式.

除了用参数的方法描述不确定性信息以外,还可以利用基于拉格朗日方法的距离测度,直接从流场特性定义矢量场集合模拟数据的不确定性.例如,Guo 等人<sup>[31]</sup>用拉格朗日测度方法定义从同一个时空点出发的迹线的相互距离(如图 8(e)所示).他们用体渲染、平行坐标的可视化方法展现三维流场集合模拟数据中的差异在空间和时间中的分布与变化.他们开发的高性能系统框架 eFLAA 采用 MapReduce 风格,将差异计算与并行流线积分系统融合在一起,以支持大规模流场集合模拟数据的不确定性分析.Hummel 等人<sup>[32]</sup>针对流场集合模拟数据的研究也使用了拉格朗日方法.他们不仅考虑了集合模拟数据中单成员在时空点上的基于拉格朗日流场差异属性“单成员差异(individual variance)”,还拓展到多矢量场的情境,定义“相互差异(joint variance)”来描述多个矢量场之间的稳定性(如图 8(c)所示).他们的可视分析界面将两种差异投射到二维的平面上,并用颜色对应大小,使用户能够直接观察不同差异在集合模拟流场数据中的分布.

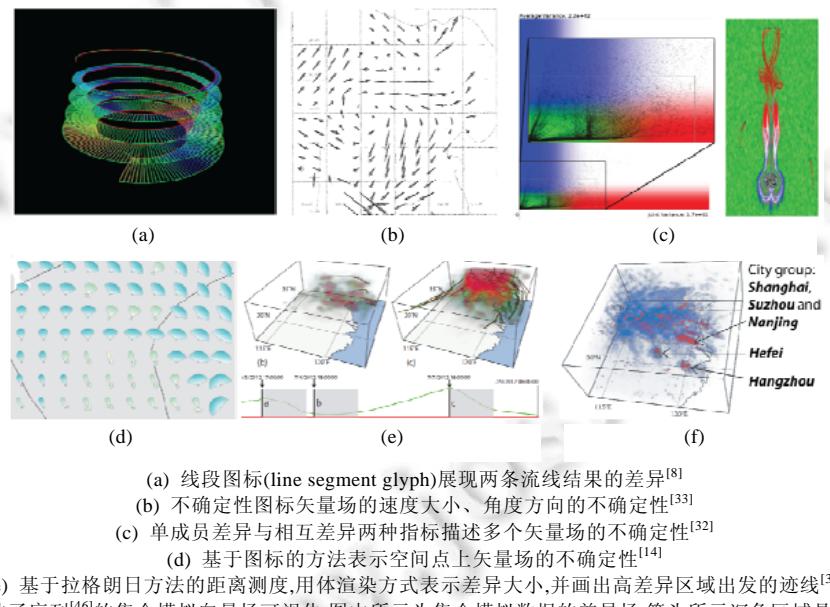


Fig.8 Uncertainty visualization for ensemble vector fields

图 8 集合模拟矢量场不确定性可视化

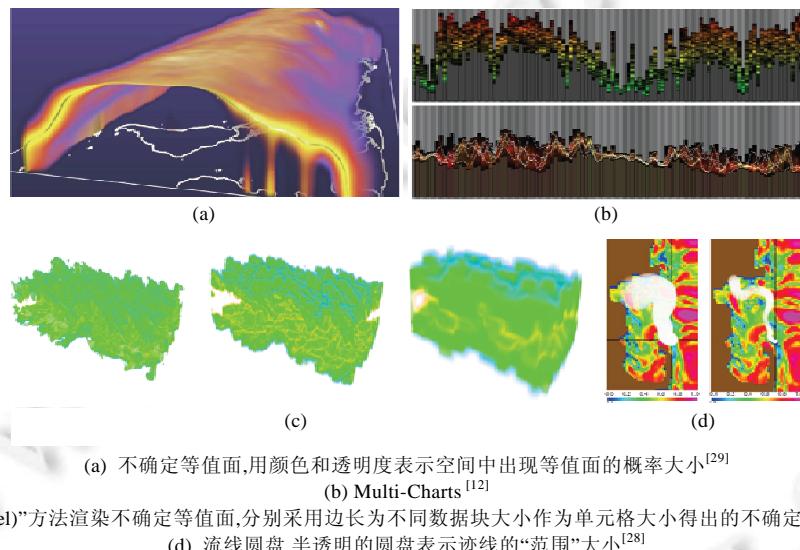
以上是对基于参数描述的不确定性可视化方法的有代表性研究工作的举例.这些方法用有限个参数来描述空间位置点上的多组输出,然后,尽可能地利用标量场可视化方法、多变量可视化方法,或者拓展、创造新的

可视化方法来表达模拟数据的不确定性信息。这些方法的共同前提是,假定所有数据分布都可以转化成有限个参数来描述。事实上,并不是所有的数据都满足这个条件,例如,有些特殊的集合模拟数据具有不规则的分布形状,仅用有限个参数来概括会忽略一些信息。而基于分布函数的描述方法则不依赖这个假设,适用于所有分布形状的数据。

#### 4.1.2 基于分布函数的不确定性描述

在基于分布函数的不确定性描述方法中,集合模拟数据的空间位置点上的多个数值被看作一个数据对象,用概率密度函数来表示。这种新颖的数据表示方法需要相应的新存储和新计算方法,使其能够兼容于已有的可视化算法流程,使现有的可视化技术可以灵活地适用。

这种通过拓展可视化方法来满足集合模拟可视化的案例有很多,例如,拓展光线追踪(ray casting)算法、marching cube 算法以适用于不确定数据场中<sup>[29,41,47]</sup>。这些工作按照空间位置点的概率分布函数来计算该位置的等值面出现的概率大小,从而得到的不确定等值面(uncertain isosurface),用颜色和透明度表示空间的等值面大小(如图 9(a)所示)。



(a) 不确定等值面,用颜色和透明度表示空间中出现等值面的概率大小<sup>[29]</sup>

(b) Multi-Charts<sup>[12]</sup>

(c) “直方像素(hixel)”方法渲染不确定等值面,分别采用边长为不同数据块大小作为单元格大小得出的不确定等值面渲染结果<sup>[30]</sup>

(d) 流线圆盘,半透明的圆盘表示迹线的“范围”大小<sup>[28]</sup>

Fig.9 Location-Based uncertainty visualization for ensemble simulation data,  
uncertainty information described by probability distribution function

图 9 基于位置的集合模拟数据不确定性可视化,用概率分布函数描述的不确定性信息

另外,还有一些关于用分布函数表示不确定性的工作。比如,包括 Bordoloi 等人<sup>[48]</sup>根据分布来聚集空间点,把层次聚类算法成功应用在集合模拟数据上。他们计算了每个空间点上的数据分布之间的相似性,然后基于此把空间点分成几种不同的分布类别,用不同的颜色标识,由此画出不同位置上数据分布的相似性。又如, Thompson 等人<sup>[30]</sup>用直方图的方式表示每个点上的数据分布,称为直方像素(hixel)。他们利用每个空间点上的直方图来统计空间点出现等值(isovalue)的可能性,再用不确定等值面来绘制(如图 9(c)所示)。此外,Love 等人在他们的研究中<sup>[28]</sup>,用信息论的 KL 散度(Kullback-Leibler divergence)来衡量两个空间点上数据分布的相似程度,再用染色等值面的方法画出与目标分布形状相似的数据分布。在他们的可视化结果中,等值面的位置是与目标分布最相近的位置,颜色则对应该处数据的标准差。这种方法可以让用户通过分布形状来查询数据。他们的研究中还包括另一个例子,即在每一步流线与迹线积分算法中加入所有模拟成员的积分结果统计,以此计算所有集合模拟成员的迹线平均值和范围,最后的可视化结果称为流线圆盘(streamline circle),用半透明圆盘表示迹线范围的大小(如图 9(d)所示),以此表示综合了所有模拟成员迹线的不确定性信息。

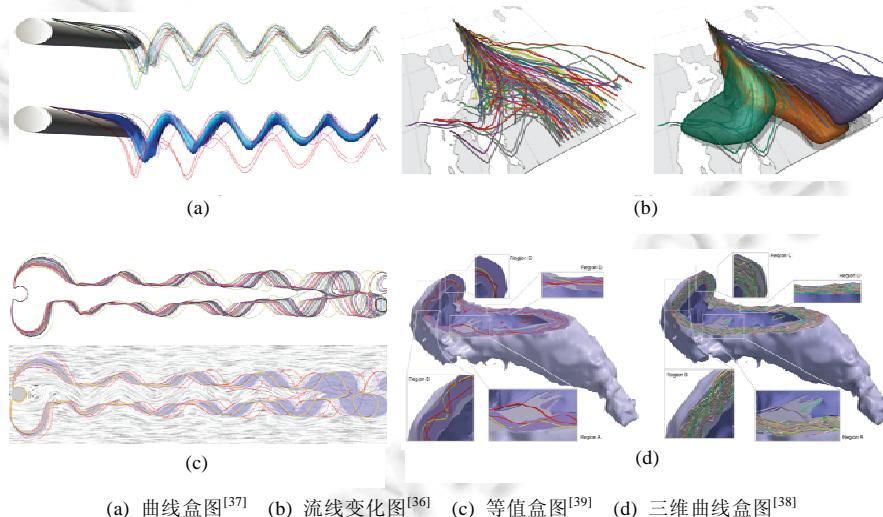
针对三维标量集合模拟可视化的例子包括 Demir 等人的工作:Multi-Charts<sup>[12]</sup>(如图 9(b)所示)。他们巧妙地

利用降维方法把空间信息和数据分布同时投影到二维屏幕空间中。该方法首先将在三维最小体素单元按照希尔伯特空间填充曲线的方式编码,统计单元格内数据分布,并把分布信息投影在二维空间中。屏幕横向表示空间位置,屏幕纵向则展示每个空间块内集合模拟的数值分布。数值分布可以用直方图来表示每个单元格内的分布形状,或者用平行坐标(parallel coordinate)展示每个成员在单元格内的平均值。用户可以交互式地探索集合模拟数据的不确定性。除了按照希尔伯特曲线编码以外,他们的方法还支持按照分布的相似程度来聚集区域,以便用户发现集合模拟数据中数值分布接近的区域位置与形状。

#### 4.2 基于特征的集合模拟不确定性可视化

在基于特征的集合模拟数据不确定性可视化中,关注的是每个成员单独提取的特征的不确定信息。集合模拟数据中单个成员的特征形式,包括等值线、等值面、流线等几何形状。相对于基于位置的方法,这个方向的相关工作要少一些。这是因为屏幕空间同时容纳过多的几何图形会导致彼此干扰和视觉混乱。对于此类特征的分析任务,概括性地描述要比简单地做比较更为困难。

在一个视图中展示所有成员特征之间的不同,重点在于如何利用视觉渲染的效果传达出几何形状的不同。最直接和简单的方案包括同时展现所有集合模拟成员的等高线。这种可视化方法被称为意大利面图,因其外形如一条条意大利面而得名。许多研究和可视化系统都用到了这项技术,例如 Noodles<sup>[23]</sup>、Ensemble-Vis<sup>[1]</sup>、Ovis<sup>[10]</sup>等。然而在集合模拟成员个数非常多的情况下,意大利面图会导致视觉上的混乱。为此,近年来有一些工作把盒图可视化方法拓展到曲线集合中,旨在概括其形状分布,比如,曲线盒图(curve boxplot)<sup>[37]</sup>(如图 10(a)所示)、等值盒图(contour boxplot)<sup>[39]</sup>(如图 10(c)所示)及其三维版本(如图 10(d)所示)。进一步地,更有流线变化图(streamline variability plot)<sup>[36]</sup>(如图 10(b)所示)利用聚类方法,把杂乱而繁多的流线简洁地总结成几股趋势。该方法可以很好地应用于神经影像、飓风模拟以及流体力学模拟数据中<sup>[36]</sup>。图 10(b)中展示的是对集合模拟飓风数据用聚类方法总结出的 3 股趋势,用线条、包络线以及包络面来表示每一股趋势的中值与范围,由此可以概括出集合模拟数据的整体成员的分布。这一系列受盒图启发的工作是近两年最新的研究动态,不久后一定会有更多基于特征的集合模拟不确定性可视化研究成果出现。



(a) 曲线盒图<sup>[37]</sup> (b) 流线变化图<sup>[36]</sup> (c) 等值盒图<sup>[39]</sup> (d) 三维曲线盒图<sup>[38]</sup>

Fig.10 Feature-Based uncertainty visualization for ensemble simulation data

图 10 基于特征的集合模拟数据不确定性可视化

## 5 讨论及总结

近年来,随着计算能力和存储技术的提高,集合模拟方法在气象、天气预报、数学、空气动力学等科学领

域中越来越普遍。集合模拟数据的特征是具有多值、多变量、时变的属性,再加上其庞大的数据量,对这类数据的分析充满挑战。集合模拟数据可视化技术把集合模拟数据更直观地展现给领域专家,用视觉和交互的手段帮助专家探索、验证、总结他们的重要领域发现。本文从比较个体成员和概括整体成员这两个基于分析任务的角度以及基于位置和基于特征这两种分析策略的角度同时出发,系统地回顾了具有代表性的集合模拟可视化工作,整理了各类方法的可视化形式、交互技术和应用案例。

在文中提及的研究中,有许多项研究相互之间存在很强的延续性与系统性。比如,有关集合流场距离的可视化工作,首先,在 Guo 等人<sup>[31]</sup>提出基于拉格朗日测度的距离度量方法与可视分析系统方法后,又有 Liu 等人针对噪声、异常值情况等鲁棒性问题进行拓展研究,并且提出了基于最长公共子序列算法的集合模拟流场可视分析方法<sup>[46]</sup>。同时,针对同类型的流场模拟数据的分析工作,在可视分析界面与交互上做了开拓性的尝试,包括基于手势的涡线特征提取方法<sup>[23,49]</sup>、基于图的模拟成员距离可视化方法<sup>[4]</sup>。类似地,又比如等值线与盒图可视化在集合模拟可视化领域的拓展研究:等值盒图<sup>[39]</sup>对集合模拟数据中的等高线、流线等曲线形状特征做可视化以展现其概率分布,曲线盒图<sup>[37]</sup>是针对曲线集合数据的通用可视化方法。之后,三维集合模拟数据的应用也有了相应的拓展研究工作<sup>[38]</sup>。此外,更出现了流线变化图<sup>[36]</sup>,同样地针对集合模拟中特征的可视化,引入了聚类方法,不仅展现了集合模拟的特征分布,也总结了所有特征的分布趋势。从这些发展历程可以看出,近年来,有关集合模拟数据的研究工作呈更加密集、系统的趋势,前人的成果不断启发后来的学者,领域内学术氛围蒸蒸日上。

集合模拟数据可视化在未来将有很多研究热点与研究挑战。

- 首先是有关时序集合模拟数据的研究。

回顾已有的方法,大多都是针对静态数据、非时序数据的情形,而越来越多的集合模拟数据都带有时序特征,很多现有方法不能满足时变的需求。科学家对时序集合模拟数据分析目标是从时空的角度理解多个模拟成员之间的相似性、不确定性和参数关联的变化,而时序的特性却带来更大量的计算,加重了分析的复杂性。所以,对时序集合模拟数据的可视化研究仍然有很大的探索空间。

- 其次是更多结合信息可视化技术的趋势。

随着近年来可视分析技术的成熟,越来越多的信息可视化技术被应用到科学数据可视化中,比如平行坐标<sup>[16]</sup>、盒图<sup>[10]</sup>、图可视化<sup>[4]</sup>等技术都已被尝试运用在集合模拟数据可视化研究中。因为集合模拟数据本身具有多维、多变量、多值和时变的属性,复杂的数据特征需要更强大的分析技术,而传统的地图、折线图、体渲染等单一的可视化方式已经不能支持对这种复杂数据的深层分析,需要更多交互式、直观的方法来帮助科学家理解集合模拟的特征。所以,将信息可视化应用在集合模拟数据的尝试会更加普遍,未来将会有更多引入信息可视化和交互的可视分析手段,以满足领域科学家的具体分析需求。

- 与此同时,集合模拟数据的不确定性分析也有很多值得研究的方向。

对于集合模拟数据的分析与可视化,不确定性不仅存在于集合模拟的成员之间,而且整个数据处理与可视化流程中均有一定的不确定性被引入。例如,集合模拟数据在生成、过滤、存储以及计算时,都会被引入不确定性。对这些不确定性的研究,也将是一个热点与挑战。

- 此外,另一个不容忽视的研究方向是通过优化对数据存储和预处理来提升可视化的效果。

比起多变量数据和时变数据,集合模拟数据可视化的特点在于其多成员属性,这使得集合模拟数据比多变量数据、时变数据多一个数据维度。多一个数据维度,意味着数据存储大一个数量级。然而从另一个角度考虑集合模拟的多成员特性,亦即成员与成员之间存在一定的数据关联和相似性。利用相似性的特性,研究者可以从数据压缩、数据预处理着手,通过消除冗余度来提高可视化和交互的性能,例如针对多个模拟成员的相似性进行数据压缩,以提高渲染时的性能。随着可视化和可视分析技术的发展,一定会有更多适用于大规模高性能计算平台的可视化技术、工具库、系统出现,甚至是原位可视分析框架,以满足集合模拟数据更快、更高效的分析需求。

**References:**

- [1] Potter K, Wilson AT, Bremer PT, Williams DN, Doutriaux CM, Pascucci V, Johnson CR. Ensemble-Vis: A framework for the statistical visualization of ensemble data. In: Proc. of the IEEE Int'l Conf. on Data Mining Workshops. 2009. 233–240. [doi: 10.1109/ICDMW.2009.55]
- [2] Allen M, Erickson D, Kendall W, Fu J, Ott L, Pawson S. The influence of internal model variability in GEOS-5 on interhemispheric CO<sub>2</sub> exchange. *Journal of Geophysical Research: Atmospheres*, 2012,117(D10):1–20. [doi: 10.1029/2011JD017059]
- [3] Bauer P, Thorpe A, Brunet G. The quiet revolution of numerical weather prediction. *Nature*, 2015,525(7567):47–55. [doi: 10.1038/nature14956]
- [4] Shu Q, Guo H, Che L, Yuan X, Liu J, Liang J. EnsembleGraph: Interactive visual analysis of spatiotemporal behaviors in ensemble simulation data. In: Proc. of the IEEE Pacific Visualization Symp. 2016. 56–63. [doi: 10.1109/PACIFICVIS.2016.7465251]
- [5] Waser J, Fuchs R, Ribicic H, Schindler B, Bloschl G, Groller E. World lines. *IEEE Trans. on Visualization and Computer Graphics*, 2010,16(6):1458–1467. [doi: 10.1109/TVCG.2010.223]
- [6] Li X, Liu J, Mauzerall DL, Emmons LK, Walters S, Horowitz LW, Tao S. Effects of trans-Eurasian transport of air pollutants on surface ozone concentrations over Western China. *Journal of Geophysical Research: Atmospheres*, 2014,119(21):12338–12354. [doi: 10.1002/2014JD021936]
- [7] Luo A, Kao D, Pang A. Visualizing spatial distribution data sets. In: Proc. of the Data Visualization Symp. 2003. 29–38.
- [8] Lodha SK, Pang A, Sheehan RE, Wittenbrink CM. UFLOW: Visualizing uncertainty in fluid flow. In: Proc. of the IEEE Visualization. 1996. 249–254. [doi: 10.1109/VISUAL.1996.568116]
- [9] Höllt T, Magdy A, Chen G, Gopalakrishnan G, Hoteit I, Hansen C, Hadwiger M. Visual analysis of uncertainties in ocean forecasts for planning and operation of off-shore structures. In: Proc. of the IEEE Pacific Visualization Symp. 2013. 185–192. [doi: 10.1109/PacificVis.2013.6596144]
- [10] Höllt T, Magdy A, Zhan P, Chen G, Gopalakrishnan G, Hoteit I, Hansen C, Hadwiger M. Ovis: A framework for visual analysis of ocean forecast ensembles. *IEEE Trans. on Visualization and Computer Graphics*, 2014,20(8):1114–1126. [doi: 10.1109/TVCG.2014.2307892]
- [11] Obermaier H, Joy KI. Future challenges for ensemble visualization. *IEEE Computer Graphics and Applications*, 2014,34(3):8–11. [doi: 10.1109/MCG.2014.52]
- [12] Demir I, Dick C, Westermann R. Multi-Charts for comparative 3D ensemble visualization. *IEEE Trans. on Visualization and Computer Graphics*, 2014,20(12):2694–2703. [doi: 10.1109/TVCG.2014.2346448]
- [13] Hao L, Healey CG, Bass SA. Effective visualization of temporal ensembles. *IEEE Trans. on Visualization and Computer Graphics*, 2016,22(1):787–796. [doi: 10.1109/TVCG.2015.2468093]
- [14] Jarema M, Demir I, Kehrer J, Westermann R. Comparative visual analysis of vector field ensembles. In: Proc. of the IEEE Visual Analytics Science and Technology. IEEE, 2015. 81–88. [doi: 10.1109/VAST.2015.7347634]
- [15] Piringer H, Pajer S, Berger W, Teichmann H. Comparative visual analysis of 2D function ensembles. *Computer Graphics Forum*, 2012,31(3.3):1195–1204. [doi: 10.1111/j.1467-8659.2012.03112.x]
- [16] Poco J, Dasgupta A, Wei Y, Hargrove W, Schwalm C, Cook R, Bertini E, Silva C. SimilarityExplorer: A visual inter-comparison tool for multifaceted climate data. *Computer Graphics Forum*, 2014,33(3):341–350. [doi: 10.1111/cgf.12390]
- [17] Malik MM, Heinzl C, Groeller ME. Comparative visualization for parameter studies of dataset series. *IEEE Trans. on Visualization and Computer Graphics*, 2010,16(5):829–840. [doi: 10.1109/TVCG.2010.20]
- [18] Nocke T, Flechsig M, Bohm U. Visual exploration and evaluation of climate-related simulation data. In: Proc. of the Simulation Conf. 2007. 703–711. [doi: 10.1109/WSC.2007.4419664]
- [19] Phadke MN, Pinto L, Alabi O, Harter J, Taylor II RM, Wu X, Petersen H, Bass SA, Healey CG. Exploring ensemble visualization. *Visualization and Data Analysis*, 2012,8294(82940B):1–12. [doi: 10.1111/12.912419]
- [20] Zhang C, Schultz T, Lawonn K, Eisemann E, Vilanova A. Glyphbased comparative visualization for diffusion tensor fields. *IEEE Trans. on Visualization and Computer Graphics*, 2016,22(1):797–806. [doi: 10.1109/TVCG.2015.2467435]

- [21] Gosink L, Bensema K, Pulsipher T, Obermaier H, Henry M, Childs H, Joy KI. Characterizing and visualizing predictive uncertainty in numerical ensembles through Bayesian model averaging. *IEEE Trans. on Visualization and Computer Graphics*, 2013,19(12): 2703–2712. [doi: 10.1109/TVCG.2013.138]
- [22] Potter K, Wilson A, Bremer PT, Williams D, Doutriaux C, Pascucci V, Johhson C. Visualization of uncertainty and ensemble data: Exploration of climate modeling and weather forecast data with integrated ViSUS-CDAT systems. *Journal of Physics: Conf. Series*, 2009,180(1):1–5. [doi: 10.1088/1742-6596/180/1/012089]
- [23] Sanyal J, Zhang S, Dyer J, Mercer A, Amburn P, Moorhead RJ. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE Trans. on Visualization and Computer Graphics*, 2010,16(6):1421–1430. [doi: 10.1109/TVCG.2010.181]
- [24] Wilson AT, Potter KC. Toward visual analysis of ensemble data sets. In: Proc. of the Workshop on Ultrascale Visualization. 2009. 48–53. [doi: 10.1145/1838544.1838551]
- [25] Kao DT, Luo A, Dungan JL, Pang A. Visualizing spatially varying distribution data. In: Proc. of the Int'l Conf. on Information Visualisation. 2002. 219–226. [doi: 10.1109/IV.2002.1028780]
- [26] Kehrer J, Muigg P, Doleisch H, Hauser H. Interactive visual analysis of heterogeneous scientific data across an interface. *IEEE Trans. on Visualization and Computer Graphics*, 2011,17(7):934–946. [doi: 10.1109/TVCG.2010.111]
- [27] Potter K, Kniss J, Riesenfeld R, Johnson CR. Visualizing summary statistics and uncertainty. *Computer Graphics Forum*, 2010, 29(3):823–832. [doi: 10.1111/j.1467-8659.2009.01677.x]
- [28] Luo A, Pang A, Kao D. Visualizing spatial multivalue data. *IEEE Computer Graphics and Applications*, 2005,25(3):69–79. [doi: 10.1109/MCG.2005.71]
- [29] Pothkow K, Hege HC. Positional uncertainty of isocontours: Condition analysis and probabilistic measures. *IEEE Trans. on Visualization and Computer Graphics*, 2011,17(10):1393–1406. [doi: 10.1109/TVCG.2010.247]
- [30] Thompson D, Levine JA, Bennett JC, Bremer PT, Gyulassy A, Pascucci V, Pebay PP. Analysis of large-scale scalar data using hixels. In: Proc. of the IEEE Symp. on Large Data Analysis and Visualization. 2011. 23–30. [doi: 10.1109/LDAV.2011.6092313]
- [31] Guo H, Yuan X, Huang J, Zhu X. Coupled ensemble flow line advection and analysis. *IEEE Trans. on Visualization and Computer Graphics*, 2013,19(12):2733–2742. [doi: 10.1109/TVCG.2013.144]
- [32] Hummel M, Obermaier H, Garth C, Joy KI. Comparative visual analysis of Lagrangian transport in CFD ensembles. *IEEE Trans. on Visualization and Computer Graphics*, 2013,19(12):2743–2752. [doi: 10.1109/TVCG.2013.141]
- [33] Wittenbrink CM, Pang A, Lodha SK. Glyphs for visualizing uncertainty in vector fields. *IEEE Trans. on Visualization and Computer Graphics*, 1996,2(3):266–279. [doi: 10.1109/2945.537309]
- [34] Alabi OS, Wu X, Harter JM, Phadke M, Pinto L, Petersen H, Bass S, Keifer M, Zhong S, Healey C, Taylor RM. Comparative visualization of ensembles using ensemble surface slicing. In: Proc. of the Visualization and Data Analysis. 2012. 82–94.
- [35] da Silva MJ, Zhang S, Demiralp C, Laidlaw DH. Visualizing the differences between diffusion tensor volume images. In: Proc. of the Int'l Society for Magnetic Resonance in Medicine Diffusion MRI Workshop. 2000. 1–3.
- [36] Ferstl F, Burger K, Westermann R. Streamline variability plots for characterizing the uncertainty in vector field ensembles. *IEEE Trans. on Visualization and Computer Graphics*, 2016,22(1):767–776. [doi: 10.1109/TVCG.2015.2467204]
- [37] Mirzargar M, Whitaker RT, Kirby RM. Curve boxplot: Generalization of boxplot for ensembles of curves. *IEEE Trans. on Visualization and Computer Graphics*, 2014,20(12):2654–2663. [doi: 10.1109/TVCG.2014.2346455]
- [38] Raj M, Mirzargar M, Kirby R, Whitaker RT, Preston JS. Evaluating alignment of shapes by ensemble visualization. *IEEE Computer Graphics and Applications*, 2015,36(3):100–109. [doi: 10.1109/MCG.2015.70]
- [39] Whitaker RT, Mirzargar M, Kirby RM. Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles. *IEEE Trans. on Visualization and Computer Graphics*, 2013,19(12):2713–2722. [doi: 10.1109/TVCG.2013.143]
- [40] Gleicher M, Albers D, Walker R, Jusufi I, Hansen CD, Roberts JC. Visual comparison for information visualization. *Information Visualization*, 2011,10(4):289–309. [doi: 10.1177/1473871611416549]
- [41] Pothkow K, Hege HC. Nonparametric models for uncertainty visualization. *Computer Graphics Forum*, 2013,32(3):131–140. [doi: 10.1111/cgf.12100]

- [42] Bruckner S, Moller T. Result-Driven exploration of simulation parameter spaces for visual effects design. *IEEE Trans. on Visualization and Computer Graphics*, 2010,16(6):1468–1476. [doi: 10.1109/TVCG.2010.190]
- [43] Liu R, Guo H, Yuan X. A bottom-up scheme for user-defined feature comparison in ensemble data. In: Proc. of the ACM SIGGRAPH Asia 2015 Symp. on Visualization in High Performance Computing. 2015. 1–4. [doi: 10.1145/2818517.2818531]
- [44] Shneiderman B. The eyes have it: A task by data type taxonomy for information visualizations. In: Proc. of the IEEE Symp. on Visual Languages. IEEE, 1996. 336–343. [doi: 10.1109/VL.1996.545307]
- [45] Kao DL, Kramer MG, Love AL, Dungan JL, Pang AT. Visualizing distributions from multi-return lidar data to understand forest structure. *The Cartographic Journal*, 2005,42(1):35–47. [doi: 10.1179/000870405X57257]
- [46] Liu R, Guo H, Zhang J, Yuan X. Comparative visualization of vector field ensembles based on longest common subsequence. In: Proc. of the IEEE Pacific Visualization Symp. 2016. 96–103. [doi: 10.1109/PACIFICVIS.2016.7465256]
- [47] Pfaffelmoser T, Reitinger M, Westermann R. Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields. *Computer Graphics Forum*, 2011,30(3):951–960. [doi: 10.1111/j.1467-8659.2011.01944.x]
- [48] Bordoloi UD, Kao DL, Shen HW. Visualization techniques for spatial probability density function data. *Data Science Journal*, 2004, 3:153–162. [doi: 10.2481/dsj.3.153]
- [49] Liu R, Guo H, Yuan X. User-Defined feature comparison for vector field ensembles. *Journal of Visualization*, 2017,20(2):217–229. [doi: 10.1007/s12650-016-0388-0]



舒清雅(1990—),女,湖北武汉人,硕士,主要研究领域为科学可视化.



刘日晨(1984—),男,博士,讲师,主要研究领域为科学可视化,计算机视觉,图形图像处理.



洪帆(1990—),男,博士生,主要研究领域为科学数据可视化.



张江(1991—),男,博士生,CCF 学生会员,主要研究领域为大数据可视化,高性能计算.



袁晓如(1975—),男,博士,研究员,博士生导师,CCF 杰出会员,主要研究领域为科学可视化,信息可视化,可视分析,计算机图形学,人机交互.