

PKU Visualization Blog 北京大学可视化与可视分析博客

基于自内嵌平行坐标的多分辨率气象集合模拟数据的参数分析 (Multi-Resolution Climate Ensemble Parameter Analysis with Nested Parallel Coordinates Plots)

作者: Richen Liu 日期: 2016年10月15日

传统的集合模拟数据本身以及参数空间都非常复杂，因为集合模拟数据往往是多成员 (multi-run)，多维度 (multi-dimensional) 以及时变的 (time-varying)。在此基础之上，领域科学家 (如气象学家) 在进行集合模拟时，还需要使用不同的分辨率来进行模拟，因为不同的分辨率的数据可以得到不同物理意义的特征 (pyhsical feature)，因此，气象学家往往会对同一个模拟进行多分辨率的多次模拟。

对于高维的数据，当维度空间比较大时，传统的方法是使用平行坐标来表示维度之间的数据分布与数据相关性。但是对于多分辨率数据，每个集合模拟数据有多个分辨率的副本，要同时将所有的副本可视化出来，可以有两种方式，第一种是重叠 (superimposition)，即将所有副本在一个平行坐标里可视化出来；第二种是并排 (Juxtaposition)，也就是将所有分辨率副本的数据分别用单独的平行坐标可视化出来，然而这两种方法都具有一定的缺点。例如图1所示，使用重叠的方式将所有三个分辨率 (3个set) 的数据都绘制到一个平行坐标中，可以比较方便地看出set之间 (inter-set) 之间的关系，但是set内部维度 (intra-set) 之间的关系比较混乱，因为存在较大的自遮挡。而对于二种并排的方式，可以很好地查看某单个set内部维度之间的关系，然而对于inter-set之间的关系，很难进行对比。因此本文[1]综合考虑这两种方式的优点，将并排的方式内嵌到重叠方式 (Nested Parallel Coordinates Plots : NPCP) 中表示集合模拟数据多分辨率之间以及单个分辨率内部的数据特征。在本文中，除了上述的NPCP设计表示参数空间中参数选择情况外，还设计了比较完整的多分辨率集合模拟数据可视分析系统。该系统提供了不同的linked view来表示时空方面的演变信息，包括使用热图和层次聚类树表示集合成员之间的关系，以及使用空间视图表示模拟的结果与观测结果之间的差异。

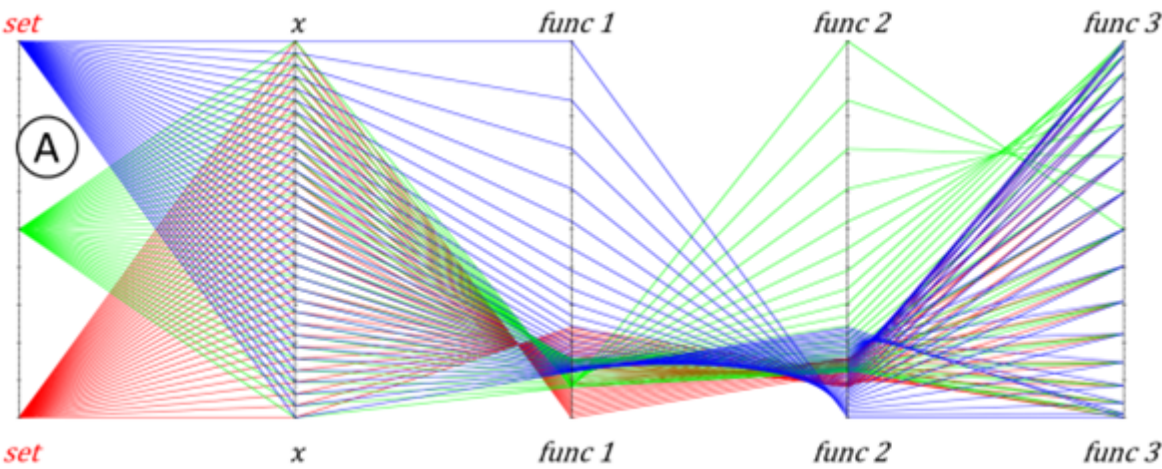


图1. 平行坐标数据重叠可视化

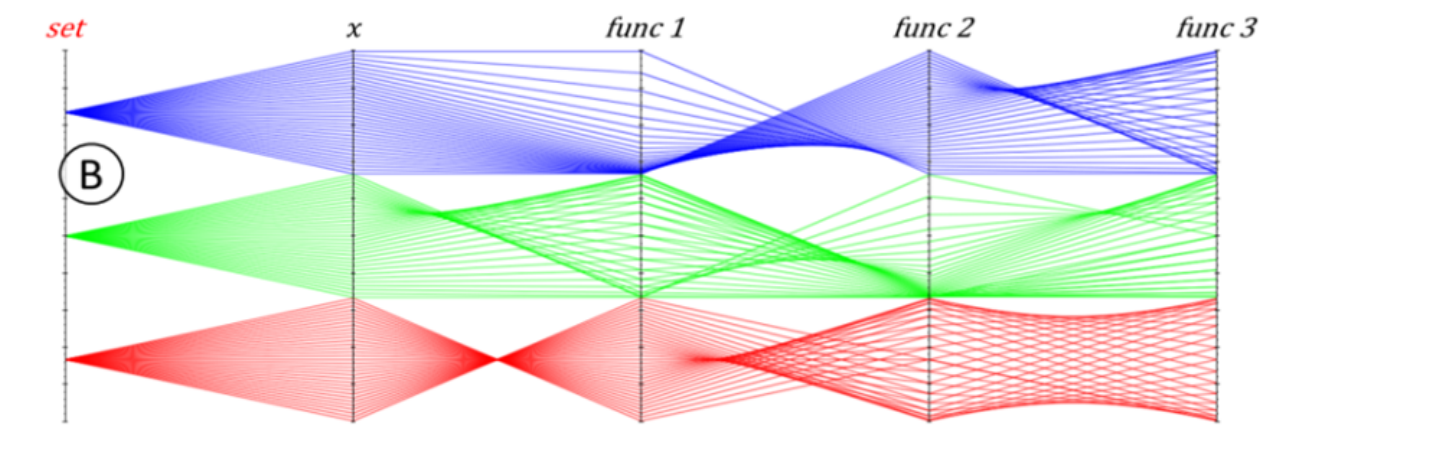


图2. 平行坐标数据并排可视化

本文的工作是与领域科学家进行密切合作的基础上进行开展的。作者在与领域科学家进行充分交流后，得到六点需求信息，从这六点需求出发，最终开发出他们的可视分析系统。六点需求如下：

- R1：系统要将各个分辨率内部的参数关联关系可视化出来 (Understanding Intra-Set Parameter Correlation)，也就是将单个分辨率内部维度信息表示出来。例如，图2中的并排可视化就可以很好地表示单个分辨率内部维度的信息。
- R2：系统要提供多分辨率数据集之间的对比 (Comparing Inter-Set Parameter Correlations)。例如，图1中使用重叠的方式将所有三个分辨率的数据集放在一个平行坐标中可视化出来，最终方便用户对比。
- R3：系统要提供参数空间中各个维度的取值范围查询(High-Dimensional Parameter Range Query)。
- R4：系统要提供用户集合模拟成员与观测值之间对比的功能(Ensemble Member/Item Quality Evaluation)。即将不同成员在不同空间点与真实观测值之间的差异信息可视化出来。
- R5：系统要对集合成员之间的关系可视化出来(Ensemble Member Comparison)
- R6：系统要同时提供总体空间特征信息以及空间特征随时间演变的信息 (Demonstrating Both Spatial and Temporal Facets of Ensembles)。

图3表示的是本文提出的可视分析系统的界面。A1表示本文的NPCP视图，A2表示参数空间的范围查询表达示，用户可以指定交并补等集合操作进行复杂的参数范围查询。B1，B2与B3表示层次聚类树图与热图，表示集合成员之间的层次关系信息。C1表示模拟结果的空间视图，C2表示观测数据的空间视图，C3表示模拟结果与观测数据之间的差异信息。用户可以在C4与C5控制面版中对差异计算方法进行表达式定义。

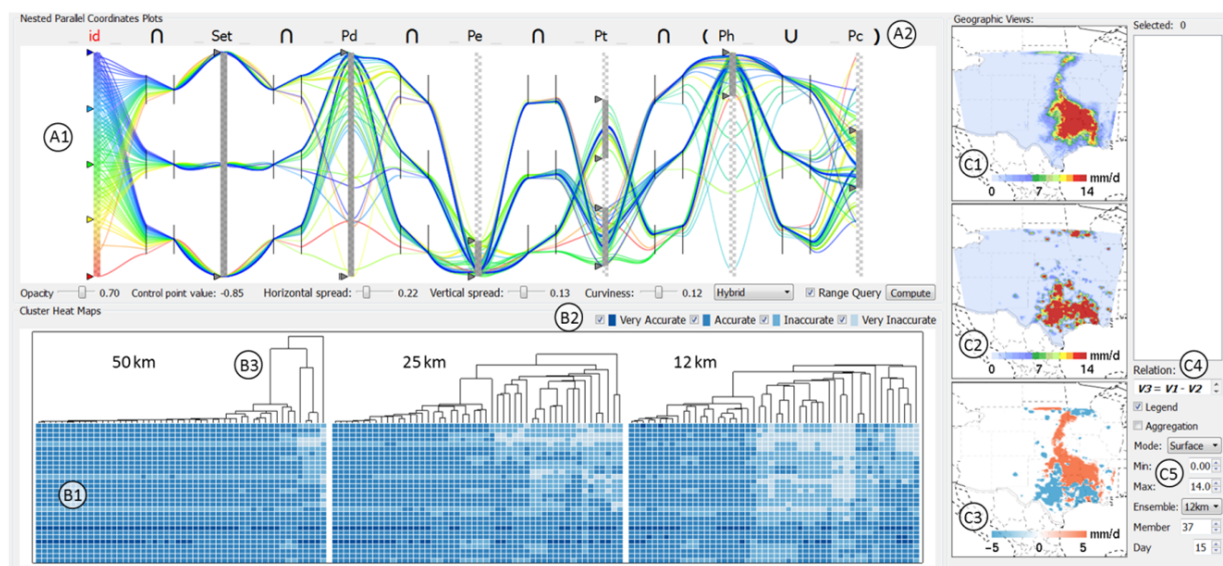


图3. 本文可视分析系统界面

对于R1与R2, 本文提出了上述的NPCP的方式, 将并排的平行坐标内嵌到重叠的平行坐标中, 同时将多分辨率数据集之间与内部的信息可视化出来。如图4所示, 为了表示本文提出的NPCP的有效性, 文中使用一些简单的基本函数来给出示例。图中的func1与func2即为 $\sin x$, $\cos x$ 或 e^x 等函数。

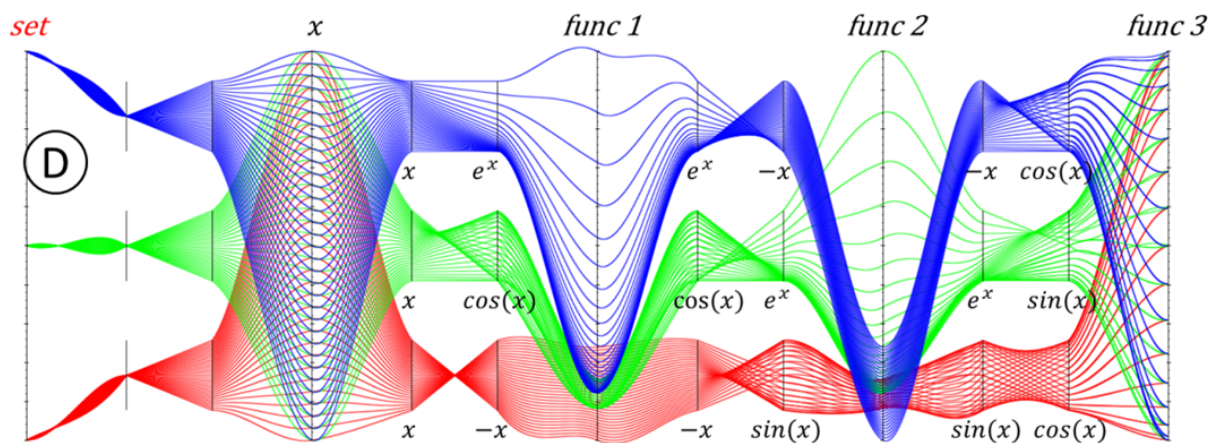


图4. 本文的NPCP设计

对于R3, 本文提供参数空间集合查询操作对参数进行进一步选择。如图5上述的维度之间的集合操作表示选择参数空间的子集, 主要是使用维度之间的交并补进行复杂的范围查询, 对于并的操作, 该维度中除了轴本身指定的范围之外, 还需要与其他轴得到的范围进行并操作。

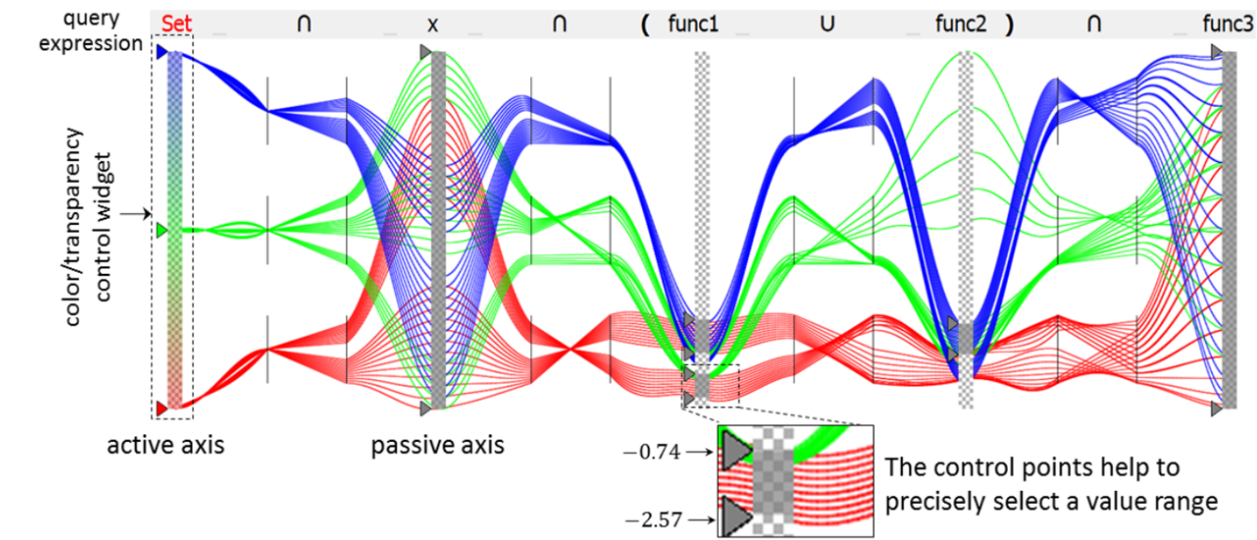


图5. 维度之间的集合操作进行进一步参数选择

对于R4，本文提供的三个空间视图C1，C2和C3可以得到模拟结果与真实观测结果之间的差异信息，此外，用户还可以通过热图（图3中的B3）表示每个集合成员与真实观测数据之间的差异信息。横轴表示不同的集合成员，纵轴表示时间，越蓝表示与观测结果越相小，越淡表示差异越大。

对于R5，通过层次结构树以及将树的叶子结点用热图表示（图3中的B3）来表示集合成员之间的层次关系。

对于R6，系统还提供了空间上的时间演变信息，如图6所示，表示其中一个成员随时间演变的信息。

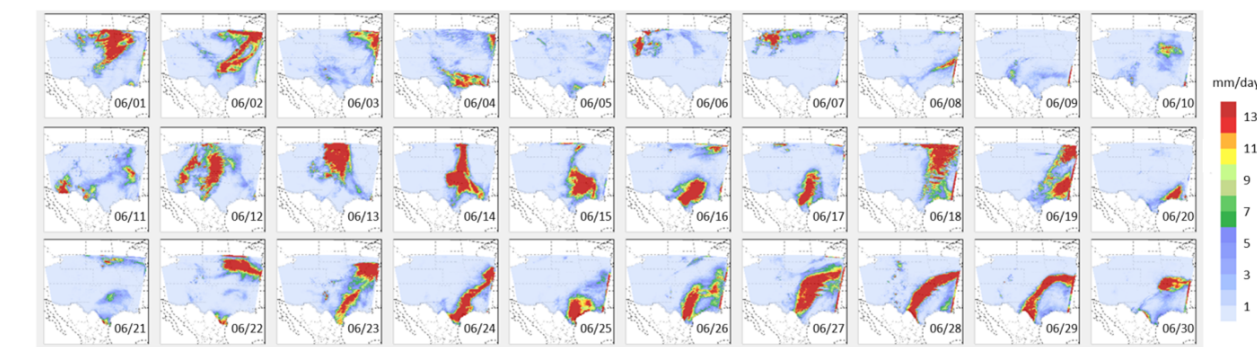


图6. 12km分辨率数据集中第33个成员于2007七月共30个时间步的时间演变信息

此外，系统还提供了一些交互功能，用户可以设置内嵌平行坐标的长宽值，以及曲线的弯曲程度等来调整NPCP的显示效果。最后，本文作者结合领域专家的反馈意见，得到NPCP比传统的平行坐标更有效的结论。

[1] Junpeng Wang, Xiaotong Liu, Han-Wei Shen, and Guang Lin. Multi-Resolution Climate Ensemble Parameter Analysis with Nested Parallel Coordinates Plots. IEEE Transactions on Visualization and Computer Graphics, 2017.