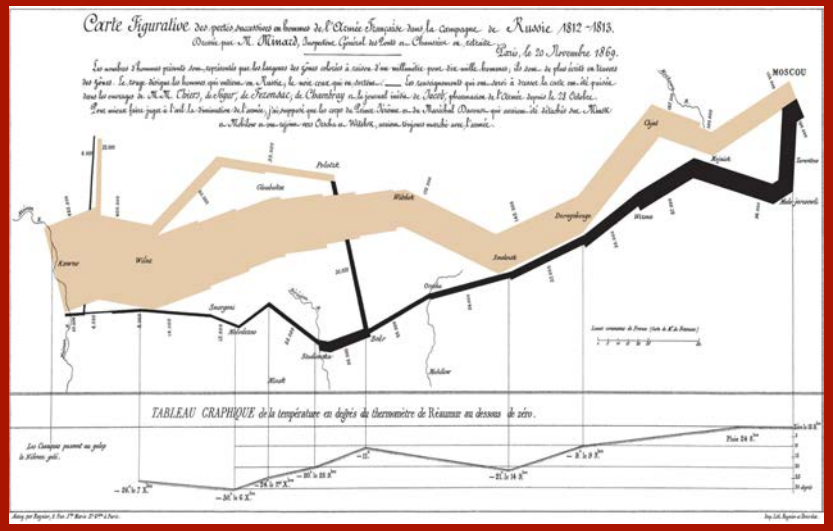
High Performance Visual Analysis

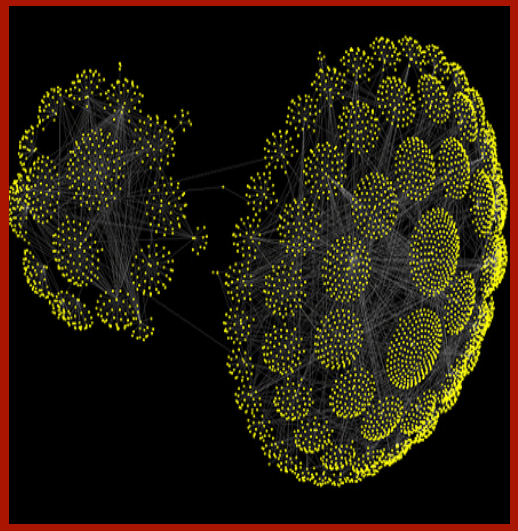
- the Challenges and Opportunities

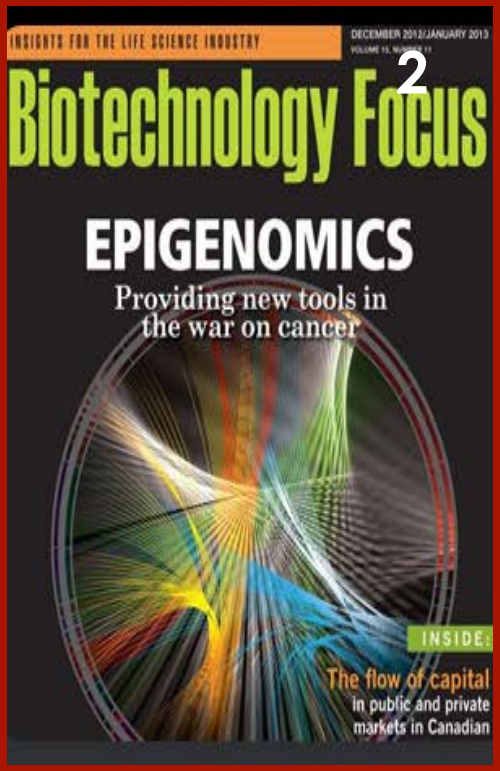
面向科学数据的高性能可视分析

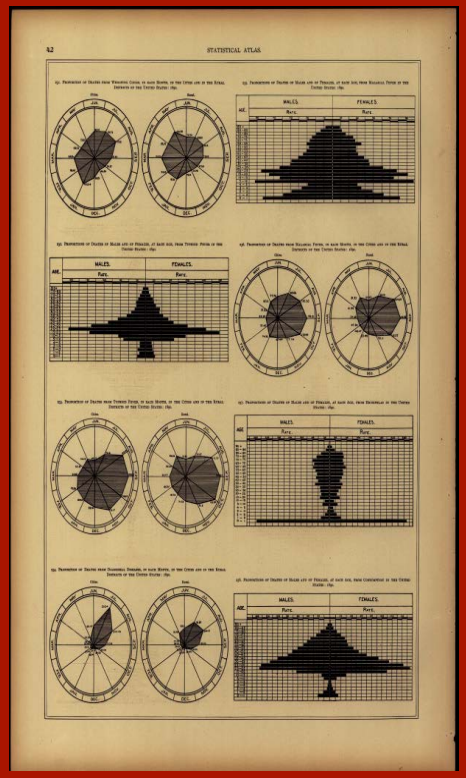
袁晓如 北京大学

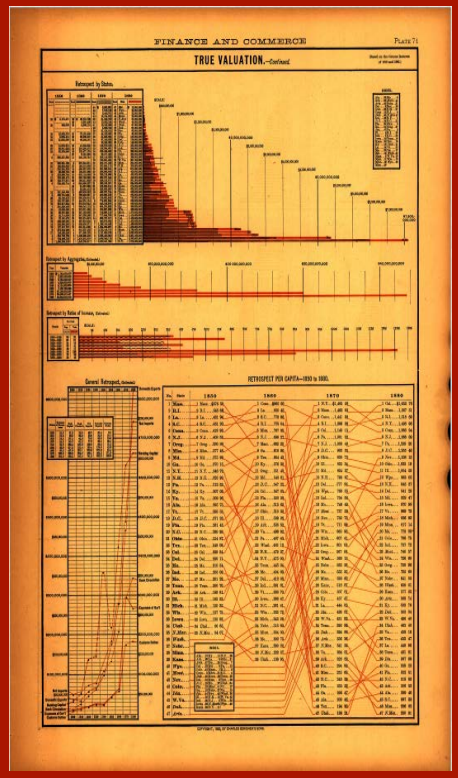
2013.11.14



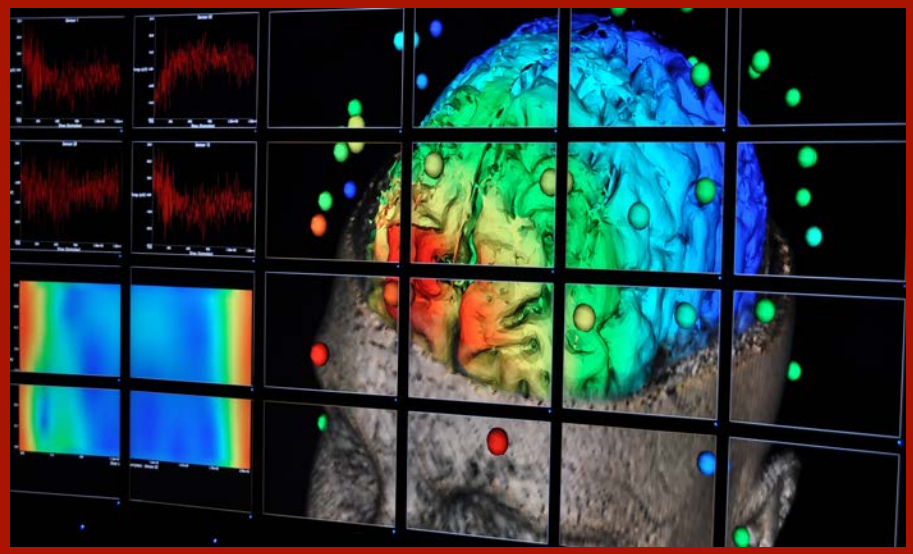






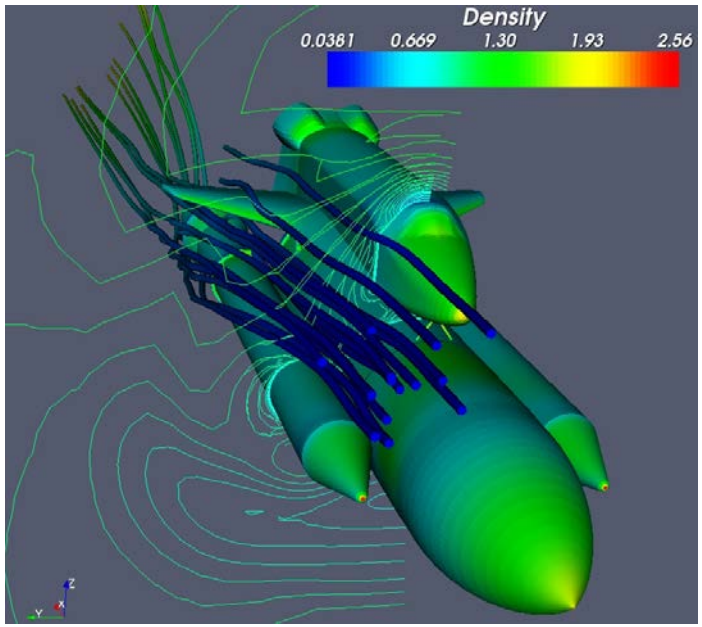


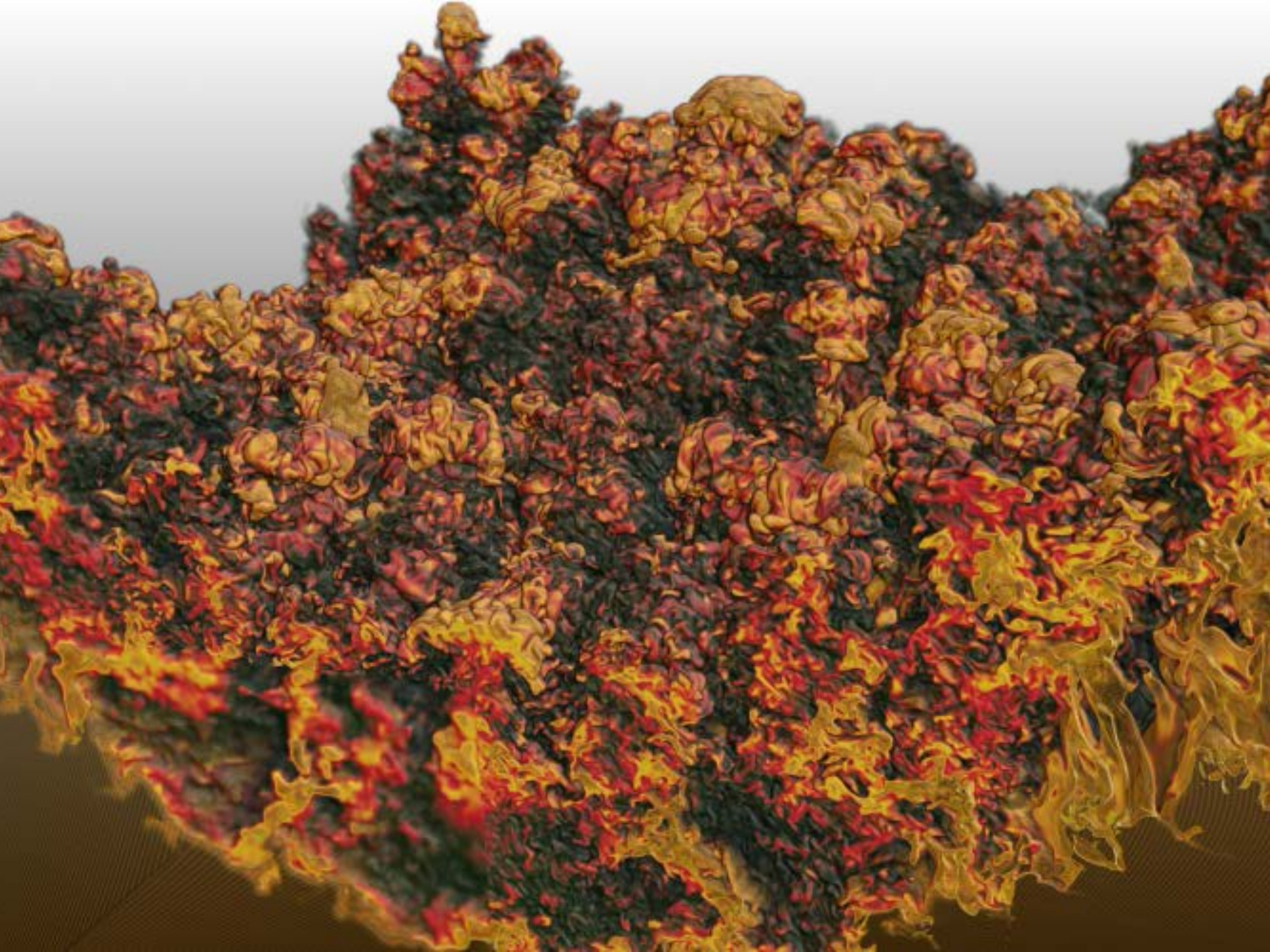
Visualization



From Data to Visualization

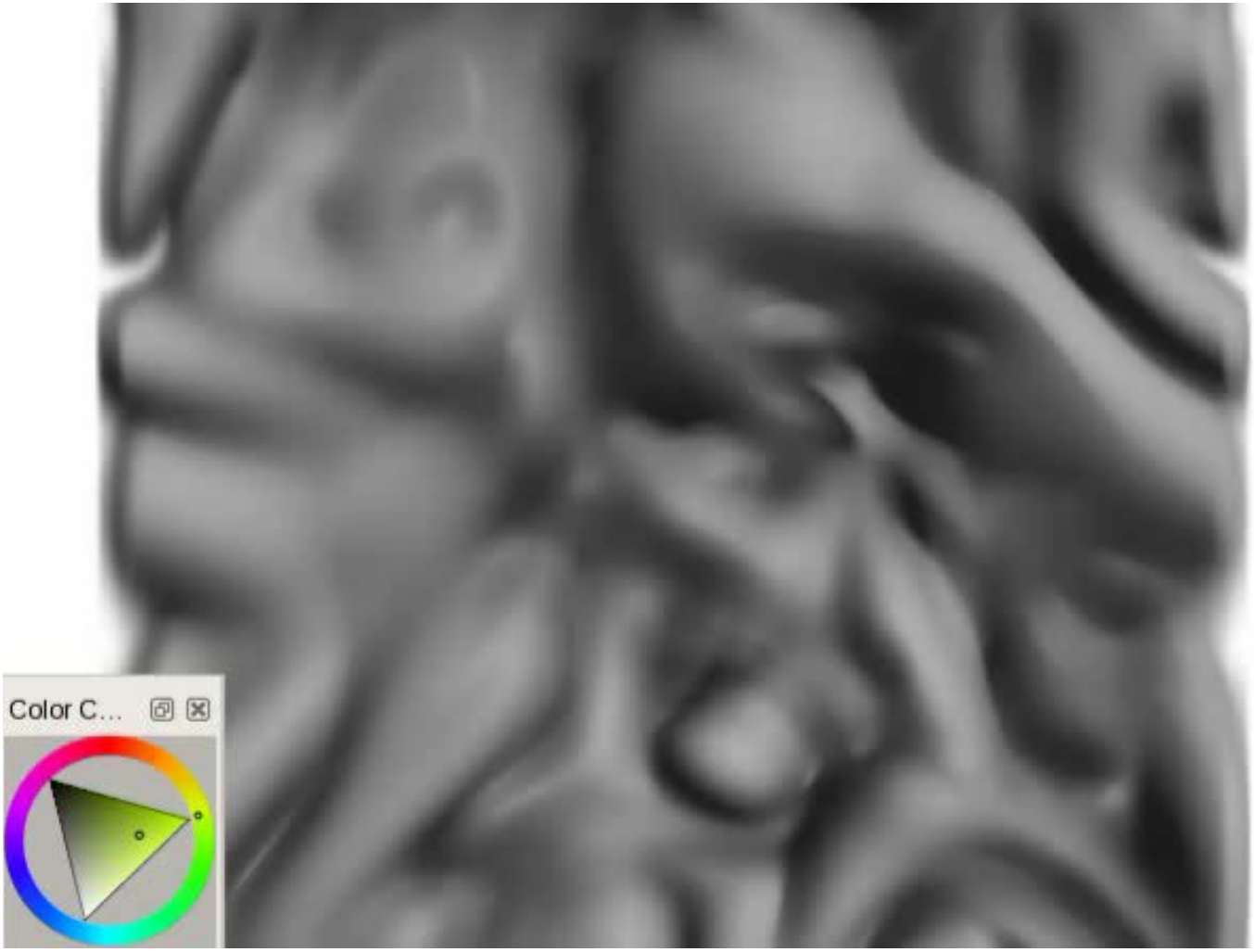
```
0265640 132304 133732 032051 037334 024721 015013 052226 001662
0265660 025537 064663 054606 043244 074076 124153 135216 126614
0265700 144210 056426 044700 042650 165230 137037 003655 006254
0265720 134453 124327 176005 027034 107614 170774 073702 067274
0265740 072451 007735 147620 061064 157435 113057 155356 114603
0265760 107204 102316 171451 046040 120223 001774 030477 046673
0266000 171317 116055 155117 134444 167210 041405 147127 050505
0266020 004137 046472 124015 134360 173550 053517 044635 021135
0266040 070176 047705 113754 175477 105532 076515 177366 056333
0266060 041023 074017 127113 003214 037026 037640 066171 123424
0266100 067701 037406 140000 165341 072410 100032 125455 056646
0266120 006716 071402 055672 132571 105645 170073 050376 072117
0266140 024451 007424 114200 077733 024434 012546 172404 102345
0266160 040223 050170 055164 164634 047154 126525 112514 032315
0266200 016041 176055 042766 025015 176314 017234 110060 014515
0266220 117156 030746 154234 125001 151144 163706 136237 164376
0266240 137055 062276 161755 115466 005322 132567 073216 002655
0266260 171466 126161 117155 065763 016177 014460 112765 055527
0266300 003767 175367 104754 036436 172172 150750 043643 145410
0266320 072074 000007 040627 070652 173011 002151 125132 140214
0266340 060115 014356 015164 067027 120206 070242 033065 131334
0266360 170601 170106 040437 127277 124446 136631 041462 116321
0266400 020243 005602 004146 121574 124651 006634 071331 102070
0266420 157504 160307 166330 074251 024520 114433 167273 030635
0266440 133614 106171 144160 010652 007365 026416 160716 100413
0266460 026630 007210 000630 121224 076033 140764 000737 003276
0266500 114060 042647 104475 110537 066716 104754 075447 112254
0266520 030374 144251 077734 015157 002513 173526 035531 150003
0266540 146207 015135 024446 130101 072457 040764 165513 156412
0266560 166410 067251 156160 106406 136770 030516 064740 022032
0266600 142166 123707 175121 071170 076357 037233 031136 015232
0266620 075074 016744 044055 102230 110063 033350 052765 172463
```





Large Volume Data Visualization

Interaction in Vis

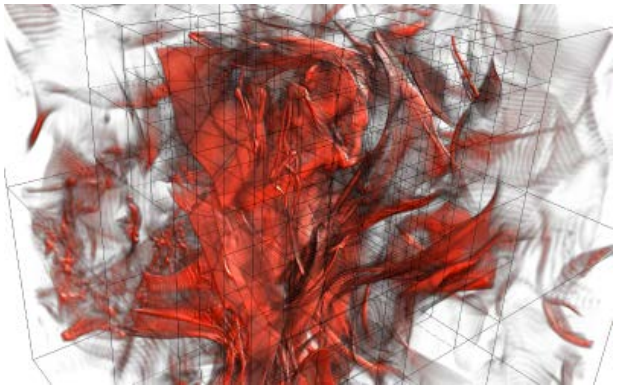


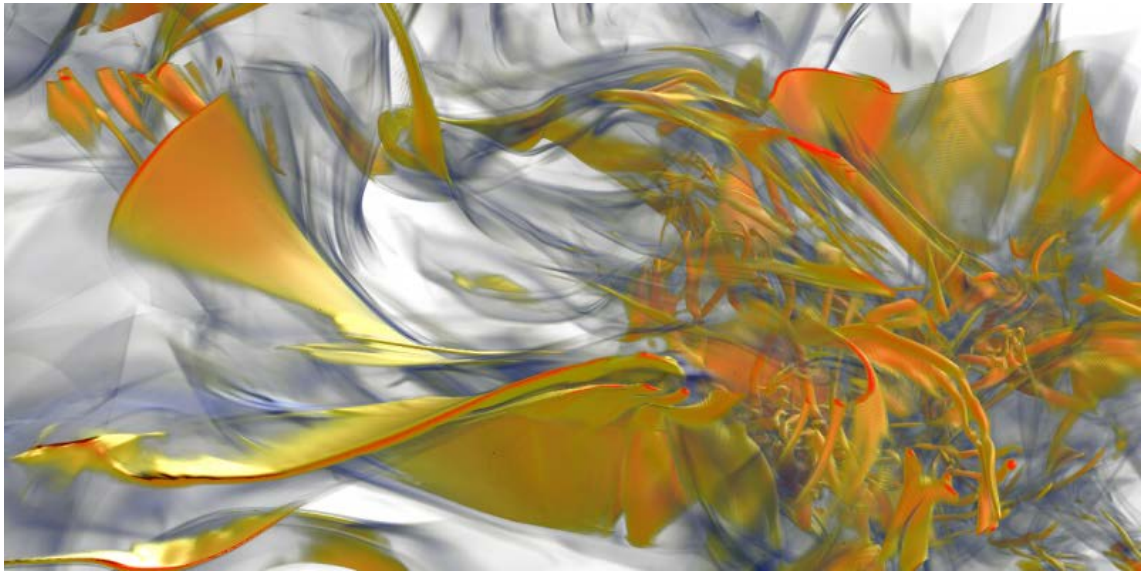
Collaborative Vis with Advanced UI

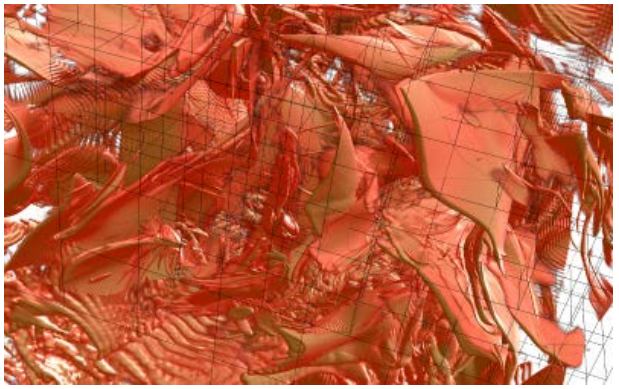

Work in Progress

Large Volume Visualization

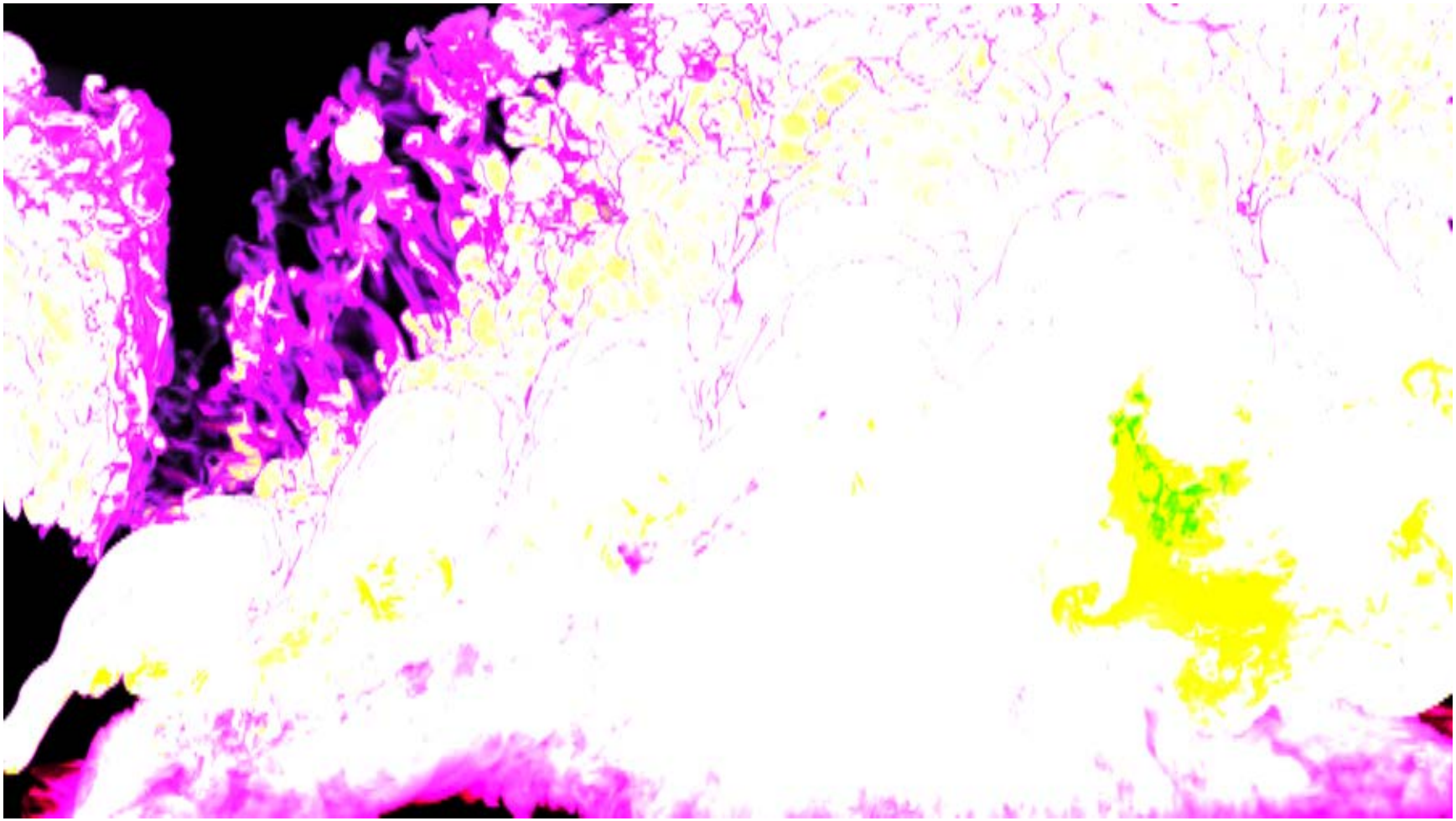
- Level of Details
- Out of Core
- Parallel Visualization





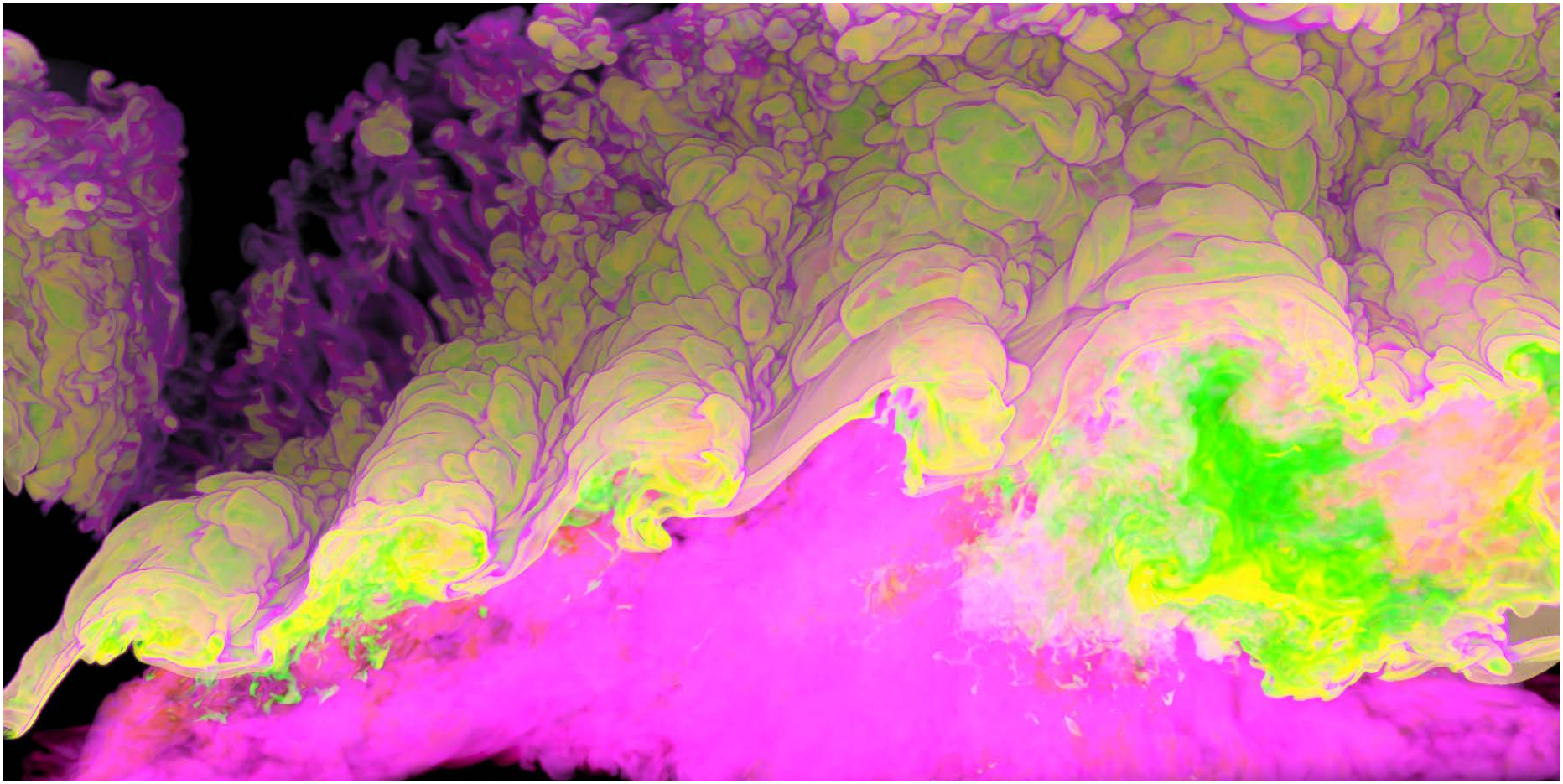


HDR VolVis – Turbulent Mixing of Air/SF_6



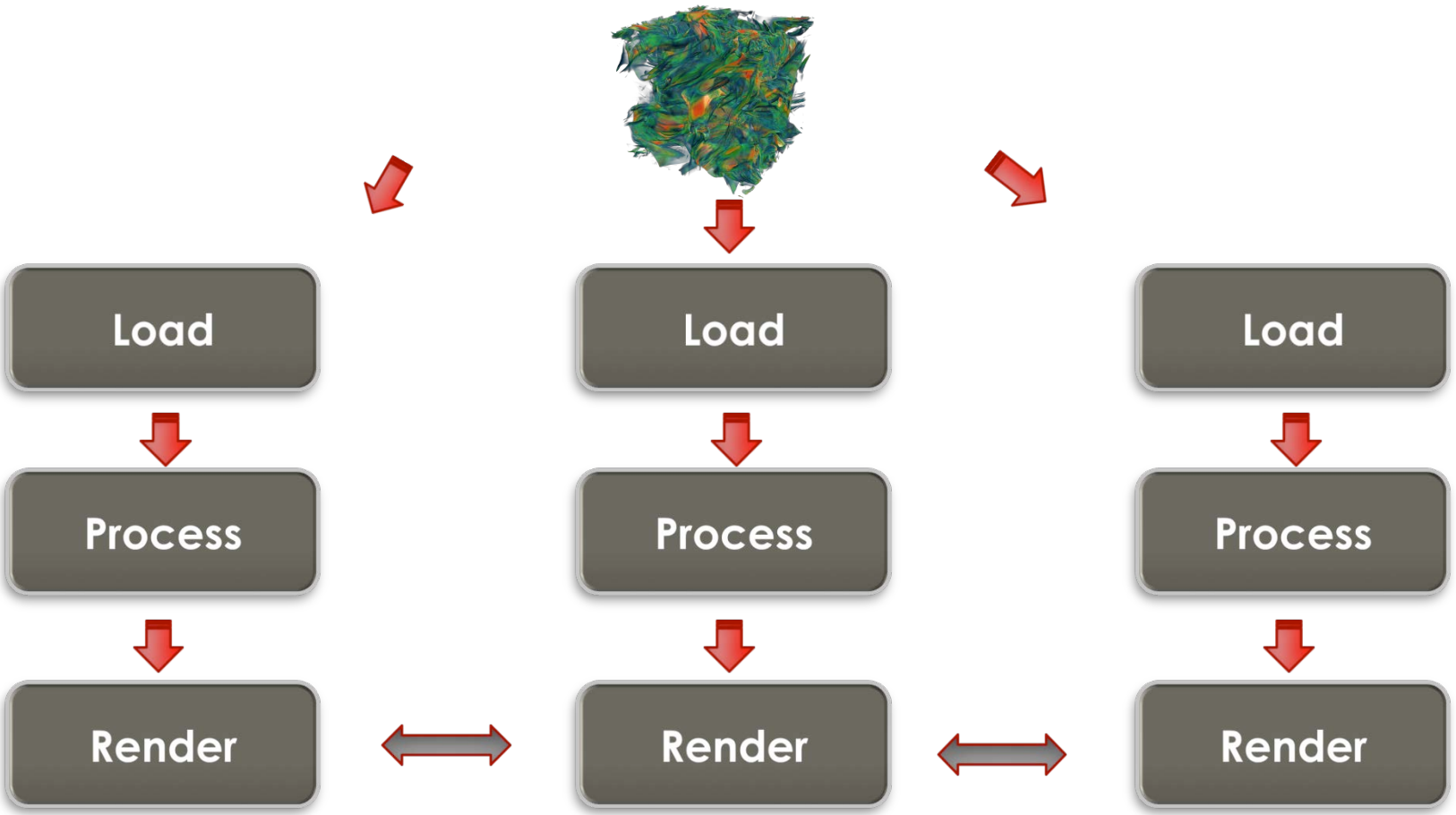
Without High dynamic range rendering results

HDR VolVis – Turbulent Mixing of Air/SF_6

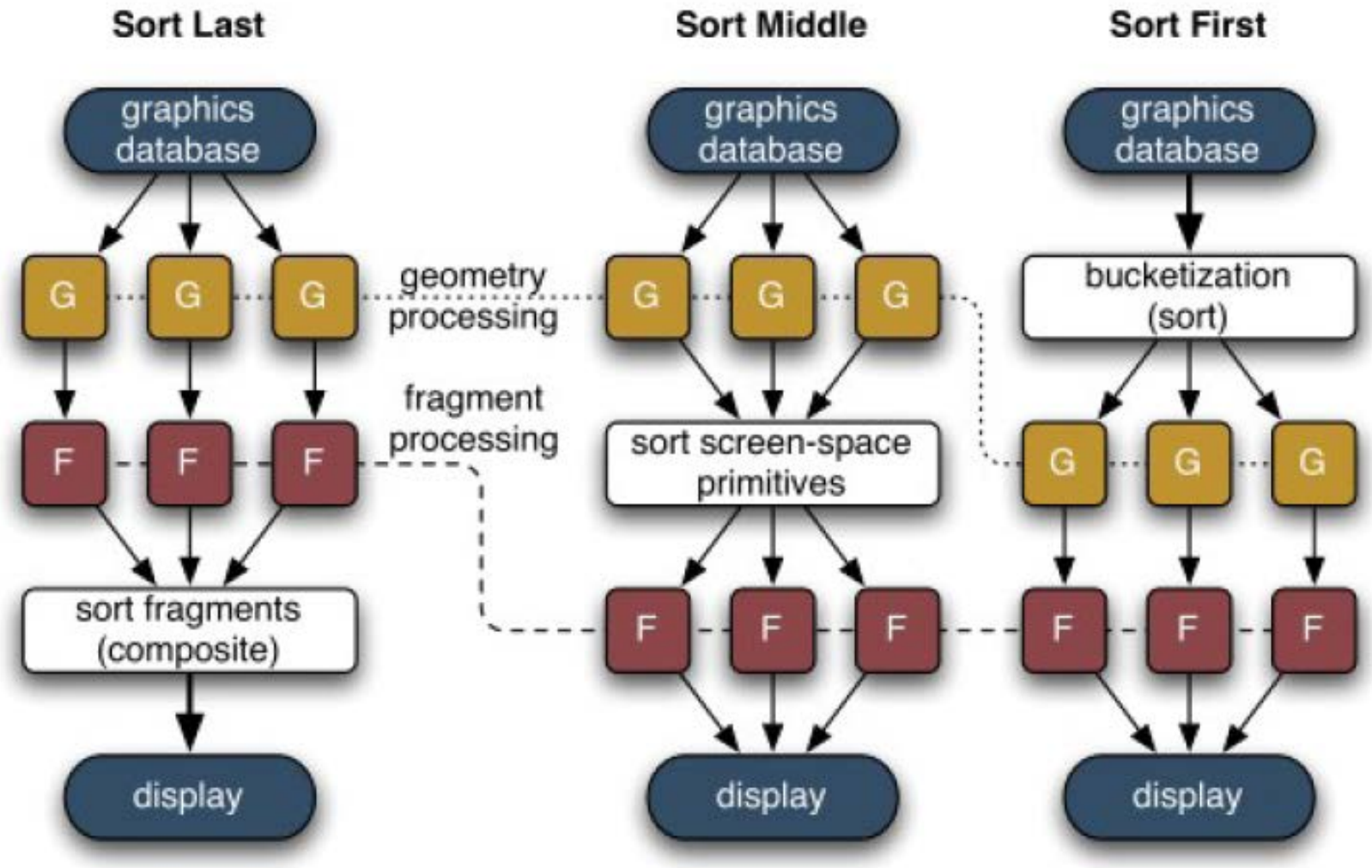


High dynamic range rendering results

Parallel Data (Volume) Vis.

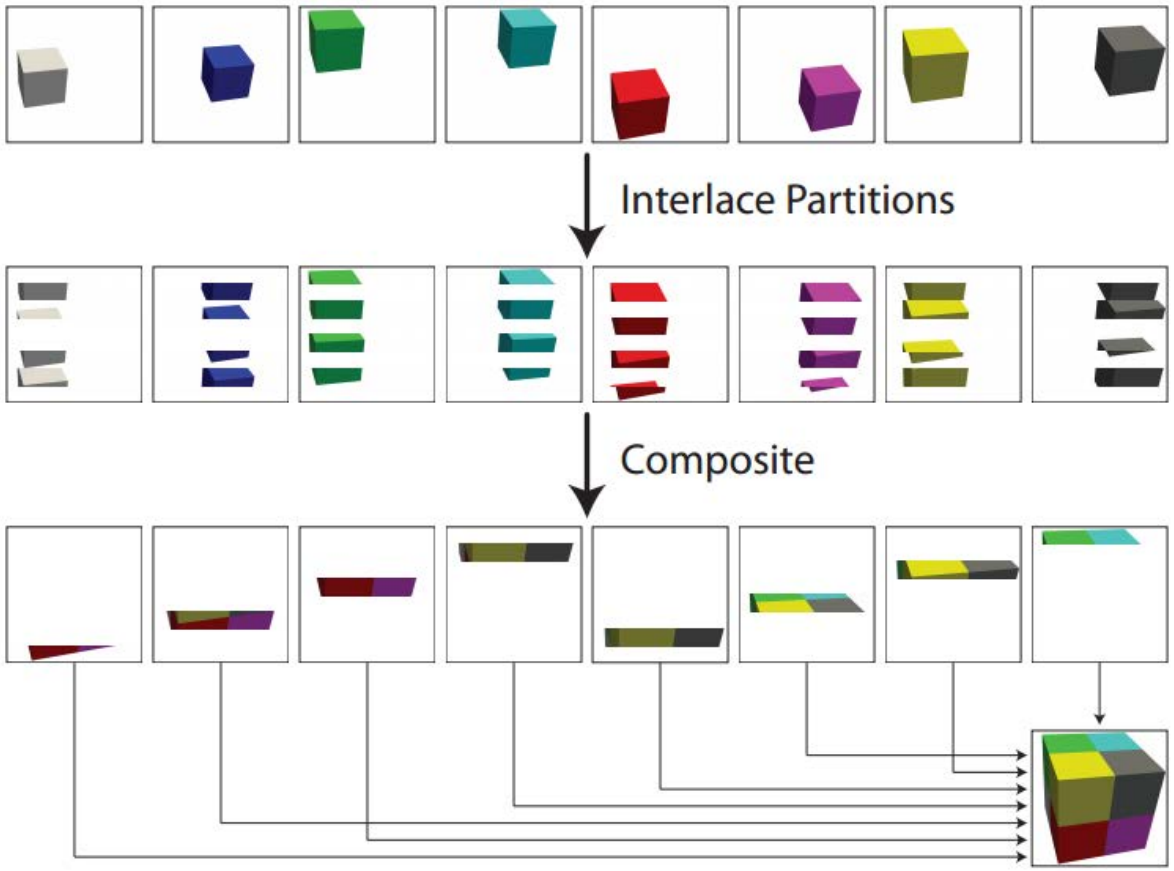


Parallel Image Compositing

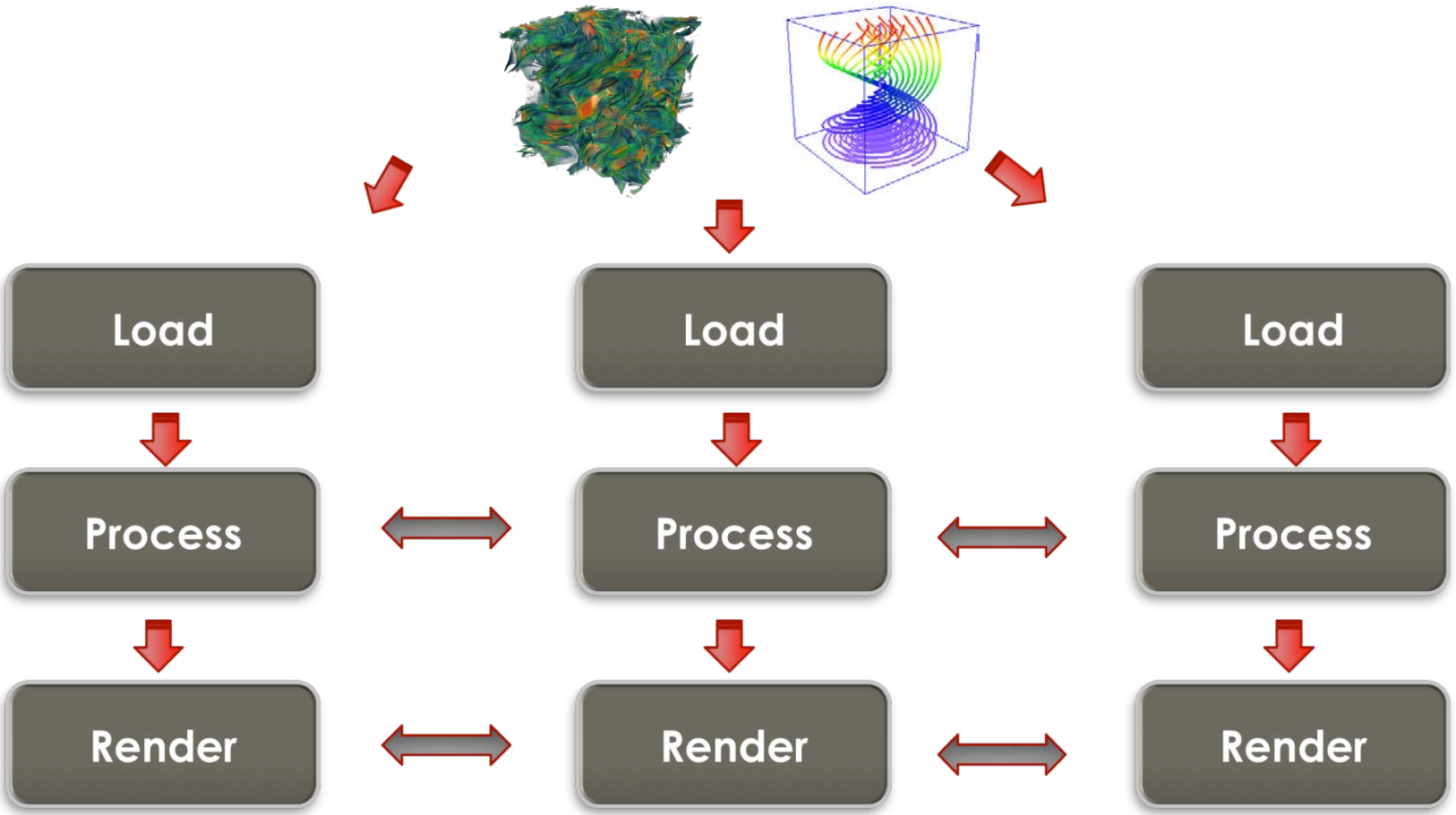


Sort-Last Image Compositing

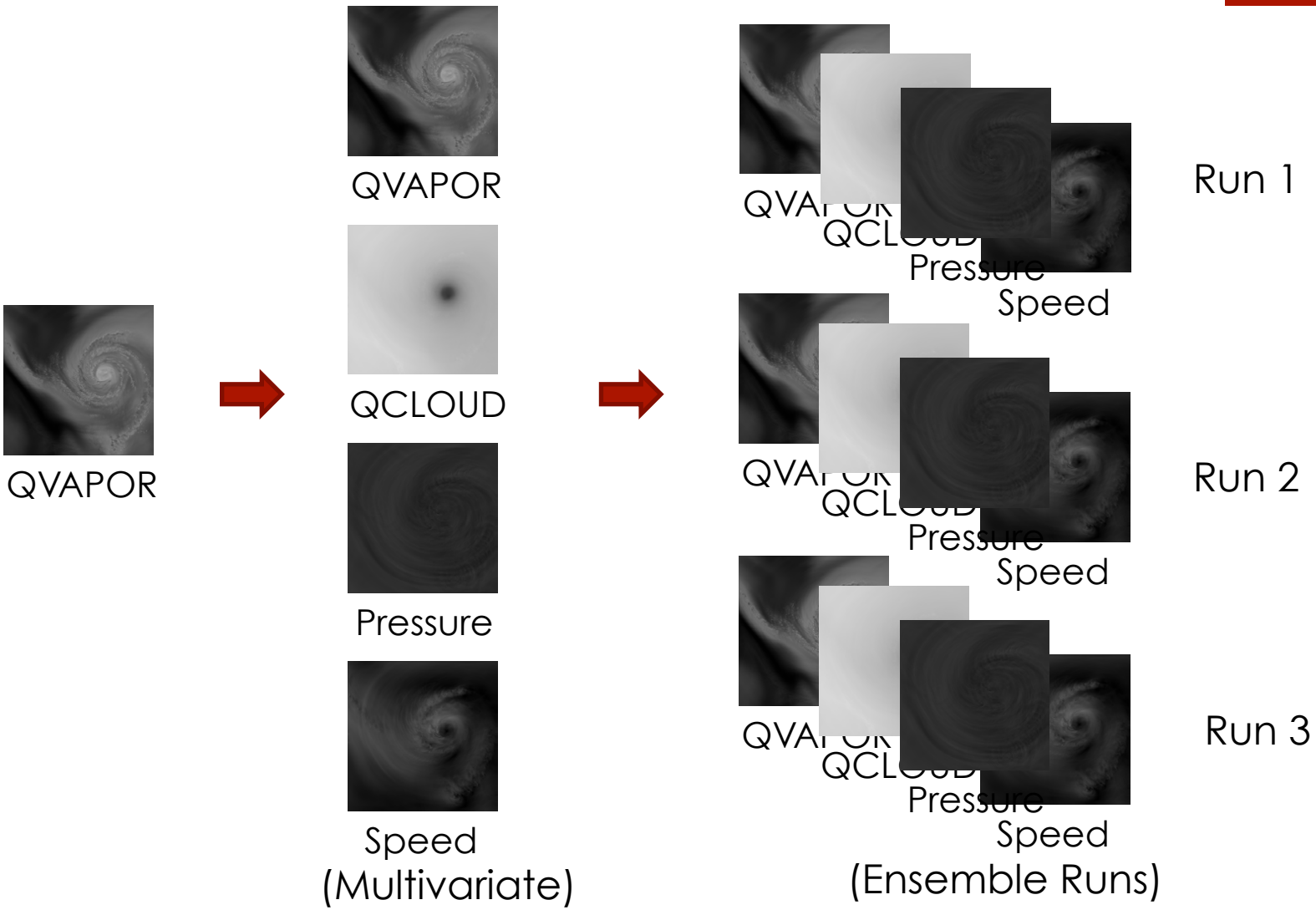
- Stages in final parallel visualization algorithms
  - Data partitioning
  - Local data rendering
  - Result image compositing



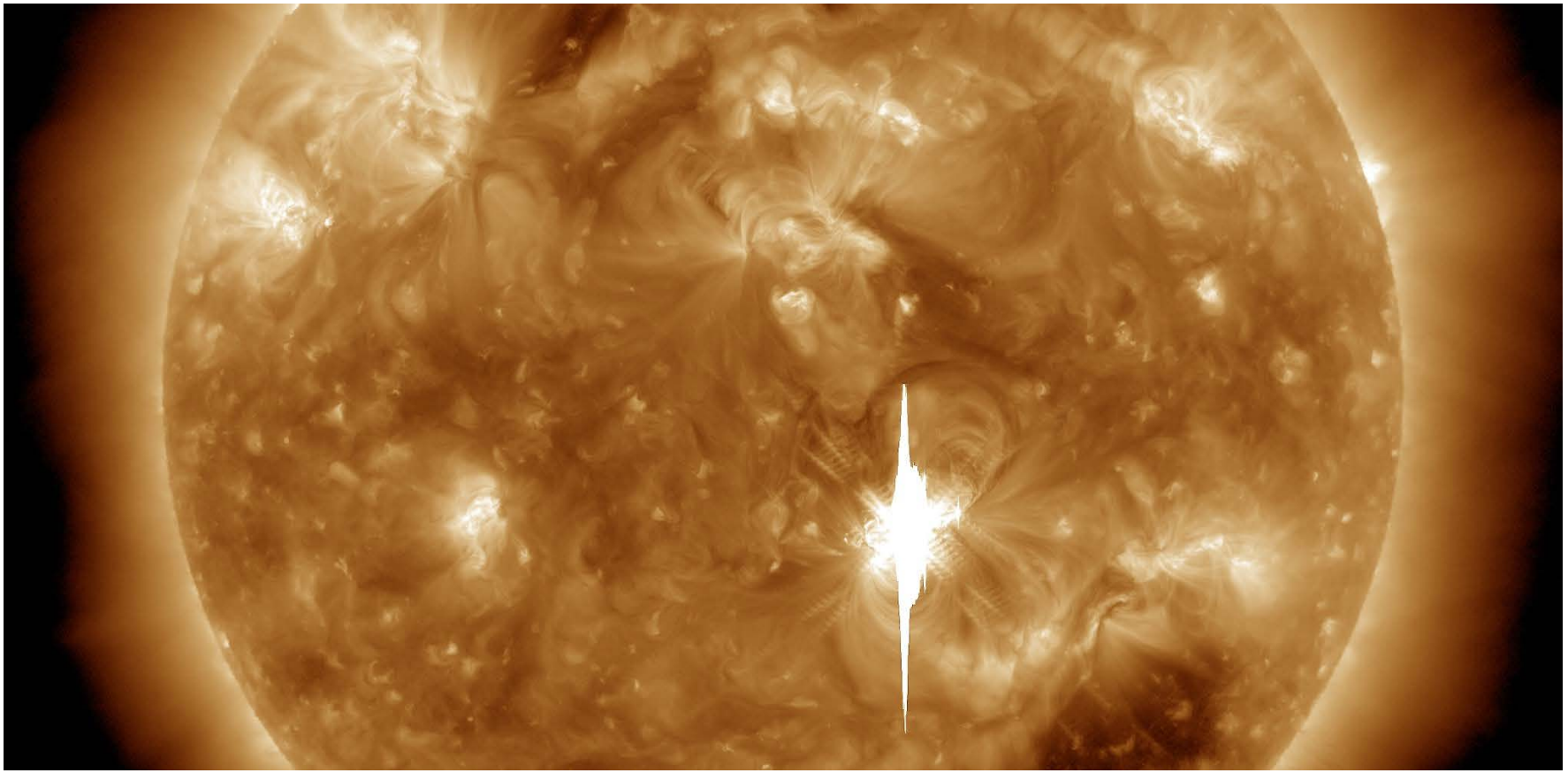
Parallel Data Vis.



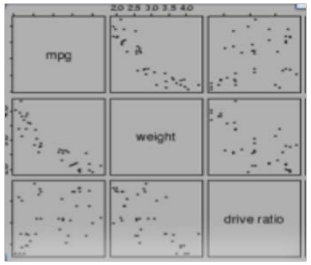
Multivariate to Multi-Run Visual Analysis



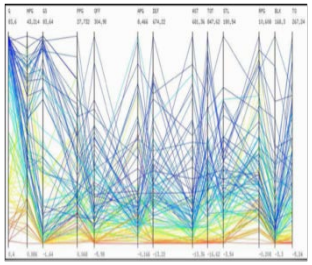
Multivariate Data



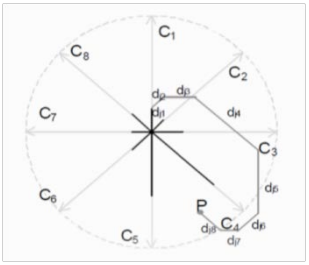
Multi-Dimensional Data Vis



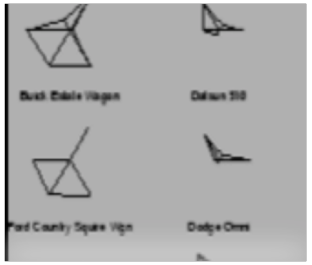
Scatter Plot Matrix



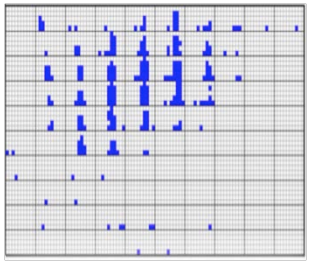
Parallel Coordinates



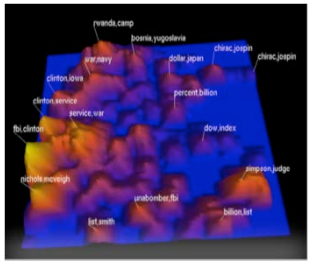
Star Coordinates



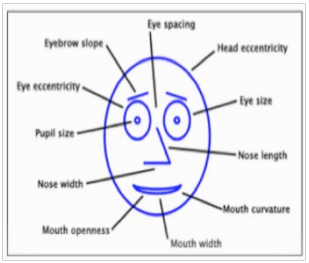
Star Glyphs



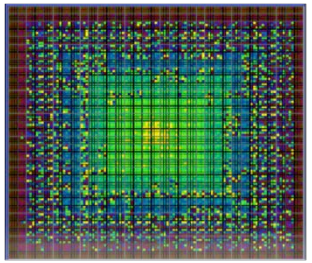
Dimension Stacking



Landscapes



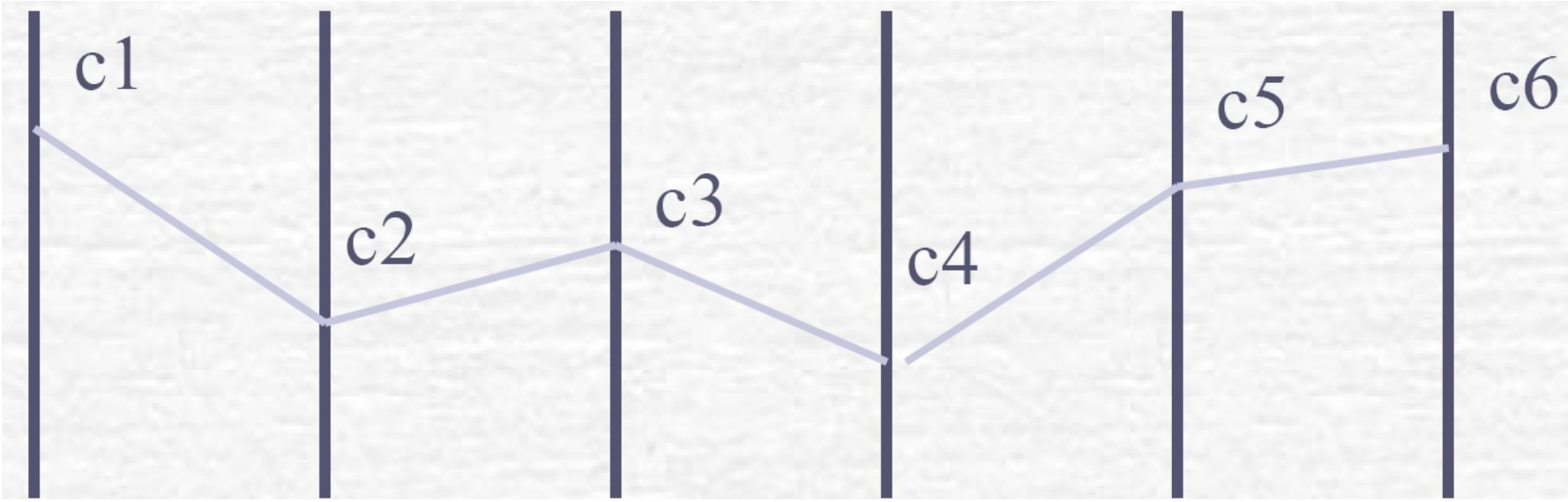
Chernoff faces



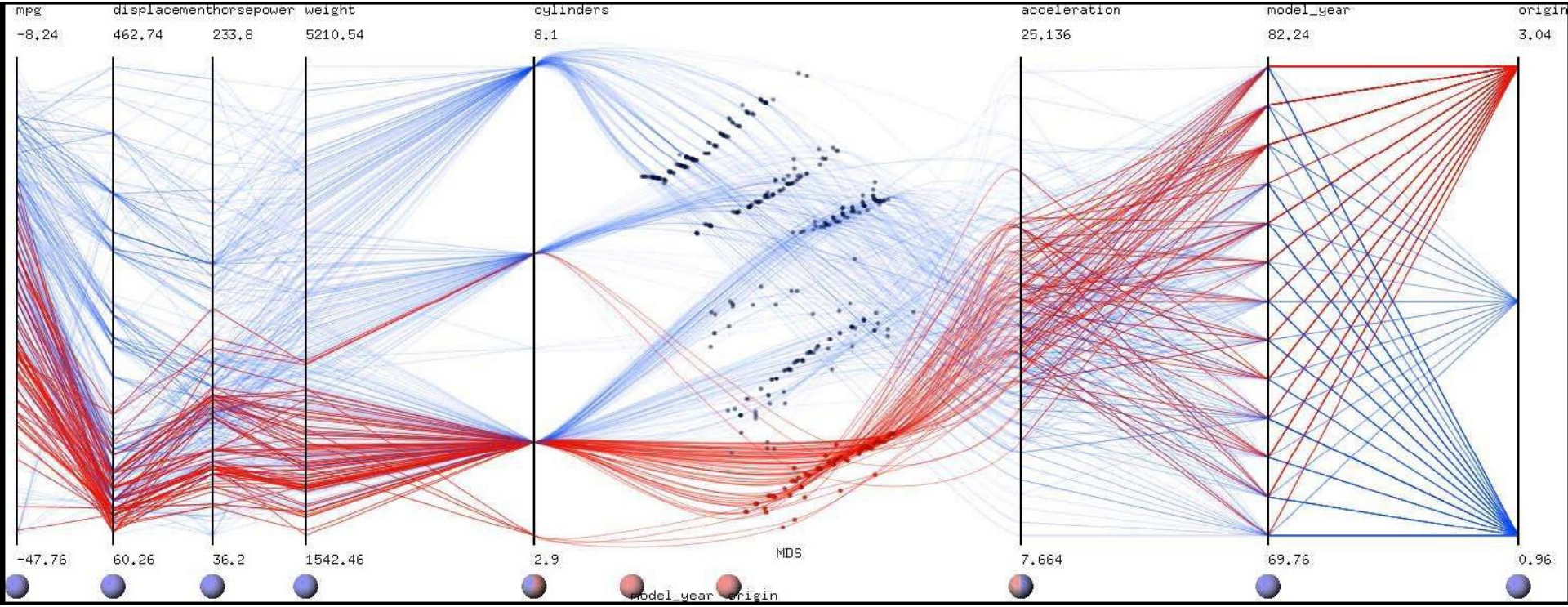
Pixel-based

Parallel Coordinates Basics

- To represent N dimensional data
  - Set N vertical axes in parallel
  - Put data to intersects on corresponding axes
  - Connect intersects

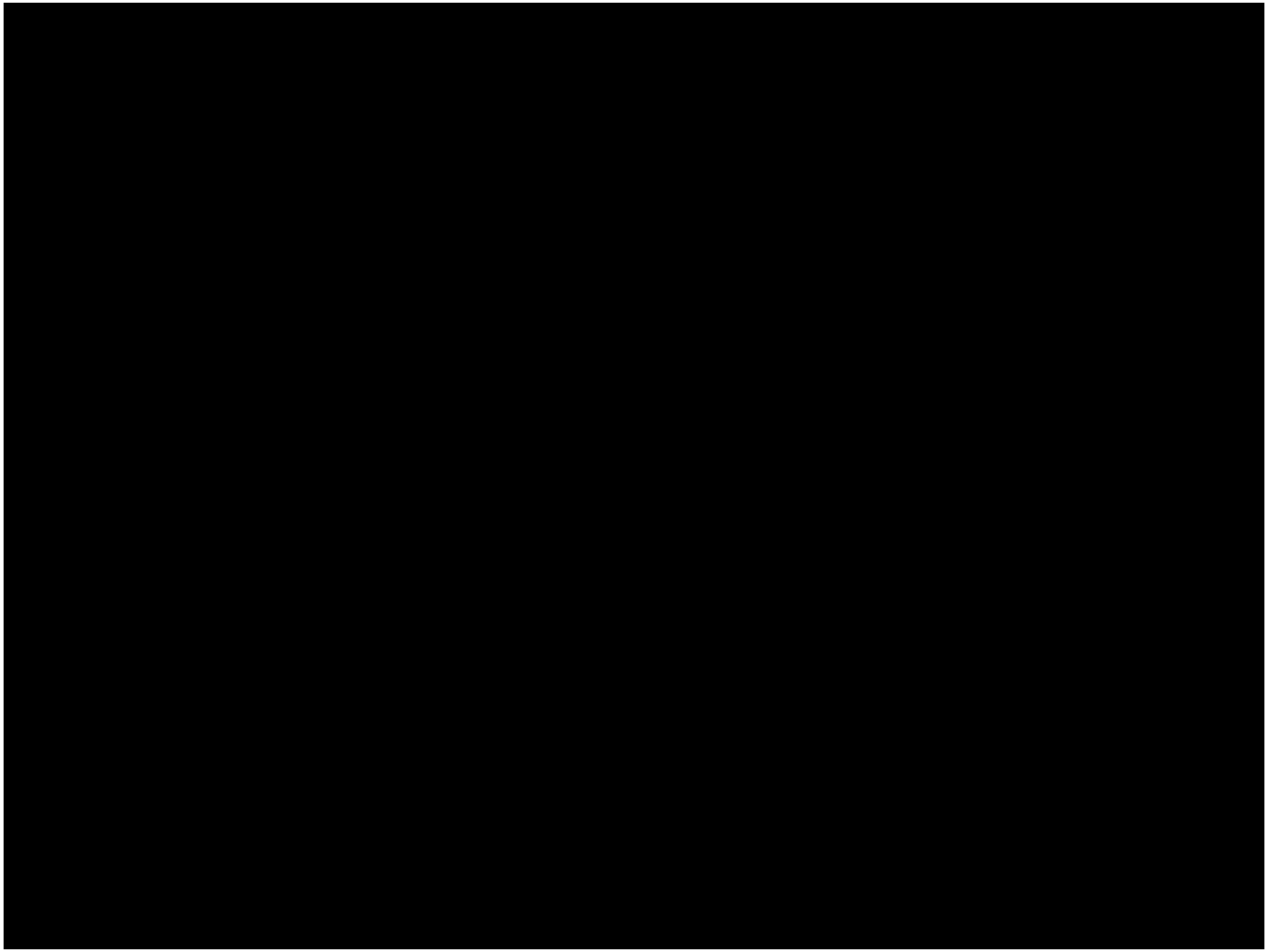


Data Exploration with SPPC

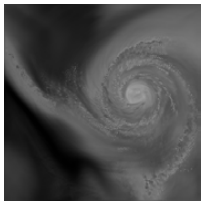


[Yuan et al. TVCG 2009 (InfoVis)]

Data Exploration with SPPC

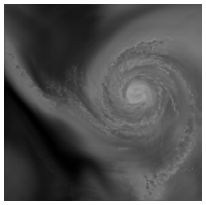


Multivariate Visual Analysis

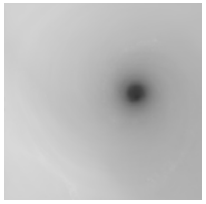


QVAPOR

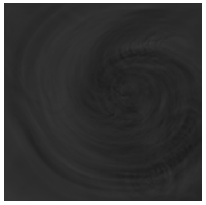
(Scalar)



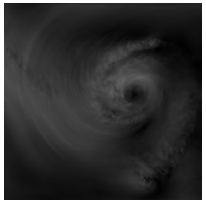
QVAPOR



QCLOUD



Pressure



Speed

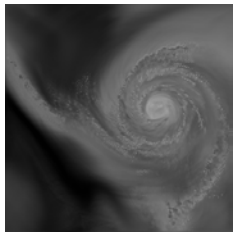
(Multivariate)

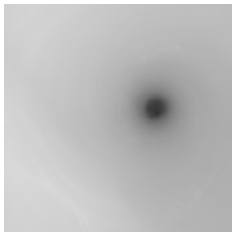
Variable	Description	Min / Max
QCLOUD	Cloud water	0.00000 / 0.00332
QGRAUP	Graupel	0.00000 / 0.01638
QICE	Cloud ice	0.00000 / 0.00099
QRAIN	Rain	0.00000 / 0.01132
QSNOW	Snow	0.00000 / 0.00135
QVAPOR	Water vapor	0.00000 / 0.02368
CLOUD	Total cloud (QICE + QCLOUD)	0.00000 / 0.00332
PRECIP	Total precipitation (QGRAUP+QRAIN+QSNOW)	0.00000 / 0.01672
P	Pressure: weight of the atmosphere above a grid point	-5471.85791 / 3225.42578
TC	Tempreature	-83.00402/31.51576
U	X wind component: west-east wind component in model coordinate; positive means winds blow from west to east	-79.47297/85.17703
V	Y wind component: south-north wind component in model coordinate; positive means winds blow from south to north	-76.03391/82.95293
W	Z wind component: vertical wind component in model coordinate; positive means upward motion	-9.06026/28.61434

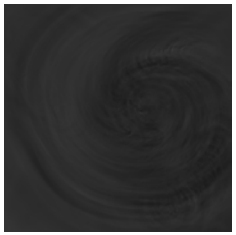
Isabel Hurricane

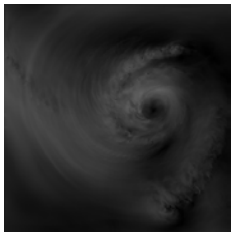
Transfer Functions

- Multivariate TFs for multi-modal data

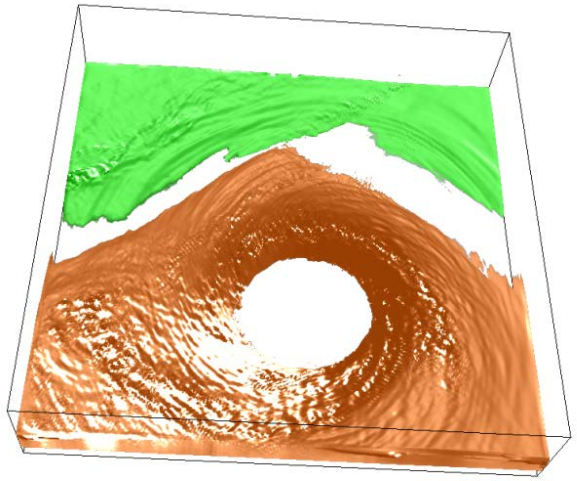

QVAPOR

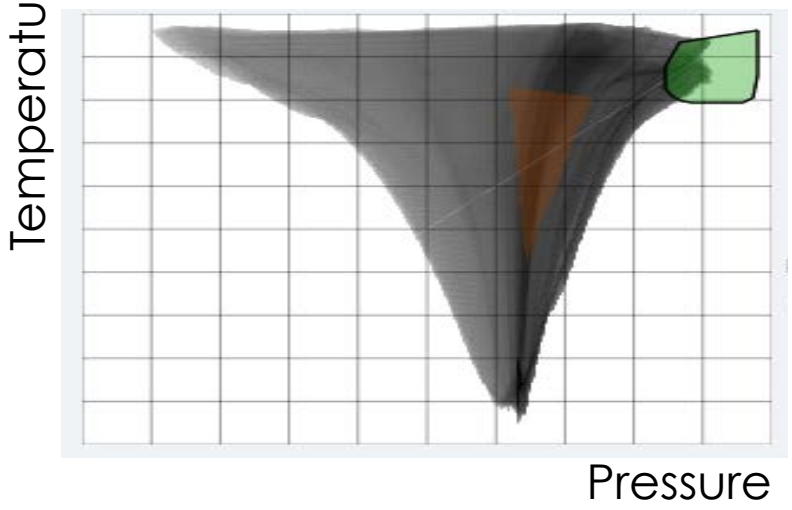

QCLOUD


Pressure


Speed

...

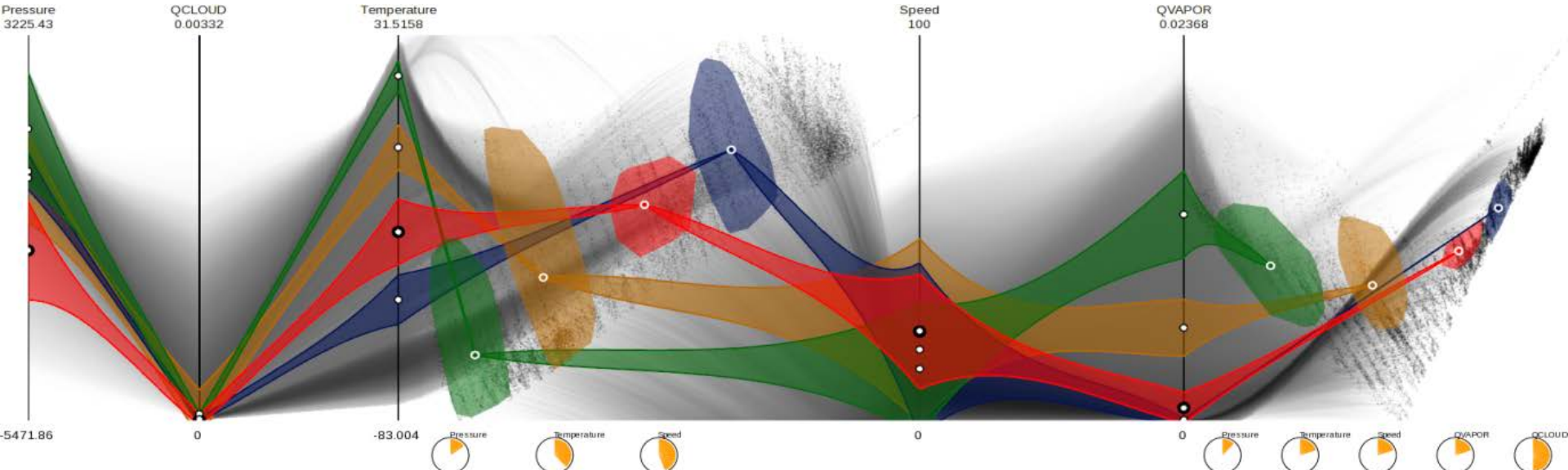


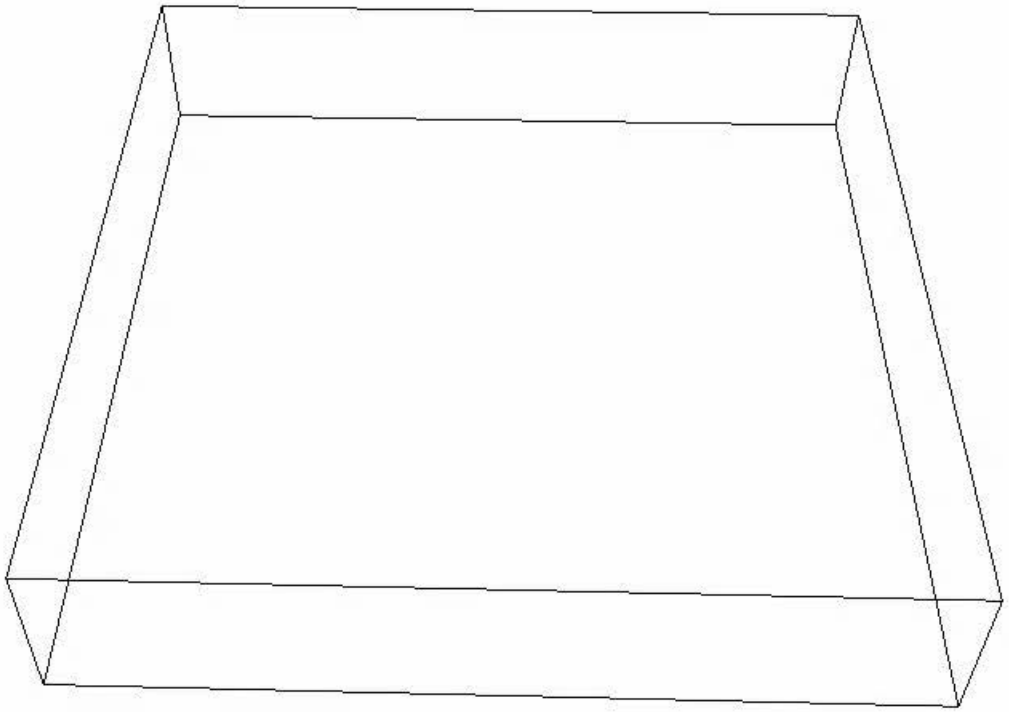


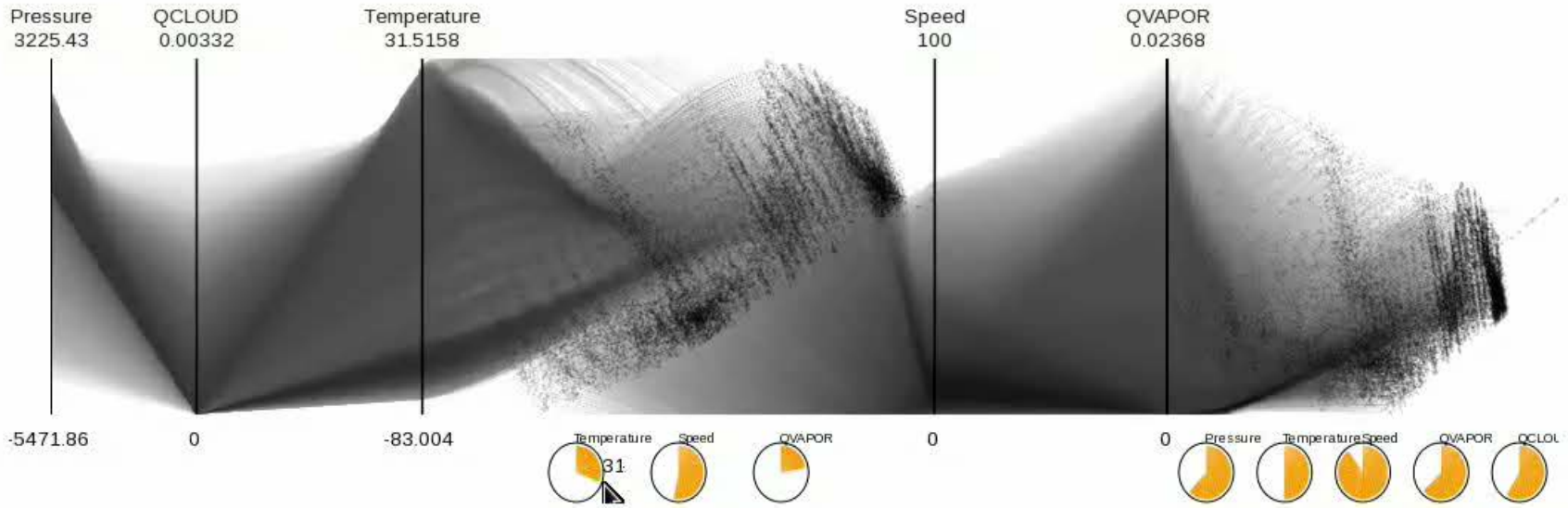
High Dimensional Transfer Function Design

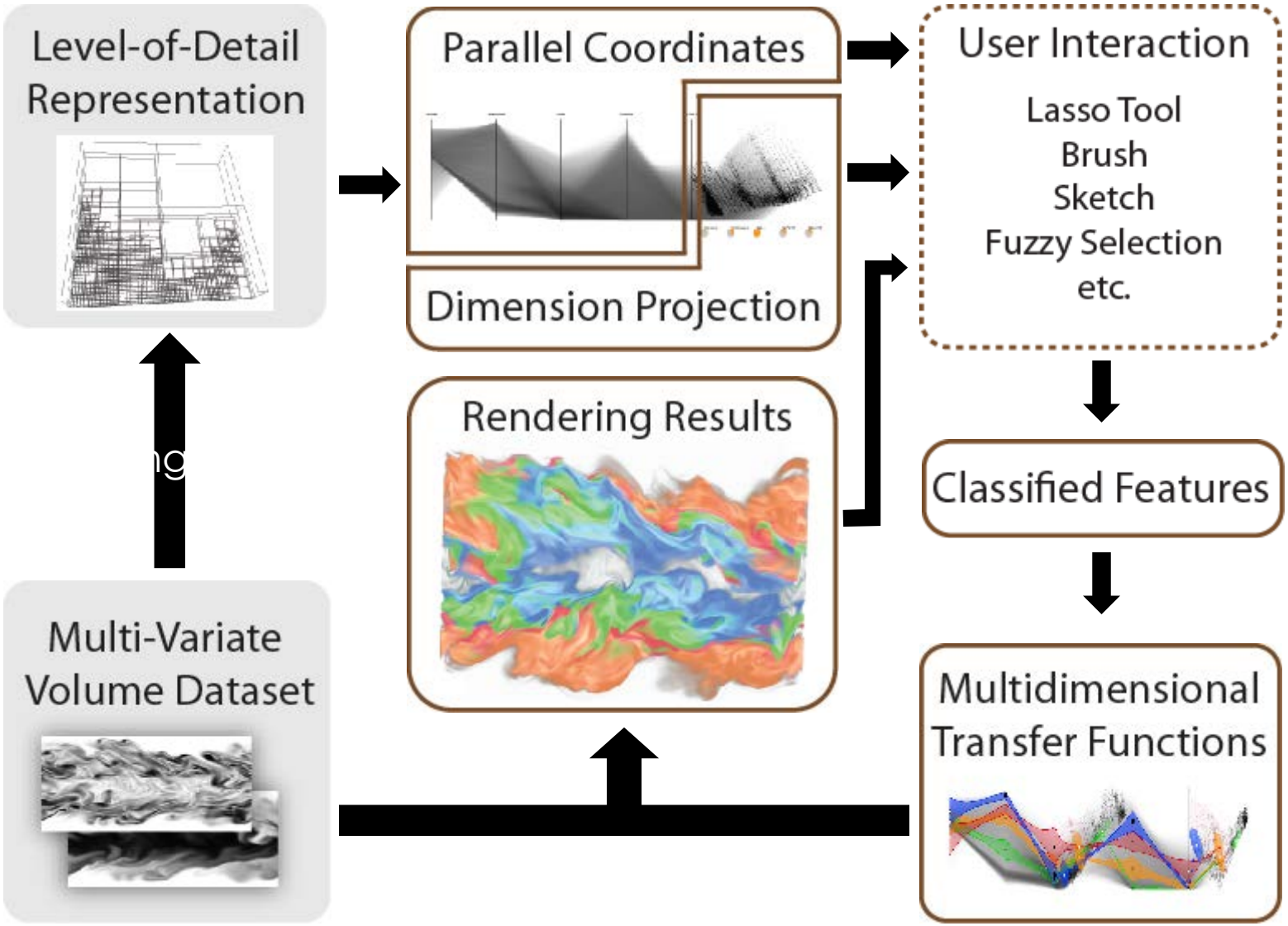
- Integrate MDS into PCP to facilitate multivariate TF design.
  - Avoids context jumps between polyline and point regions when exploring data clusters
  - Provides multiple perspective views upon the data, supporting linked queries

[Guo, Xiao, and Yuan. PacificVis 2011]

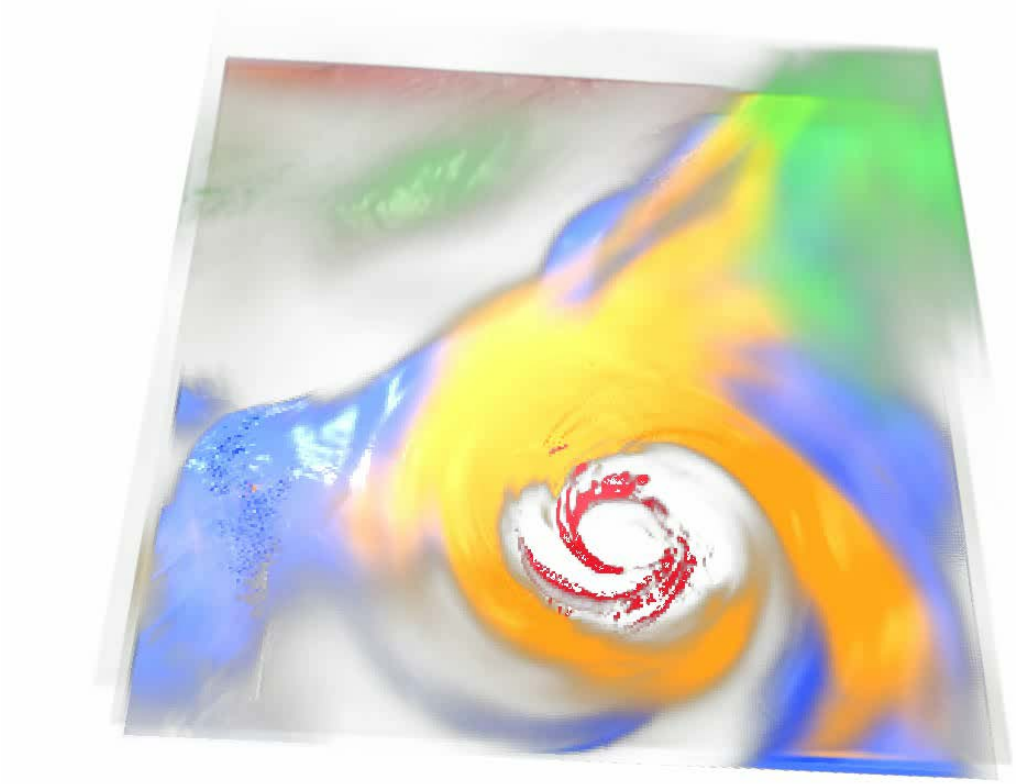


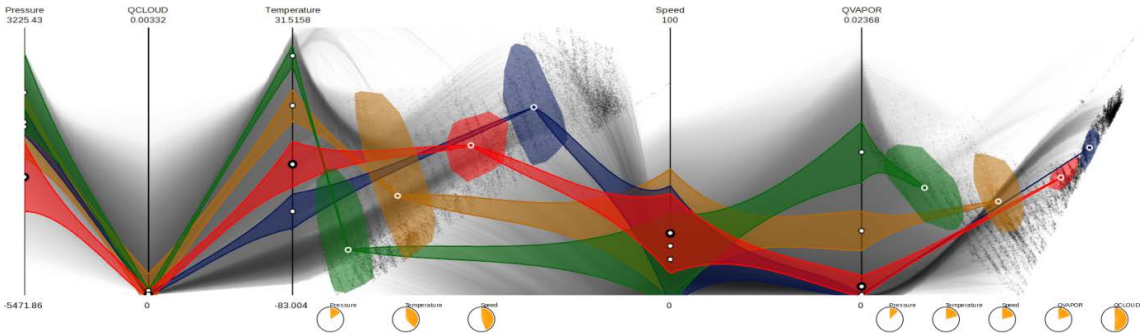



Pressure
3225.43
QCLOUD
0.00332
Temperature
31.5158
Speed
100
QVAPOR
0.02368
-5471.86
0
-83.004
0
0
Temperature
Speed
QVAPOR
31
Pressure
Temperature
Speed
QVAPOR
QCLOU


Level-of-Detail Representation
Parallel Coordinates
Dimension Projection
User Interaction
Lasso Tool
Brush
Sketch
Fuzzy Selection
etc.
Rendering Results
Classified Features
Multi-Variate Volume Dataset
Multidimensional Transfer Functions

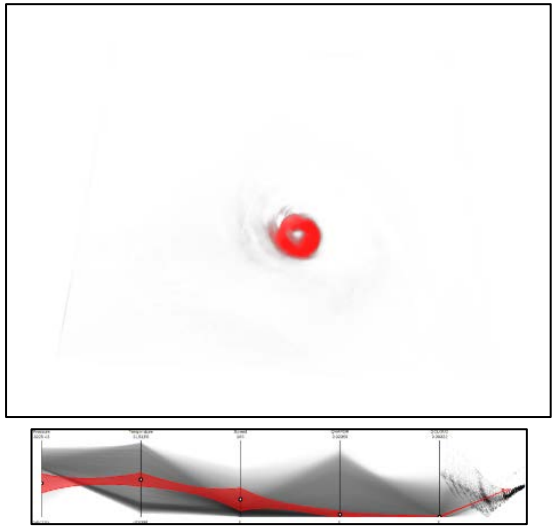
Case Study – Hurricane Isabel

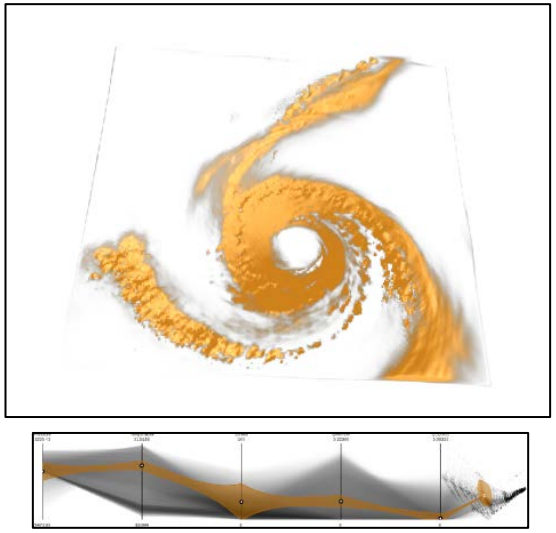


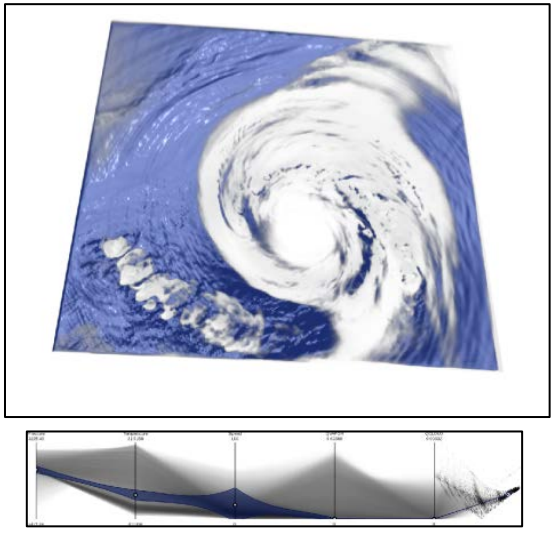


Case Study – Hurricane Isabel

- Different components

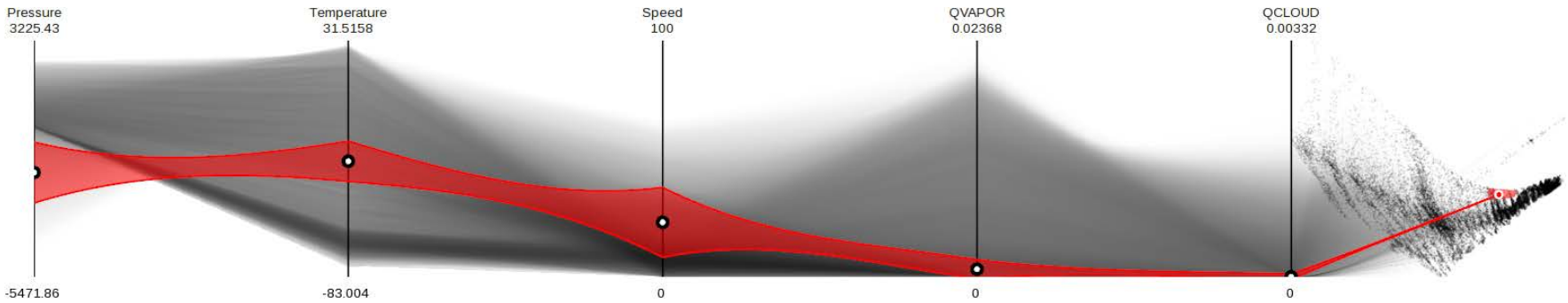






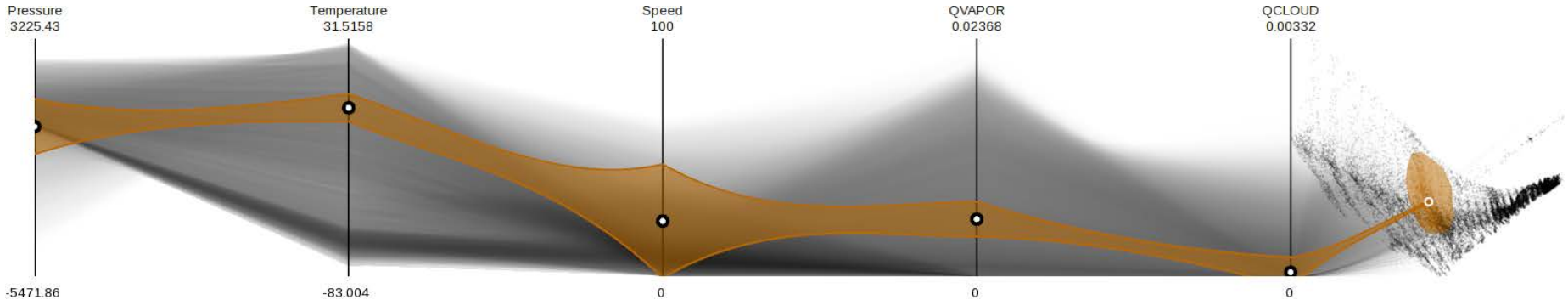
Case Study – Hurricane Isabel

Red part (hurricane eye):
low pressure,
medium temperature,
lower QCLOUD,
high wind speed

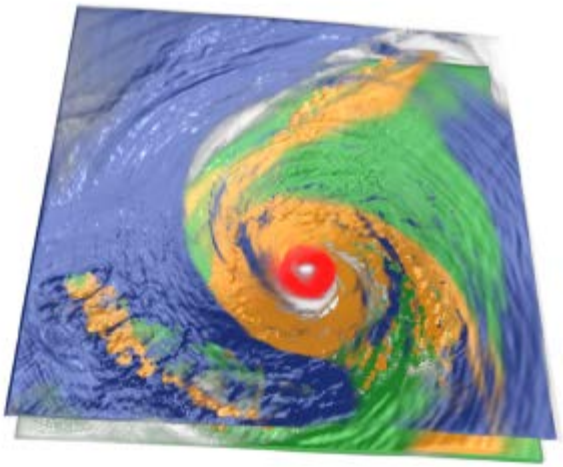


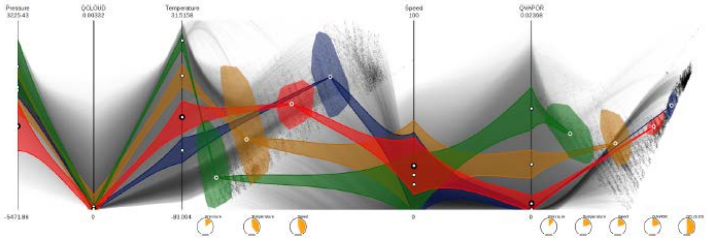
Case Study – Hurricane Isabel

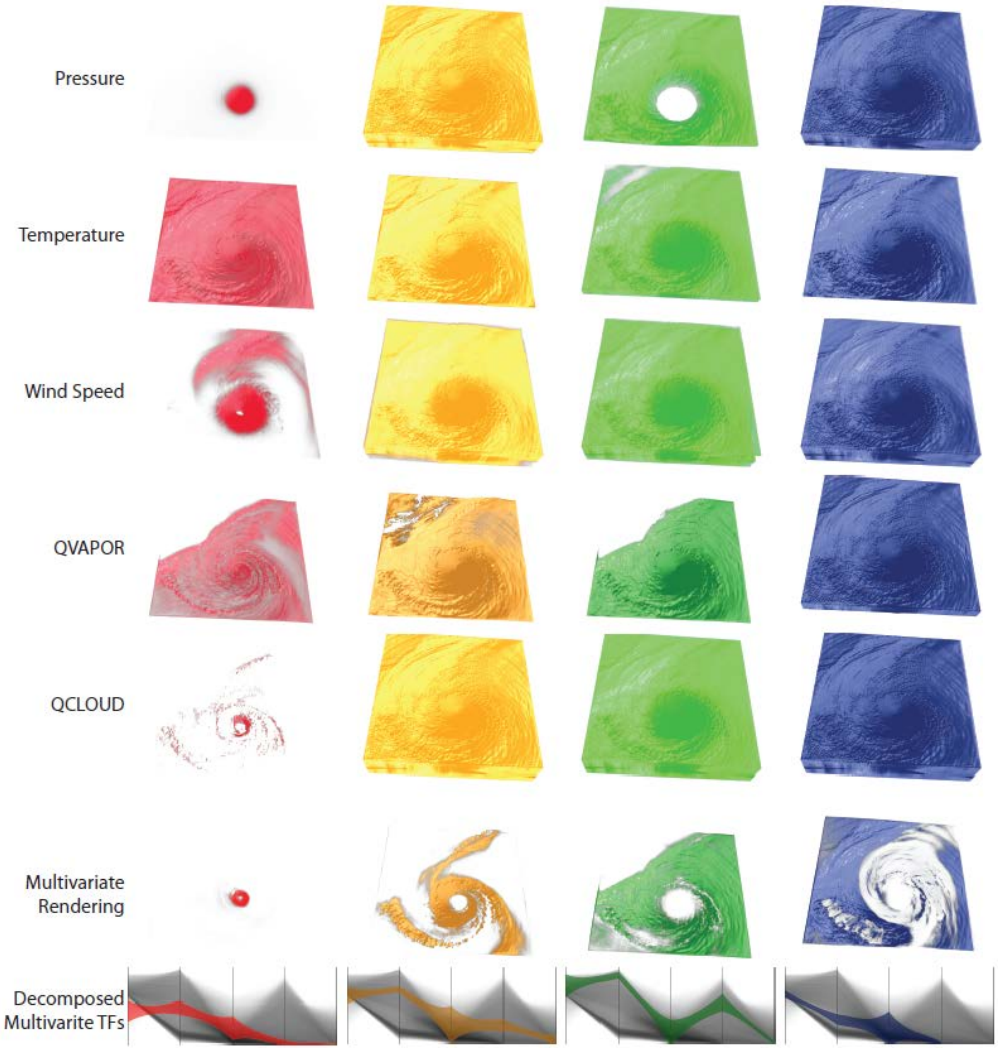
Yellow part:
higher pressure
lower wind speed



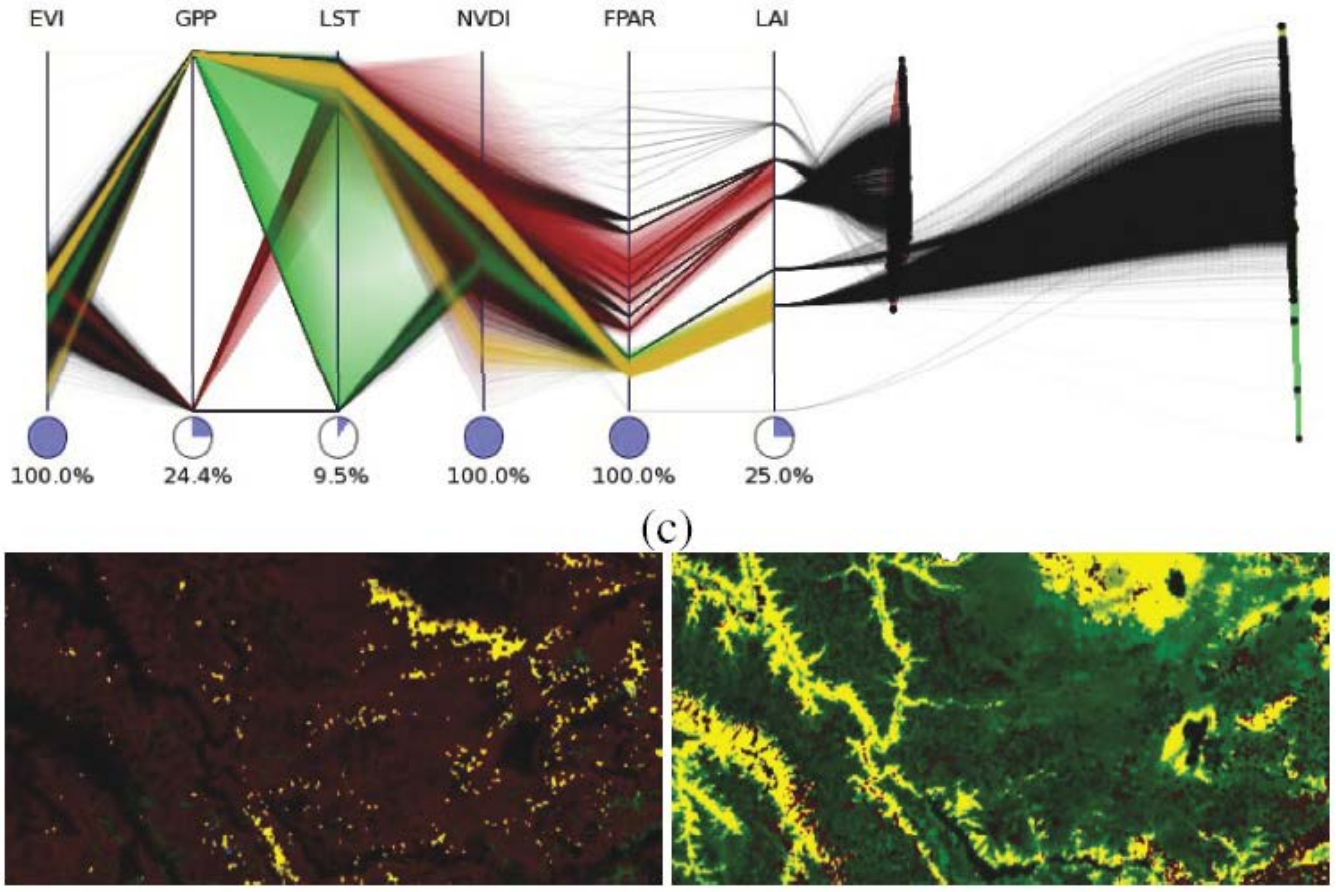
Hurricane



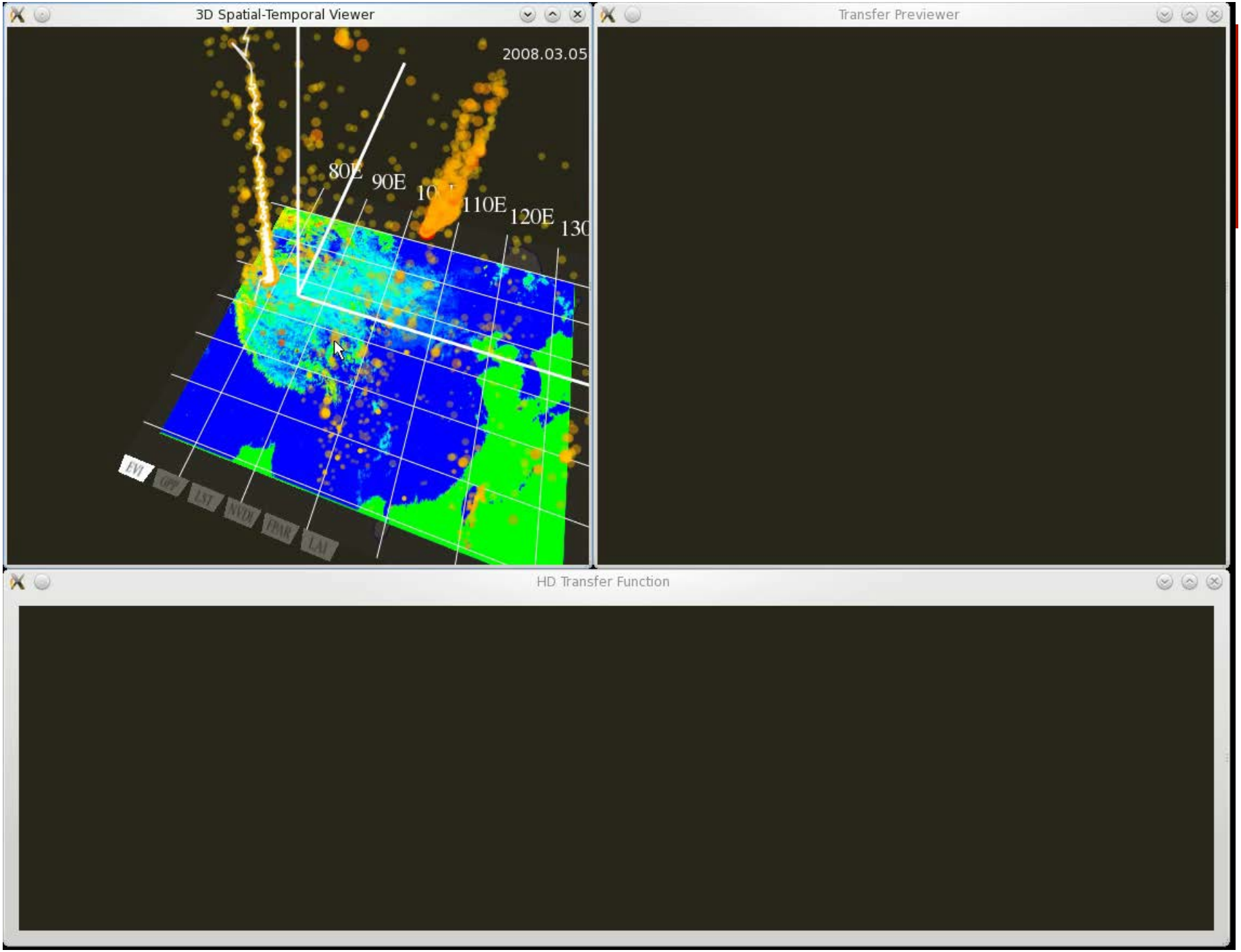



Pressure
Temperature
Wind Speed
QVAPOR
QCLOUD
Multivariate Rendering
Decomposed Multivarite TFs

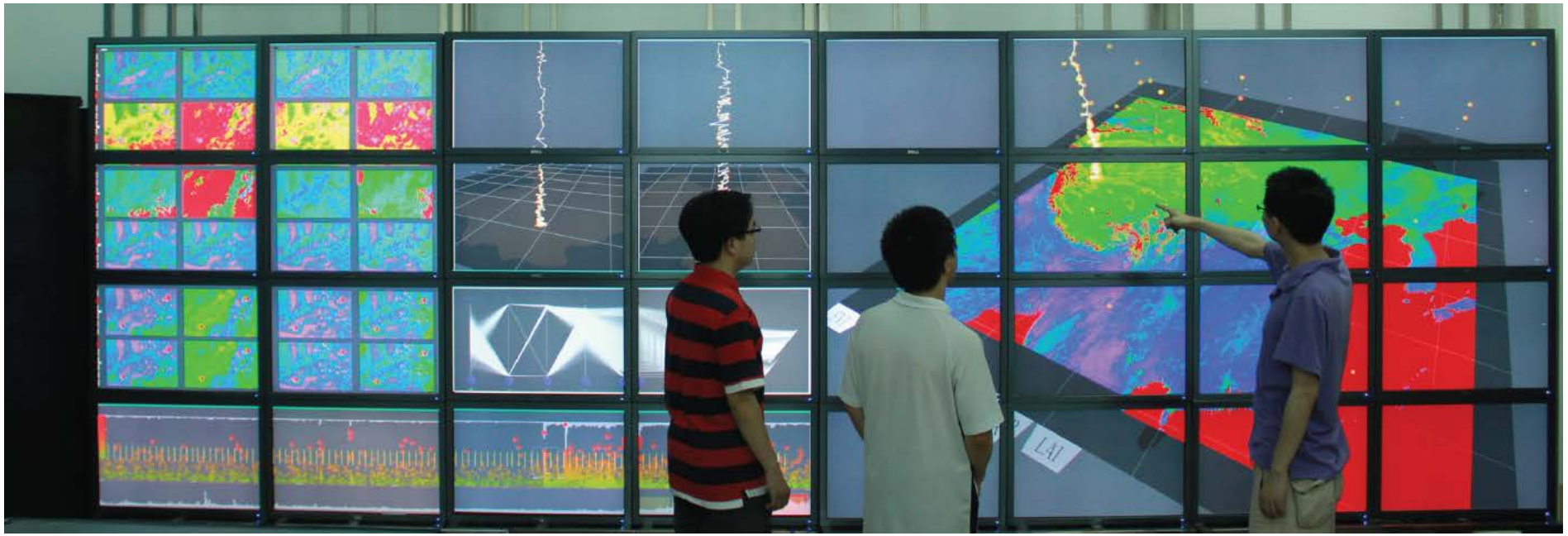
Multivariate Analytics for Images



[Yuan et al. TVCG 2010 (Vis)]


3D Spatial-Temporal Viewer
2008.03.05
80E
90E
110E
120E
Transfer Previewer
HD Transfer Function

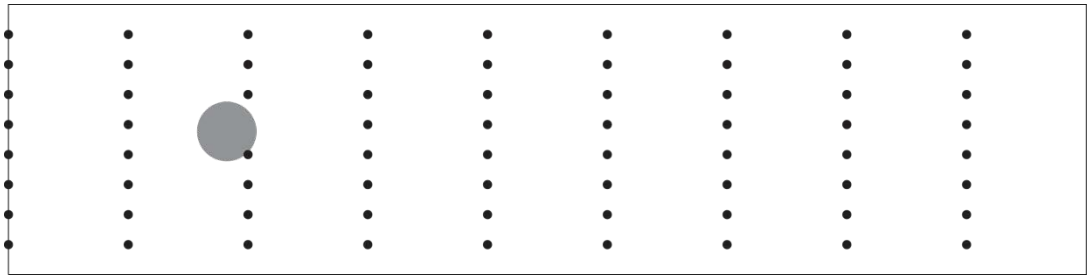
Scalable Multivariate Analytics of Seismic and Satellite-based Observational Data



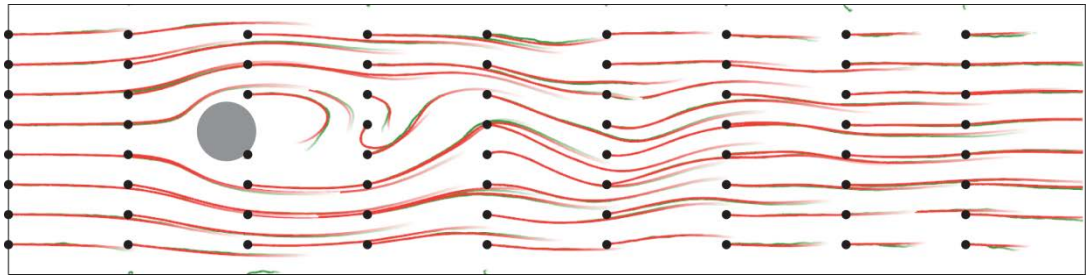
[Yuan et al. TVCG 2010 (Vis)]

Eulerian and Lagriangian Specifications

- Eulerian: $\mathbf{v} = \mathbf{v}(\mathbf{x},t), p = p(\mathbf{x},t), T = T(\mathbf{x},t),$



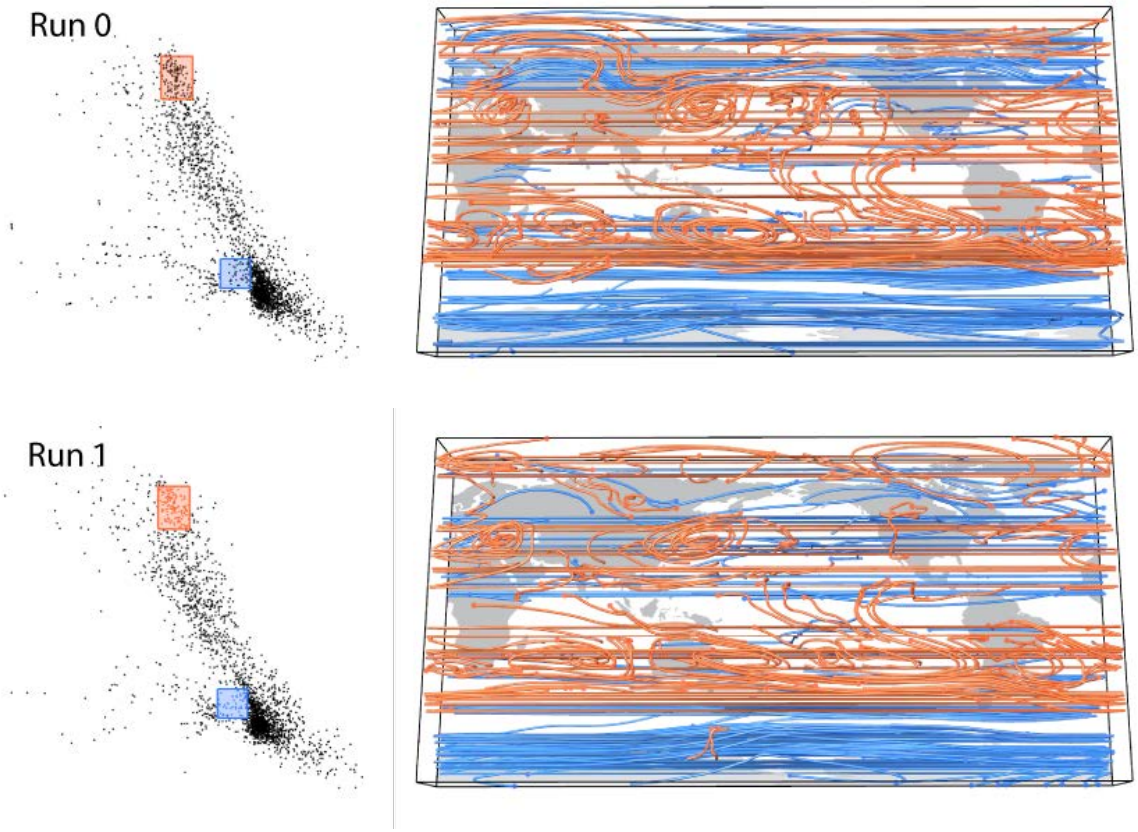
- Lagriangian: $\mathbf{X} = \mathbf{X}(\mathbf{a},t), p = p(\mathbf{a},t), T = T(\mathbf{a},t),$



- Relationships between two specifications (flow map):

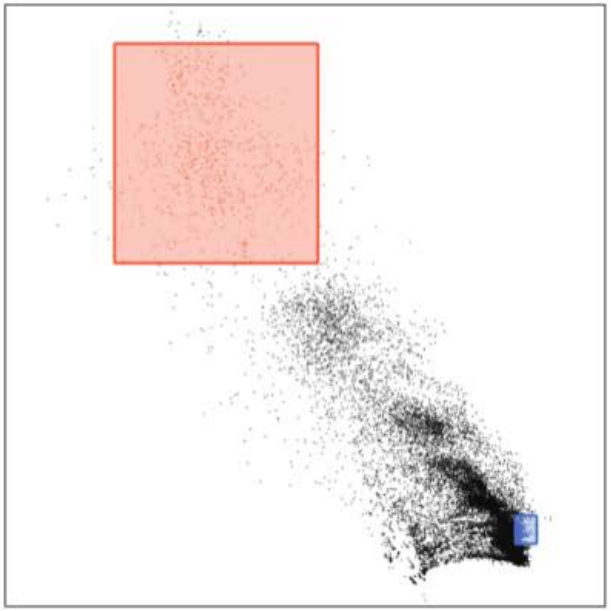
$$\mathbf{v}(\mathbf{X}(\mathbf{a},t),t) = \frac{\partial \mathbf{X}(\mathbf{a},t)}{\partial t}. \qquad \Phi : \mathbf{x} \mapsto \Phi_{t_0}^{t_0+t}(\mathbf{x})$$

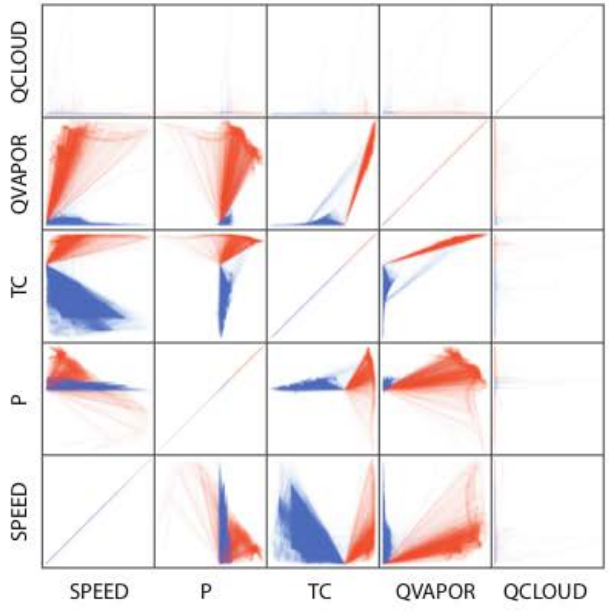
GEOS-5 Simulations

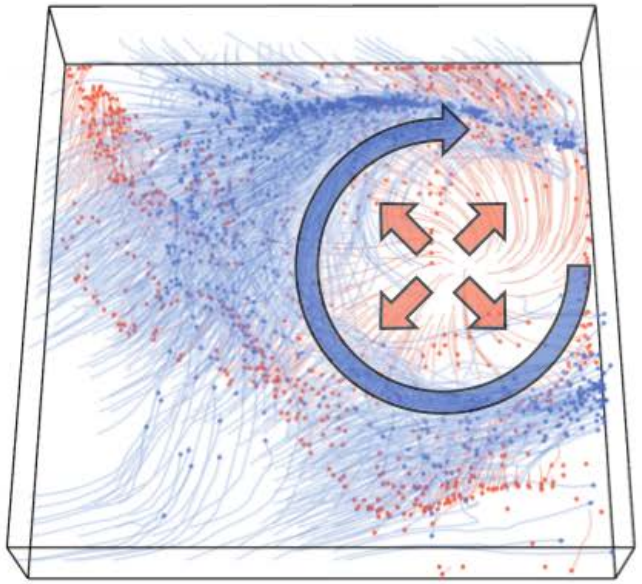


Hurricane Isabel Simulation

- Attributes along pathlines are significantly different in the two group of selected features

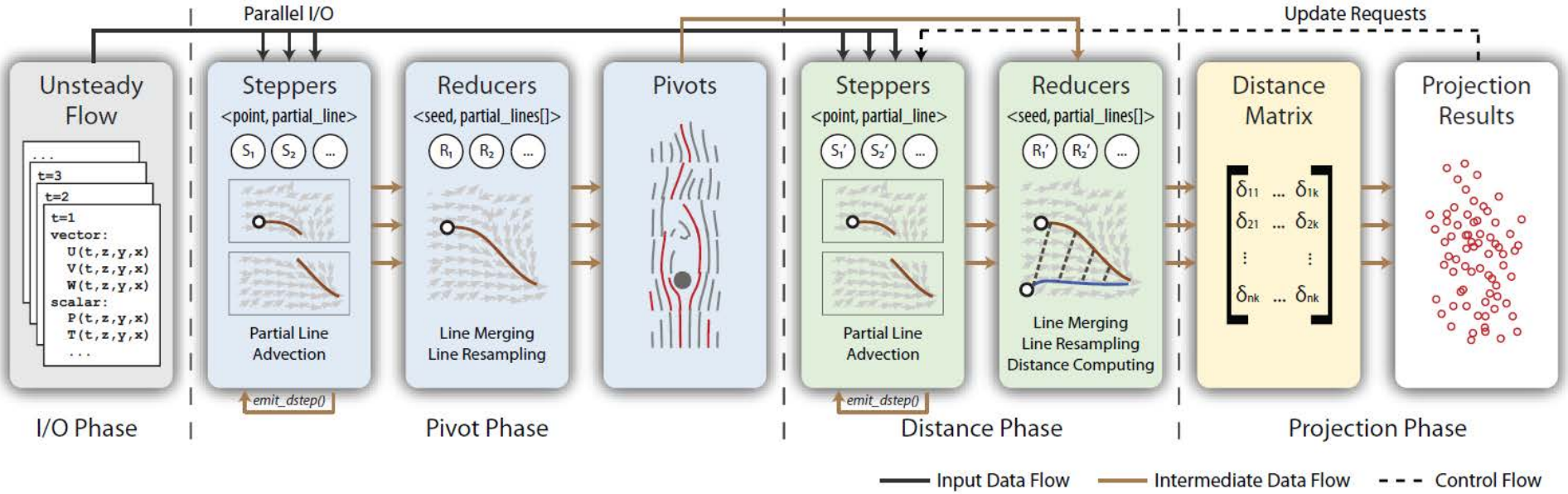




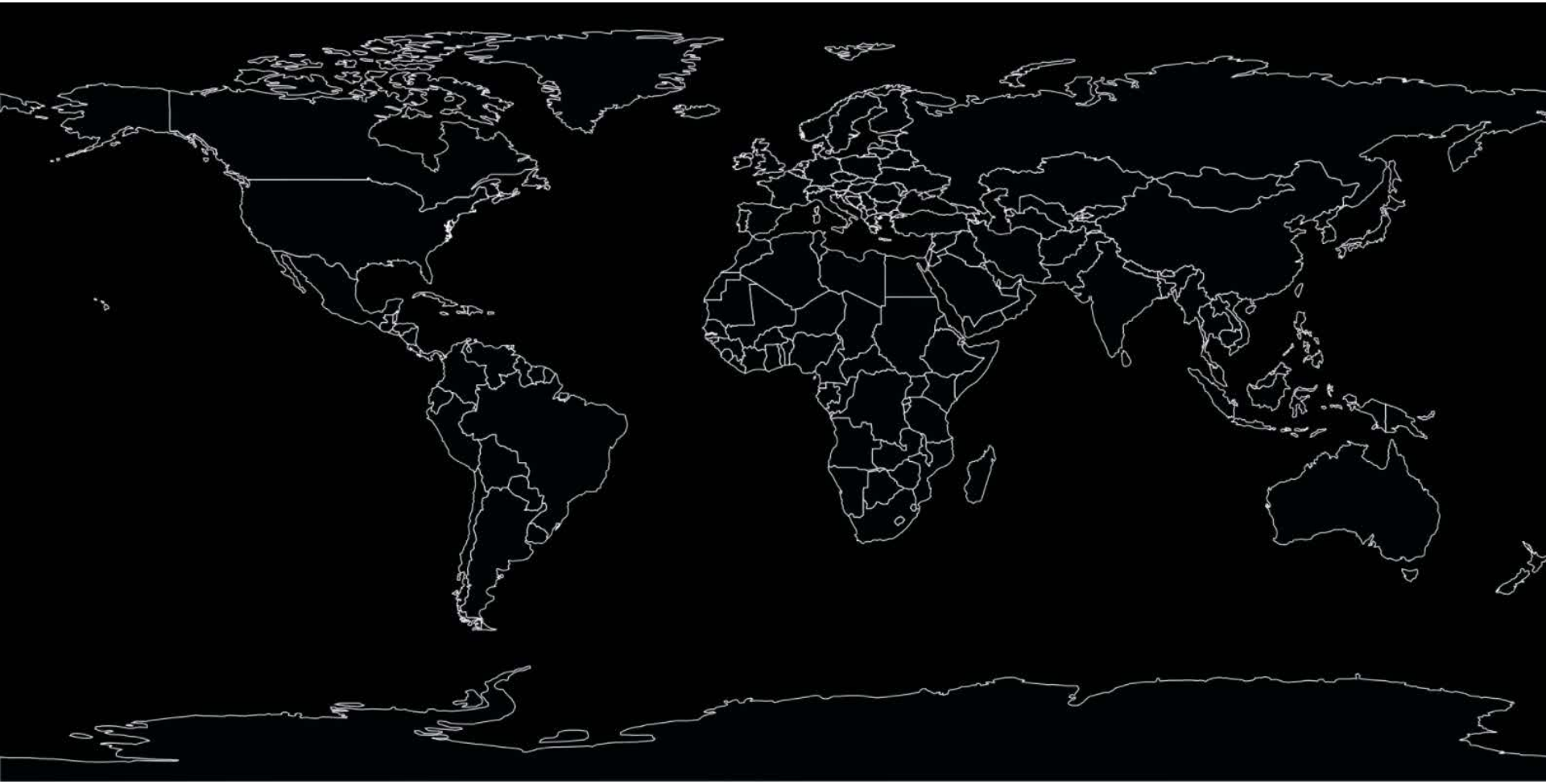


The Workflow of the Parallel System LASP

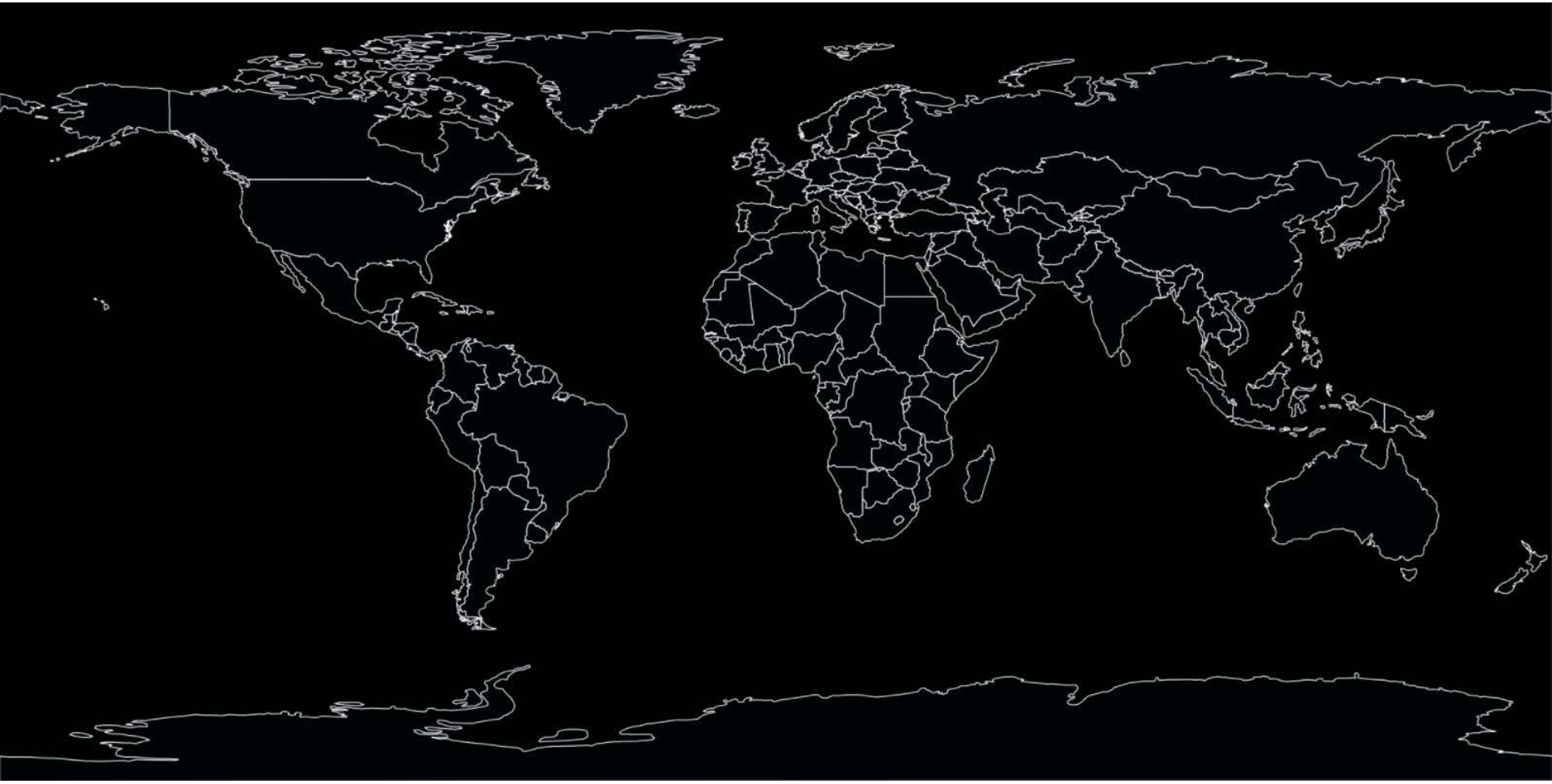
- I/O Phase: loading blocked data in parallel
- Pivot Phase: computing a small number of pivots
- Distance Phase: computing huge number of pathlines, and their distances to pivots
- Projection Phase: projecting with SPMDS in reducers



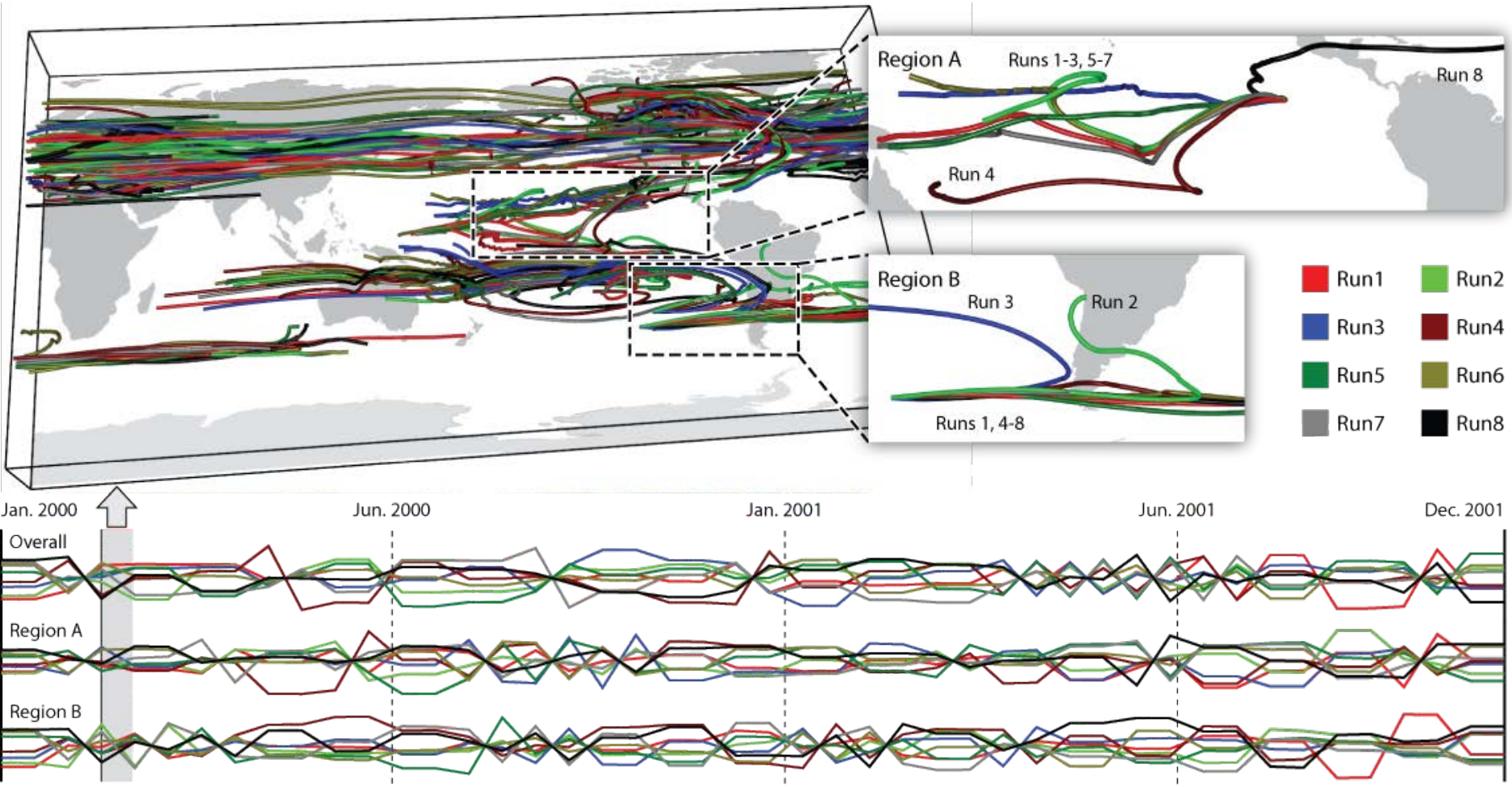
Pollution From China



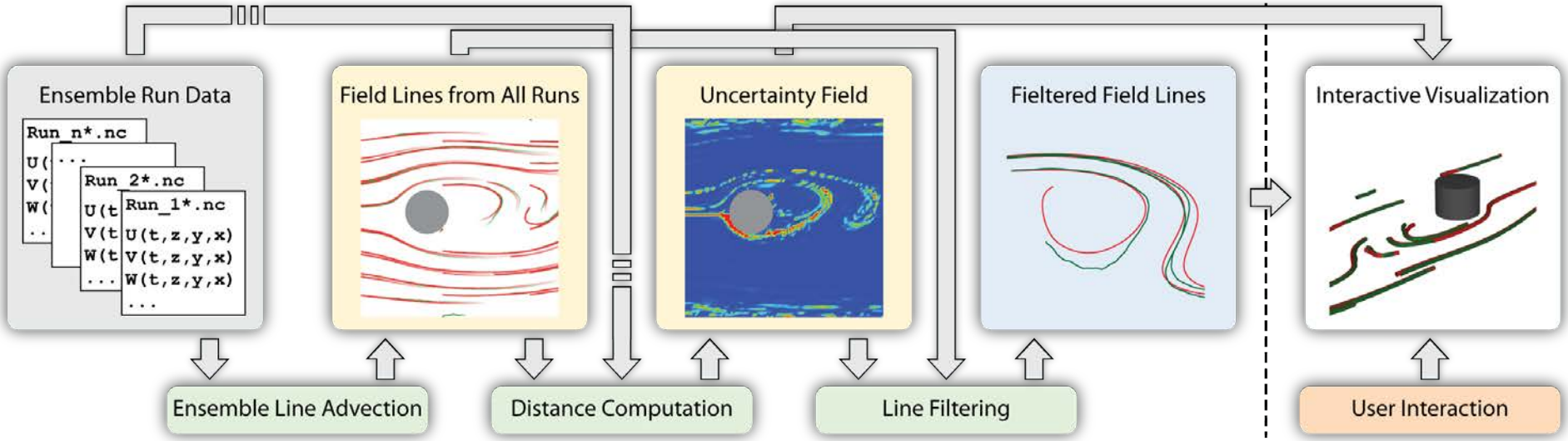
Pollution from USA



Case: GEOS-5 Simulation



Conceptual Flow Chart



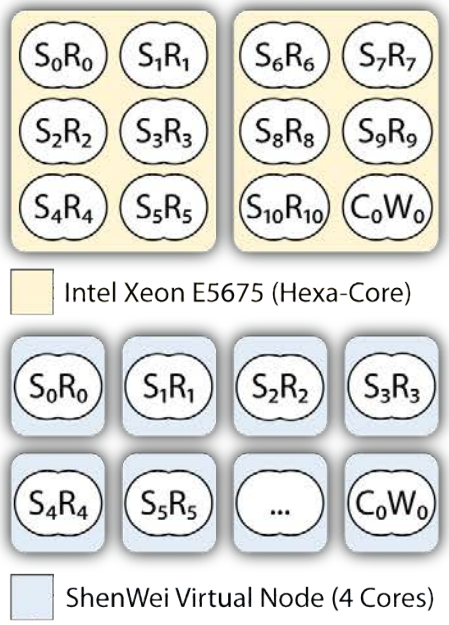
- Ensemble data (large)
- Field line data (much larger than ensemble data)
- Variation field (small)
- Filtered lines (even smaller)

Benchmark Platform: NCSSJN

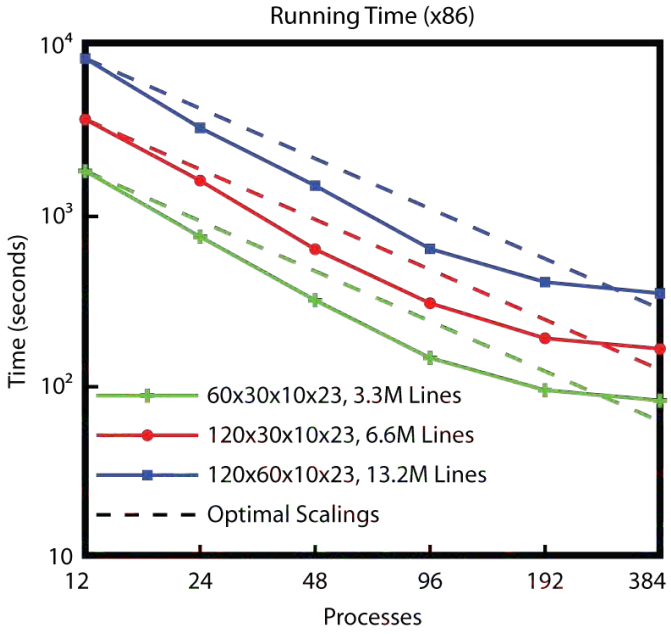
- ShenWei-based supercomputer
  - SW1600 processor, 1.0~1.1GHz
  - 1GB memory for each core
  - 40Gbps high-speed interconnection
- x86-based supercomputer
  - Intel Xeon E5675 hexa-core processor, 3.06GHz
  - 4GB memory for each core
  - QDR Infiniband interconnection
- Shared global filesystem: SWGFS

Scalability

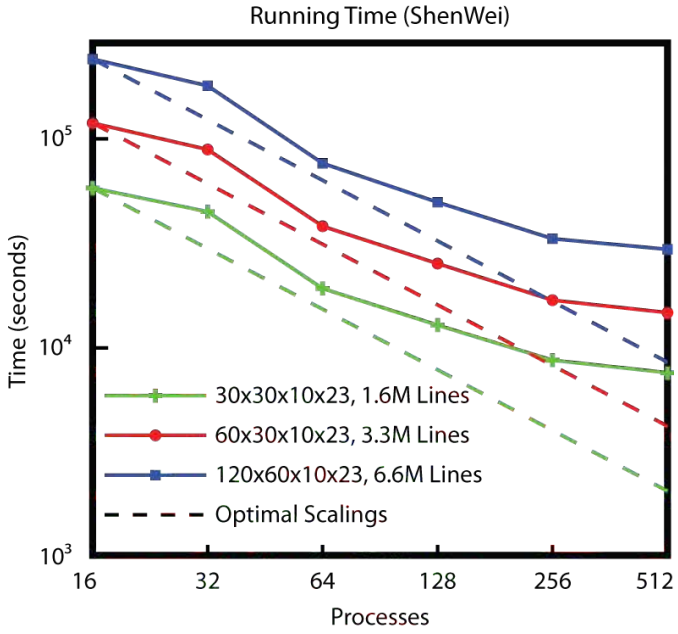
- Strong scalability test in National Super Computer Center in Jinan (ShenWei and x86 architectures)



(a)



(b)



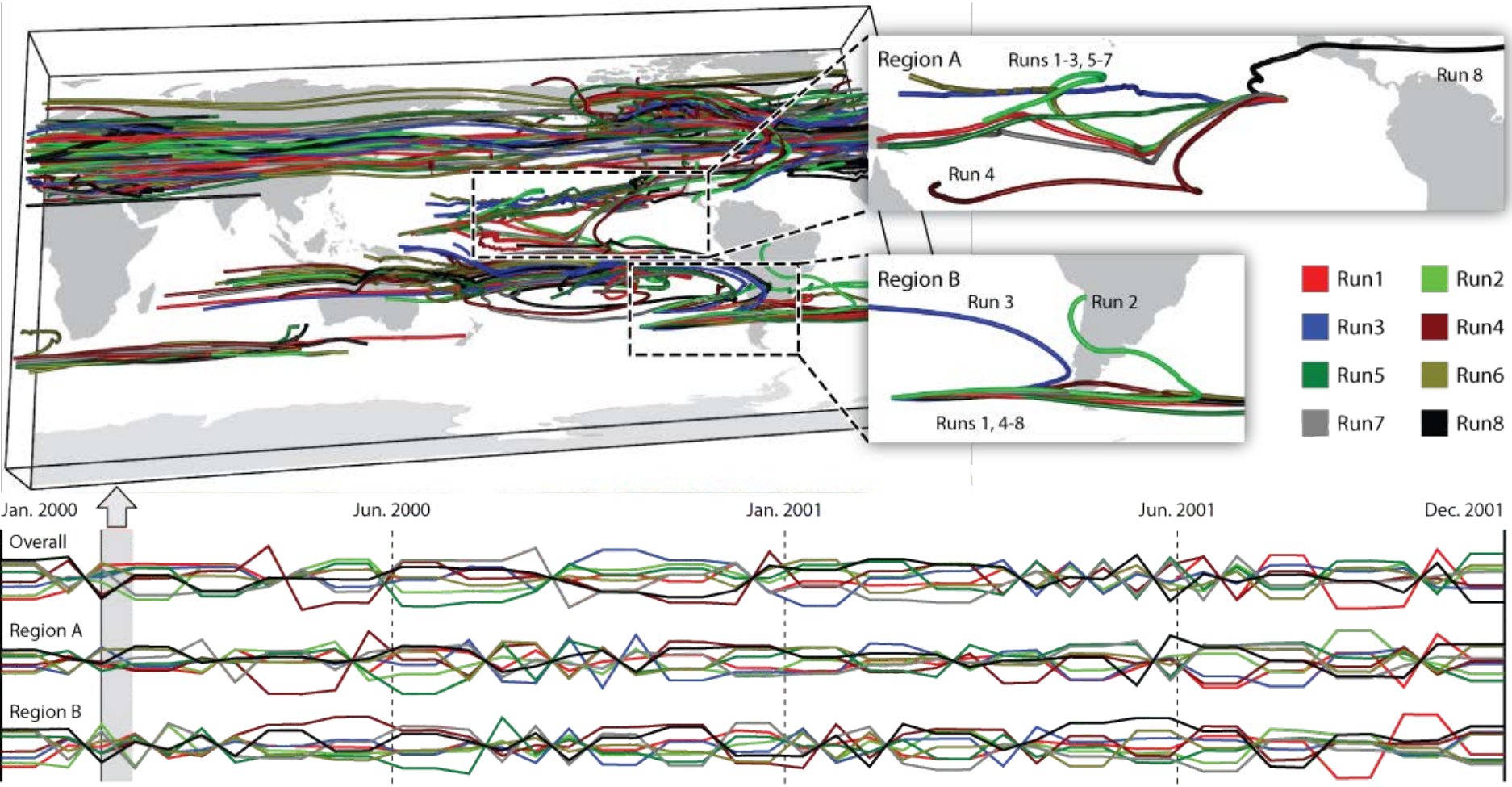
(c)

Applications

- GEOS-5 global climate model from NASA Goddard Space Flight Center
  - 8 run ensemble simulation
  - 1°x1.5° resolution, with 72 pressure levels
  - From Jan 2000 to Dec 2011

- WRF ARW model
  - **`base`** and **`no_urban`** runs to investigate the impact of urbanization
  - 100x100x27 resolution, East China
  - From 2012-7-1 00:00:00 UTC to 2012-7-10 18:00:00 UTC

GEOS-5 Simulation



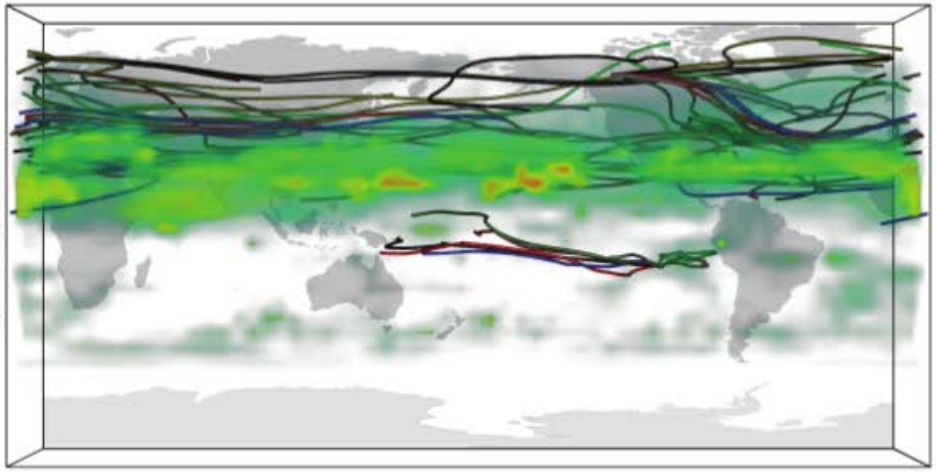
GEOS-5 Simulation

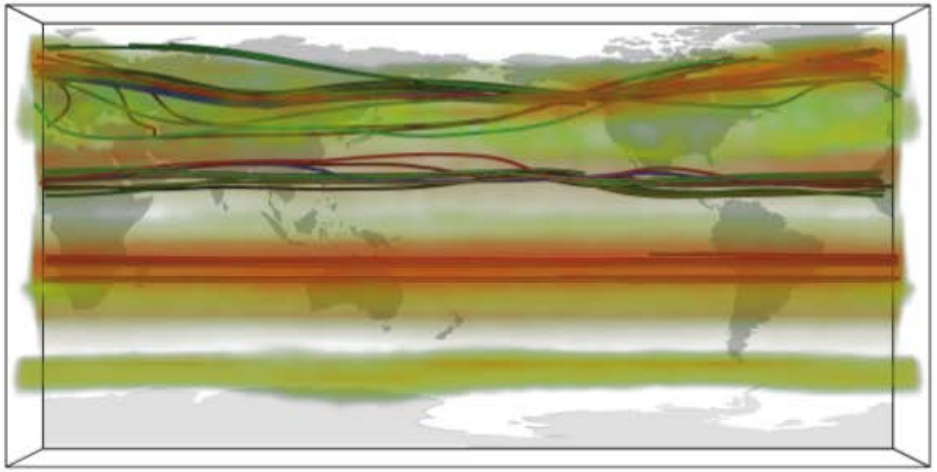
8-Run GEOS-5 Simulation Data

March, 2003

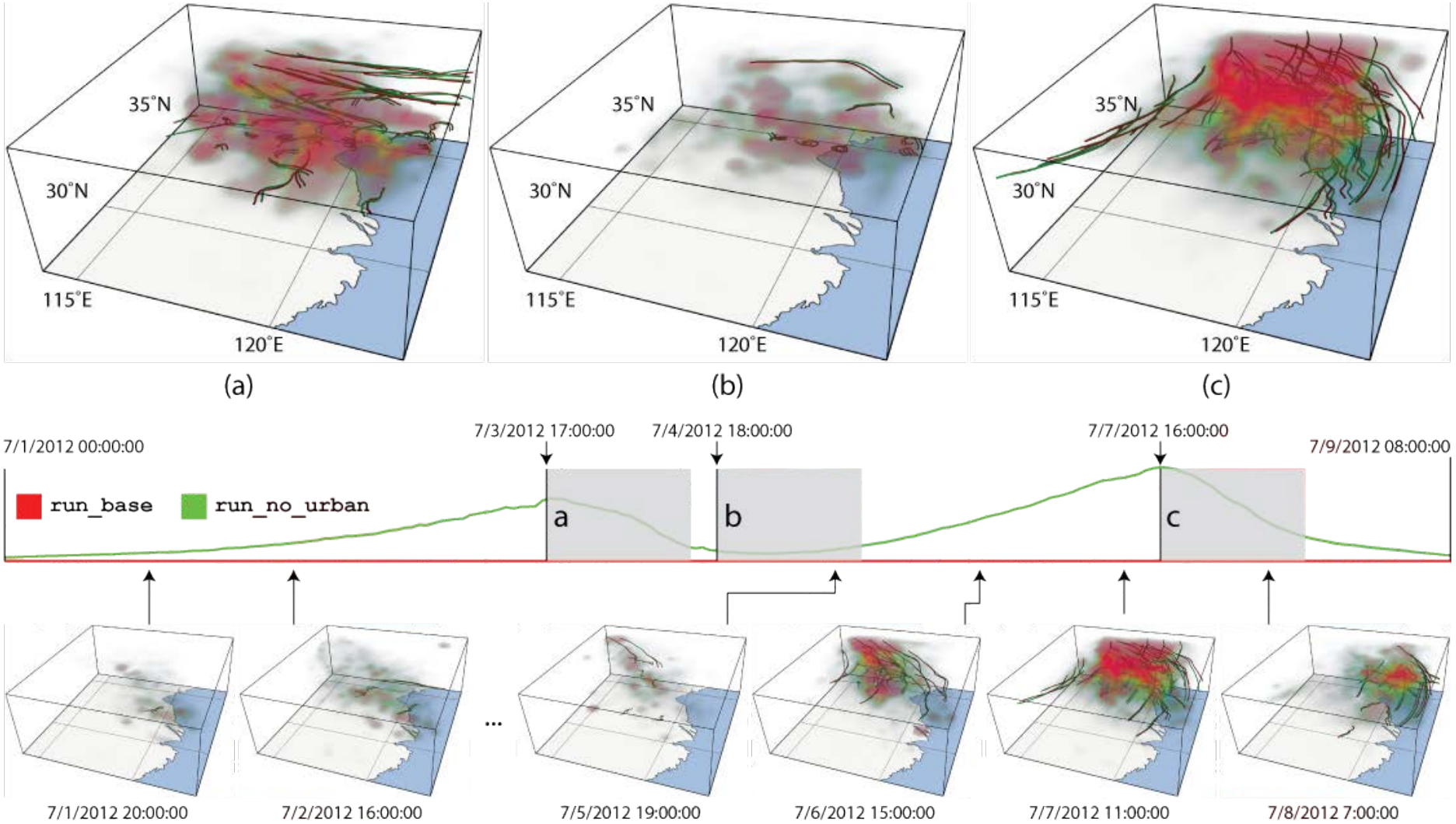
GEOS-5 Simulation: CO_2-based Metric

- The metric: the differences of locations / CO_2 concentration along the pathline
- Findings
  - The variation of the wind field is high in the north hemisphere
  - However, The CO_2 difference is higher in south hemisphere and some places in the north
  - CO_2 concentration is not sensitive to wind in above regions





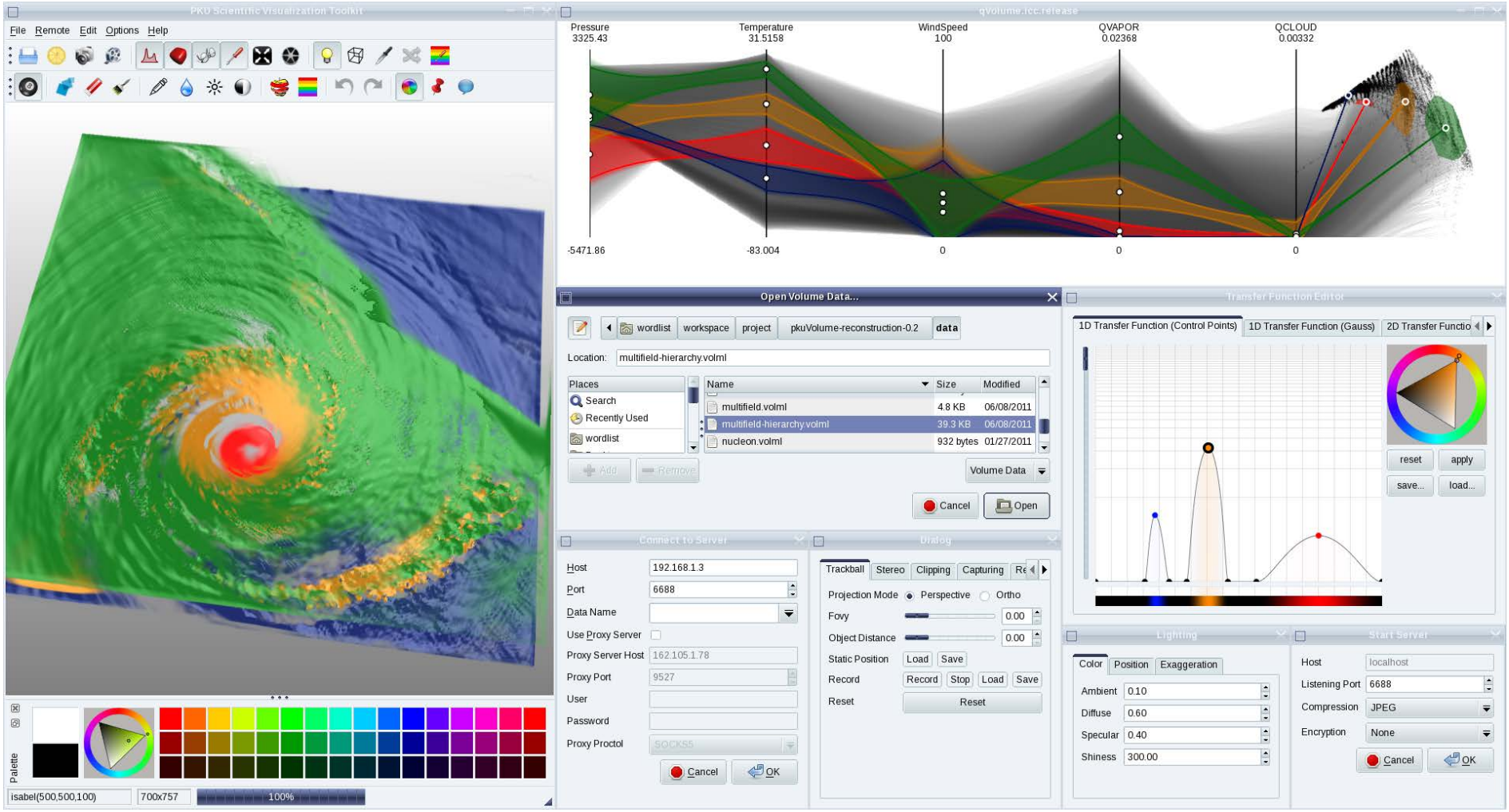
WRF Simulation



WRF Simulation

2-Run WRF Simulation Data

VisPKU PKU Scientific Visualization System 北京大学科学可视化系统



Challenges in VIS and VA

- Scalability in Data/Task complexity
  - Science imposes more computational challenges methods for visualization and visual analysis on large scientific data
- Scalability in User
  - Collaborative Visualization and Analysis on large data
  - Can scientist create novel visualization without programming

Double Gulf

Visualization User

Visualization Designer

Evaluation

Representation

Conceptual Model

Visualization

Data

Execution

Manipulation

Double Gulf

Visualization User

Visualization Designer

Evaluation

Representation

Conceptual Model

Visualization

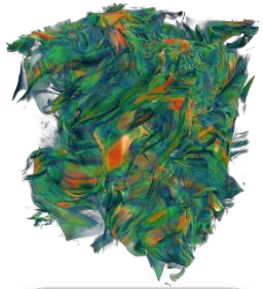
Data

Execution

Manipulation

From Data to User

Visualization User



Visualization Designer

User

Evaluation

Execution

VIS

Representation

Manipulation

Data

Scalability In Users

Visualization User

Visualization Designer

Evaluation

Representation

Conceptual Model

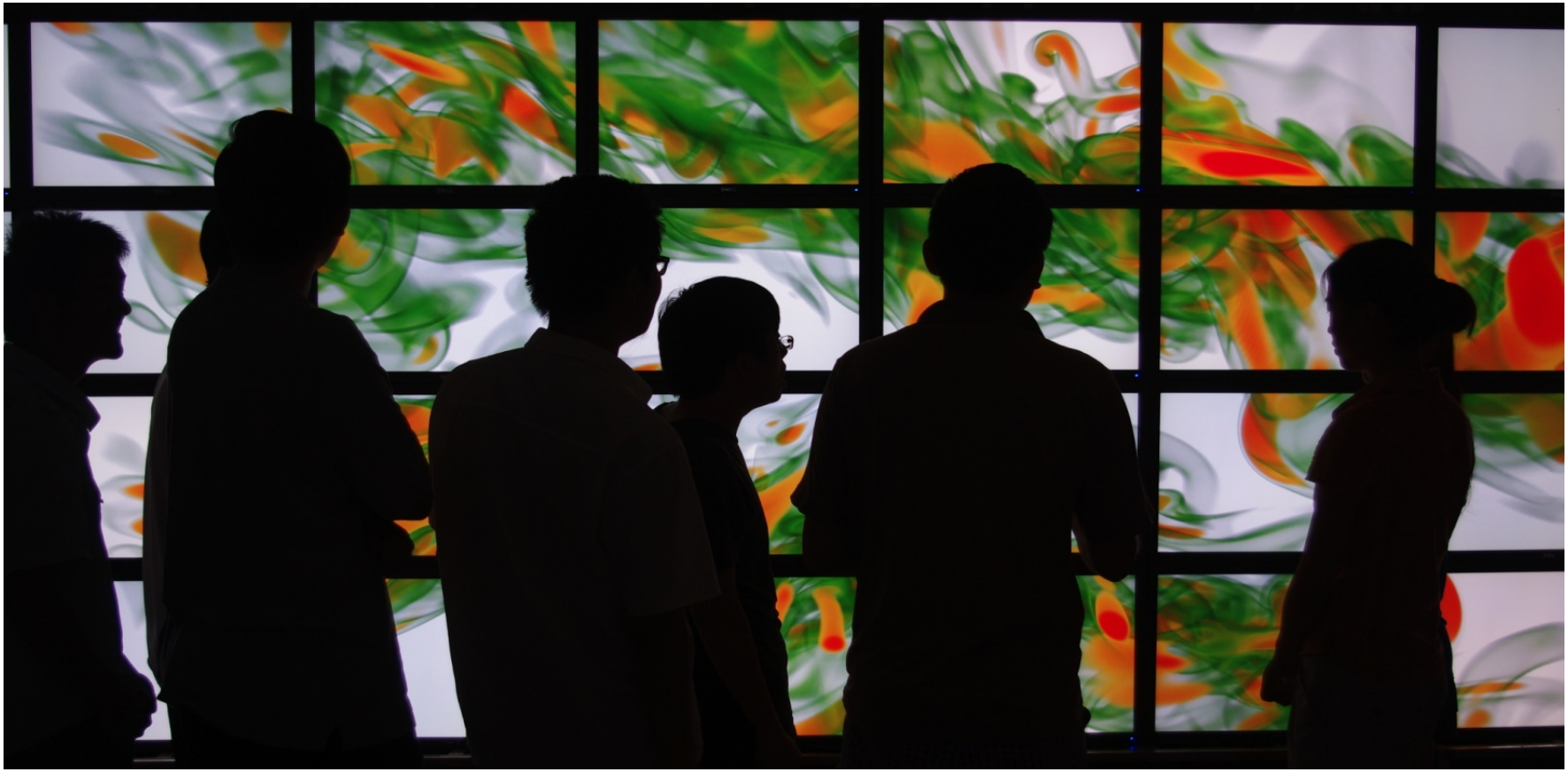
Visualization

Data

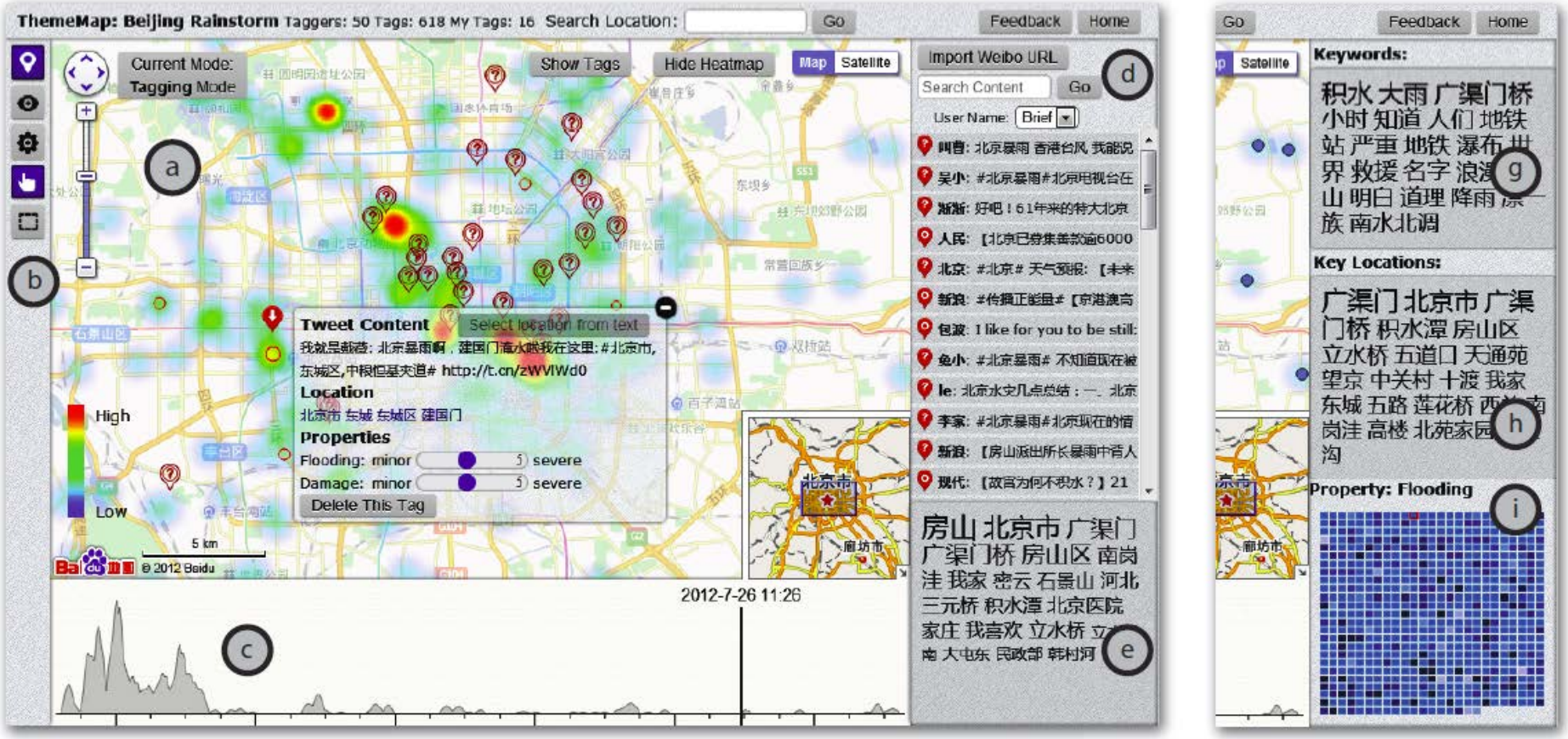
Execution

Manipulation

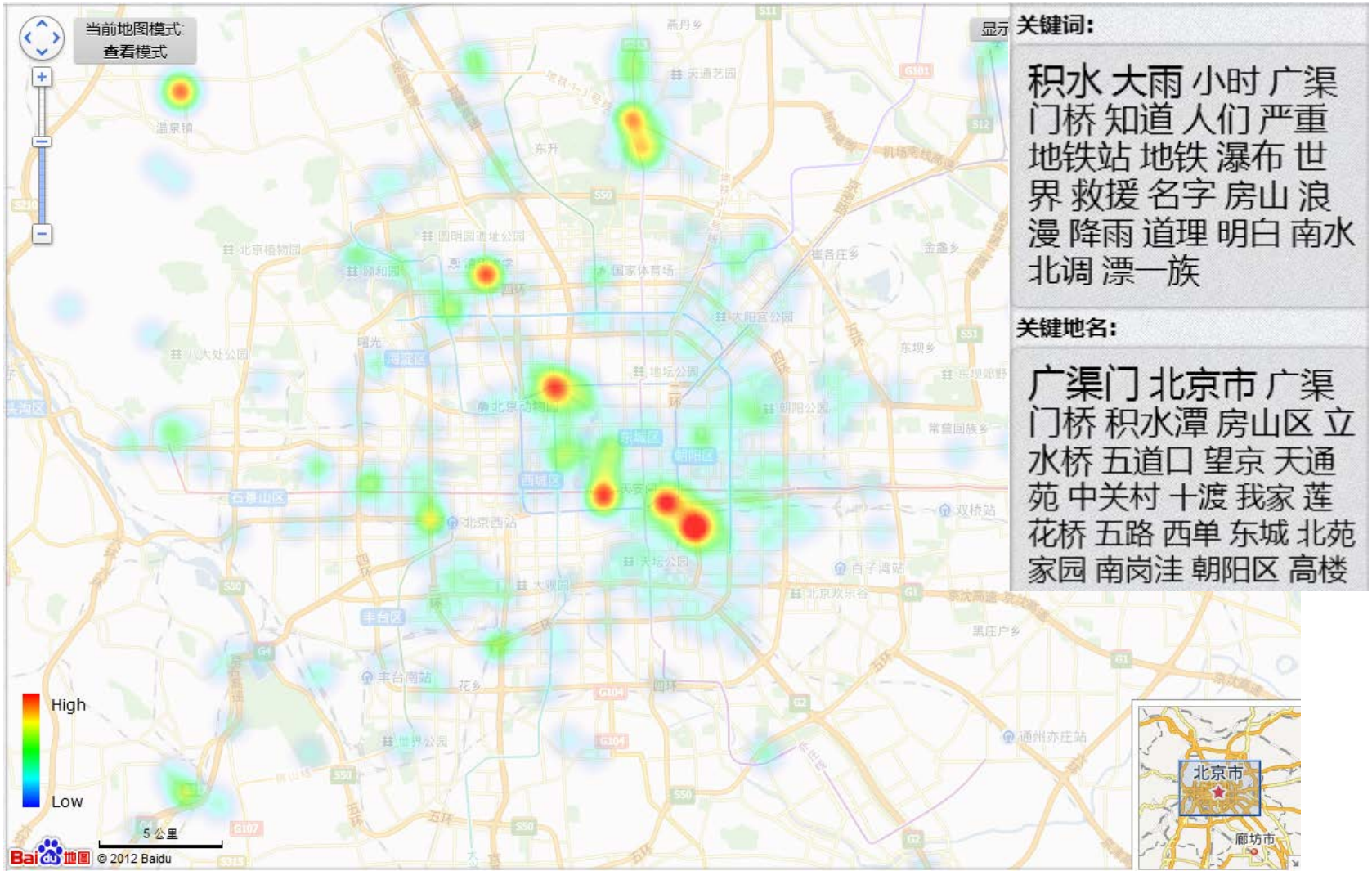
Scalability In Users – Collaborative Visualization

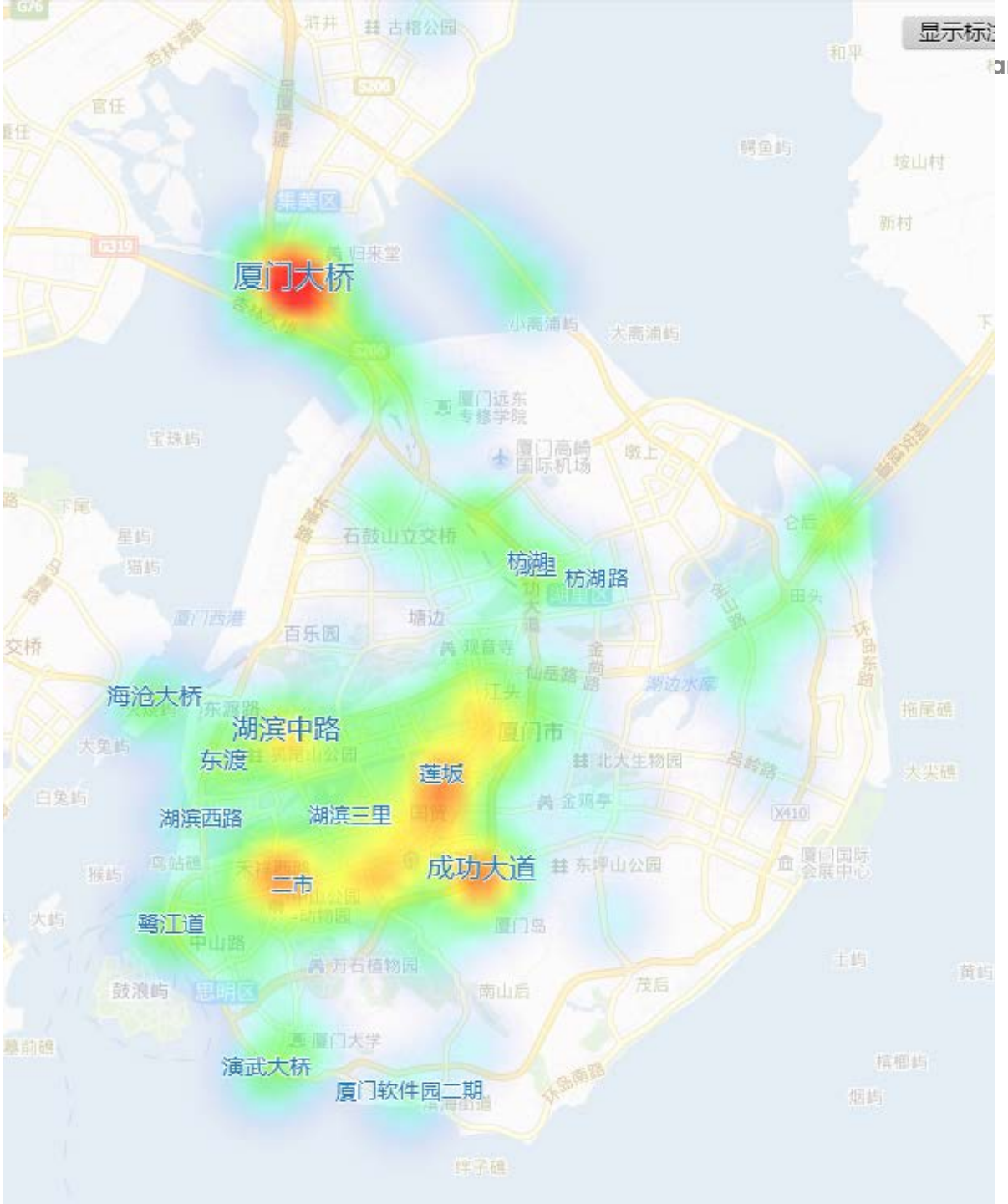


ThemeMap – Crowd Sourcing



ThemeMap – Crowd Sourcing





关键词：

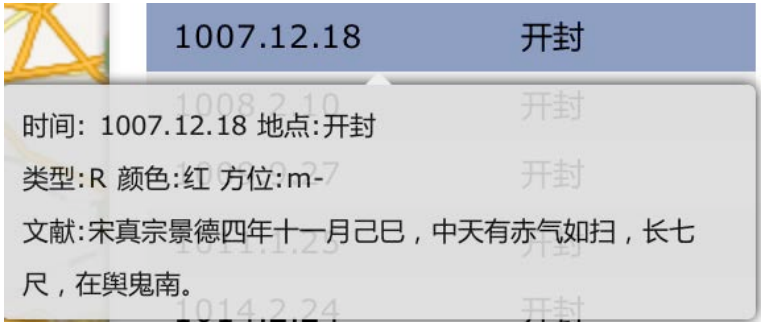
大道 严重 隧道 红绿灯 仙岳 路口 火车站 司机 路段 海沧 时间 上班 知道 轮渡 东渡 方向 中路 开车 公园 车道

关键地名：

厦门大桥 成功大道 湖滨中路 桥上 地下停车场 东渡 海沧大桥 演武大桥 莲坂 鹭江道 集美 南湖公园 湖滨三里 湖里 厦门软件园二期 二楼 二市 枋湖 湖滨西路 枋湖路

Aurora Observation Vis

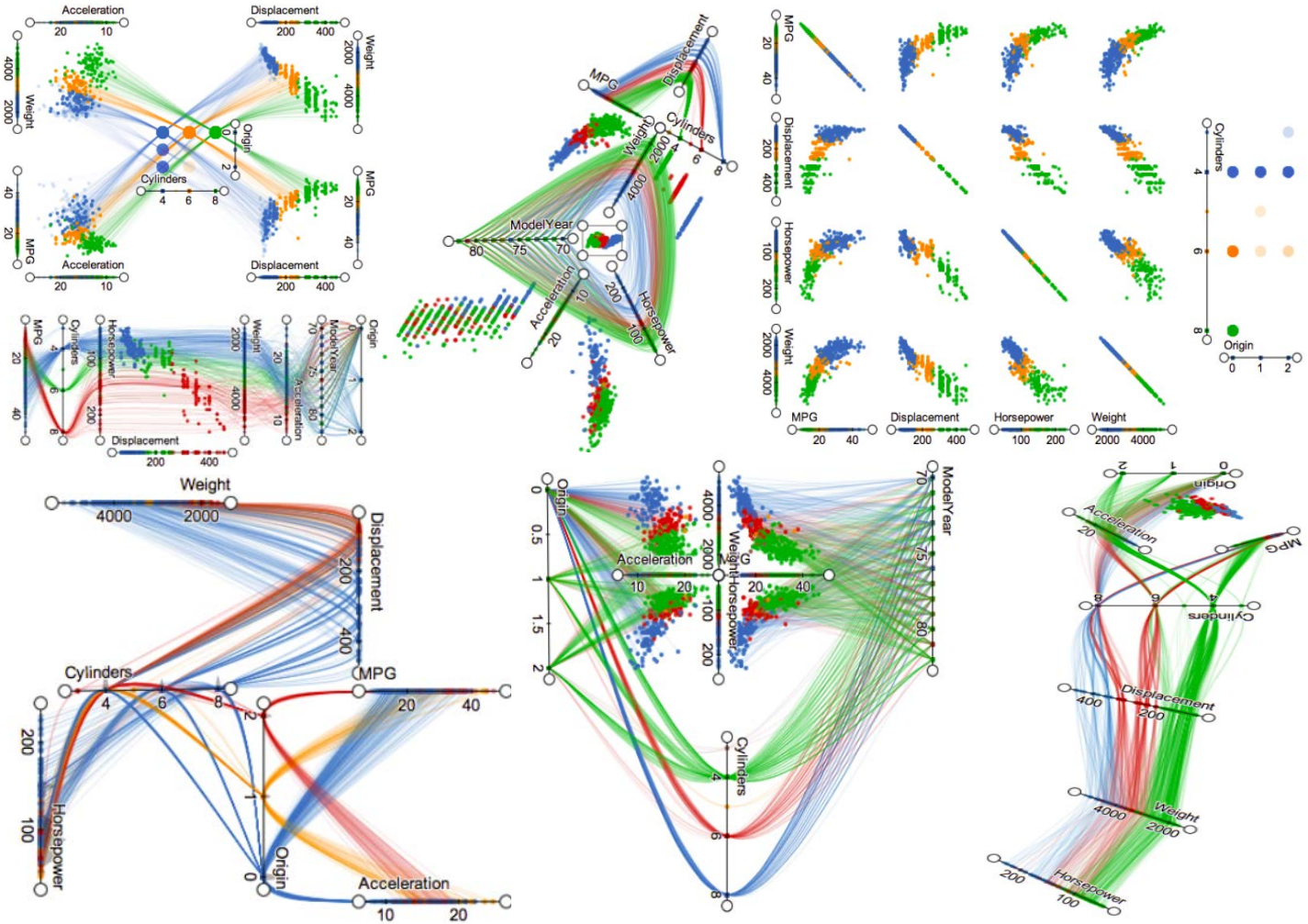
Aurora Observation Vis



开封 西安
杭州 北京
内蒙巴林右旗 湖北江陵
陕西潼关 湖北西北部
洛阳 四川仁寿井研
嘉兴 洞庭湖

北京空气质量可视化 mapview periodview

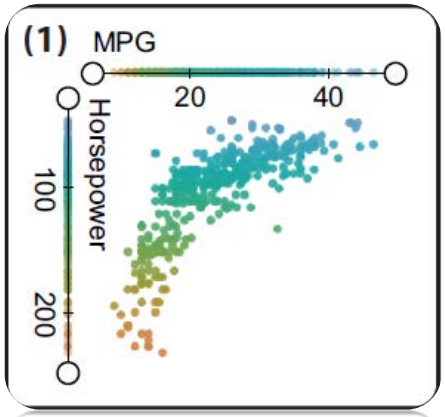
Scalability In Users – User - Visualization Expert

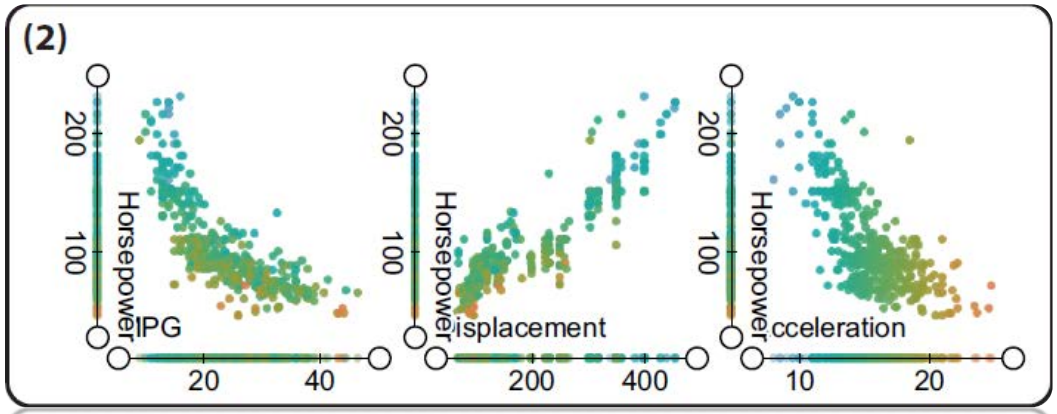


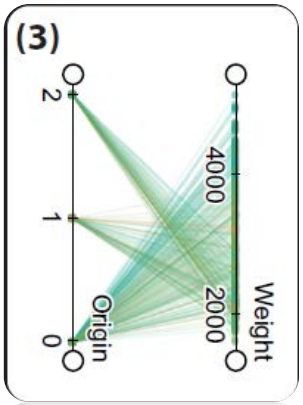
Visualization Assembly Line

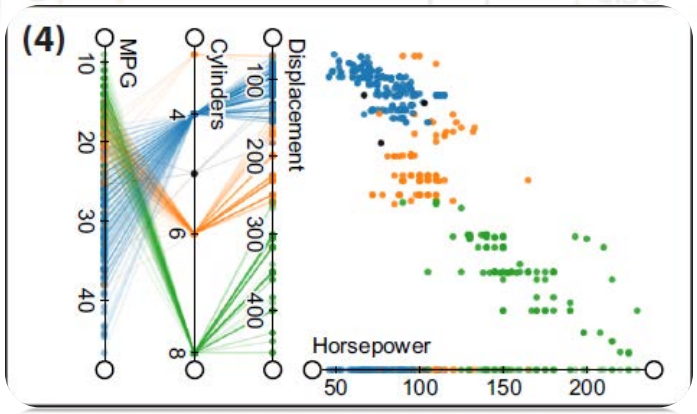
http://vis.pku.edu.cn/mddv/val/

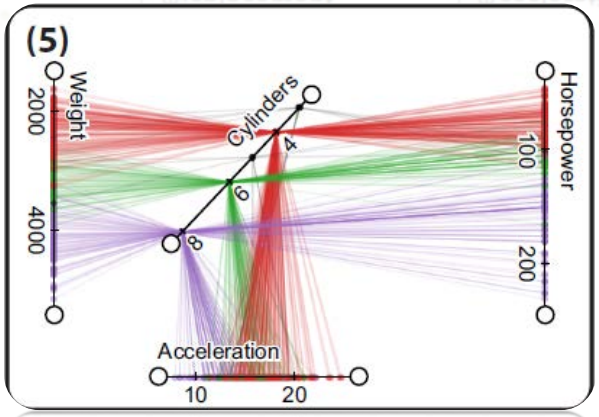
Visualization Assembly Line

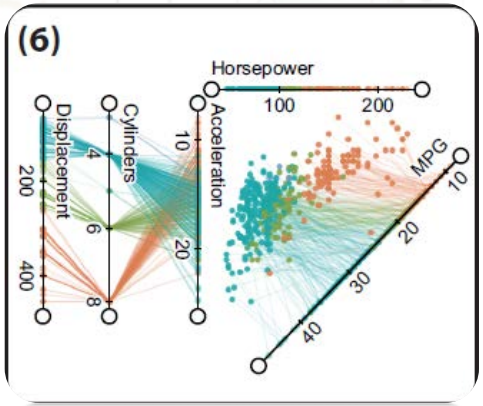




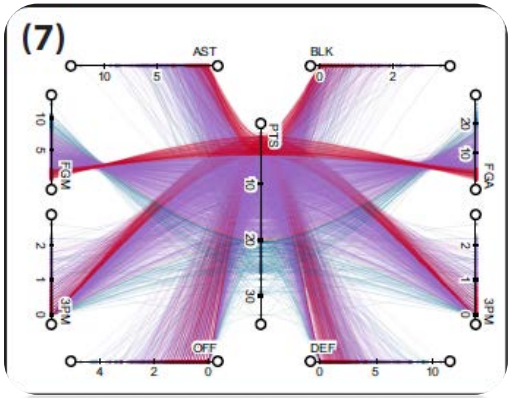


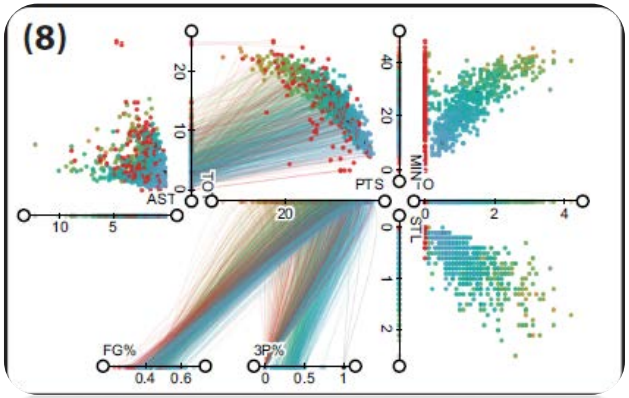


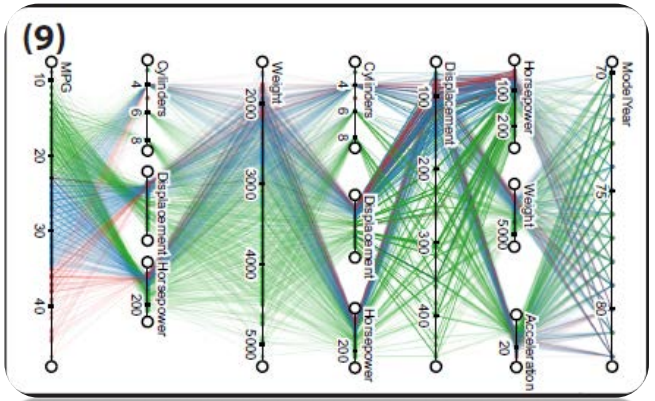


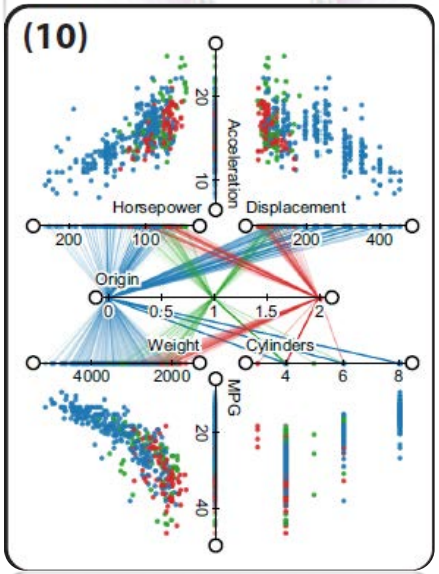


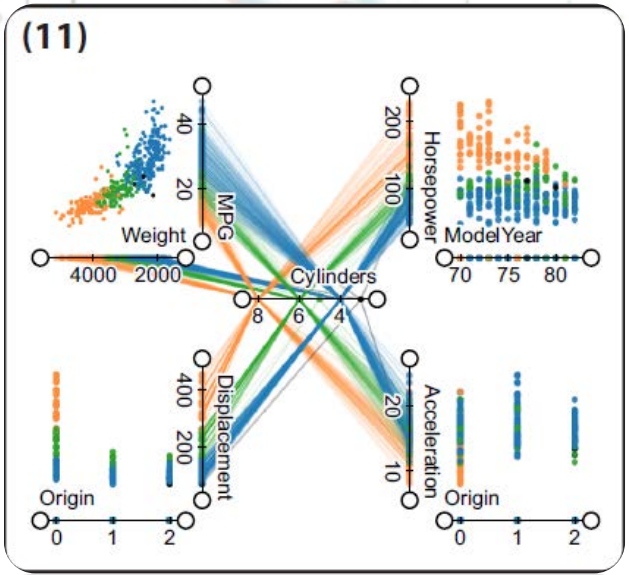
Visualization Assembly Line

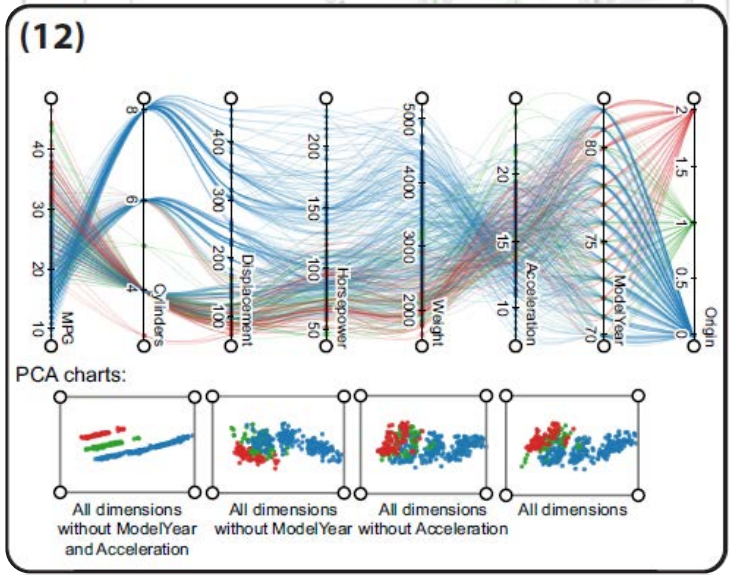












Challenges in VIS and VA

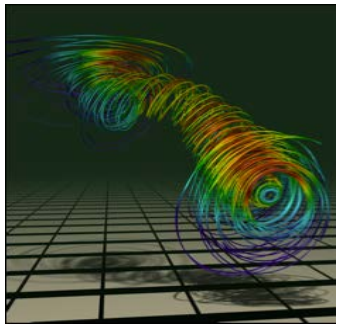
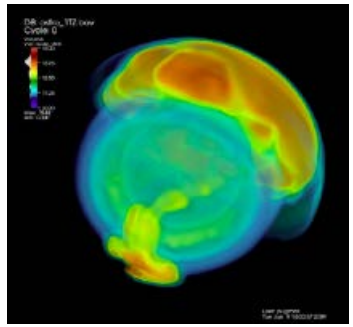
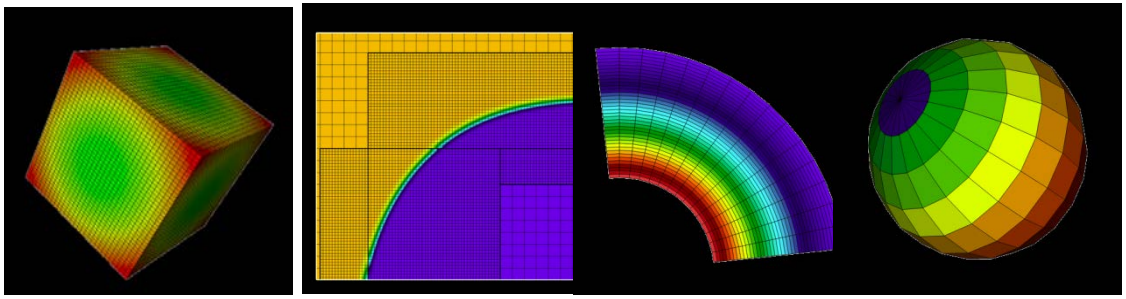
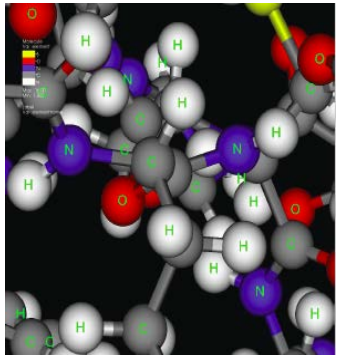
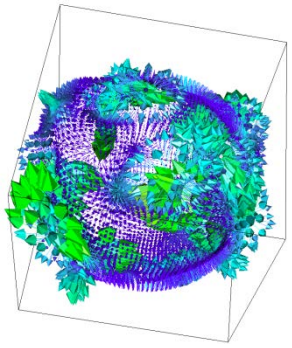
- Scalability in Data/Task complexity
  - Science imposes more computational challenges methods for visualization and visual analysis on large scientific data
- Scalability in User
  - Collaborative Visualization and Analysis on large data
  - Can scientist create novel visualization without programming
- System Development
  - *Domain and Development Libraries, Frameworks, and Tools*
  - *Social, Community, and Government Engagements*

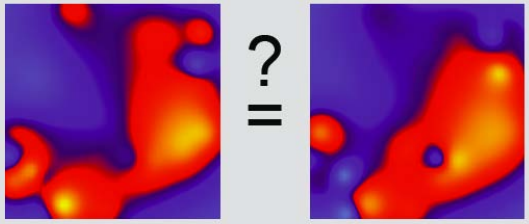
SCIVIS Visualization Systems

- VisIt - LLNL
  https://wci.llnl.gov/codes/visit
- ParaView- Kitware/SNL/LANL
  http://www.paraview.org
- IceT (Image Composition Engine for Tiles) - Sandia
  http://icet.sandia.gov
- Daxtoolkit - Data Analysis at Extreme
  http://www.daxtoolkit.org
- PISTON - Portable Data-Parallel Visualization and Analysis Library - LANL
  http://viz.lanl.gov/projects/PISTON.html

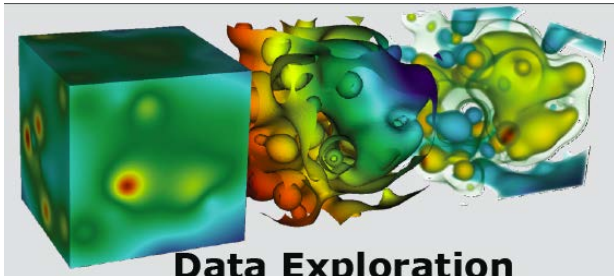
VisIt

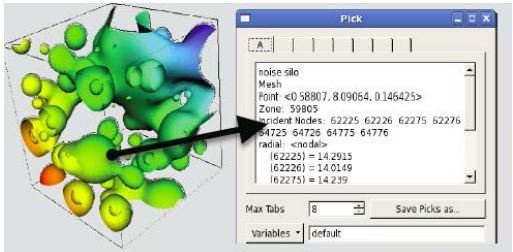
- Production end-user tool supporting scientific and engineering applications.
- Parallel post-processing that scales from desktops to massive HPC



Comparative Analysis

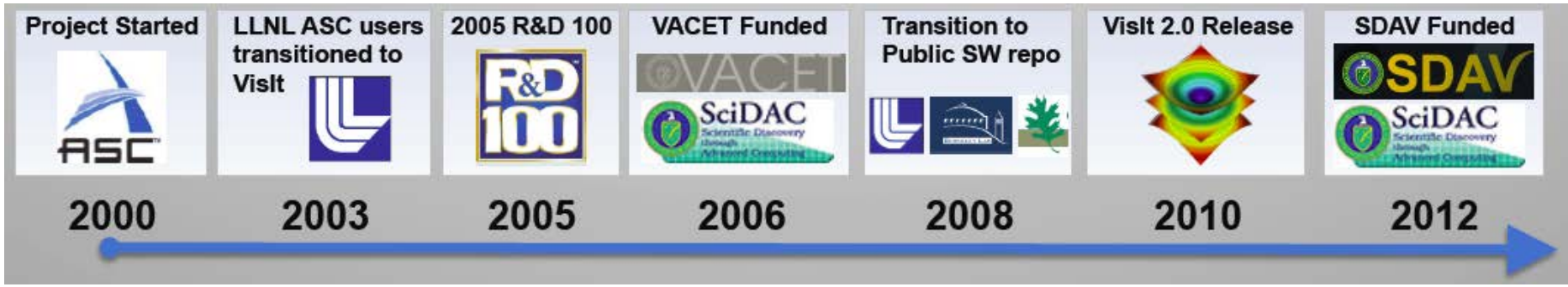

Data Exploration



Visual Debugging

Development of VisIt

- The VisIt project started in 2000 to support LLNL's large scale ASC physics codes.
- Supported by multiple organizations: LLNL, LBNL, ORNL, UC Davis, Univ. of Utah, …
- Over 75 person years effort.
- 1.5+ million lines of code.



Based on SC'11 Tutorial

ParaView is an open-source, multi-platform data analysis and visualization application. ParaView users can quickly build visualizations to analyze their data using qualitative and quantitative techniques. The data exploration can be done interactively in 3D or programmatically using ParaView's batch processing capabilities.

ParaView was developed to analyze extremely large datasets using distributed memory computing resources. It can be run on supercomputers to analyze datasets of terascale as well as on laptops for smaller data.

News More News >

05.07.2013 Kitware to Enhance the Visualization Toolkit (VTK) to Accelerate ...

03.06.2013 KiwiViewer 2.0 Released and Available for Download in iTunes

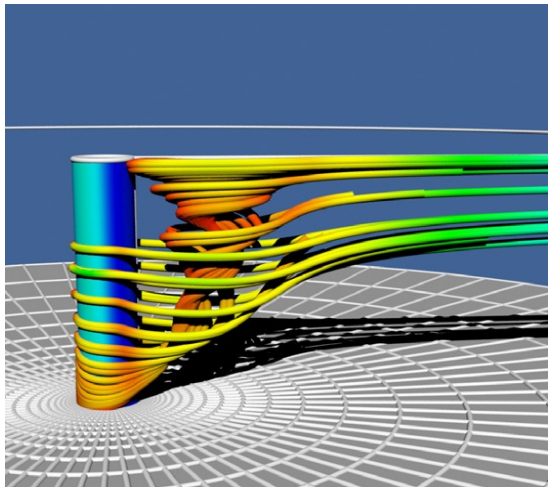
02.12.2013 ParaView 3.98.1 Now Available

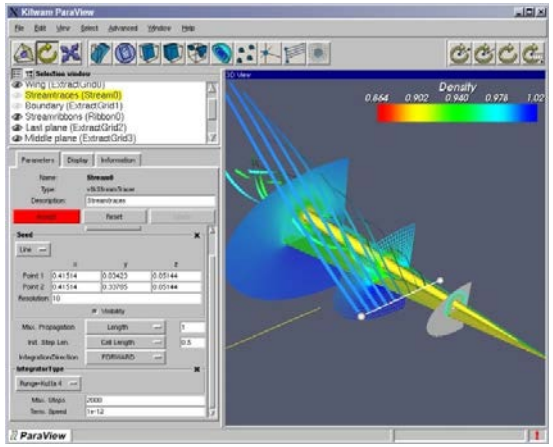
12.03.2012 ParaView 3.98.0 Now Available

11.14.2012 Kitware Receives Honors in 2012 HPCwire Readers' and Editors' Cho...

Kitware ASC Sandia National Laboratories Los Alamos NATIONAL LABORATORY ARL U.S. Army Research Laboratory

VTK





1993年，3名来自GE Corporate R&D 中心的可视化研究员开始开发一个开源的可视化系统Visualization Toolkit (VTK)。

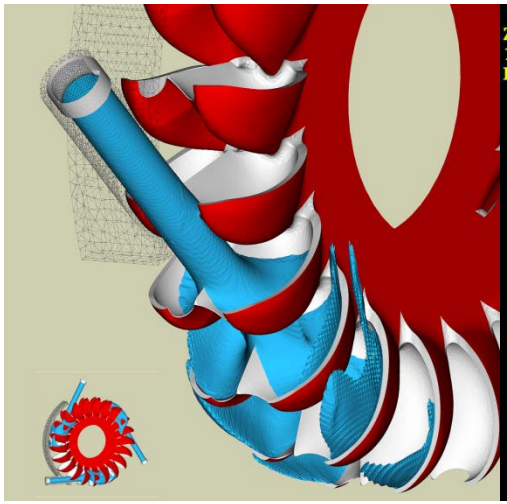
开始被创作的意图是用于教学和研究的（因此它是开源的）。这个软件很快的被接受，不光因为它的面向对象设计和软件过程的理念，还因为它得到使用者们的改善。

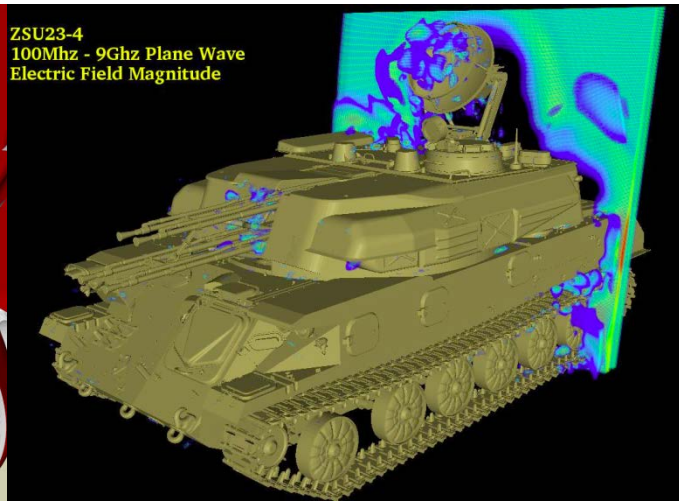
VTK现在被广泛的使用，许多公司和相关产品由于使用它而诞生，比如Kitware公司于1998年建立后编写了基于VTK的一款开源可视化系统ParaView和一款商业的体绘制应用程序VolView。VTK持续被学术人员，美国国家实验室和商家完善，并且用于许多商业软件中。

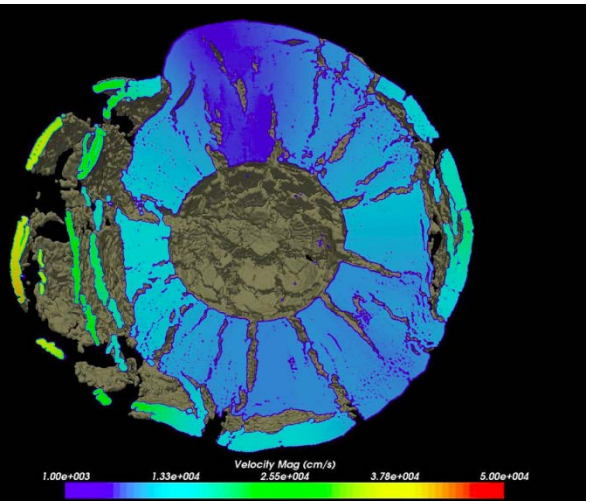
W.J. Schroeder, K. Martin, and W. Lorensen, *The Visualization Toolkit: An Object Oriented Approach to Computer Graphics, Third Edition*, Kitware, Inc., ISBN-1-930934-12-2 (2004).
S. E. Rogers, D. Kwak, and U. K. Kaul, A numerical study of three-dimensional incompressible flow around multiple post. In *Proceedings of AIAA Aerospace Sciences Conference*. AIAA Paper 86-0353. Reno, Nevada, 1986.

ParaView

- 2000 Los Alamos National Laboratories and Kitware Inc.
- 2005 Sandia National Laboratories and Kitware Inc.
- Used by academic, government, and commercial institutions worldwide.
- Downloaded ~100K times per year.

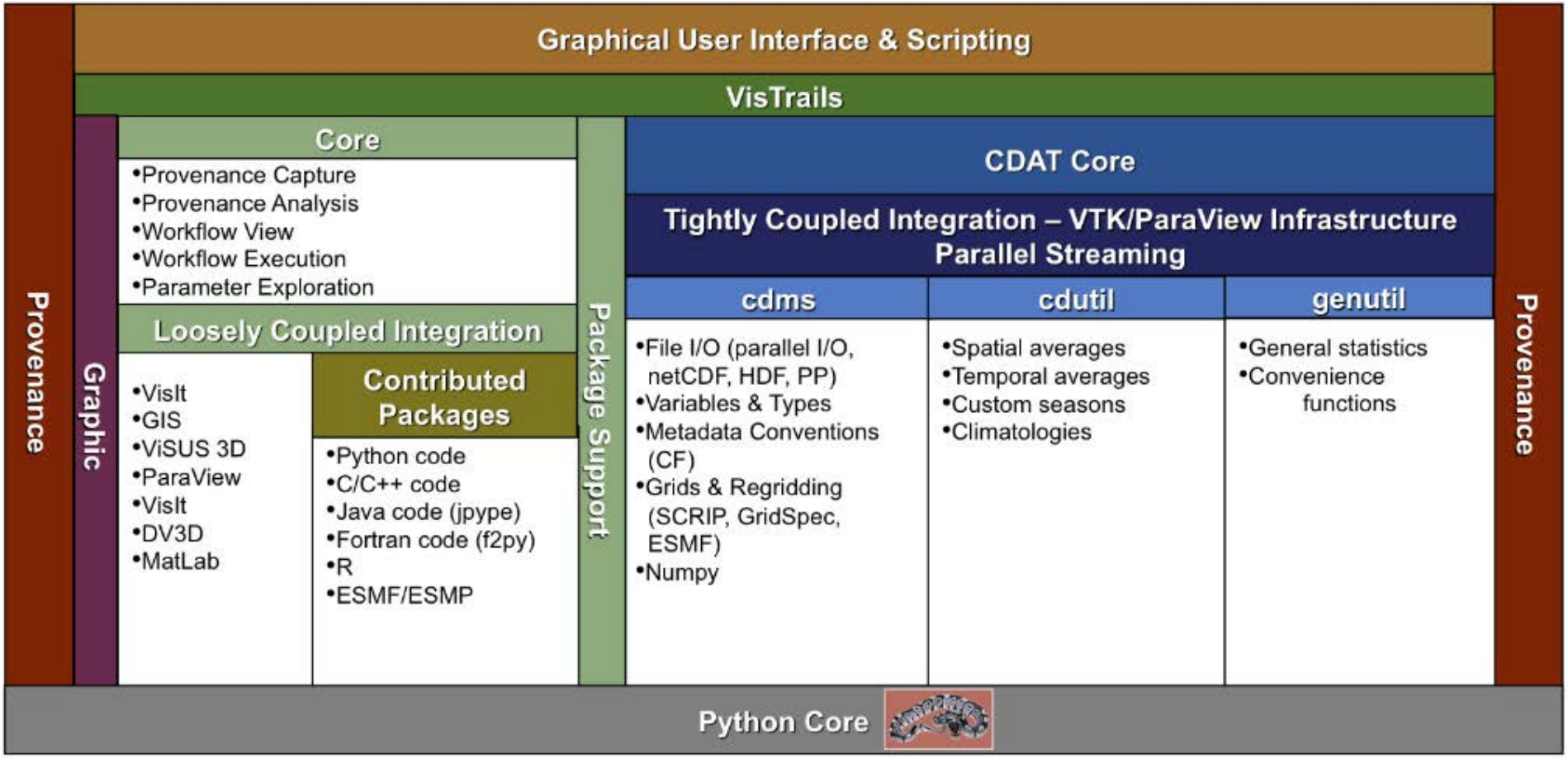






UV-CDAT Project

Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT) Architectural Layers



Build a successful vis system

- System Design
- Domain User – Visualization Scientist "Co-design"
- Stable Development Team
- Funding Mechanism

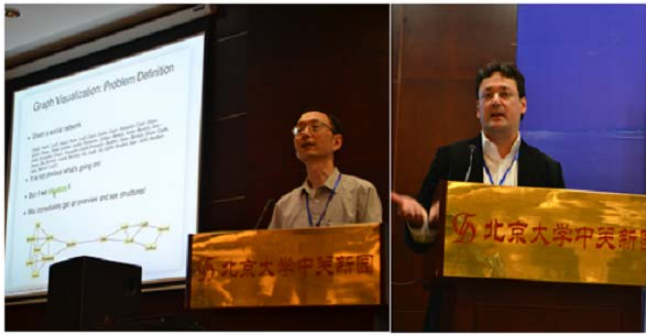
Social, Community, and Government Engagements

- Universities
  - University of Tennessee in Knoxville
  - Ohio State University
  - SCI Institute, University of Utah
  - University of California, Davis
  - University of California, San Diego
  - University of Nebraska-Lincoln
  - Michigan Technological University
  - Drexel University
- Supercomputer centers
  - San Diego Supercomputer Center (SDSC)
  - Texas Advanced Computing Center (TACC)
  - National Center for Supercomputing Applications at the University of Illinois (NCSA)
- DoE Labs
  - Argonne National Laboratory (ANL)
  - Lawrence Berkeley National Laboratory (LBNL)
  - Lawrence Livermore National Laboratory (LLNL)
  - Los Alamos National Laboratory (LANL)
  - Pacific Northwest National Laboratory (PNNL)
  - Oak Ridge National Laboratory (ORNL)
  - Sandia National Laboratories (SNL)
  - National Renewable Energy Laboratory (NREL)
- Companies
  - Kitware

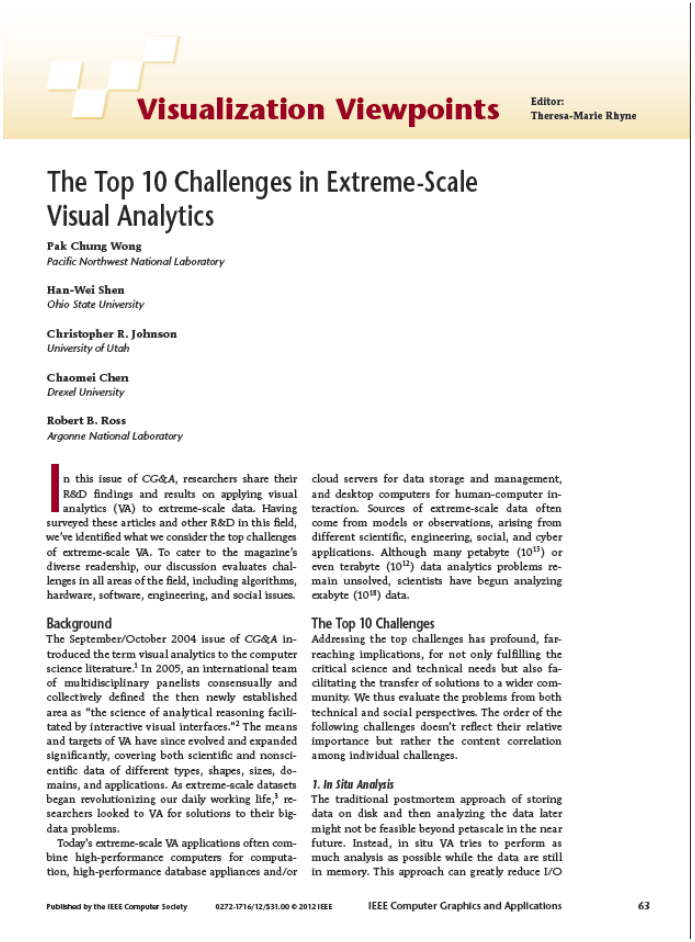
Vis Workshop 2013 @ PKU

- 2013.7.12-13

会场



Top 10 Challenges in Extreme-Scale Data Visual Analytics



Visualization Viewpoints

Editor:
Theresa-Marie Rhyne

The Top 10 Challenges in Extreme-Scale Visual Analytics

Pak Chung Wong
Pacific Northwest National Laboratory

Han-Wei Shen
Ohio State University

Christopher R. Johnson
University of Utah

Chaomei Chen
Drexel University

Robert B. Ross
Argonne National Laboratory

In this issue of *CG&A*, researchers share their R&D findings and results on applying visual analytics (VA) to extreme-scale data. Having surveyed these articles and other R&D in this field, we've identified what we consider the top challenges of extreme-scale VA. To cater to the magazine's diverse readership, our discussion evaluates challenges in all areas of the field, including algorithms, hardware, software, engineering, and social issues.

Background

The September/October 2004 issue of *CG&A* introduced the term visual analytics to the computer science literature.[1] In 2005, an international team of multidisciplinary panelists consensually and collectively defined the then newly established area as "the science of analytical reasoning facilitated by interactive visual interfaces."[2] The means and targets of VA have since evolved and expanded significantly, covering both scientific and nonscientific data of different types, shapes, sizes, domains, and applications. As extreme-scale datasets began revolutionizing our daily working life,[3] researchers looked to VA for solutions to their big-data problems.

Today's extreme-scale VA applications often combine high-performance computers for computation, high-performance database appliances and/or cloud servers for data storage and management, and desktop computers for human-computer interaction. Sources of extreme-scale data often come from models or observations, arising from different scientific, engineering, social, and cyber applications. Although many petabyte (10^{15}) or even terabyte (10^{12}) data analytics problems remain unsolved, scientists have begun analyzing exabyte (10^{18}) data.

The Top 10 Challenges

Addressing the top challenges has profound, far-reaching implications, for not only fulfilling the critical science and technical needs but also facilitating the transfer of solutions to a wider community. We thus evaluate the problems from both technical and social perspectives. The order of the following challenges doesn't reflect their relative importance but rather the content correlation among individual challenges.

1. In Situ Analysis

The traditional postmortem approach of storing data on disk and then analyzing the data later might not be feasible beyond petascale in the near future. Instead, in situ VA tries to perform as much analysis as possible while the data are still in memory. This approach can greatly reduce I/O

Published by the IEEE Computer Society 0272-1716/12/$31.00 © 2012 IEEE

Pak Chung Wong (PNNL)

Han-Wei Shen (OSU)

Chris Johnson (Utah)

Chaomei Chen (Drexel)

Robert Ross (Argonne)

Top 10 Challenges in Extreme-Scale Data Visual Analytics

- ***In Situ Analysis***
  - Perform as much analysis as possible while the data are still in memory
- ***Interaction and User Interfaces***
  - Machine-based automated systems vs. Human Cognition
- ***Large-Data Visualization***
  - Data projection and dimension Reduction, display technology
- ***Databases and Storage***
  - A cloud-based solution might not meet the needs
- ***Algorithms***
  - Address both data-size and visual-efficiency issues

Top 10 Challenges in Extreme-Scale Data Visual Analytics

- ***Data Movement/Transport, & Network Infrastructure***
  - Efficiently use networking resources and provide convenient abstractions
- ***Uncertainty Quantification***
  - Cope with incomplete data
- ***Parallelism***
- ***Domain and Development Libraries, Frameworks, and Tools***
  - Affordable resource libraries, frameworks, and tools
- ***Social, Community, and Government Engagements***

Acknowledgement

- Students
- Funds
  - NSFC
  - 863
  - PKU
  - Beijing NSF
- Collaborators
  - Jian Huang University of Tennessee
  - Zhu Xiaoming, SDSCC
  - Yongxian Zhang, China Earthquake Network Center
  - Xiaoguang Ma, CAS IAP
  - More …

http://vis.pku.edu.cn/wiki
Xiaoru.yuan@pku.edu.cn