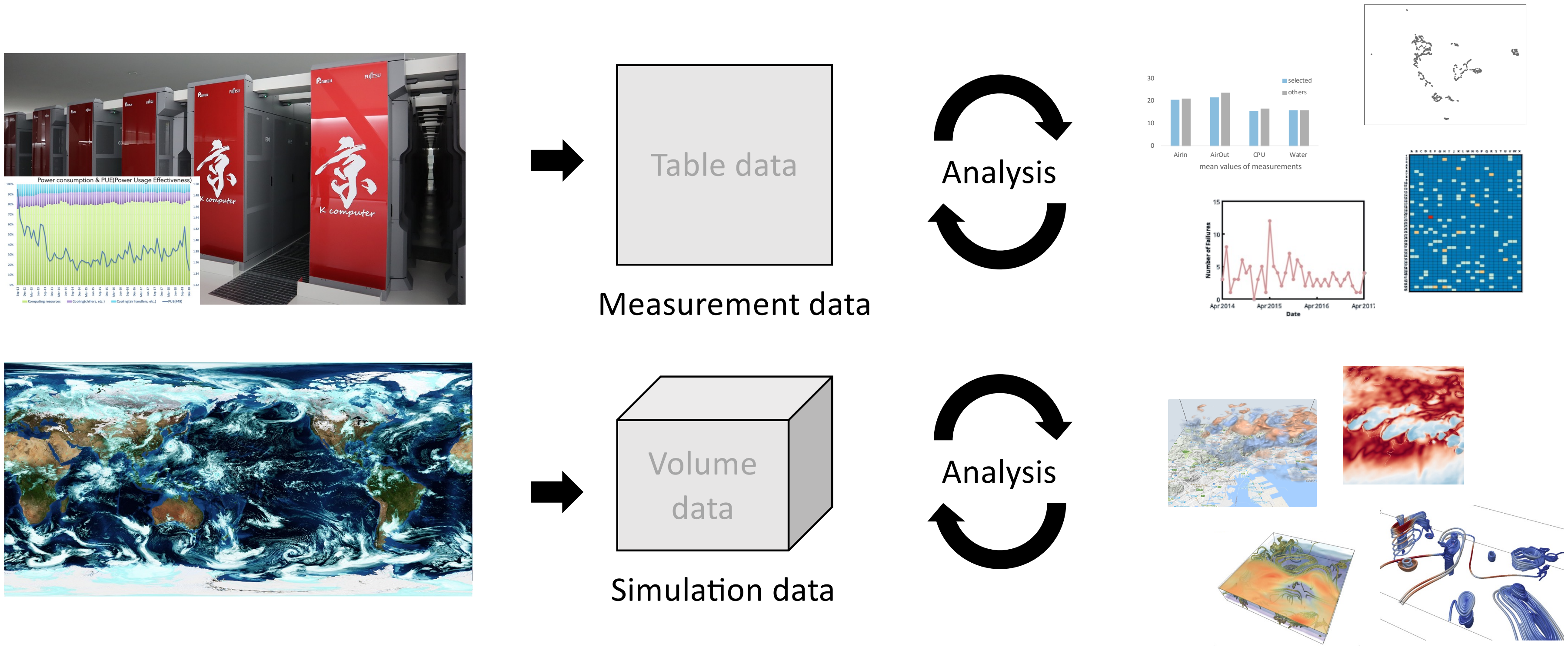
Tensor based Visual Analytics for Multidimensional Time-series Data

Naohisa Sakamoto
Graduate School of System Informatics, Kobe University, Japan

Data Science, Statistics & Visualization 2022 (DSSV 2022)
National Cheng Kung University, Tainan, Taiwan
June 28, 2022

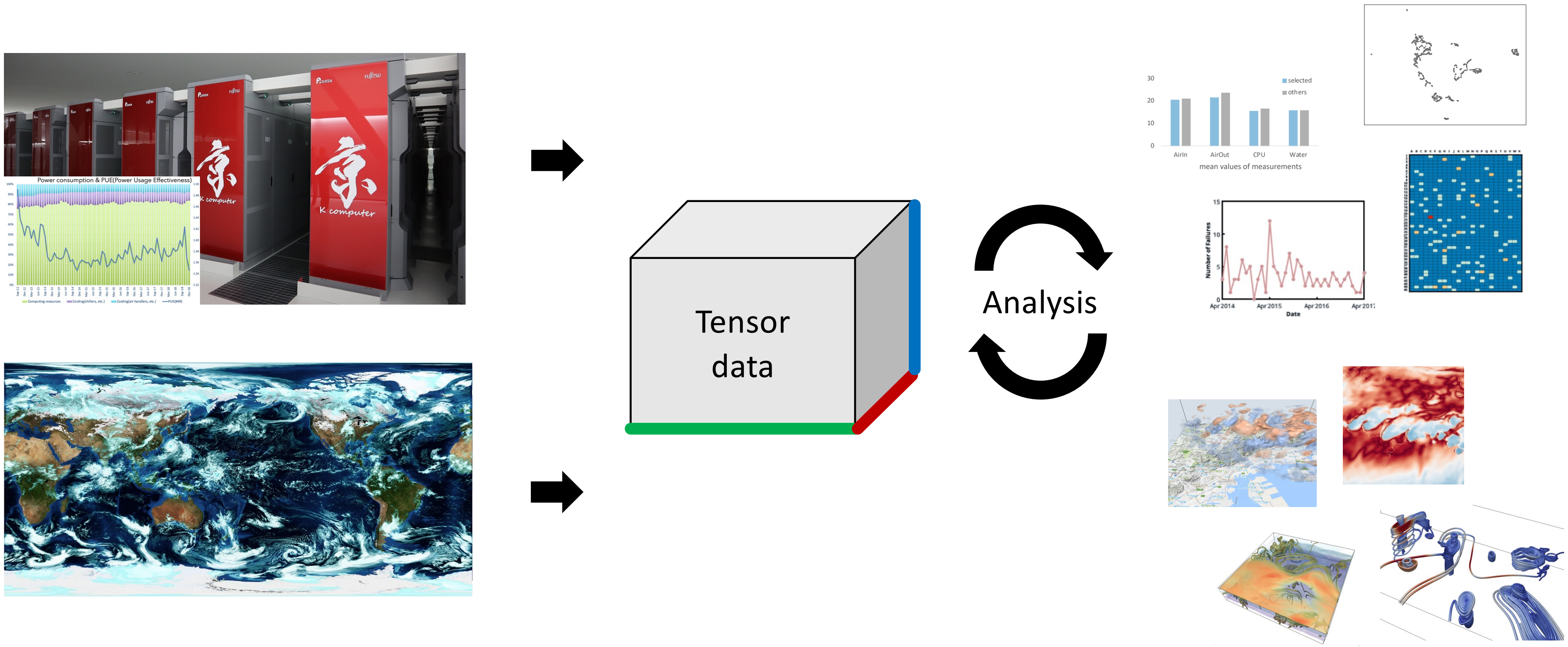
Background

- Multidimensional Time Series Data



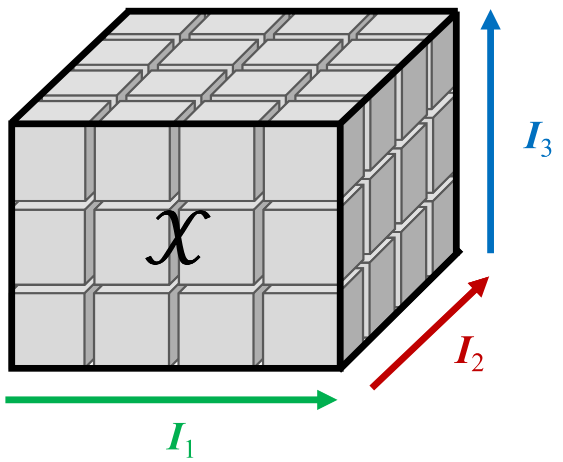
Background

- Multidimensional Time Series Data

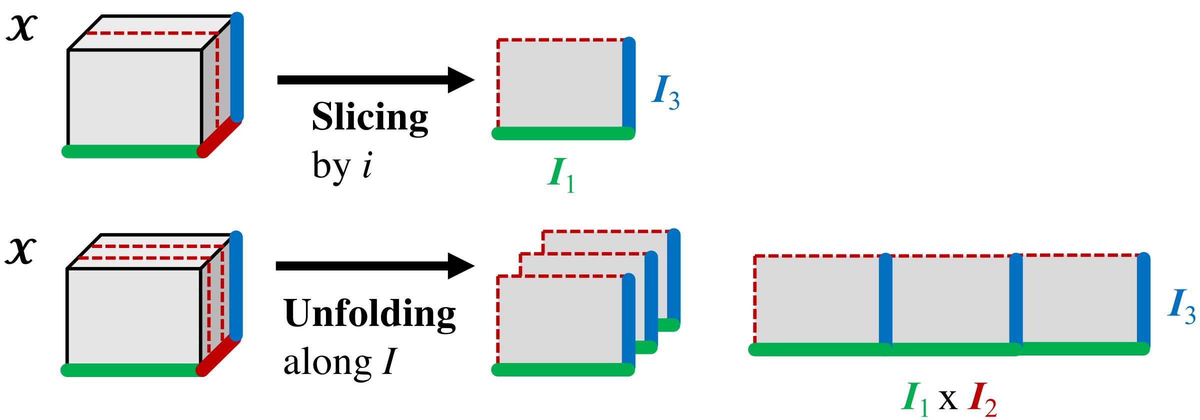


Tensor Data Model

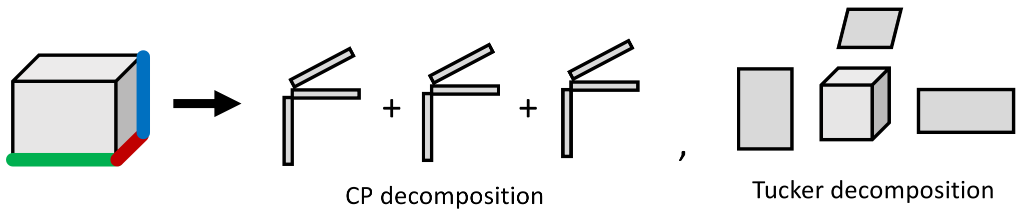
- Multidimensional Array
    - Ex. 3rd order tensor (3 dimensional array)



Basic operations

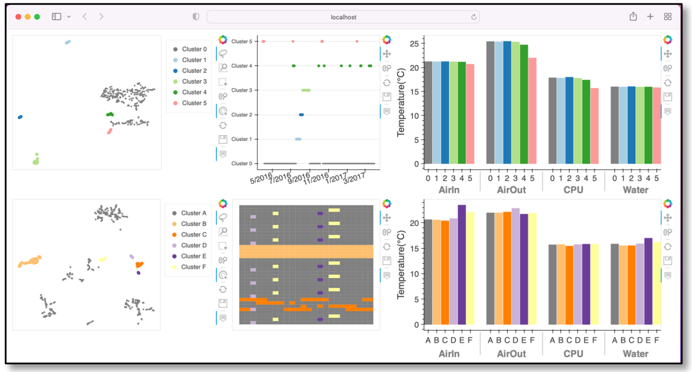


Decomposition

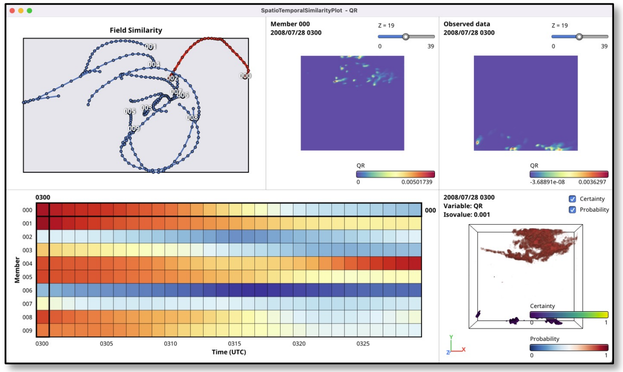


Tensor based Visual Analytics

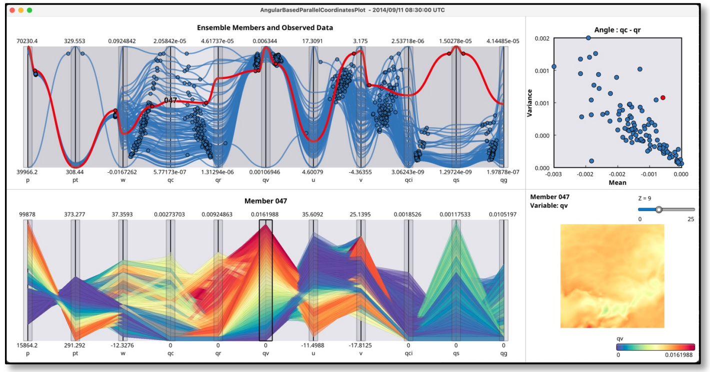
1. HPC system log data analysis
2. Weather ensemble data analysis



HPC log data analysis with multi-step dimensionality reduction



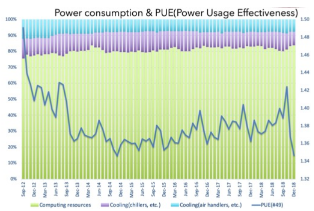
Weather ensemble data analysis based on field similarity



Weather ensemble data analysis with angular based edge bundled PCP

Visual Analytics for HPC System Log Data

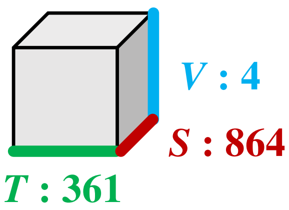
- Background
  - Recent advances on high performance computing (HPC) systems
    - Continuously increasing the size, complexity and performance
      - Supercomputer Fugaku (ex. 158,976 nodes)
    - Highly important to ensure a reliable and stable operation
  - Various environmental data from sensors
    - For monitoring the hardware system
    - There is a demand for an effective operational data analysis tool to understand its behavior

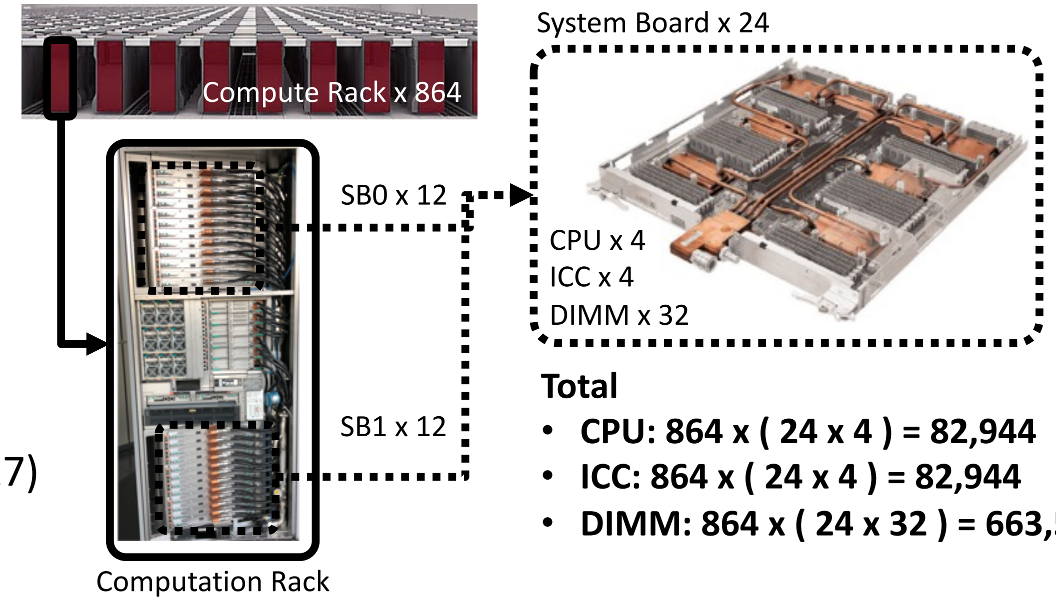


[F.Shoji, Operation of the K computer and the facility, The 1st R-CCS International Symposium, 2019]

Environmental Log Data

- Log data from sensors (3rd order tensor data)
  - **Variable (T)**
    - Temperature
      - AirIn, AirOut, CPU, Water
  - **Space (S)**
    - Rack position
    - # of racks: 864
  - **Time (T)**
    - Measurement time
    - 361 days (April 2026 – March 2017)



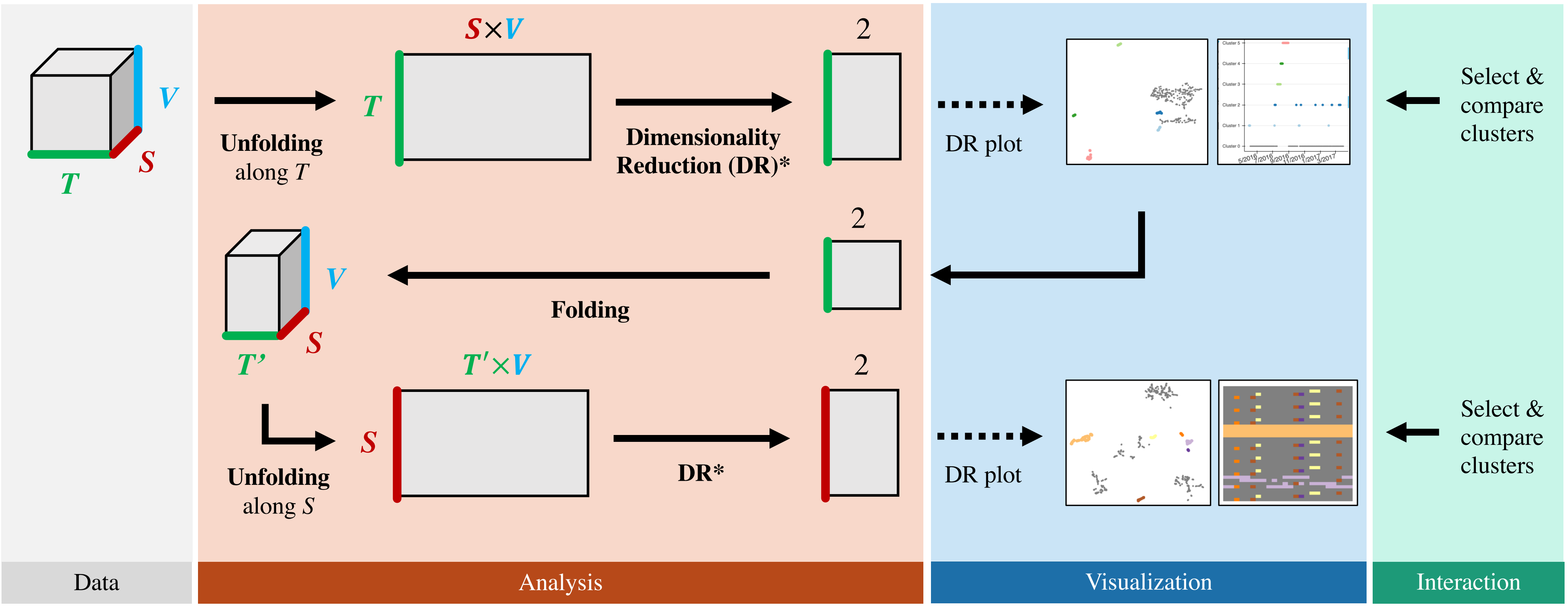


Total

- **CPU: 864 x (24 x 4) = 82,944**
- **ICC: 864 x (24 x 4) = 82,944**
- **DIMM: 864 x (24 x 32) = 663,552**

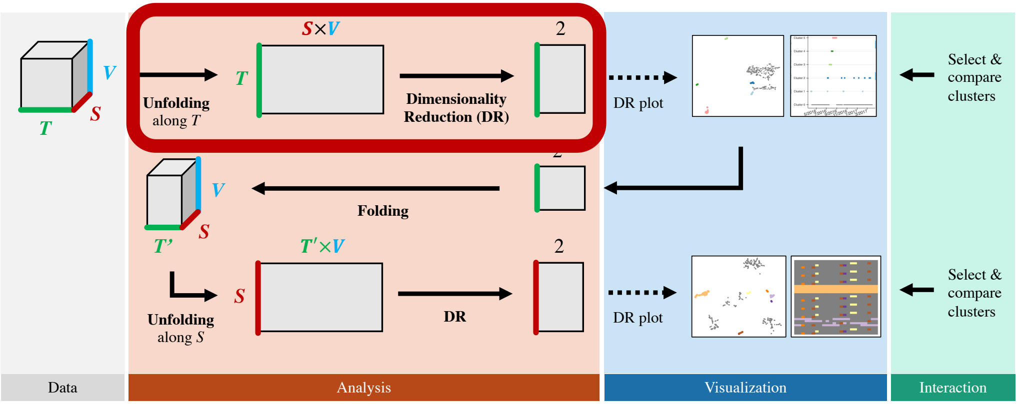
System configuration of the K computer

Workflow



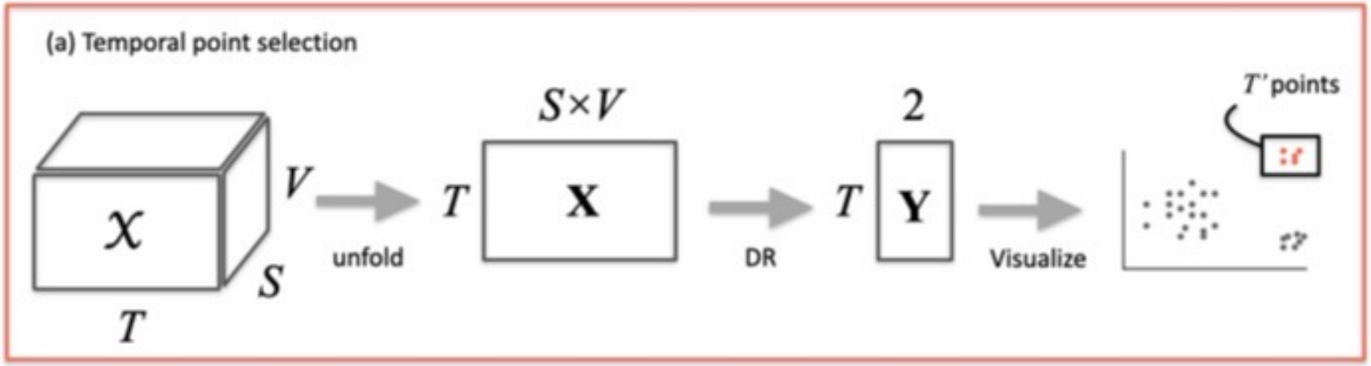
* Inspired by Multi-step DR: T.Fujiwara, Shilpika, N. Sakamoto, J. Nonaka, K. Yamamoto, and K.-L. Ma, A Visual Analytics Framework for Reviewing Multivariate Time-Series Data with Dimensionality Reduction, TVCG, 2020.

Workflow

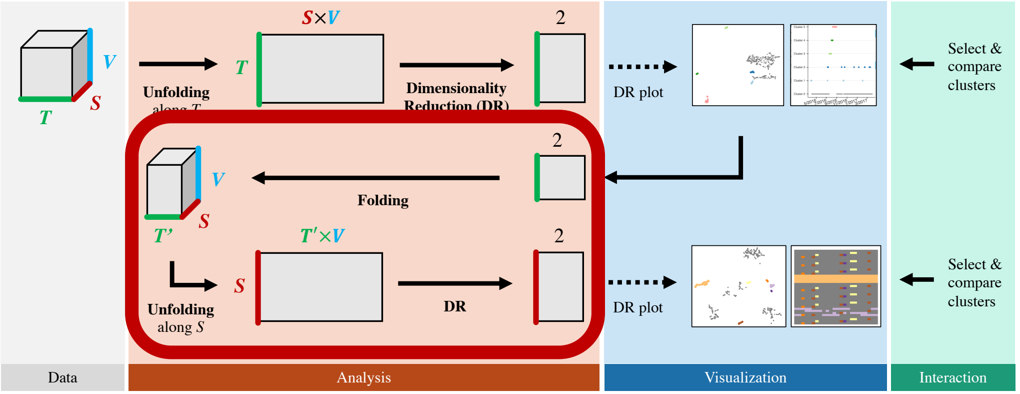


Finding temporal point clusters that show characteristic behaviors

1. Unfold the tensor $\mathcal{X}$ into a matrix $\mathbf{X}$ with T rows x $(\mathrm{S}\mathrm{x}\mathrm{V})$ columns
2. Apply DR to $\mathbf{X}$ and obtain a matrix $\mathbf{Y}$ with T rows x 2 columns
3. Plot $\mathbf{Y}$ in a 2D space and select the temporal points of interest

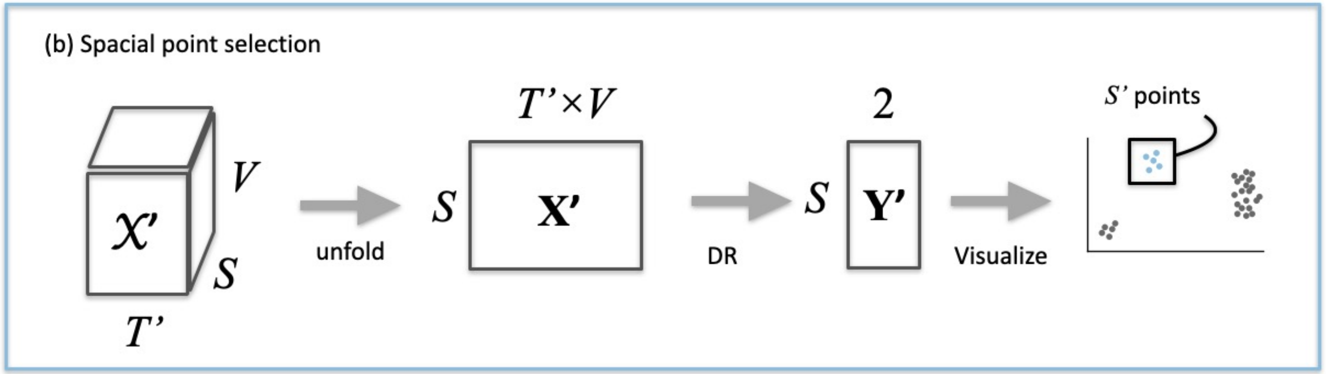


Workflow



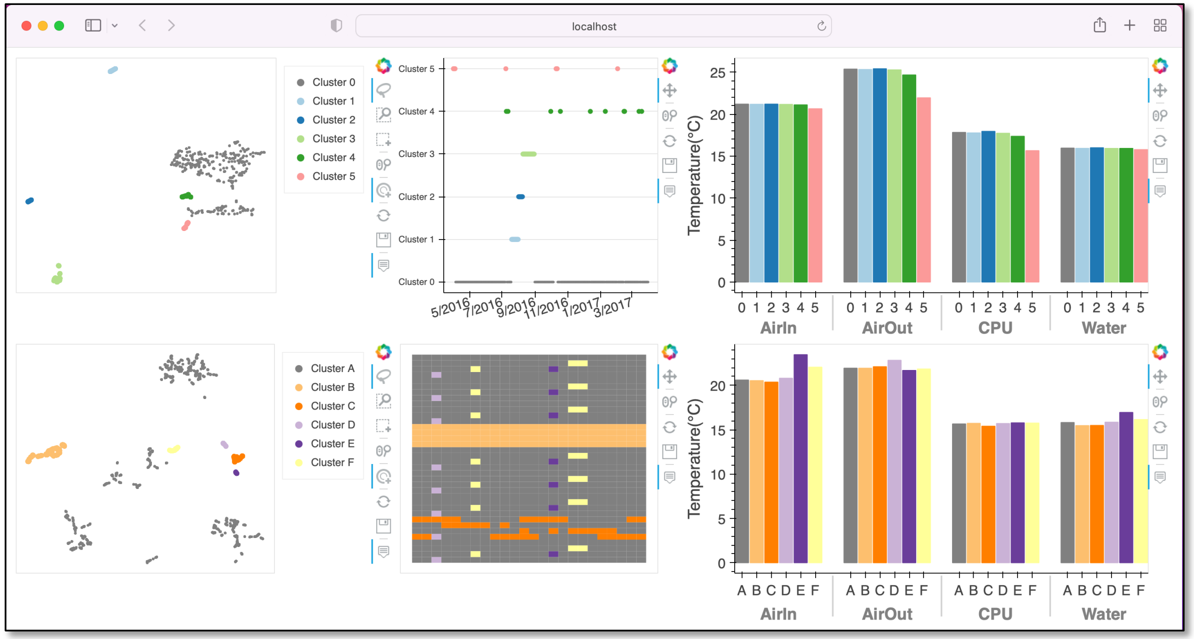
Finding space point clusters from the data corresponding to the selected temporal points

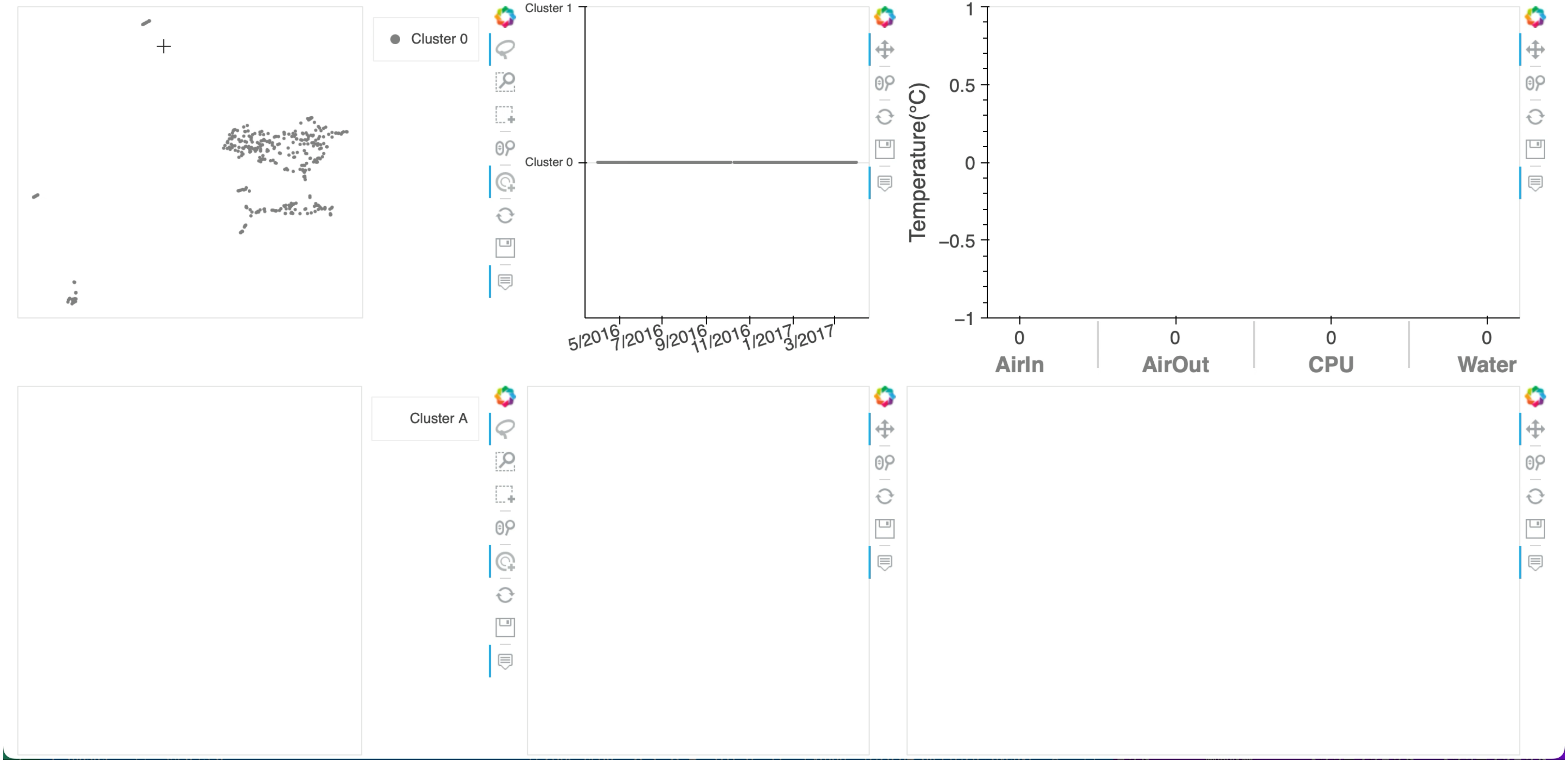
1. Extract $\mathcal{X}'$ corresponding to T' points from $\mathcal{X}$
2. Unfold $\mathcal{X}'$ into a matrix $\mathbf{X}'$ with S rows x T'xV columns
3. Apply DR to $\mathbf{X}'$ and plot the obtained $\mathbf{Y}'$ in a 2D space



Visual Analytic System

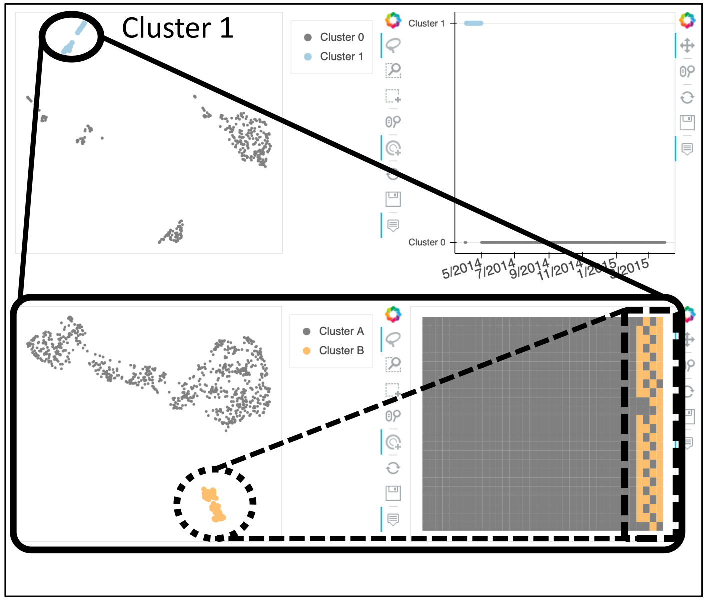
- HPC log data analysis



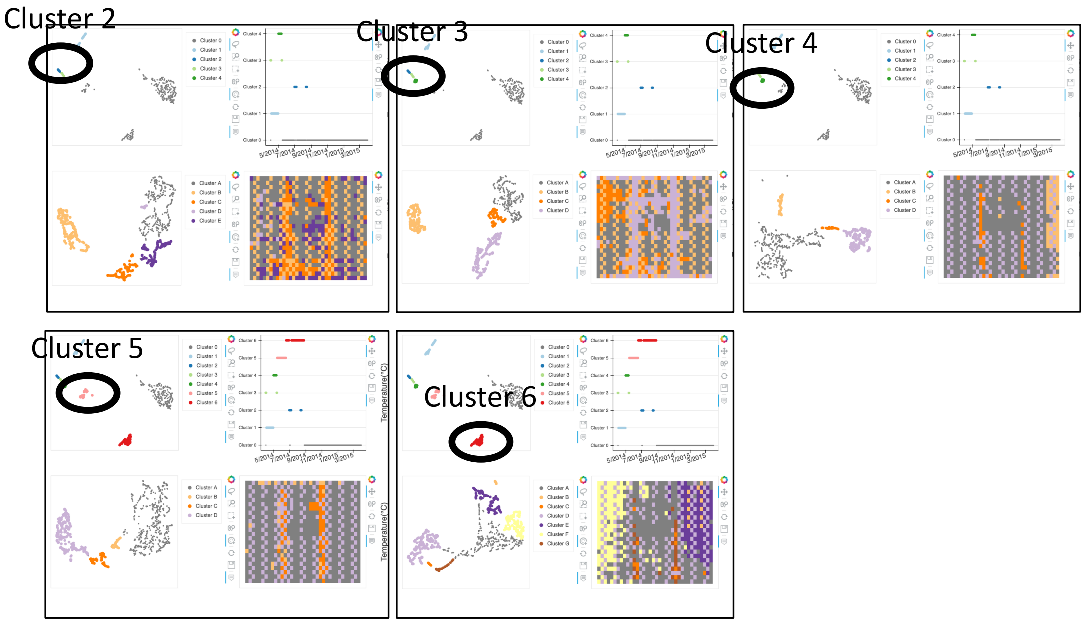

Cluster 0
Cluster 1
Cluster 0
5/2016 7/2016 9/2016 11/2016 1/2017 3/2017
Temperature(°C)
1
0.5
0
−0.5
−1
0 AirIn
0 AirOut
0 CPU
0 Water
Cluster A

Example of analysis

- Temporal & Spatial point selection



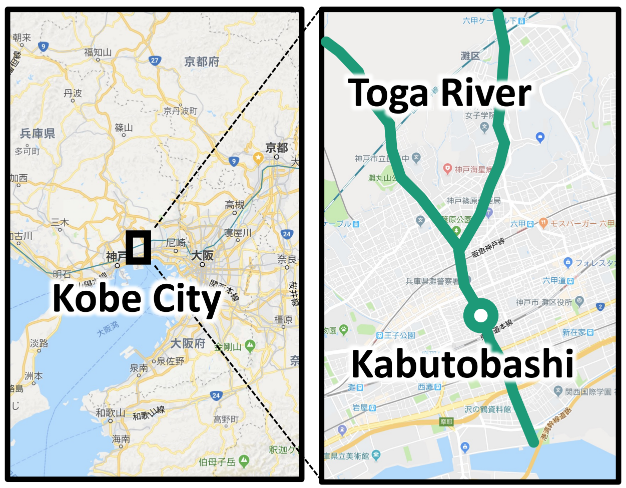
→ Effected by the Gordon Bell Challenge

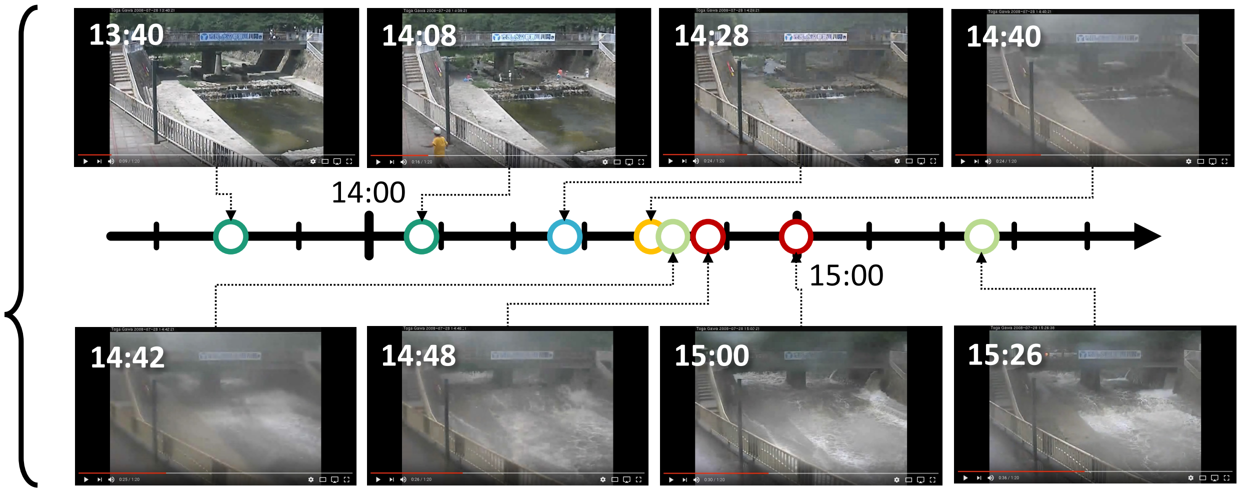


→ No characteristic patterns

Visual Analytics for Weather Ensemble Data

- Background
  - Guerrilla Rainstorm
    - Flood accident in Toga river, Kobe, 2008.7.28
  - Prediction using supercomputers for disaster prevention and mitigation from sudden and unexpected torrential rain

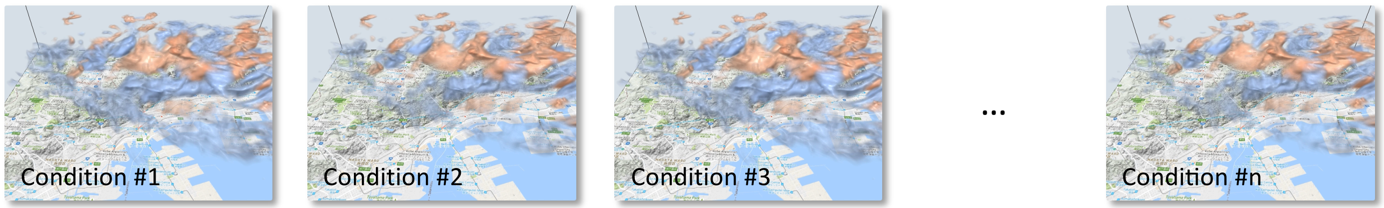




https://www.youtube.com/watch?v=gwrwHNVsFe8

Visual Analytics for Weather Ensemble Data

- Weather ensemble simulation
  - Reproduce future atmospheric conditions by running multiple weather simulations with different initial conditions
  - Probabilistic forecasts that account for uncertainties in weather models



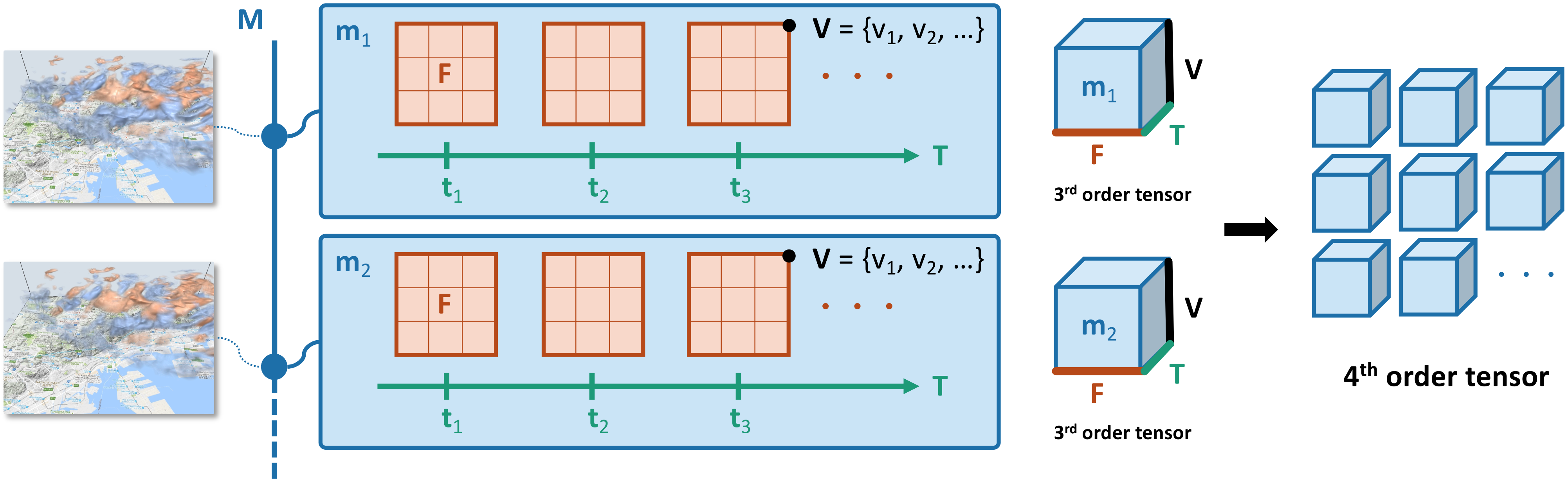
Parallel world

Statistical analysis

- Multiple ensemble members
- Multi-variate
- Multi-timesteps (time-varying)

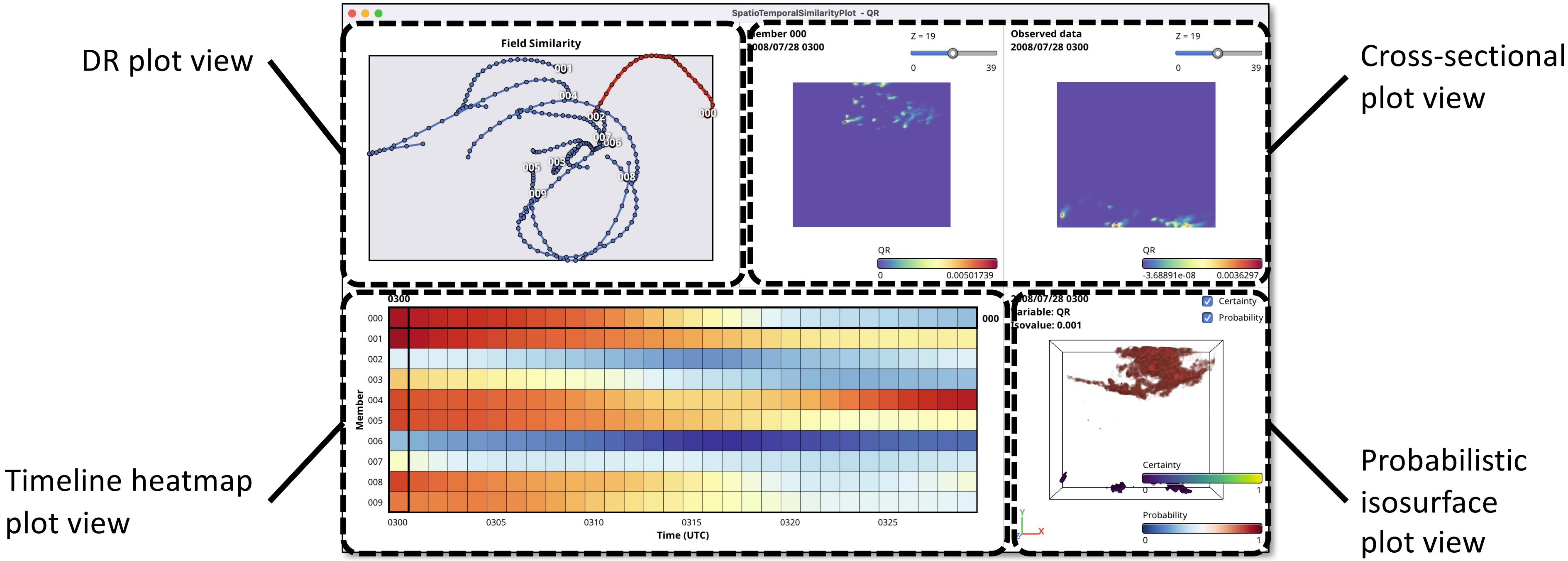
Weather Ensemble Data

- 4^{th} order tensor data
  - Member (M), Field (F), Variable (V), Time (T)



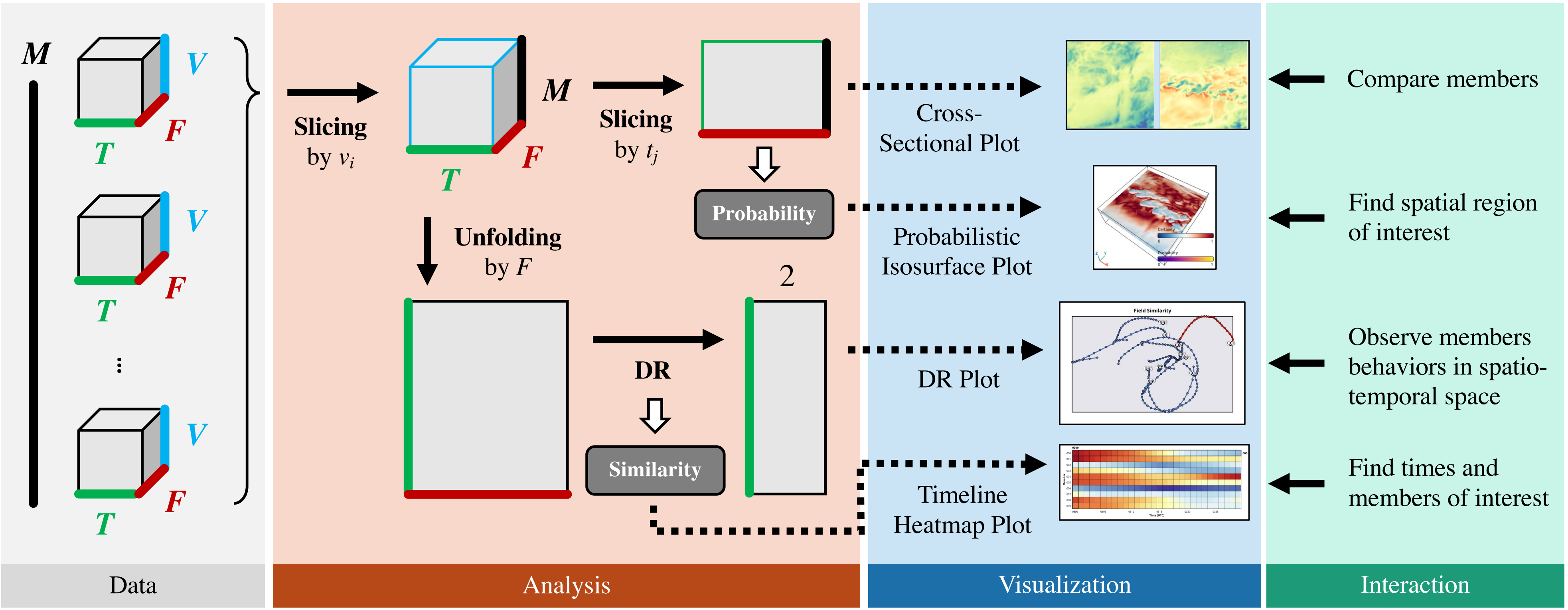
Visual Analytic System

- Weather ensemble data analysis



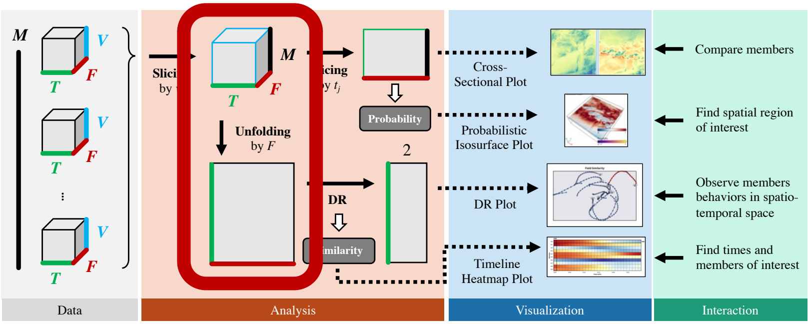
Workflow

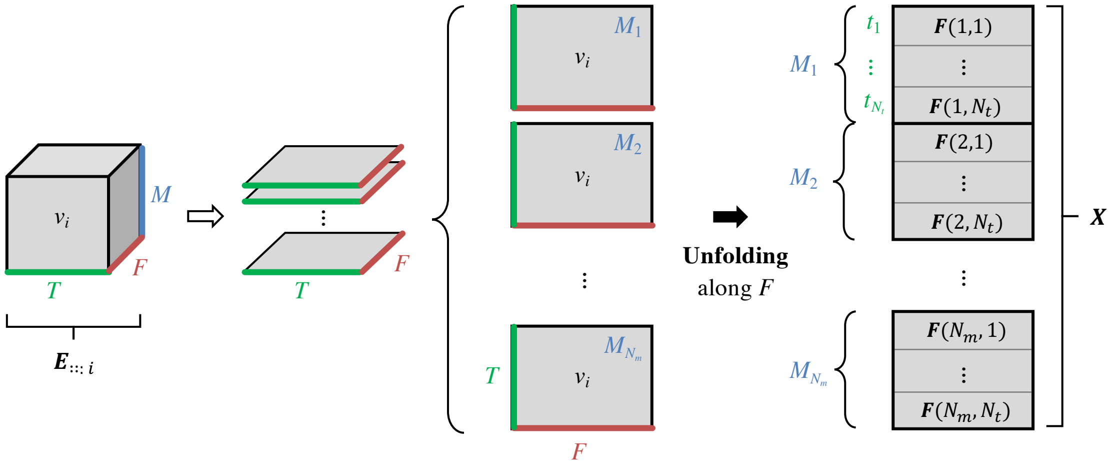
- Analysis with the fixed variable point (v_i)



Workflow

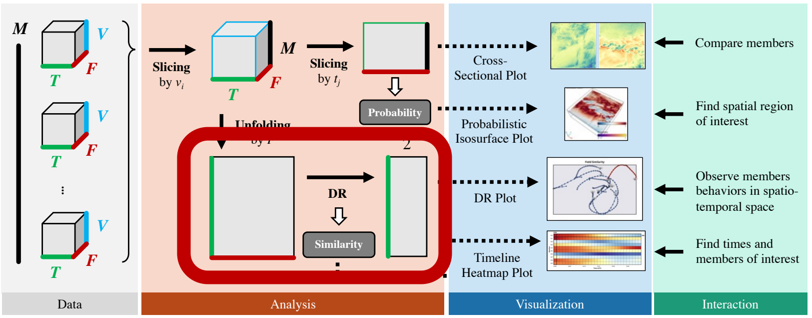
- Analysis with the fixed variable point (v_i)

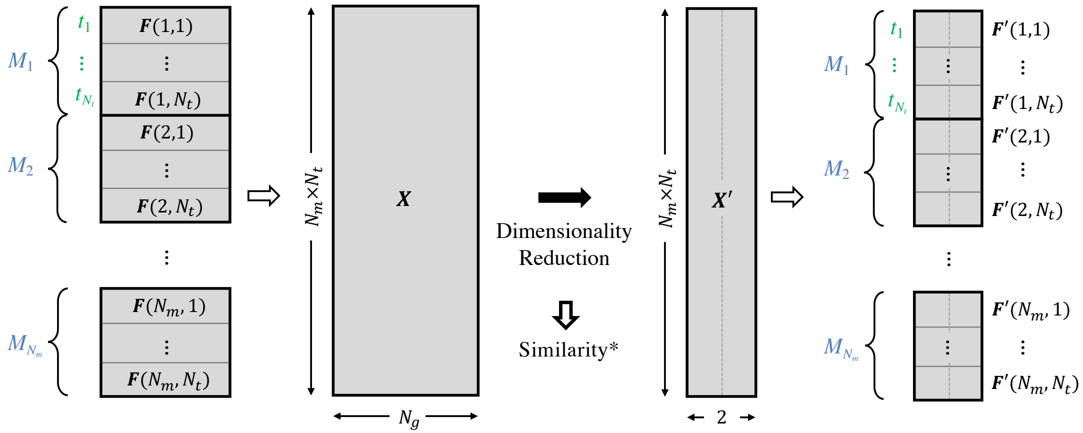




Workflow

- Analysis with the fixed variable point (v_i)

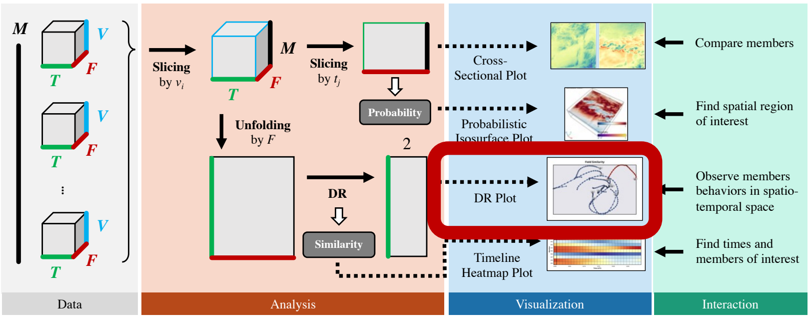


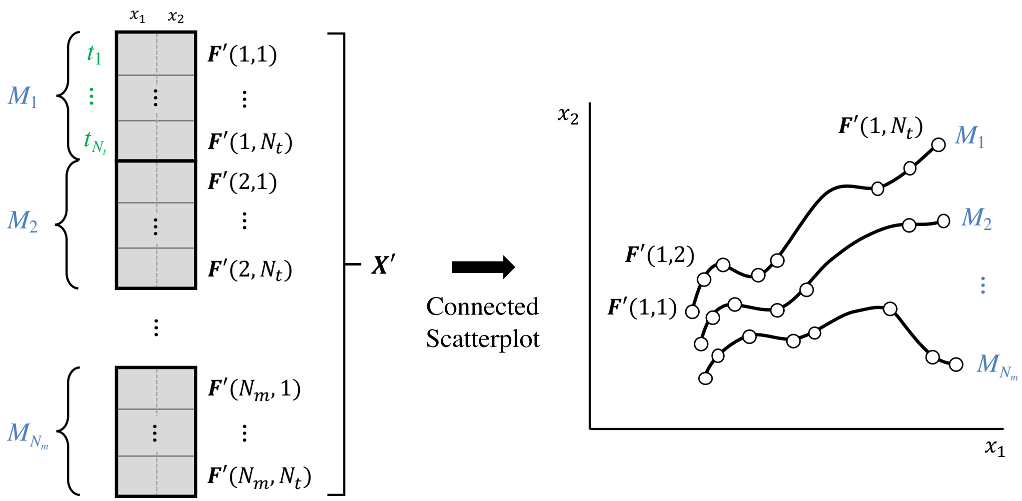


* Similarity: A.Fofonov et.al.. Projected field similarity for comparative visualization of multi-run multi-field time-varying spatial data. CGF, 2019

Workflow

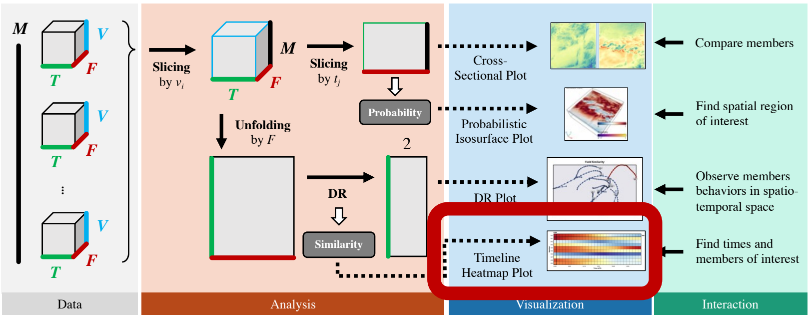
- Analysis with the fixed variable point (v_i)

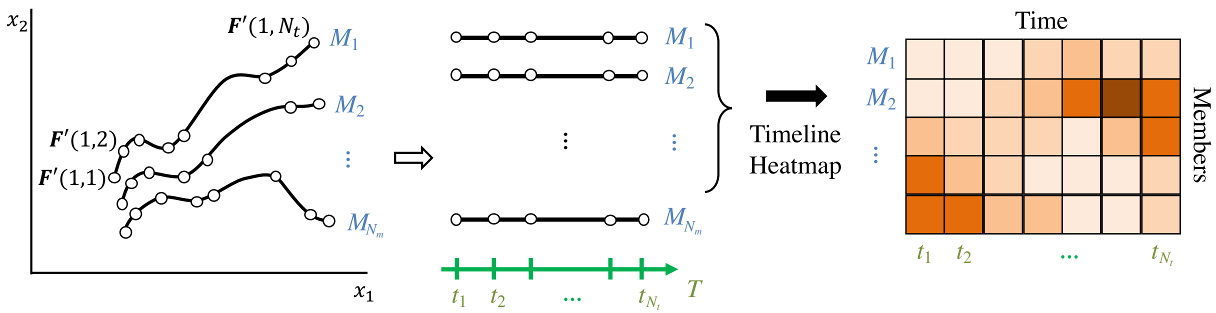




Workflow

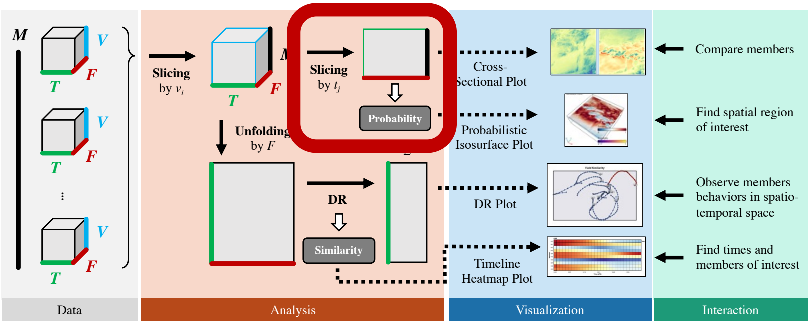
- Analysis with the fixed variable point (v_i)

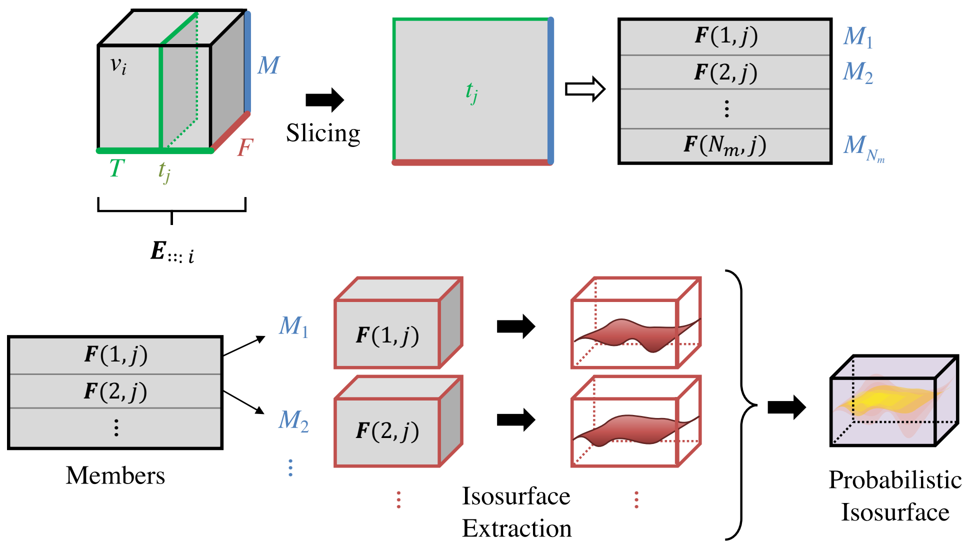




Workflow

- Analysis with the fixed variable point (v_i)

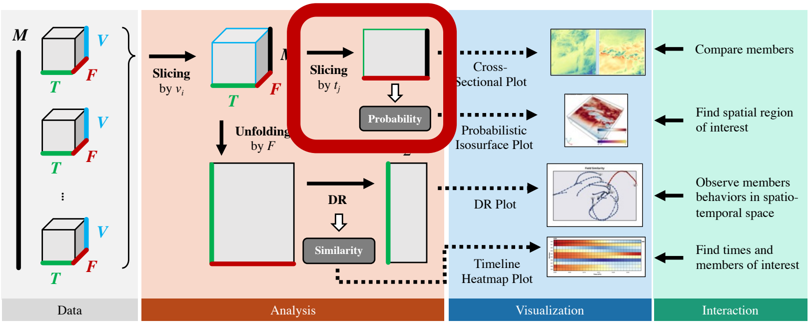


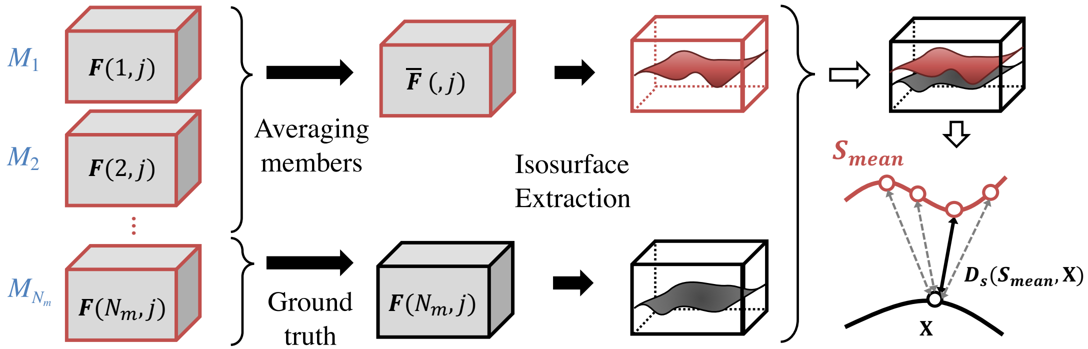


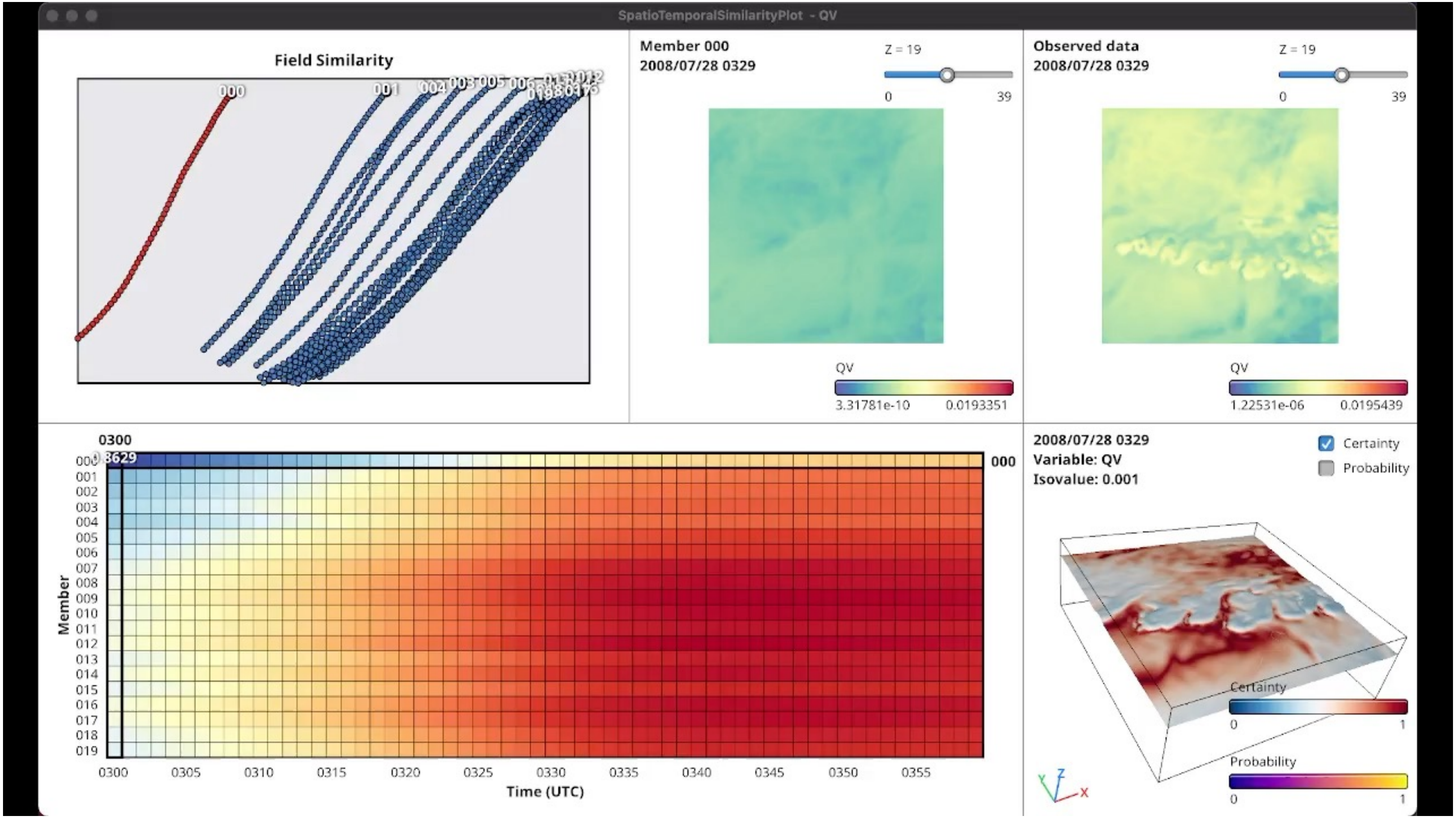
K. Pöthkow, etl.al.. Probabilistic marching cubes. CGF, 2011.

Workflow

- Analysis with the fixed variable point (v_i)



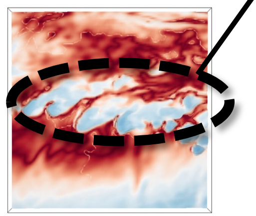




SpatioTemporalSimilarityPlot - QV
Field Similarity
Member 000
2008/07/28 0329
Z = 19
0
39
QV
3.31781e-10
0.0193351
Observed data
2008/07/28 0329
Z = 19
0
39
QV
1.22531e-06
0.0195439
0300
Member
Time (UTC)
000
2008/07/28 0329
Variable: QV
Isovalue: 0.001
Certainty
Probability
Certainty
0
1
Probability
0
1

Example of analysis

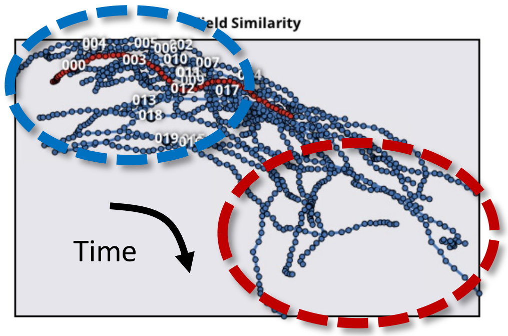
- Case study
  - Variable: QR (mixture ratio of rainwater)
  - Ensemble members: 20
  - Timesteps: 60
  - Resolution: 301 x 301 x 50



Ascending air current and clouds, with strong non-linearity

- Producing high variability in the simulation results
- Reducing the accuracy of the simulation model

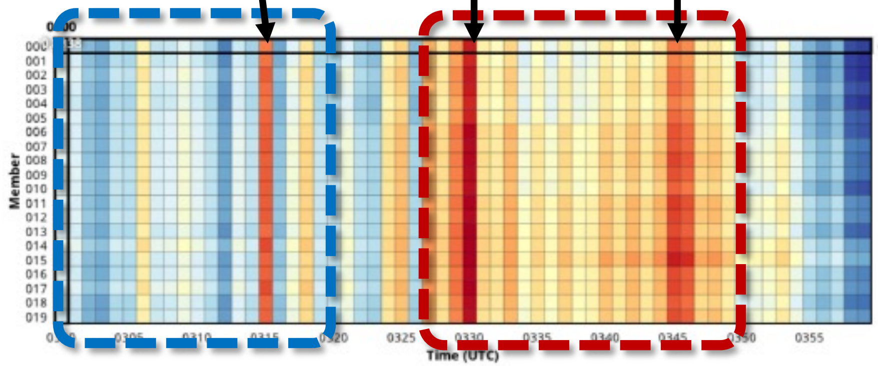
Densely Plotted (Similar)



Separately plotted (Not similar)

Affected by the period of cumulonimbus cloud development?

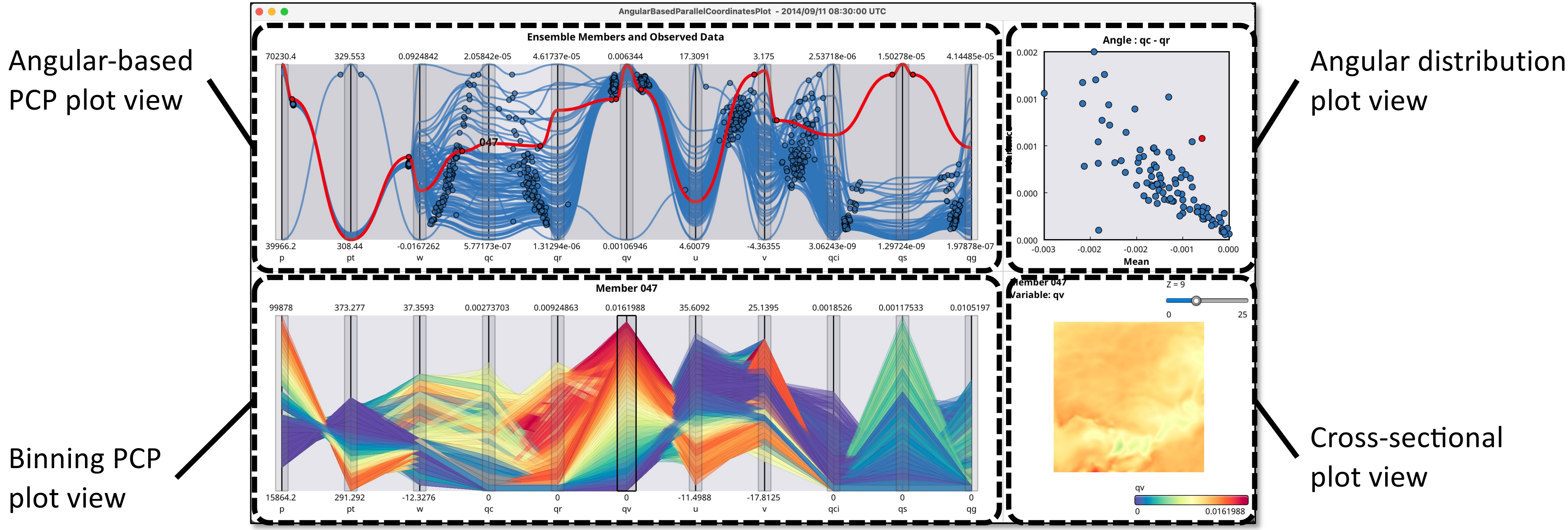
3:15 3:30 3:45



Similar → Time → Not similar

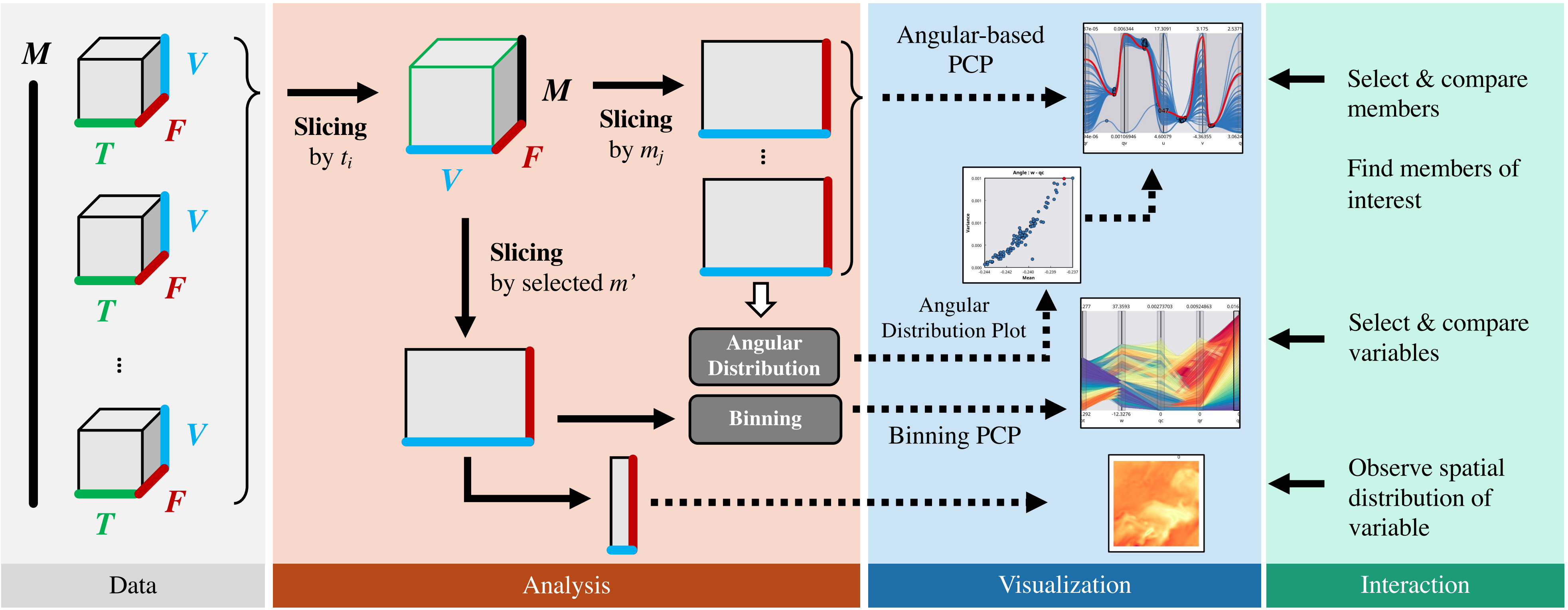
Visual Analytic System

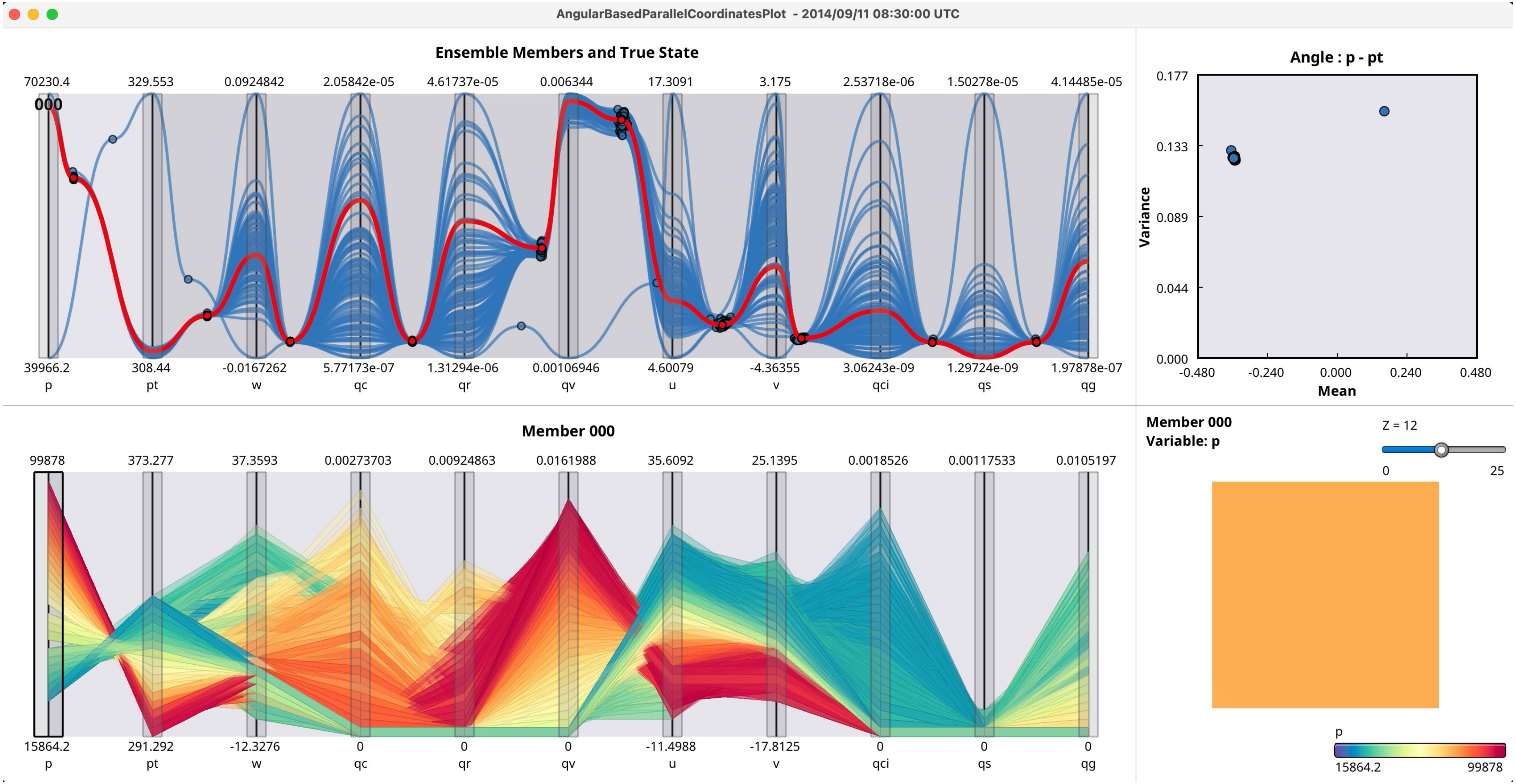
- Weather ensemble data analysis



Workflow

- Analysis with the fixed time point




AngularBasedParallelCoordinatesPlot - 2014/09/11 08:30:00 UTC
Ensemble Members and True State
70230.4 329.553 0.0924842 2.05842e-05 4.61737e-05 0.006344 17.3091 3.175 2.53718e-06 1.50278e-05 4.14485e-05
000
39966.2 308.44 -0.0167262 5.77173e-07 1.31294e-06 0.00106946 4.60079 -4.36355 3.06243e-09 1.29724e-09 1.97878e-07
p pt w qc qr qv u v qci qs qg
Angle : p - pt
Variance
0.177 0.133 0.089 0.044 0.000
-0.480 -0.240 0.000 0.240 0.480
Mean
Member 000
99878 373.277 37.3593 0.00273703 0.00924863 0.0161988 35.6092 25.1395 0.0018526 0.00117533 0.0105197
15864.2 291.292 -12.3276 0 0 0 -11.4988 -17.8125 0 0 0
p pt w qc qr qv u v qci qs qg
Member 000
Variable: p
Z = 12
0 25
p
15864.2 99878

Conclusion

- Tensor based visual analytics
  - Concise workflow description by using tensor models
  - Easy application of data analysis techniques such as dimensionality reduction
  - Efficient data analysis by combining effective visualization techniques
  - Further improvement of data interpretability is a future challenge