

多次元時系列データの視覚的解析

坂本 尚久

神戸大学大学院システム情報学研究科

総合工学委員会科学的知見の創出に資する可視化分科会
ICT時代の文理融合研究を創出する可視化小委員会（第25期・第3回）
2021年12月13日

内容

前半

- これまでの研究
 - 自己紹介
 - 研究内容

後半

- 現在の研究
 - 多次元時系列データの視覚的解析
 - 事例 1 : スパコンログデータ解析
 - 事例 2 : アンサンブルデータ解析

内容

前半

- これまでの研究
 - 自己紹介
 - 研究内容

後半

- 現在の研究
 - 多次元時系列データの視覚的解析
 - 事例 1 : スパコンログデータ解析
 - 事例 2 : アンサンブルデータ解析

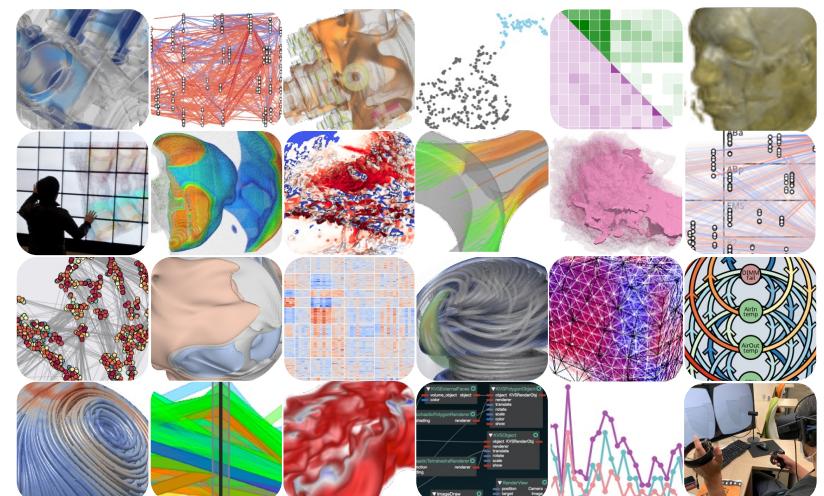
自己紹介

・坂本尚久（さかもとなおひさ）

- ・所属：神戸大学・システム情報学研究科・計算科学専攻
- ・学位：工学博士（京都大学工学研究科・小山田耕二教授）
- ・職歴：株式会社KGT（ビジュアリゼーション事業部・宮地英生教授）

・研究内容

- ・データ可視化と視覚的データ解析
 - ・Intelligent Visualization（知的な可視化）
 - ・Interactive Analytics（対話的な解析）
 - ・Immersive Exploration（没入的な探索）



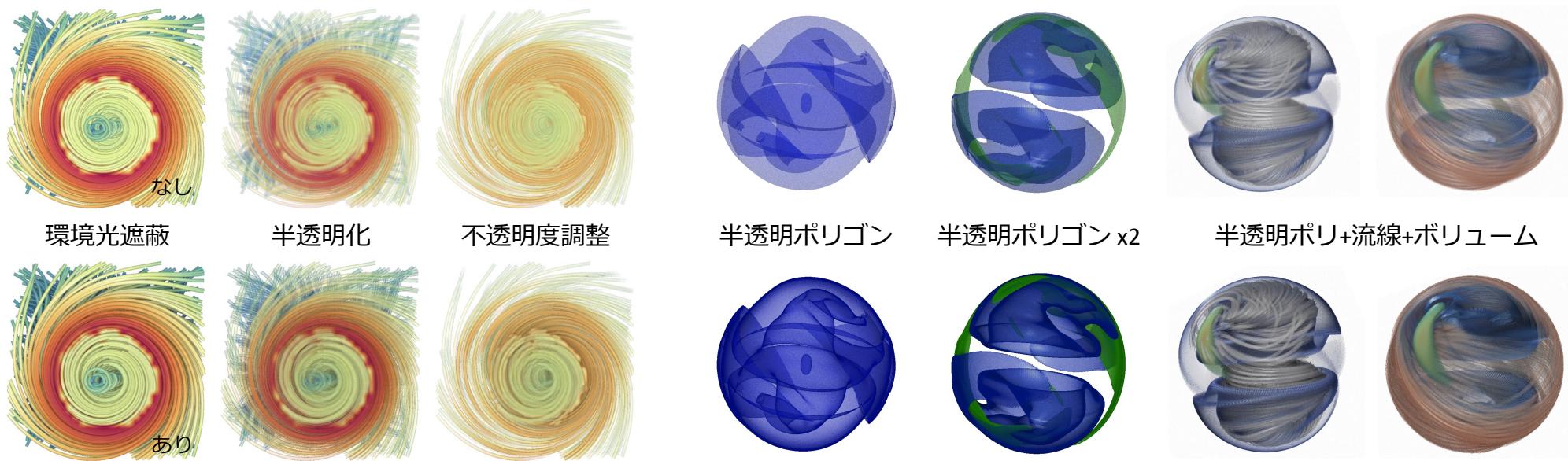
知的な可視化：Intelligent Visualization

- ・重要な変化を見逃さない賢い可視化
 - 複雑データ向け高精細半透明レンダリング
 - スパコン向けスマートIn-situ可視化

知的な可視化：Intelligent Visualization

・重要な変化を見逃さない賢い可視化

- 複雑データ向け高精細半透明レンダリング
- スパコン向けスマートIn-situ可視化



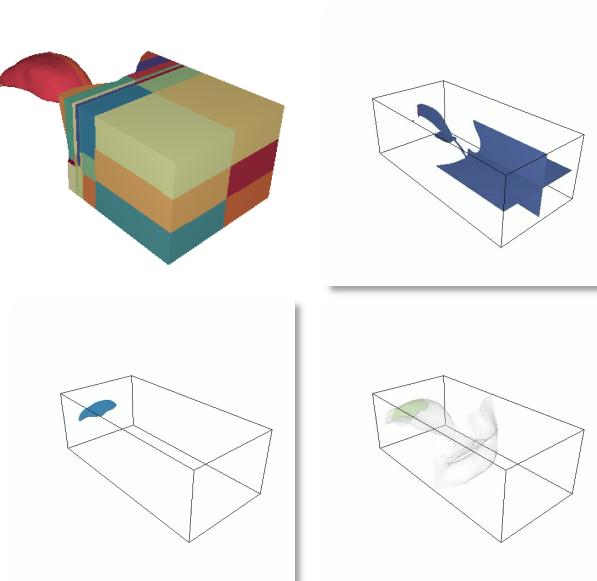
Y. Fujita, et. al., Ambient Occlusion for Semi-transparent Streamlines with Stochastic Rendering Technique, ASV15, 2019.

知的な可視化：Intelligent Visualization

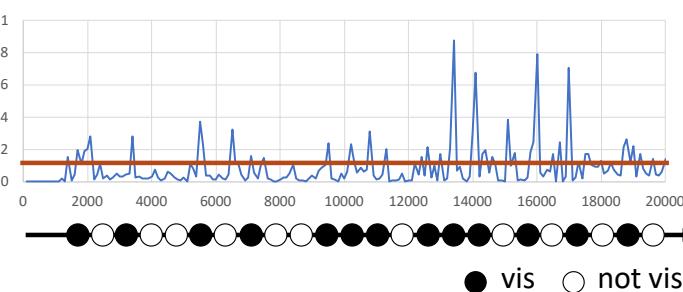
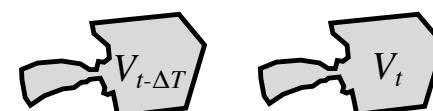
- 重要な変化を見逃さない賢い可視化

➤複雑データ向け高精細半透明レンダリング

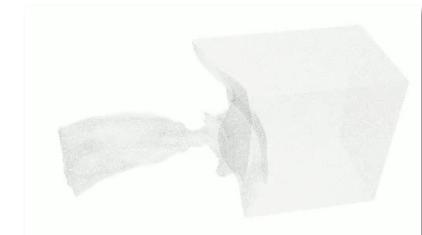
➤スパコン向けスマートIn-situ可視化



$t - \Delta T$ t Time



間引きなし (4,000フレーム)



間引きあり (2,064フレーム)

対話的な解析 : Interactive Analytics

- 対話的に試行錯誤ができる視覚的解析

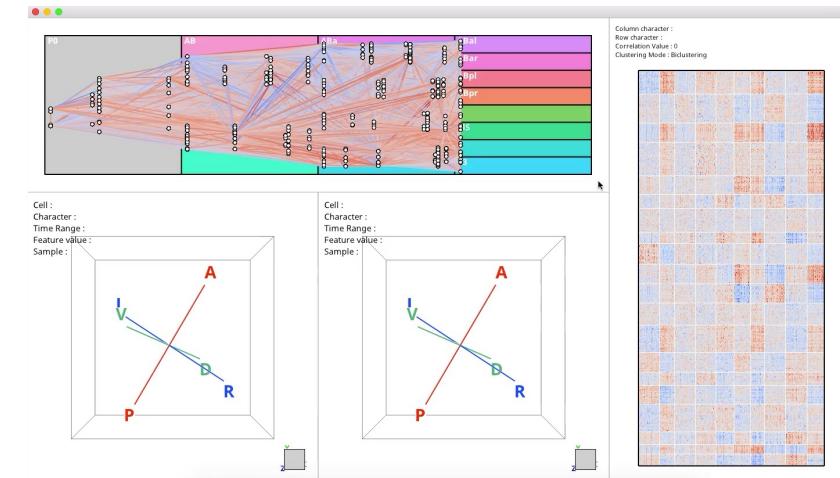
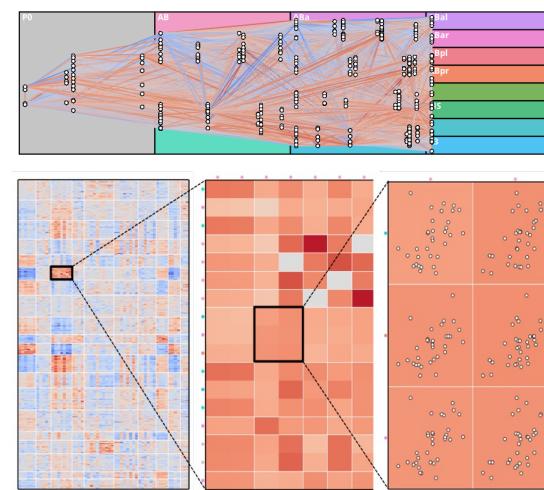
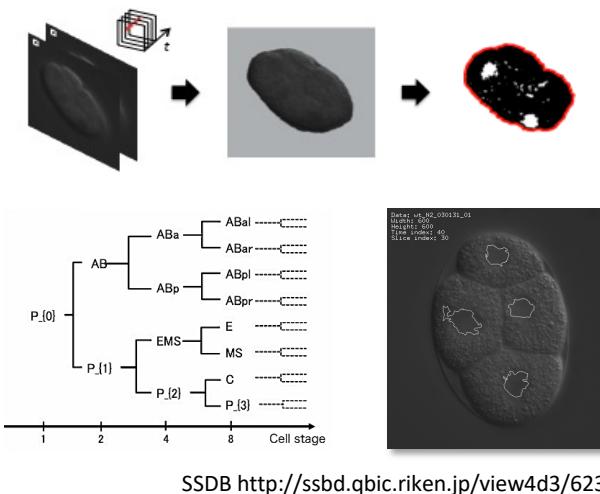
- 生物細胞の発生動態解析
- スパコンログデータを使った異常解析

対話的な解析 : Interactive Analytics

- 対話的に試行錯誤ができる視覚的解析

➤生物細胞の発生動態解析

➤スパコンログデータを使った異常解析

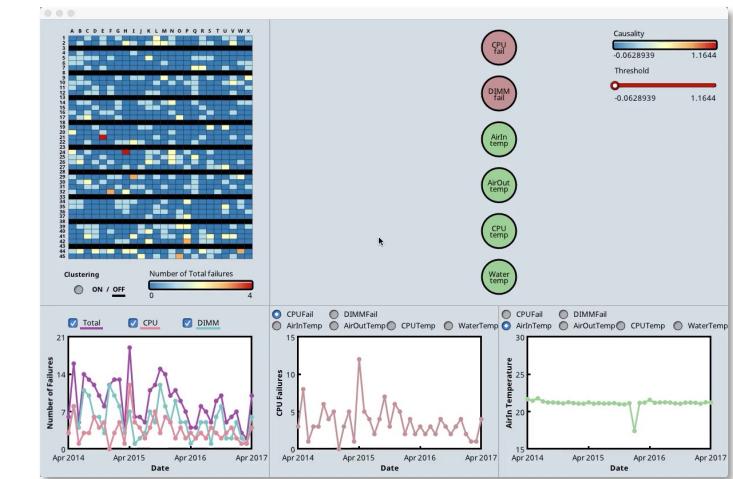
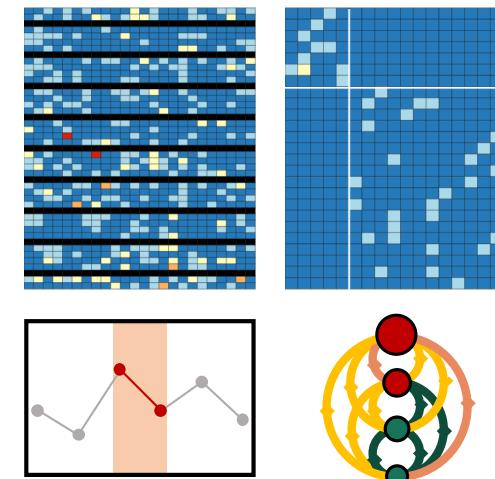
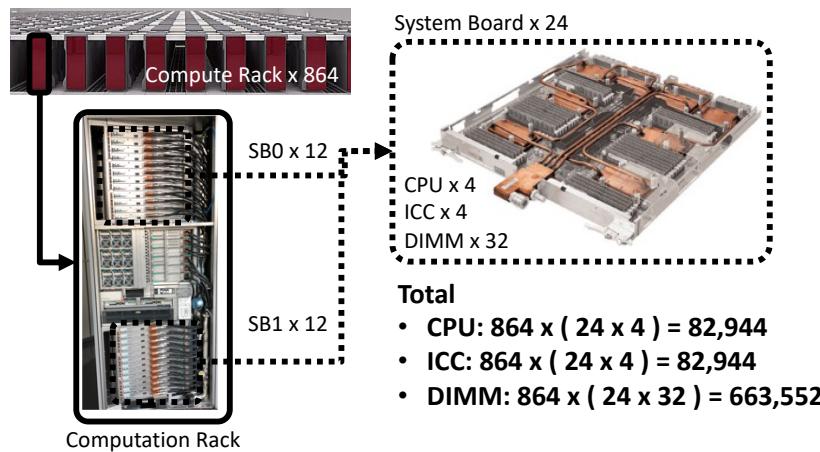


対話的な解析 : Interactive Analytics

- 対話的に試行錯誤ができる視覚的解析

➤生物細胞の発生動態解析

➤スパコンログデータを使った異常解析



没入的な探索 : Immersive Exploration

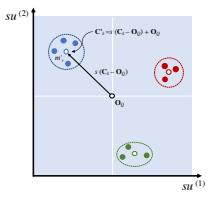
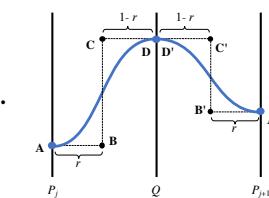
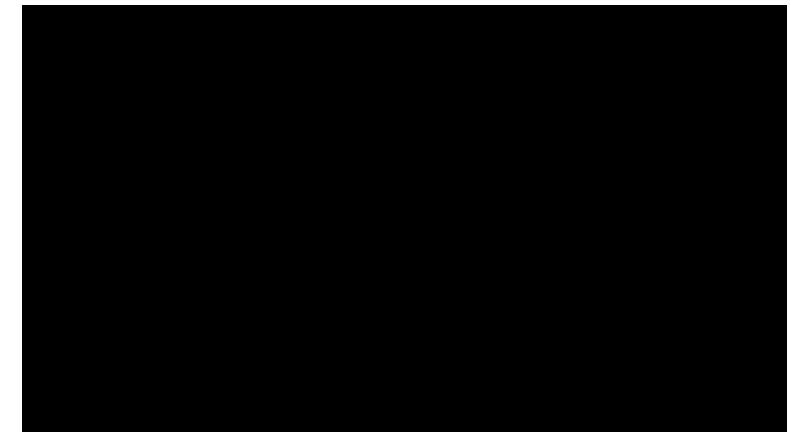
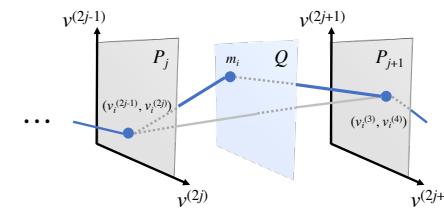
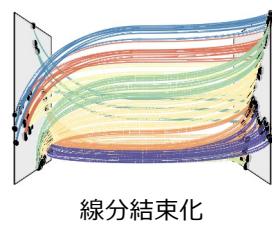
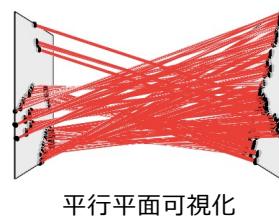
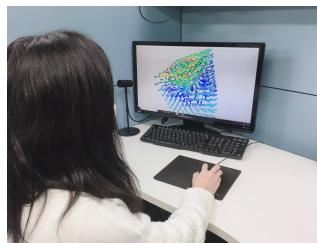
- VR技術を使った没入型データ解析空間

- VR可視化基盤の開発
- 高次元データ解析空間の構築

没入的な探索 : Immersive Exploration

- VR技術を使った没入型データ解析空間

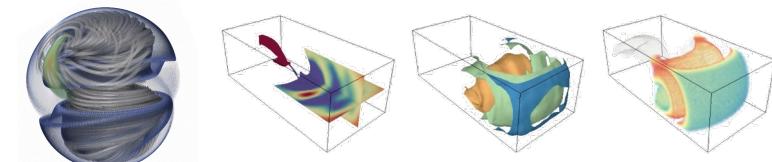
- VR可視化基盤の開発
- 高次元データ解析空間の構築



研究テーマ

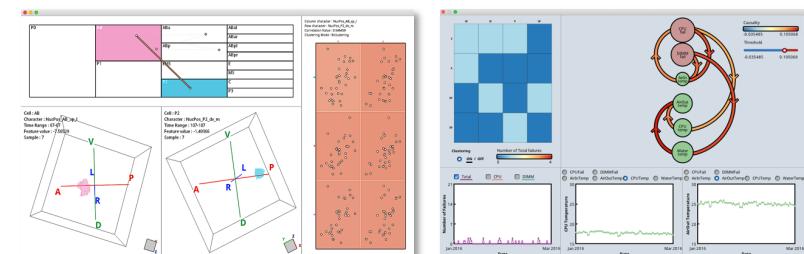
- **Intelligent Visualization : 知的な可視化**

- 重要な変化を見逃さない知的な可視化
 - ✓ 高精細半透明レンダリング
 - ✓ スパコン向けスマートIn-situ可視化



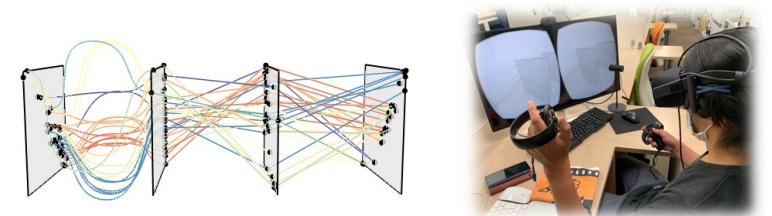
- **Interactive Analytics : 対話的な解析**

- 対話的に試行錯誤ができる視覚的解析
 - ✓ 生物細胞の発生動態解析
 - ✓ スパコンログデータを使った異常解析



- **Immersive Exploration : 没入的な探索**

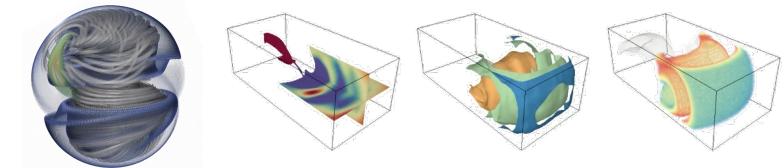
- VR技術を使った没入型データ解析空間
 - ✓ VR可視化基盤の開発
 - ✓ 高次元データ解析空間の構築



研究テーマ

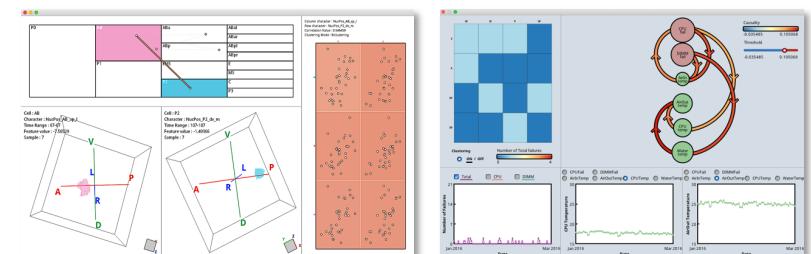
- **Intelligent Visualization : 知的な可視化**

- 重要な変化を見逃さない知的な可視化
 - ✓ 高精細半透明レンダリング
 - ✓ スパコン向けスマートIn-situ可視化



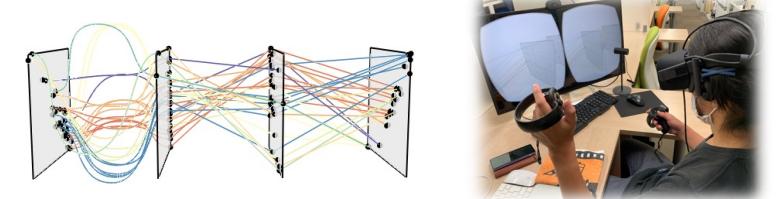
- **Interactive Analytics : 対話的な解析**

- 対話的に試行錯誤ができる視覚的解析
 - ✓ 生物細胞の発生動態解析
 - ✓ スパコンログデータを使った異常解析



- **Immersive Exploration : 没入的な探索**

- VR技術を使った没入型データ解析空間
 - ✓ VR可視化基盤の開発
 - ✓ 高次元データ解析空間の構築



内容

前半

- これまでの研究
 - 自己紹介
 - 研究内容

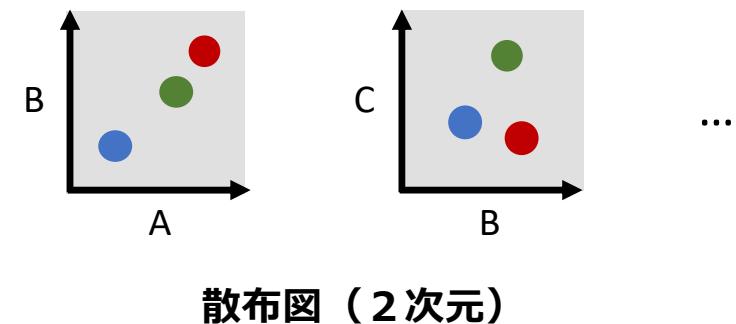
後半

- 現在の研究
 - 多次元時系列データの視覚的解析
 - 事例 1：スパコンログデータ解析
 - 事例 2：アンサンブルデータ解析

多次元時系列データ

- 対象となるデータの多くが多次元（時系列）データとして表現
 - 例）項目数: 3、属性数(次元数): 4

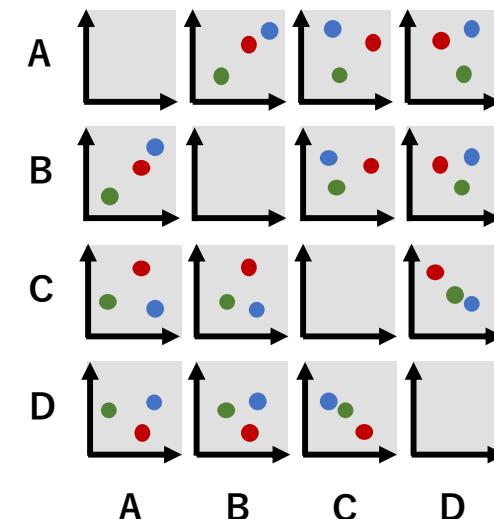
	A	B	C	D
Red	92	90	50	83
Green	74	81	82	39
Blue	23	56	61	70



多次元時系列データ

- 対象となるデータの多くが多次元（時系列）データとして表現
 - 例) 項目数: 3、属性数(次元数): 4

	A	B	C	D
Red	92	90	50	83
Green	74	81	82	39
Blue	23	56	61	70

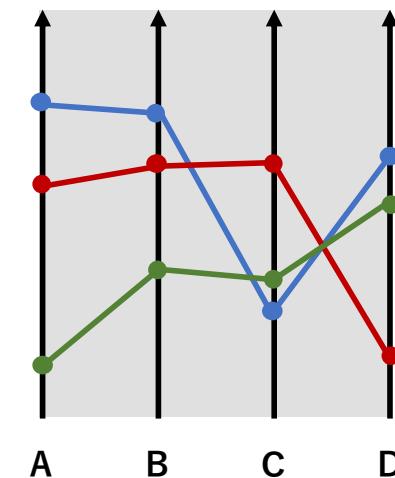


散布図行列（4次元）

多次元時系列データ

- 対象となるデータの多くが多次元（時系列）データとして表現
 - 例) 項目数: 3、属性数(次元数): 4

	A	B	C	D
Red	92	90	50	83
Green	74	81	82	39
Blue	23	56	61	70

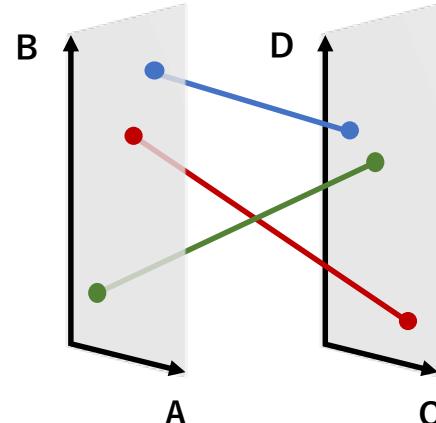


平行座標（4次元）

多次元時系列データ

- 対象となるデータの多くが多次元（時系列）データとして表現
 - 例）項目数: 3、属性数(次元数): 4

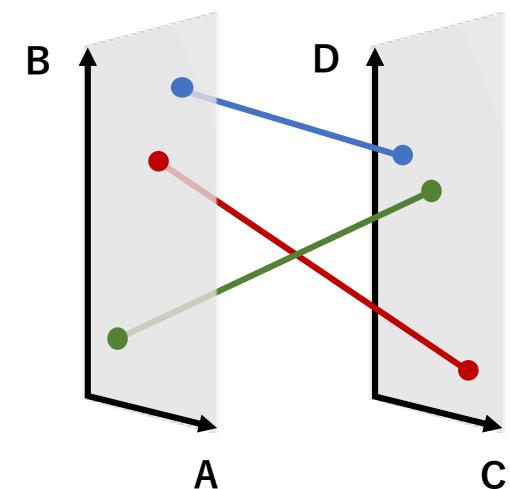
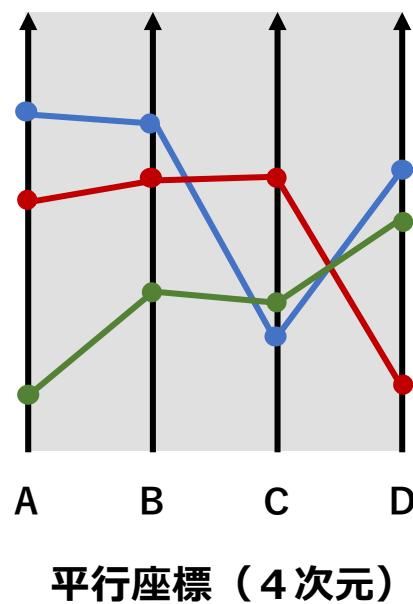
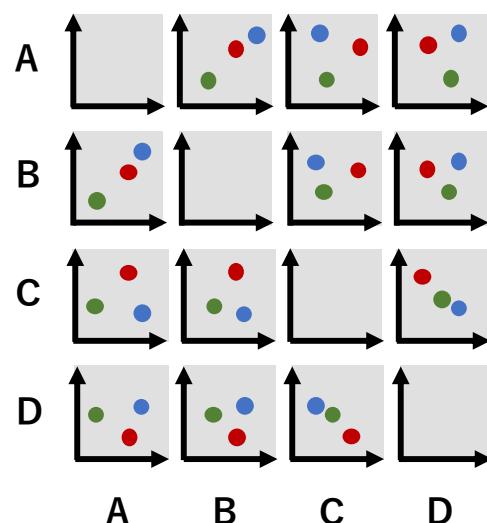
	A	B	C	D
Red	92	90	50	83
Green	74	81	82	39
Blue	23	56	61	70



平行平面（4次元）

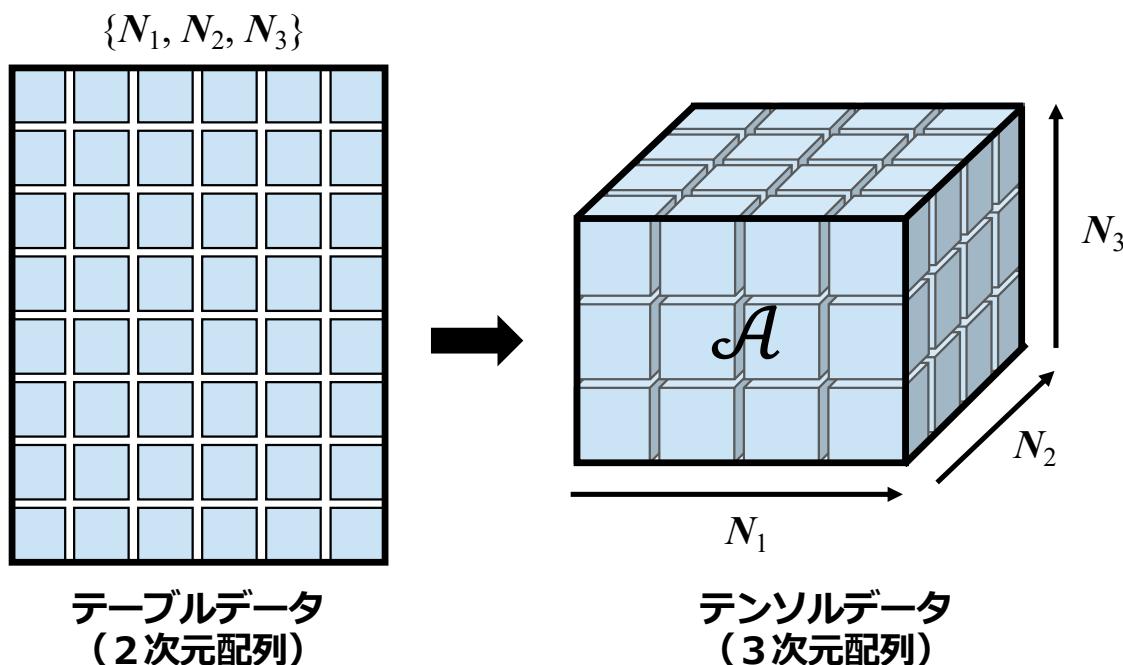
多次元時系列データ

- 対象となるデータの多くが多次元（時系列）データとして表現
 - 例) 項目数: 3、属性数(次元数): 4



テンソルデータ表現

- 多次元データをテンソルデータ（多次元配列データ）として表現し、データの低次元化や特徴抽出を行う。



$$\text{CP分解} \quad \mathcal{A} = \sum_{i=1}^R \lambda_i \mathbf{a}_i^1 \otimes \mathbf{a}_i^2 \otimes \cdots \otimes \mathbf{a}_i^m,$$

$$\begin{matrix} \mathcal{A} \\ = \\ \begin{matrix} \nearrow \\ \square \end{matrix} + \begin{matrix} \nearrow \\ \square \end{matrix} + \begin{matrix} \nearrow \\ \square \end{matrix} \end{matrix}$$

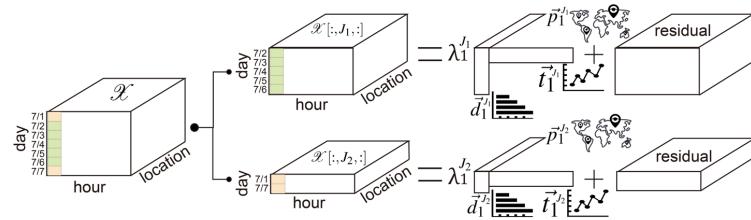
Tucker分解

$$\mathcal{A} = \sum_{j_1=1}^{N_1} \sum_{j_2=1}^{N_2} \cdots \sum_{j_m=1}^{N_d} s_{j_1, j_2, \dots, j_m} \mathbf{u}_{j_1}^1 \otimes \mathbf{u}_{j_2}^2 \otimes \cdots \otimes \mathbf{u}_{j_m}^m,$$

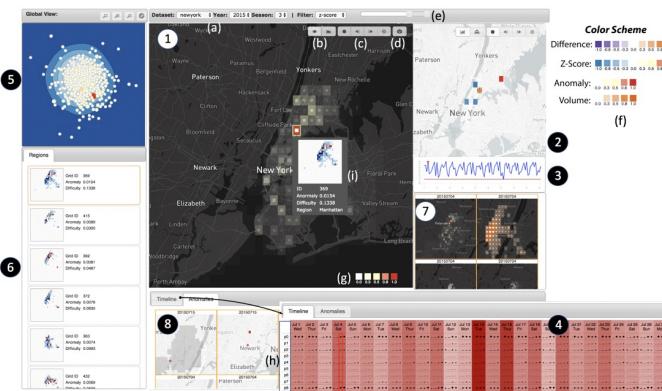
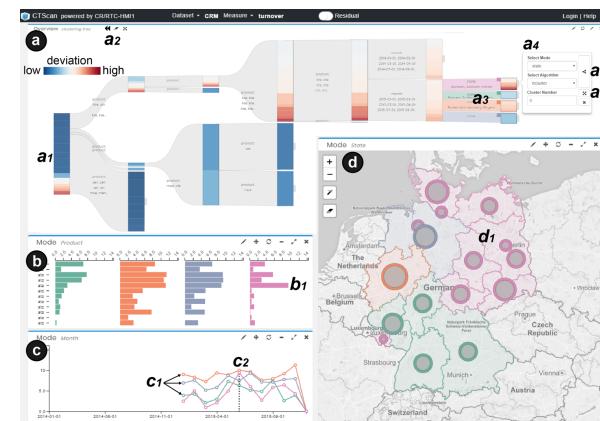
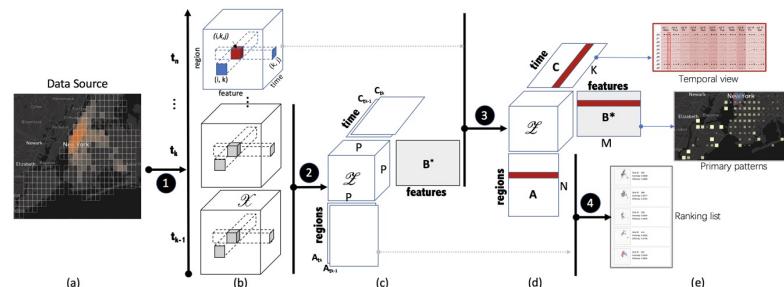
$$\begin{matrix} \mathcal{A} \\ = \\ \begin{matrix} \nearrow \\ \square \end{matrix} \quad \begin{matrix} \nearrow \\ \square \end{matrix} \quad \begin{matrix} \nearrow \\ \square \end{matrix} \quad \begin{matrix} \nearrow \\ \square \end{matrix} \end{matrix}$$

テンソルデータ表現

- テンソルデータ向け視覚的解析
 - TPFlow [D. Liu et al., , TVCG, 2019]



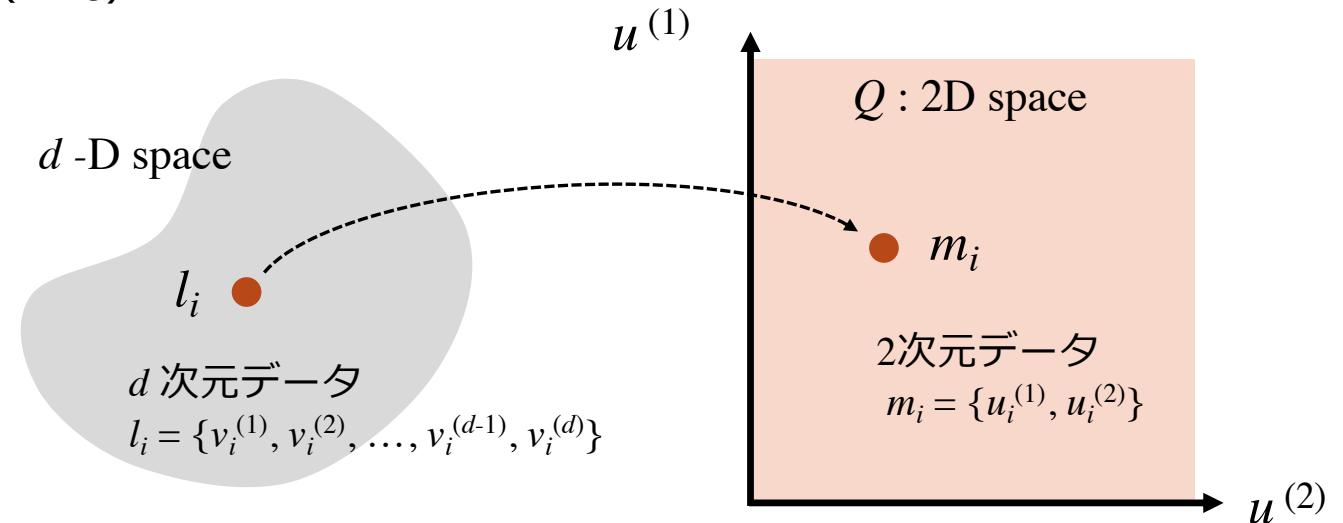
- Voila [N.Cao et al., TVCG 2017]



次元削減を利用したテンソルデータ解析

- 次元削減

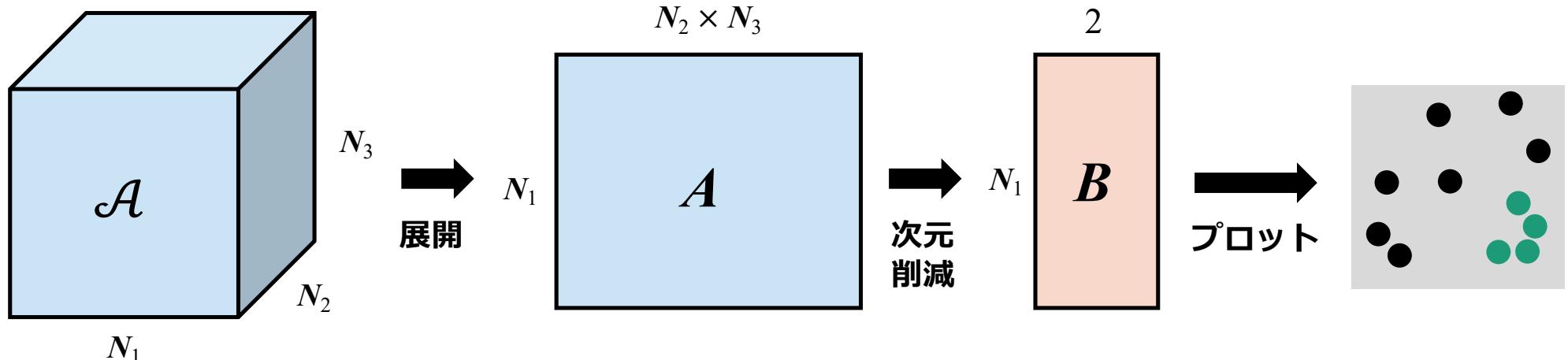
- 多次元データ（高次元空間上のデータ）の持つ特徴をなるべく保持したまま低次元空間に埋め込む操作
 - 主成分分析 (PCA)
 - 多次元尺度構成法 (MDS)
 - t-SNE
 - UMAP
 - ...



次元削減を利用したテンソルデータ解析

- ・**多重次元削減**

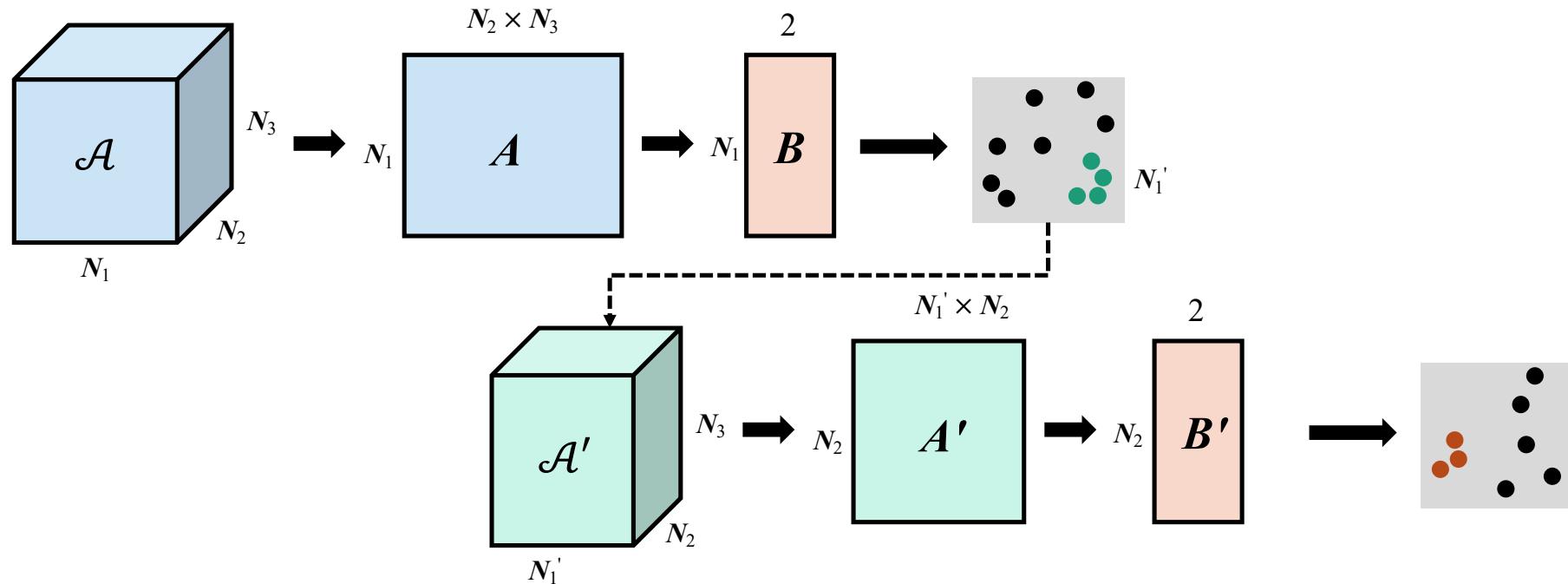
- ✓ テンソルデータの基準軸を決め次元削減を適用する
- ✓ 複数の基準軸を選択し次元削減結果をプロットする



次元削減を利用したテンソルデータ解析

- 多段階次元削減

✓ 次元削減を段階的に適用する



内容

前半

- これまでの研究
 - 自己紹介
 - 研究内容

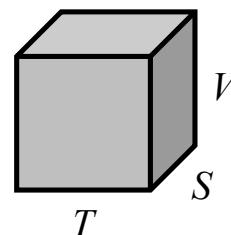
後半

- 現在の研究
 - 多次元時系列データの視覚的解析
 - 事例 1：スパコンログデータ解析
 - 事例 2：アンサンブルデータ解析

スパコンログデータ解析

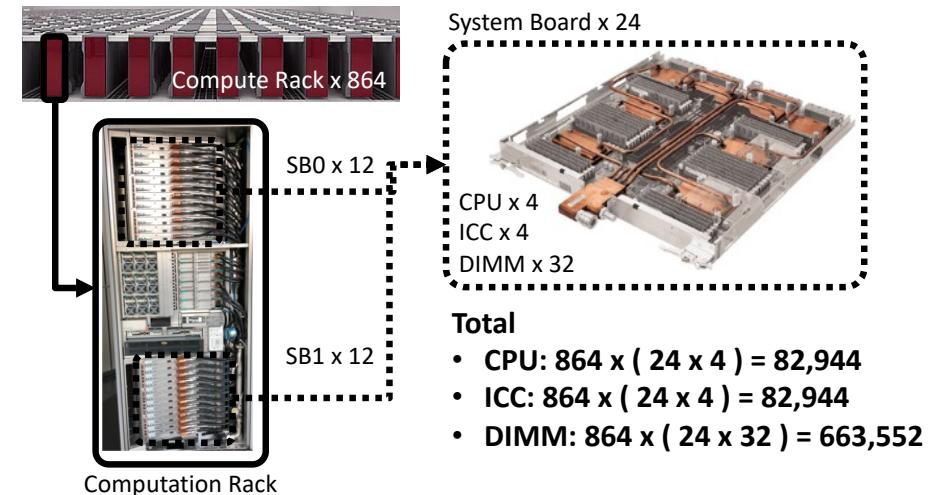
- 長期安定運用のためシステムの挙動理解が重要かつ課題

- 京コンピュータのデータを利用する。
- センサリング技術の向上により様々なログデータ（環境データ）が計測されている。
- 環境データは、**時間・空間・測定値**の属性を持つ。
- 測定値には、複数種類のデータ（CPUの温度など）が含まれる。



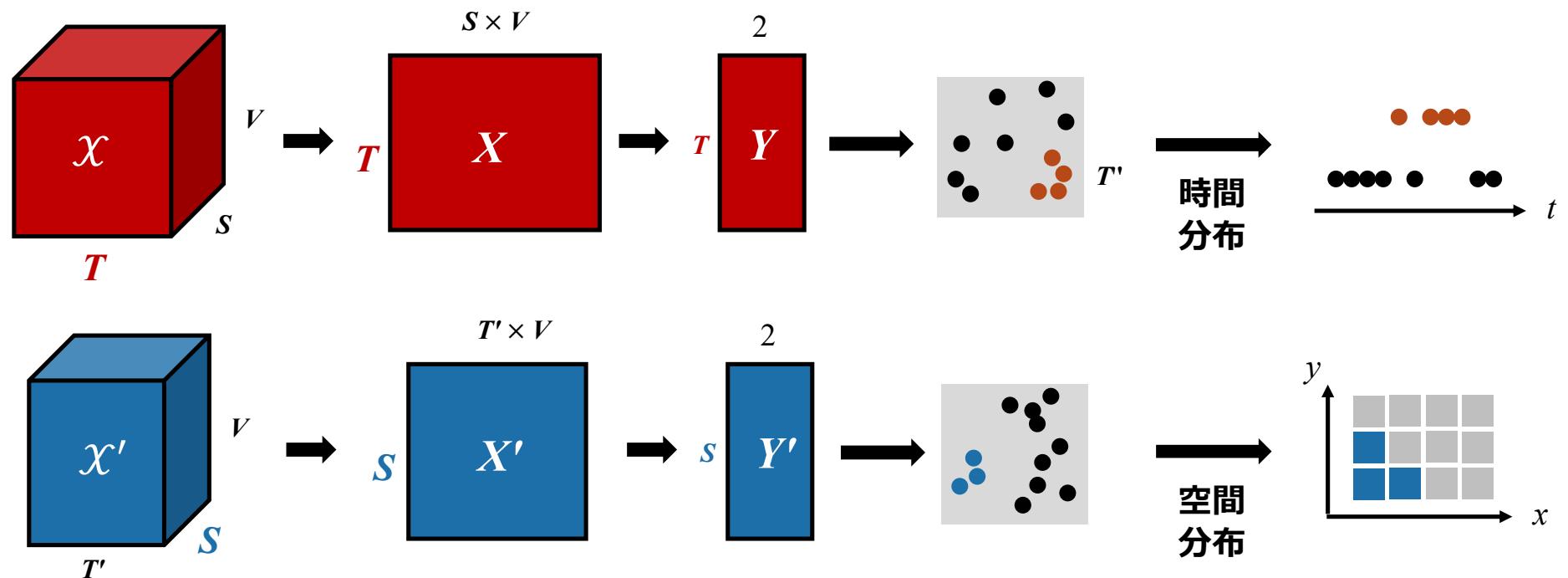
3次テンソルデータ

- 時間 (T)
- 空間 (S)
- 測定値 ($V: V_1, V_2, \dots$)



スパコンログデータ解析

- 多段階次元削減（2段階：時間 $T \rightarrow$ 位置 S ）



スパコンログデータ解析

・実験

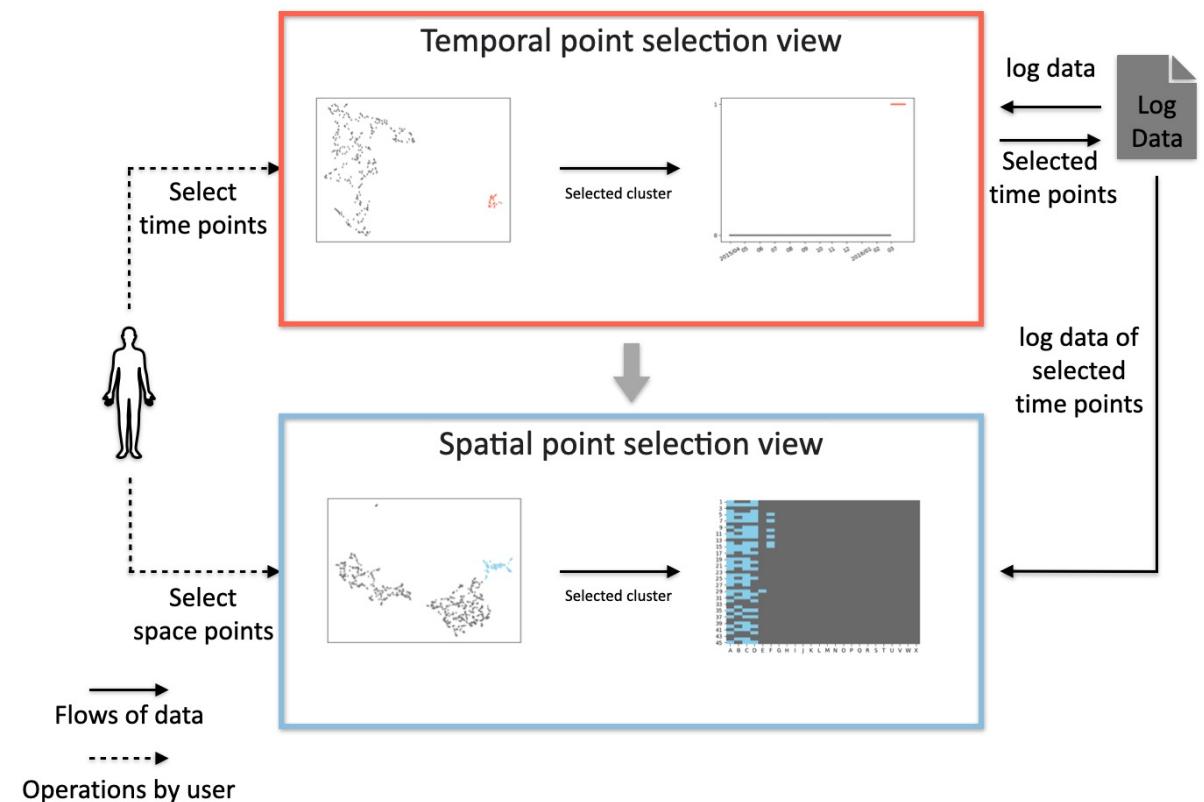
- ・スーパーコンピュータ「京」の環境ログデータ
- ・**時間（期間）：365日**（2014年4月～2015年3月）
 - ・5分ごとに測定されたデータを1日ごとに平均化
- ・**空間（計算ラック）：864ラック**
- ・**測定値（温度）：4種類**
 - ・空冷装置の吸気温度（AirIn temp）
 - ・空冷装置の排気温度（AirOut temp）
 - ・ラック中のCPU温度平均（CPU temp）
 - ・水冷装置の冷却水温度（Water temp）



<https://www.r-ccs.riken.jp/jp/k/facility.html>

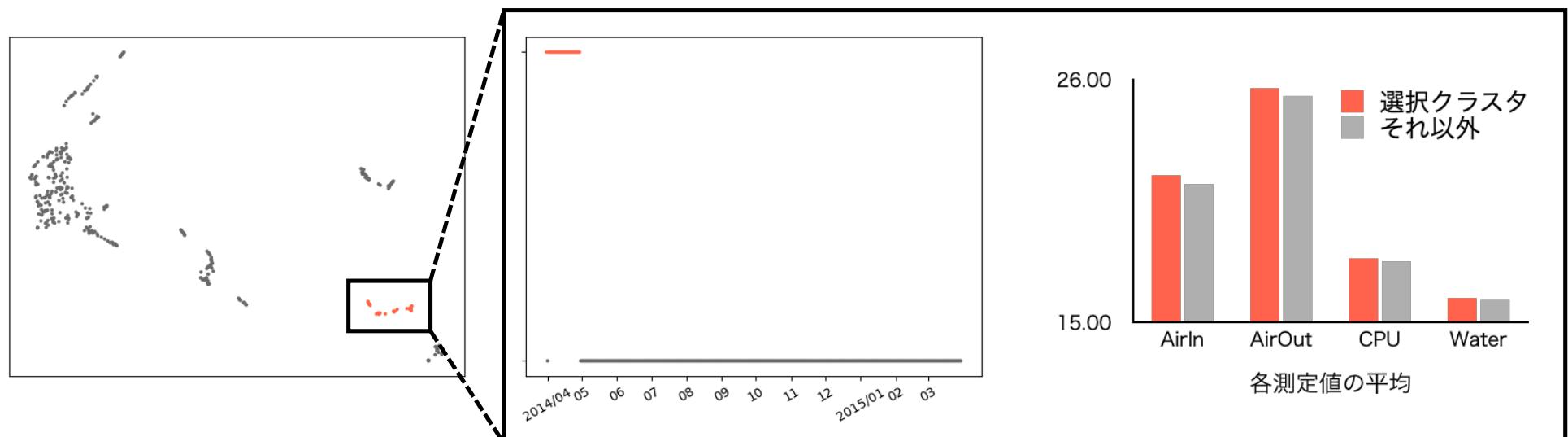
スパコンログデータ解析

- プロトタイプシステム
 - データ処理
 - Pandas
 - Scikit-learn
 - 可視化
 - Matplotlib
 - Bokeh
 - 次元削減
 - PCA
 - UMAP



スパコンログデータ解析

- ・時間を基準にした次元削減結果



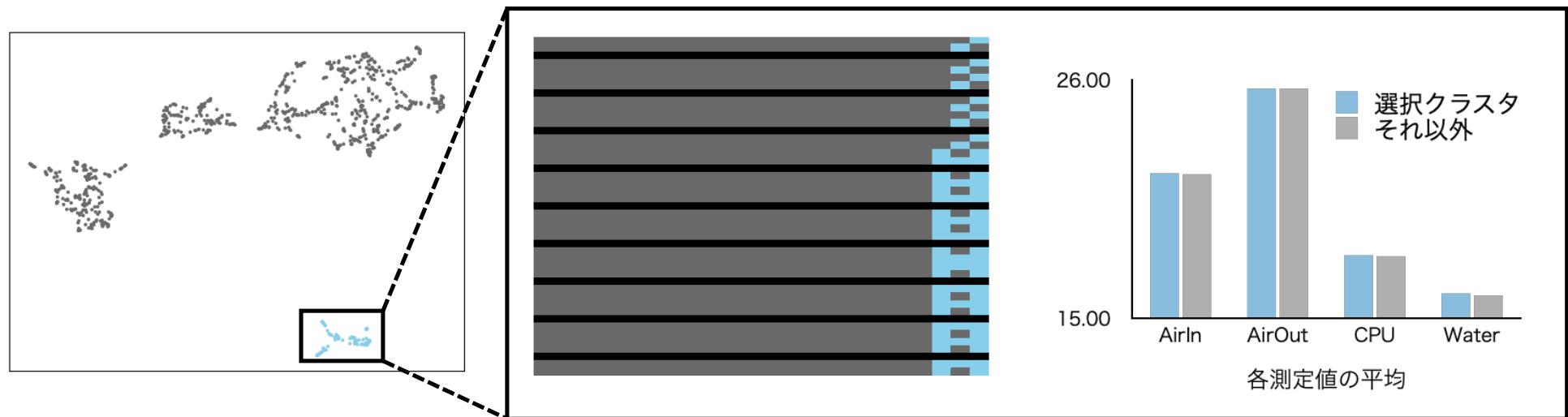
選択した期間：2014年4月2日～2014年4月30日

測定値の比較：吸気温度（AirIn）と排気温度（AirOut）が高い

→ ゴードンベルチャレンジの影響

スパコンログデータ解析

- 空間を基準にした次元削減結果



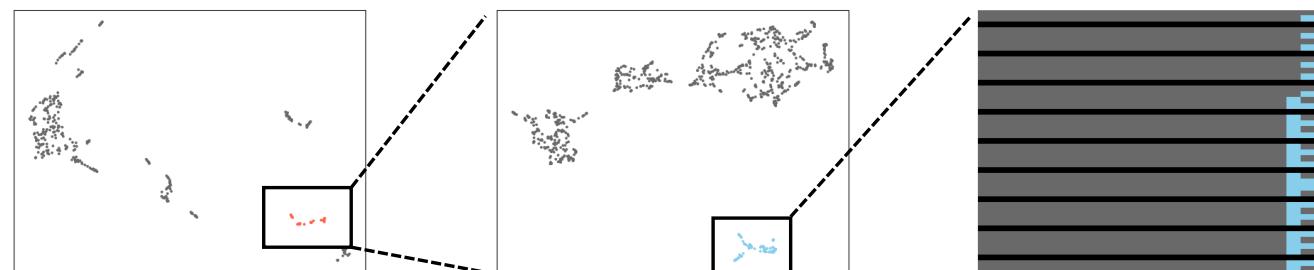
選択したラック位置：右端のラック群が選択された
測定値の比較：いずれの測定値も僅かに高い

→ システム全体の負荷が増加している？
計算ラックの論理空間の接続と関係している？

スパコンログデータ解析

・考察

- ・多次元時系列データとして表現されるスパコンログデータを、**時間・空間・測定値を基準軸とするテンソルデータに変換**
- ・テンソルデータに対して**多段階次元削減**を適用することで、特徴的な時空間パターンを効率的に探索
- ・システムの状態理解につながる可能性がある特徴パターンを見つけることができたが、より深い理解を得るために、**より多くの測定値を考慮した解析が必要**



内容

前半

- これまでの研究
 - 自己紹介
 - 研究内容

後半

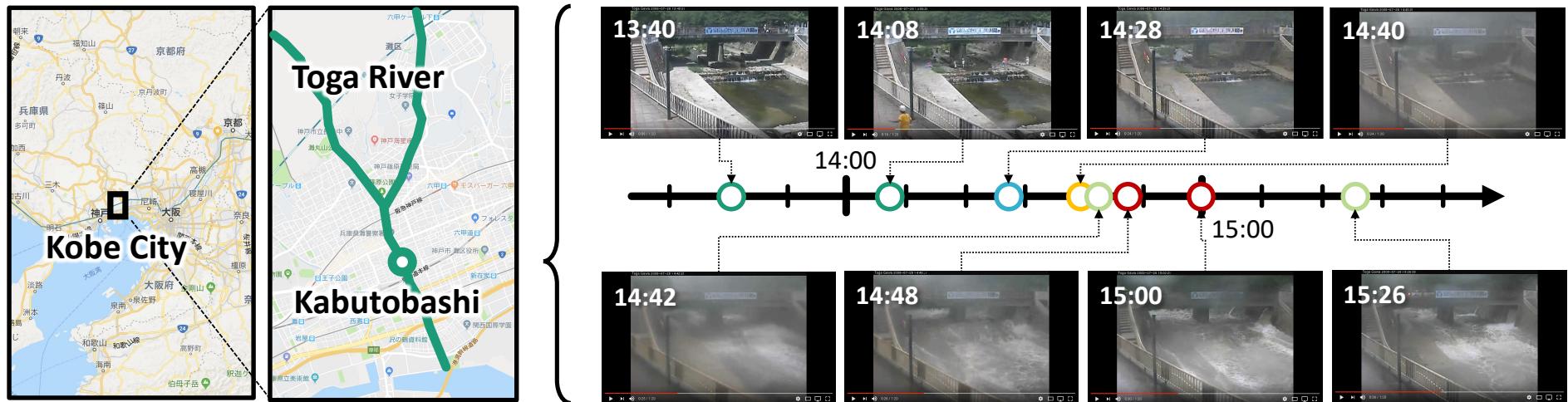
- 現在の研究
 - 多次元時系列データの視覚的解析
 - 事例 1：スパコンログデータ解析
 - 事例 2：アンサンブルデータ解析（現在研究中）

アンサンブルデータ解析

・極端気象現象の予測

- ・集中豪雨や台風などの極端気象現象による激甚災害が頻発
- ・極端気象現象からの防災・減災のためのスパコンを使った予測

都賀川水難事故（2008年7月28日）



<https://www.youtube.com/watch?v=gwrvHNvFe8>

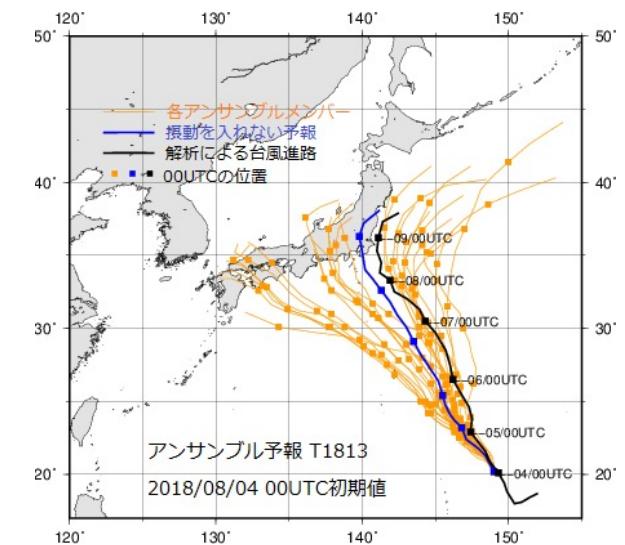
アンサンブルデータ解析

・気象アンサンブルシミュレーション

- ・計算条件が異なる複数の気象シミュレーションを実行し、将来の大気の状態（パラレルワールド）を再現
- ・気象モデルの不確実性を考慮した確率的な予測

・アンサンブルデータの可視化と解析

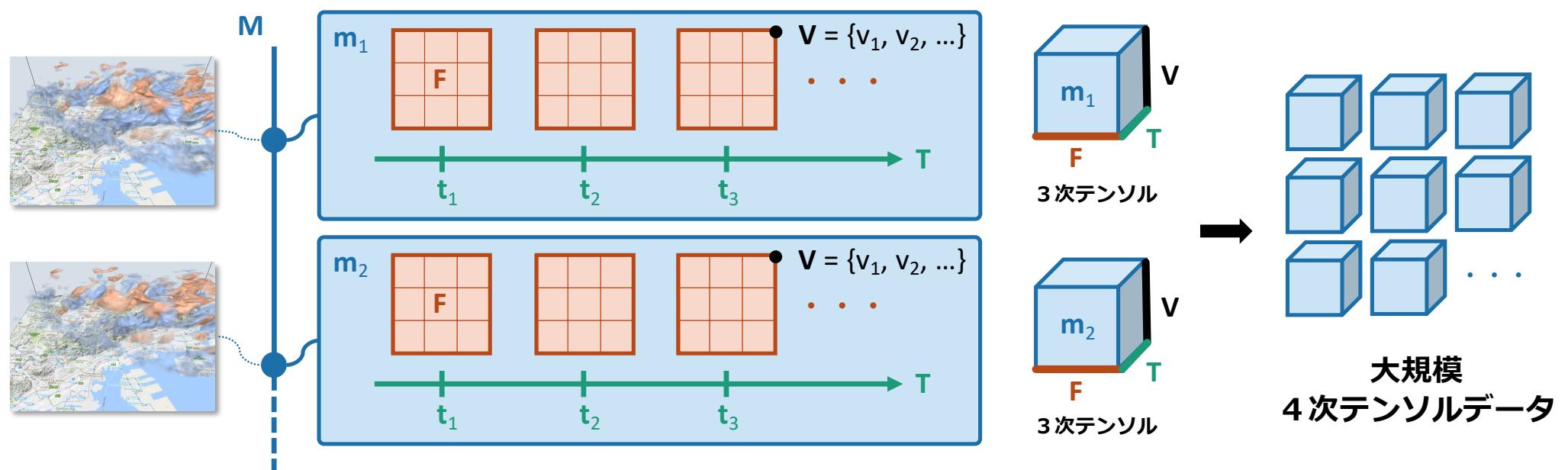
- ・計算機性能の向上に伴う大規模化
 - ・複数のアンサンブルメンバー
 - ・複数の変数
 - ・時系列
- ・詳細な時空間特徴解析
 - ・4次テンソルデータ向け解析技術が必要



台風進路のアンサンブル予報の例 (気象庁)
<https://www.jma.go.jp/jma/kishou/know/whitep/1-3-8.html>

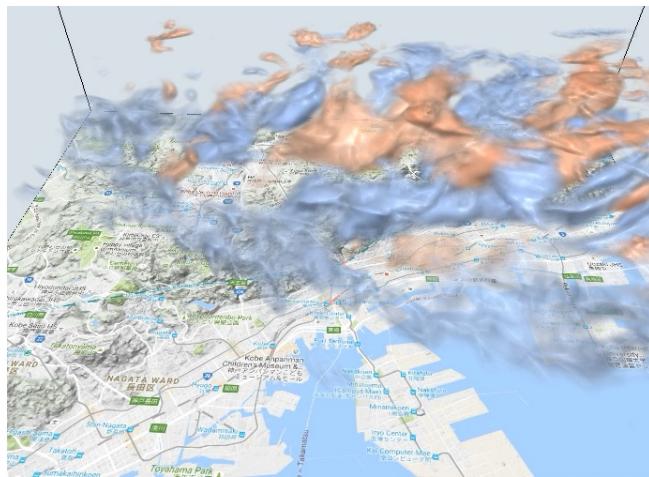
アンサンブルデータ解析

- ・アンサンブルデータ
 - ・世界 (**M**)、空間 (**F**)、時間 (**T**)、変数 (**v**)

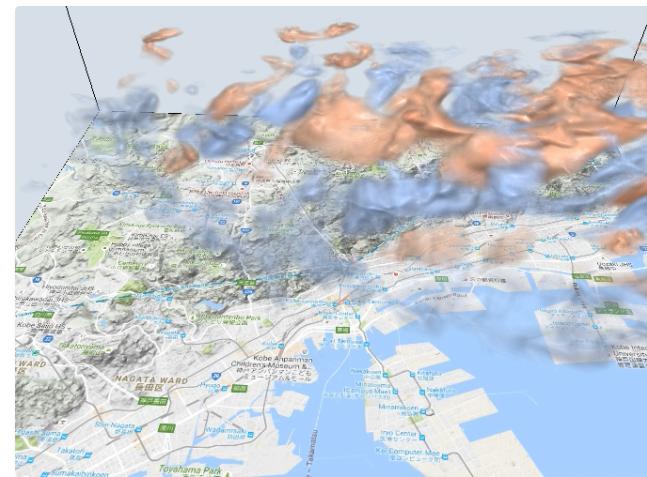


アンサンブルデータ解析

- 変数を固定した複数メンバ間の時空間特徴解析
 - 着目変数の空間分布を可視化
 - 時間変化をアニメーション表示



m_1

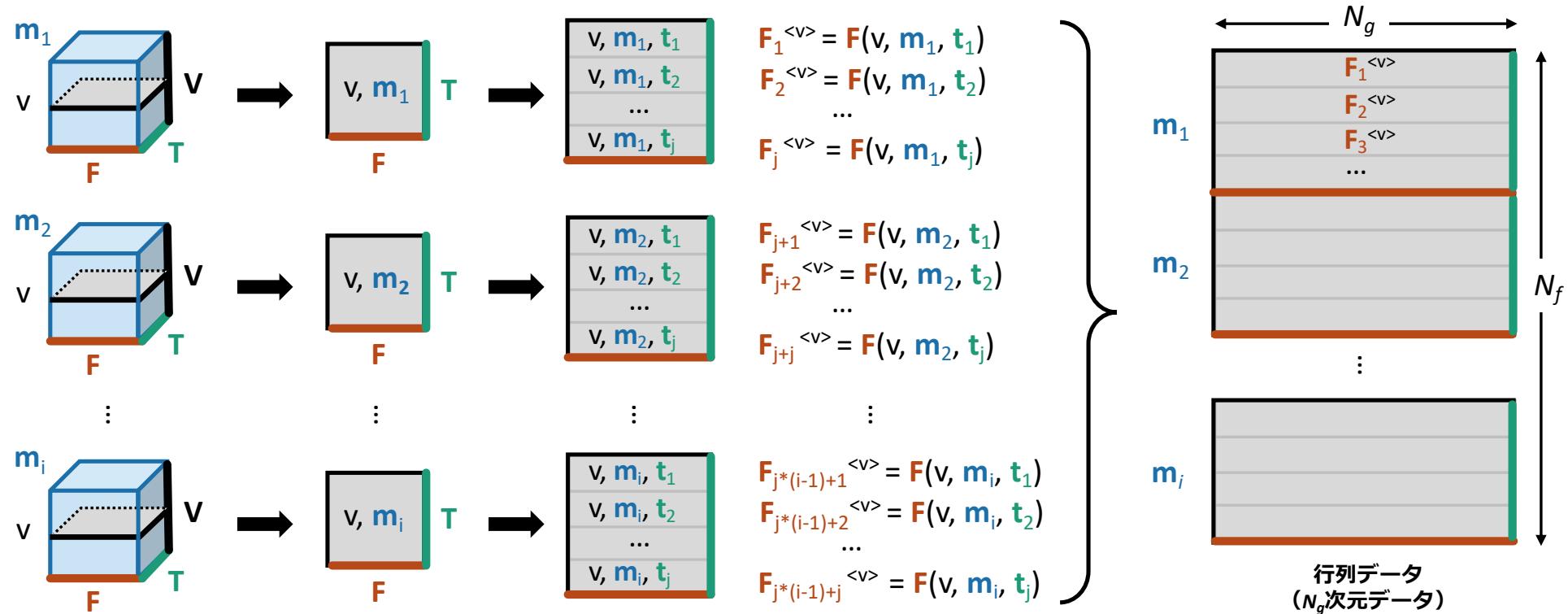


m_2

...

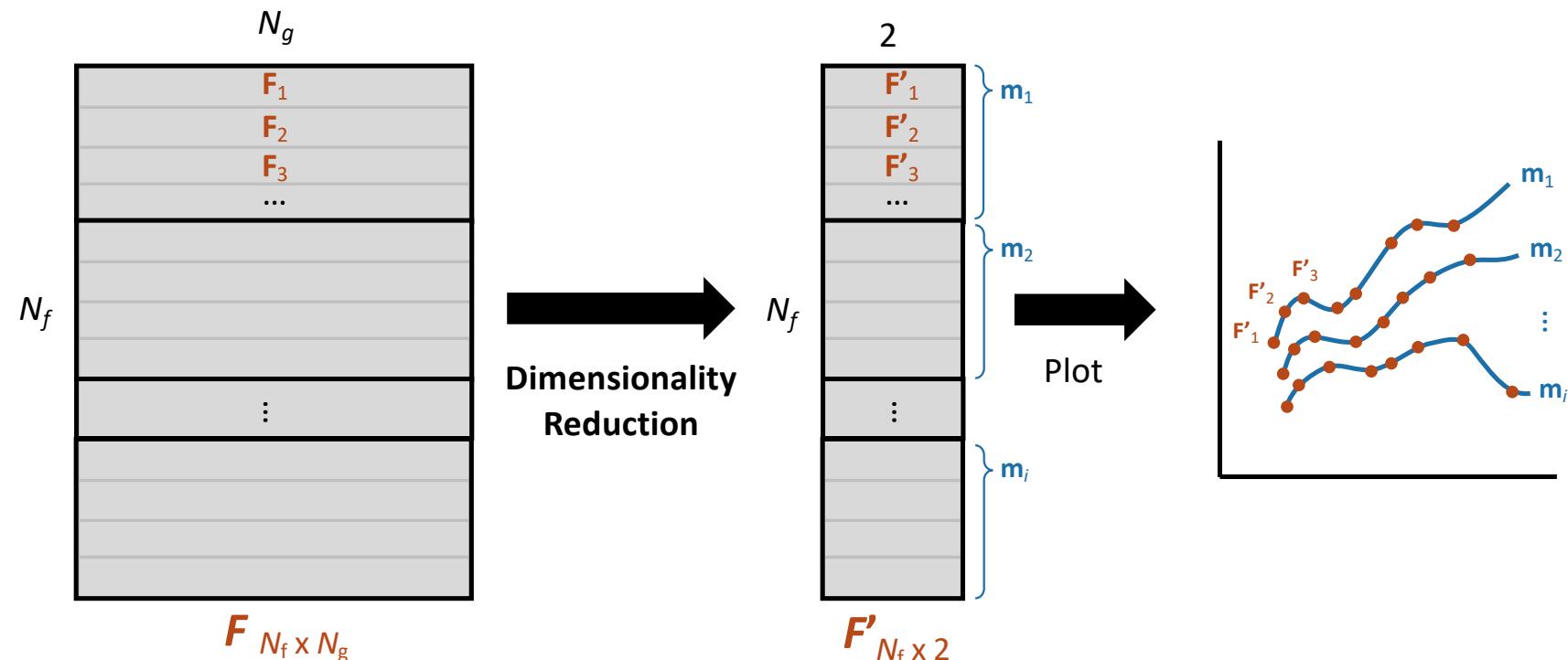
アンサンブルデータ解析

- 変数を固定した複数メンバ間の時空間特徴解析



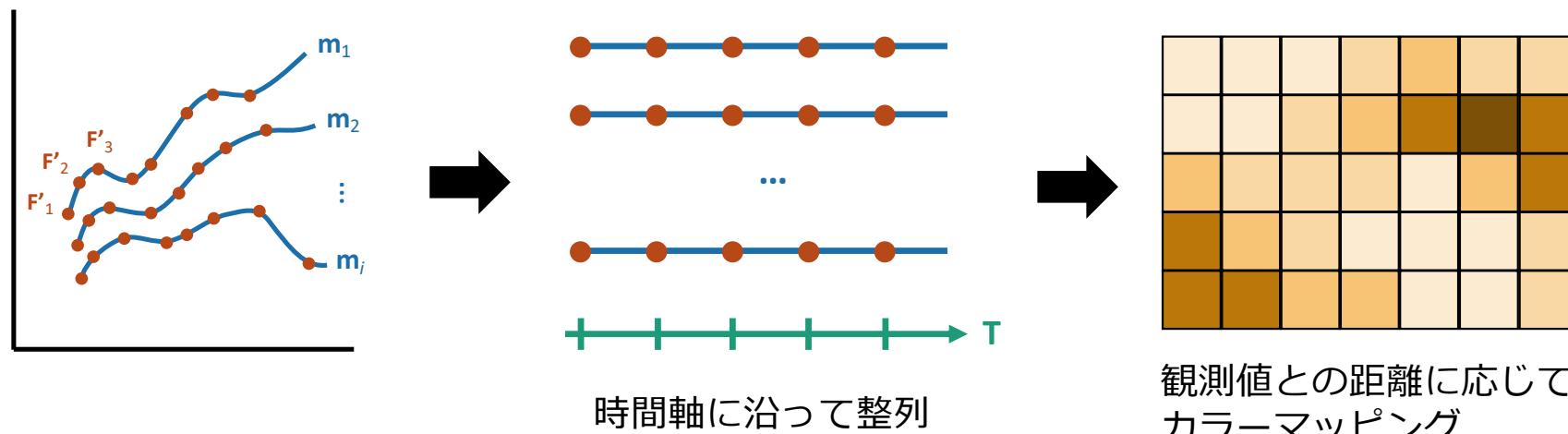
アンサンブルデータ解析

- 全メンバの時間発展の様子を2次元平面にプロット



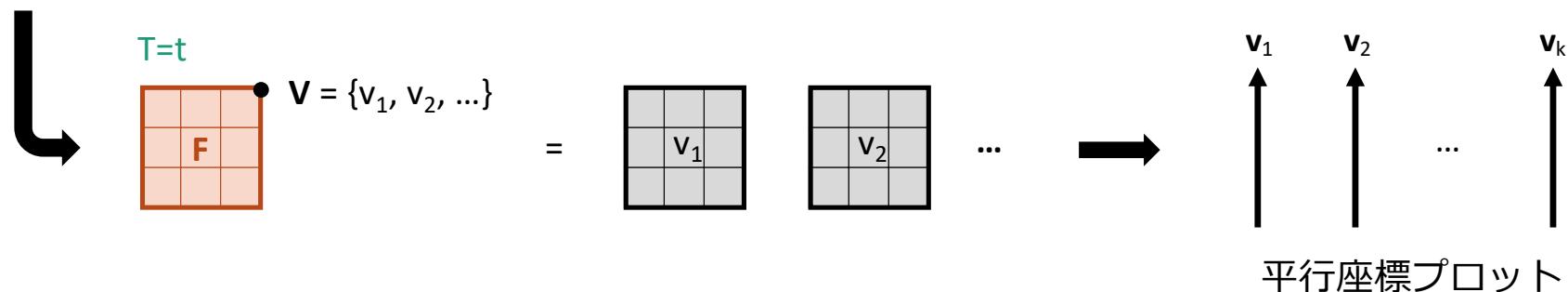
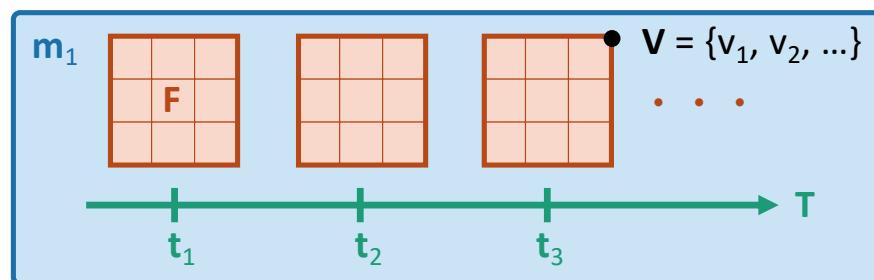
アンサンブルデータ解析

- 全メンバの時間発展の様子を2次元平面にプロット
 - 時間軸を基準にした時間変化の比較



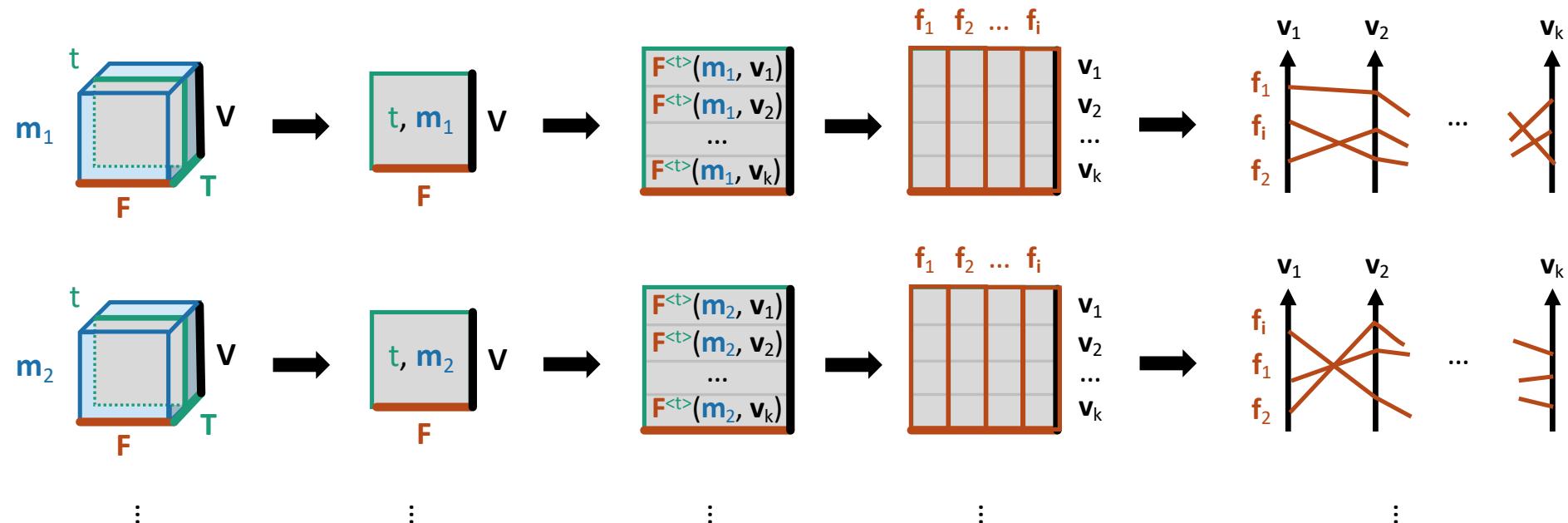
アンサンブルデータ解析

- 時刻を固定した変数間の関係解析



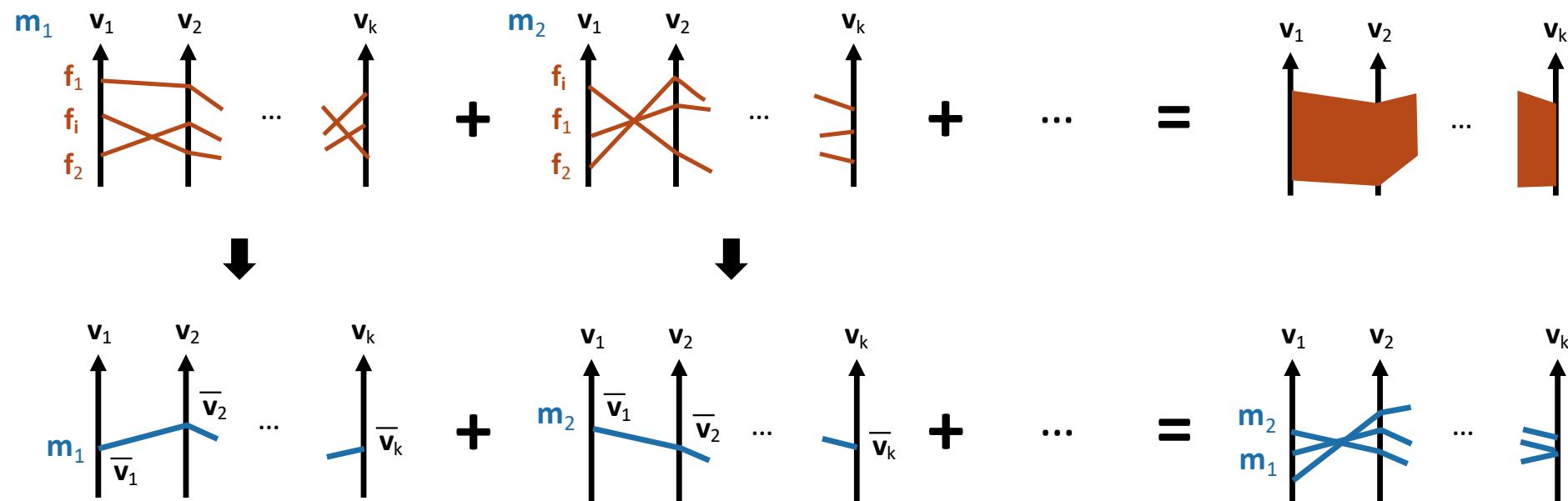
アンサンブルデータ解析

- 平行座標プロットを使った複数変数間の関係可視化
 - 時間を固定しメンバの全変数をプロット



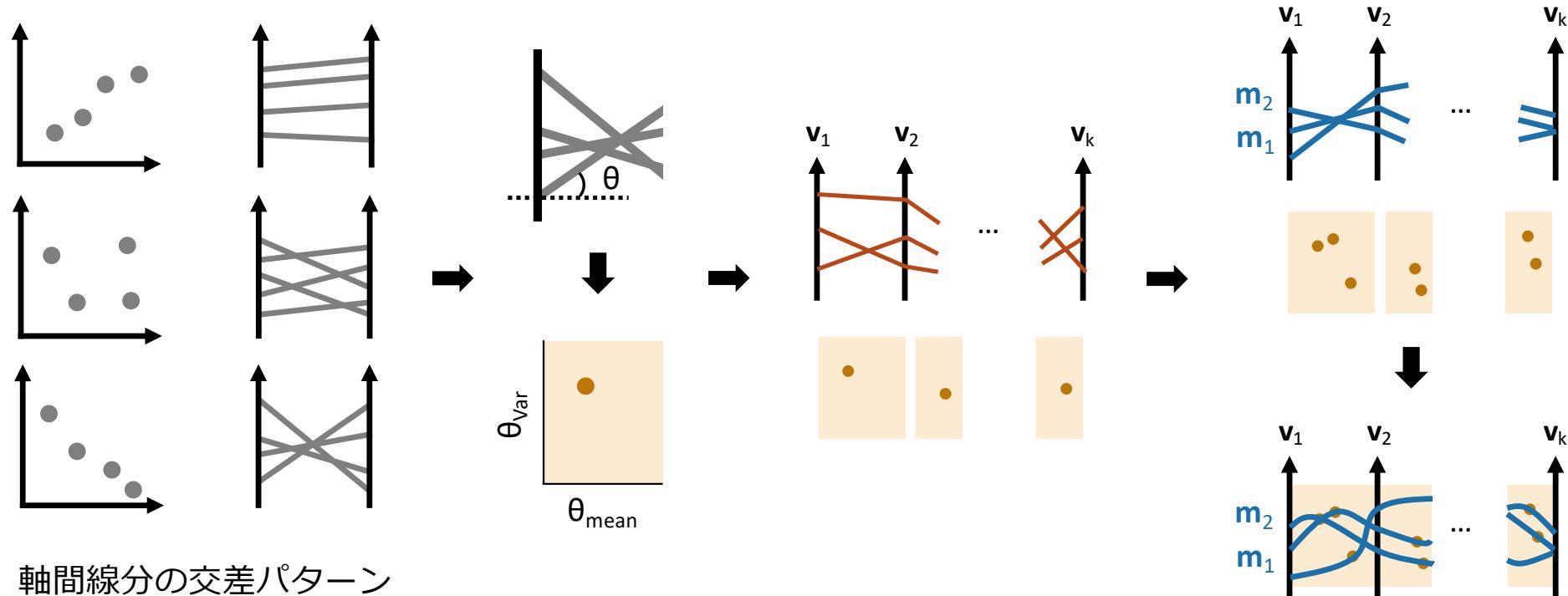
アンサンブルデータ解析

- 複数メンバの比較
 - 各変数の平均値による代表線分プロット



アンサンブルデータ解析

- 角度分布情報に基づく平行座標プロットの結束化



アンサンブルデータ解析

・ 実験

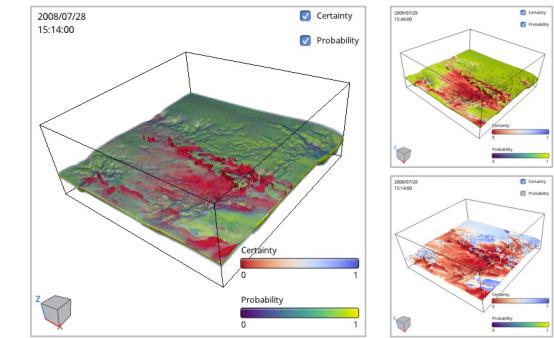
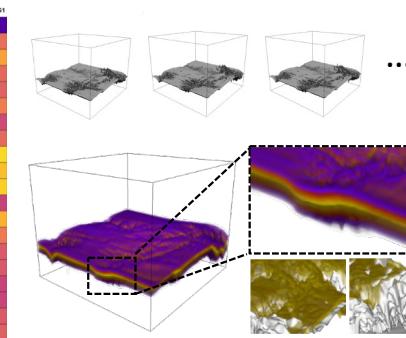
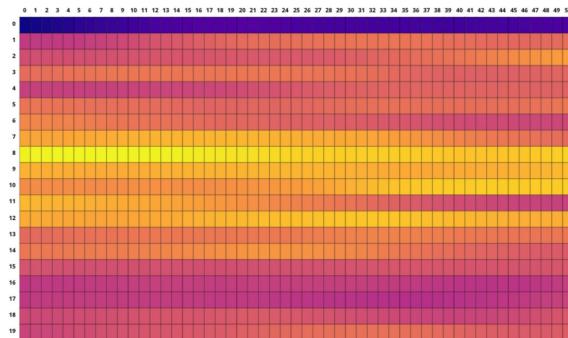
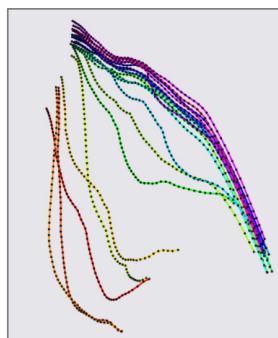
- ・ 神戸市灘区都賀川周辺の気象アンサンブルシミュレーション
 - ・ 格子解像度 : **301 x 301 x 50**
 - ・ 水平方向: 1[km]間隔
 - ・ 鉛直方向: -20[m] ~ 22,244[m]を非等間隔で分割
 - ・ 時間ステップ数 : **60**
 - ・ 2008年7月28日 03時00分～03時59分までの1分間隔
 - ・ アンサンブル数: **20**
 - ・ 変数 : **11**
 - ・ 東西風, 南北風, 鉛直風, 気温, 気圧, 水蒸気混合比, 雲水混合比, 雨混合比, 雲氷混合比, 雪混合比, あられ混合比



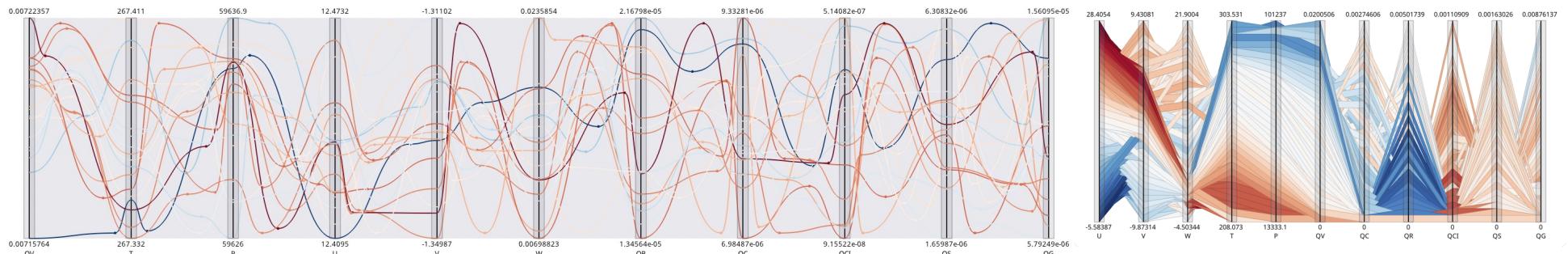
データ提供：理研R-CCSデータ同化研究チーム前島康光氏

アンサンブルデータ解析

- 変数を固定した複数メンバ間の時空間特徴解析



- 時刻を固定した変数間の関係解析



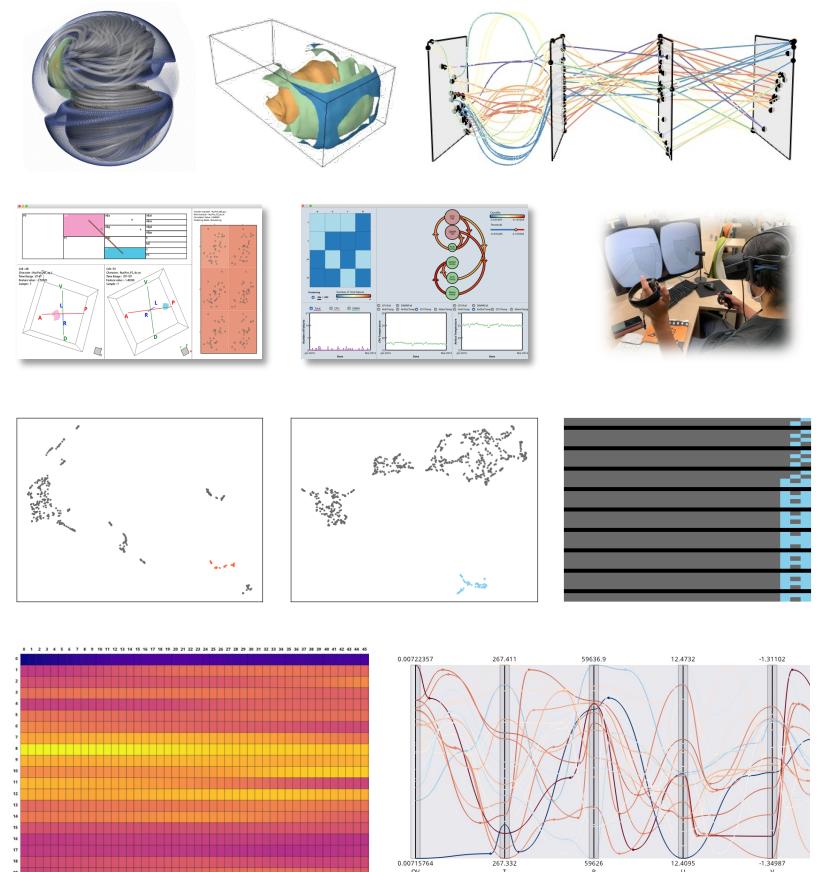
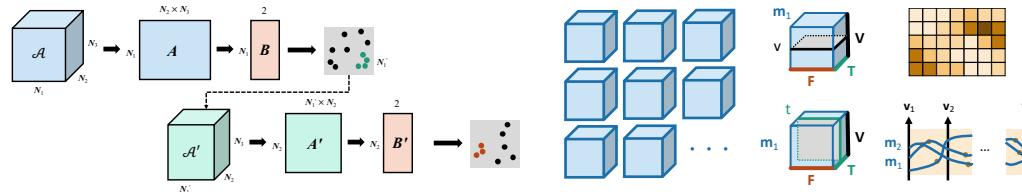
まとめ

・データ可視化と視覚的データ解析

- 知的な可視化 (Intelligent Visualization)
 - 対話的な解析 (Interactive Analytics)
 - 没入的な探索 (Immersive Exploration)

・多次元時系列データの視覚的解析

- テンソルデータ表現とその解析
 - 多段階次元削減
 - 事例 1 : スパコンログデータ解析
 - 事例 2 : アンサンブルデータ解析



最後に

- ・数値データ (x,y,z で表現できるデータ) だけでなく**多次元データ** (表形式データ) として表現できるデータが解析 (可視化) 対象
- ・**まず解析そして可視化**：対象データを解析し視覚的に表現する。そして、重要な部分を確認後、対話的なデータ探索（**試行錯誤**）を行う。必要に応じて可視化を行い、直感的な現象理解につなげる。

