

データサイエンスを支える 可視化技術

Jan. 30, 2016

京都大学
学術情報メディアセンタービジュアルリゼーション分野
小山田耕二

内容

- データサイエンスを支える可視化
 - 自己紹介
 - 可視化について
 - 科学的可視化について
 - 因果関係の探索

Wikipedia: Data Science is an interdisciplinary field about processes and systems to extract knowledge or insights from data in various forms, either structured or unstructured,[1][2] which is a continuation of some of the data analysis fields such as statistics, data mining, and predictive analytics, similar to Knowledge Discovery in Databases (KDD).

自己紹介

- 1960 神戸生まれ
- 1983 京都大学工学部卒業
- 1985 京都大学工学研究科修士課程修了
- 1985 日本アイビーエム(株)関西営業本部
- 1988 日本アイビーエム(株)東京基礎研究所
- 1998 岩手県立大学ソフトウェア情報学部助教授
- 2001 京都大学学術情報メディアセンター助教授
- 2003 京都大学高等教育研究開発推進センター教授
- 2014 自然科学研究機構客員教授
- 2015 京都大学学術情報メディアセンター教授

JST A-STEP 「融合可視化技術に関する研究開発」 (平成24～26年度)

高レベル
技術

粒子ベースボリュームレンダリングの製品化で大規模データの可視化を可能に

研究成果最適展開支援プログラム(A-STEP)

シーズ育成タイプ

課題名 融合可視化技術に関する研究開発

開発期間 平成24年10月～平成27年3月

キーワード▶ ボリュームレンダリング、大規模データ、可視化

◆プロジェクトリーダー所属機関
サイバネットシステム株式会社

◆研究者
小山田 耕二(京都大学)

粒子ベースボリュームレンダリング、および、
確率的ボリュームレンダリングをAVS/
Expressに実装し、サーフェイスレンダ

リングによる従来の可視化手法と一緒に
利用できるようにした。操作方法は、既
存製品に準拠しているため、既存ユーザ
は操作方法を学習することなく粒子ボリ
ュームレンダリングが利用可能となった。
これにより数十億点を超えるような大規
模なデータであってもデスクトップPC
で対話的な可視化が可能になった。

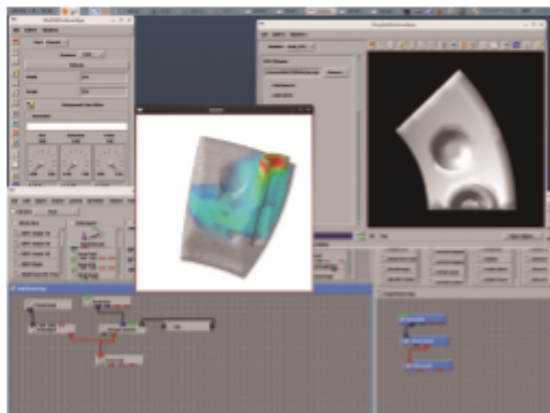
期待されるインパクト
(効果、意義、市場規模、売り上げ予測)

京をはじめとした大規模計算機による
解析結果の可視化ソリューション
として有効。HPC市場は400億程と言
われているが、観測、計測データや
オープンデータなどの市場への展開
も期待できる。

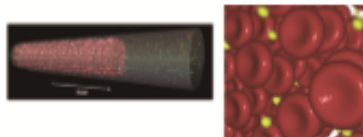
開発者の声

本研究の評価を、いろいろな分野の
研究者の方々をお願いすることで、
本研究を知ってもらえたことができたと
同時に、その人脈は大きな資産と
なった。製品化も順調に進んでおり、
今後のビジネス展開に期待している。
本制度なしに、自社で高度な技術を
開発することは難しかったと思う。

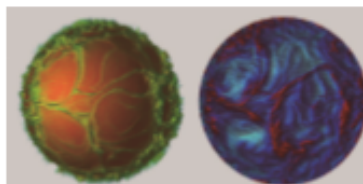
※この成果は、サイバネットシステム株式会
社からプレスリリースとして発表されてい
ます。



AVS/Express 概観

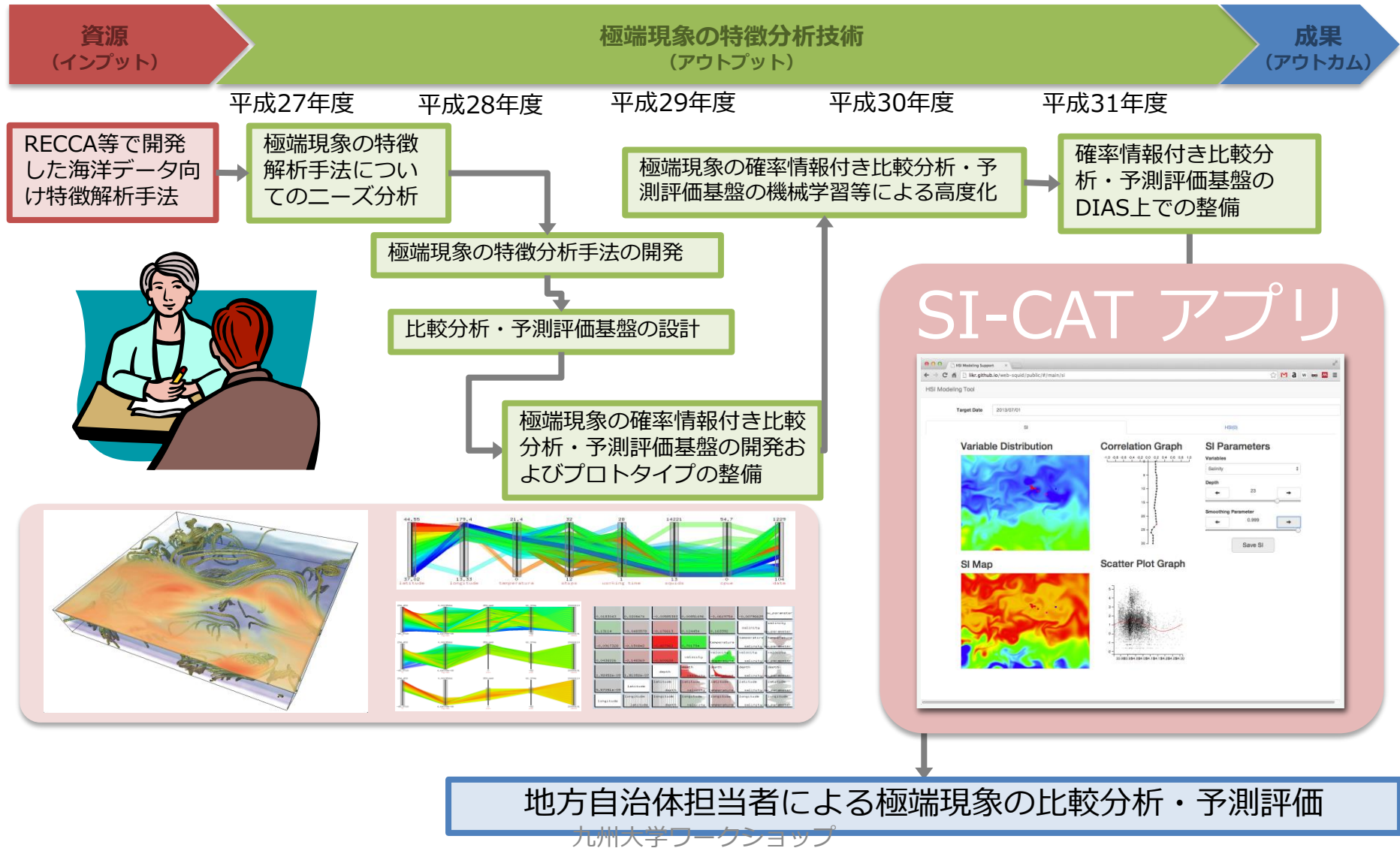


大規模血流解析データの可視化
(データ提供: 理化学研究所 杉山氏)

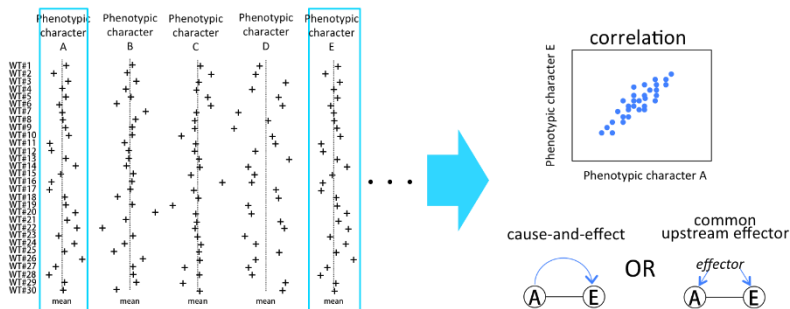


熱せられた球体を通る流れ解析の可視化
(データ提供: 京都大学 武蔵先生)

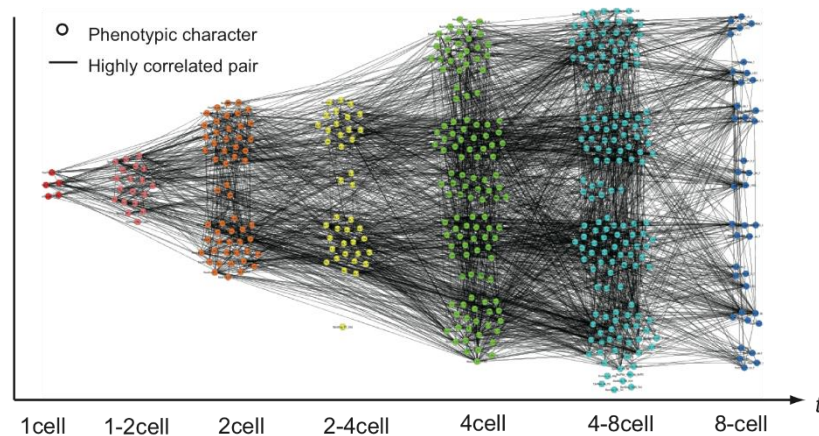
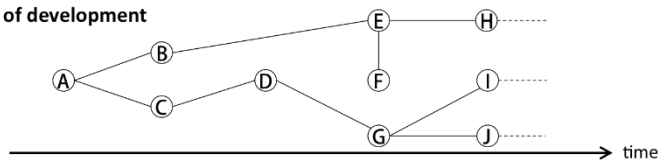
文科省 気候変動適応技術社会実装プログラム (平成27～31年度)



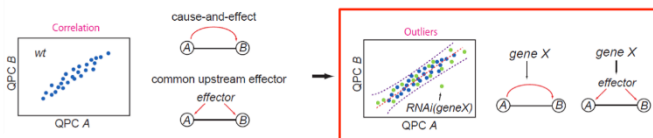
JST/CREST研究領域「ビッグデータ応用」 データ駆動型解析による多細胞生物の発生メカニズムの解明 (平成27~32年度)



Model of development



3372 significant relationships among 430 phenotypic characters from one- to eight-cell stage

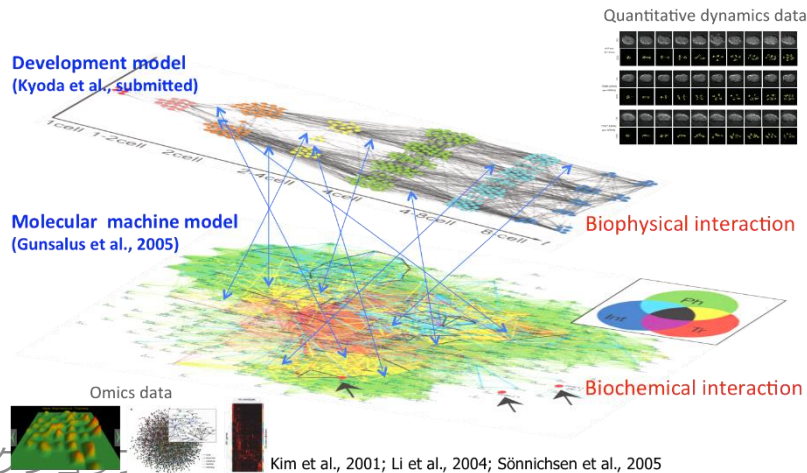


17,129 gene functions were inferred
4,625 functions are new (27% based on par-2/3 results)

| Gene ID | Gene Name | Gene Description |
|---------|-----------|------------------|
| Wt1 | Wt1 | Wt1 |
| Wt2 | Wt2 | Wt2 |
| Wt3 | Wt3 | Wt3 |
| Wt4 | Wt4 | Wt4 |
| Wt5 | Wt5 | Wt5 |
| Wt6 | Wt6 | Wt6 |
| Wt7 | Wt7 | Wt7 |
| Wt8 | Wt8 | Wt8 |
| Wt9 | Wt9 | Wt9 |
| Wt10 | Wt10 | Wt10 |
| Wt11 | Wt11 | Wt11 |
| Wt12 | Wt12 | Wt12 |
| Wt13 | Wt13 | Wt13 |
| Wt14 | Wt14 | Wt14 |
| Wt15 | Wt15 | Wt15 |
| Wt16 | Wt16 | Wt16 |
| Wt17 | Wt17 | Wt17 |
| Wt18 | Wt18 | Wt18 |
| Wt19 | Wt19 | Wt19 |
| Wt20 | Wt20 | Wt20 |
| Wt21 | Wt21 | Wt21 |
| Wt22 | Wt22 | Wt22 |
| Wt23 | Wt23 | Wt23 |
| Wt24 | Wt24 | Wt24 |
| Wt25 | Wt25 | Wt25 |
| Wt26 | Wt26 | Wt26 |
| Wt27 | Wt27 | Wt27 |
| Wt28 | Wt28 | Wt28 |
| Wt29 | Wt29 | Wt29 |

Development model
(Kyoda et al., submitted)

Molecular machine model
(Gunsalus et al., 2005)



Kim et al., 2001; Li et al., 2004; Sönnichsen et al., 2005

NII 湘南会議

- アジアにおける最初のダグストゥール形式のセミナー
- 世界トップクラスの研究者が集まり、情報学の難問を解決することを目指す
- 情報学の分野における課題を合宿形式により集中的に議論

Big Data Visual Analytics

Posted by kikakuka on Sep 30, 2014 in Seminars | Comments Off on Big Data Visual Analytics

NII Shonan Meeting:

@ [Shonan Village Center](#), November 8-11, 2015

Organizers

- Seok-Hee Hong, The University of Sydney, Australia
- Koji Koyamada, Kyoto University, Japan
- Kwan-Liu Ma, UC Davis, USA



ポストペタスケール高性能計算に資する可視化処理小委員会

- メンバー（敬称略・五十音順）
 - 行場（東北大）・小山田(京大)・齋藤（北大）・佐藤（富士通）・下條（阪大）・田中克（京大）・田中覚（立命大）・土井（岩手県大）・萩原（明大）・藤代（慶大）・美馬（東大）・宮地（サイバ）

- 審議事項
 - 可視化效能の評価のあり方についての検討
 - 科学的発見を支援するための可視化技術のあり方についての検討
 - 社会実装を行う上で可視化技術の果たす役割についての検討

日本学術会議 公開シンポジウム

可視化

ビッグデータ時代の科学を拓く

日時 平成 26 年 9 月 22 日 10:00~18:00
場所 日本学術会議講堂
(〒106-8555 東京都港区六本木 7-22-34)

開催趣旨
ビッグデータ時代において、可視化は科学の発展との向き合い方に大きな期待がかかっていますが、可視化は今後どうあるべきであるのか、いくつかの問いを提起いたします。本シンポジウムでは、可視化先進活用・研究について、第一人者の取り組み、事例、経験について報告を聞き、会場の聴衆とともにビッグデータ時代の可視化について考えていきます。

日本にとって今必要なのは「科学リテラシー」

プログラム

| | |
|-------------|---|
| 10:00-10:10 | 開会の辞 萩原 一郎 (日本学術会議委員、明治大学先端数理科学インスティテュート・副所長) |
| 10:10-11:10 | 話題提供1 「米国におけるビジュアライゼーションの動向」 寺門和夫 (一般財団法人日本宇宙フォーラム・主任研究員) |
| 11:10-11:50 | 話題提供2 「ビッグデータからの知的発見—海洋科学の研究領域において—」 荒木文明 (独) 海洋研究開発機構地球情報基盤センター・グループリーダー 高橋 桂子 (日本学術会議連携委員、(独) 海洋研究開発機構地球情報基盤センター・センター長) |
| 11:50-13:30 | 昼休み |
| 13:30-13:40 | 話題提供3 「科学と可視化」 小山田 耕二 (日本学術会議連携委員、京都大学国際高等教育院・教授)、萩原 一郎 |
| 13:40-14:00 | 話題提供4 「融合可視化技術」 宮地 英生 (サイバネットワークシステム株式会社アドバンスソリューション事業部・部長) |
| 14:00-14:30 | 話題提供5 「可視化におけるHPC活用」 中島憲宏 (独) 日本原子力研究開発機構 システム計算科学センター・次長 |
| 14:30-15:00 | 話題提供6 「大規模計算機環境における可視化について」 小野謙二 (理化学研究所 計算科学研究機構・チームリーダー) |
| 15:00-15:30 | 休憩 |
| 15:30-17:50 | 話題提供7 パネルディスカッション:日本版「可視化研究開発の課題」トピック10 パネリスト:伊藤貴之(お茶の水女子大学・教授)、瀬尾 拓史(株式会社サイアメント代表取締役社長・医師)、田中 覚(立命館大学情報理工学部・教授)、寺門和夫、小山田 耕二 司会: 藤代 一成(慶應義塾大学大学院 理工学研究科・教授) |
| 17:50-18:00 | 閉会の辞 小山田 耕二 |

お申し込みはこちらから <http://goo.gl/pSvT7E>



主催: 総合工学委員会・機械工学委員会合同計算科学シミュレーションと工学設計分科会
共催: 日本機械学会、情報処理学会、芸術科学会、画像電子学会、日本計算工学会、日本シミュレーション学会、可視化情報学会、自動車技術会

データサイエンスを支える可視化

可視化について

データと情報

データ

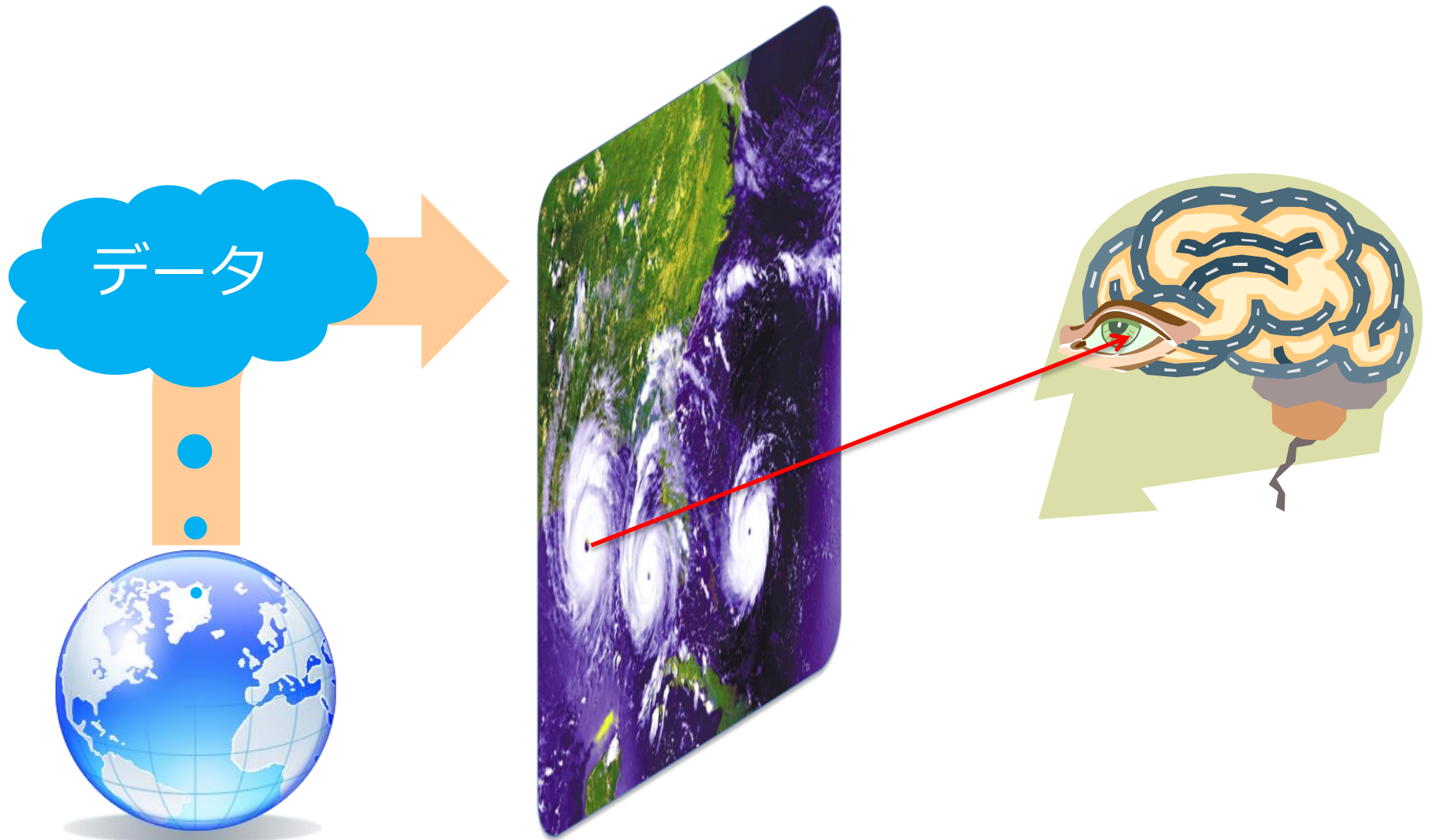
何かを符号で表現したもの

可視化

情報

人間が認識したデータ

可視化：データと脳をつなぐ



可視化研究と性能評価

医・理・工学

計測・計算

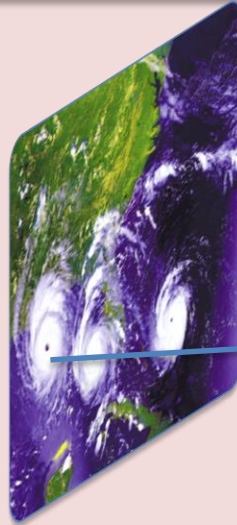
データ



どんな現象をデータ化
できたか？

情報科学

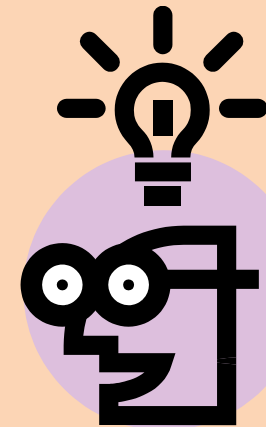
画像化



どれほど効率よく画像
化できたか？

認知科学

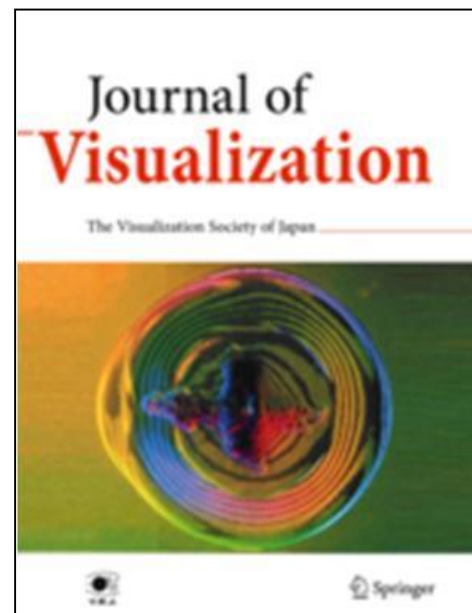
認識



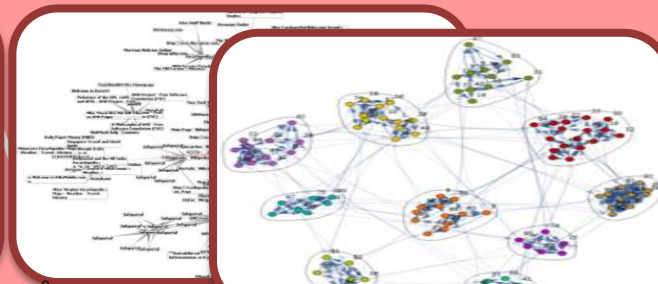
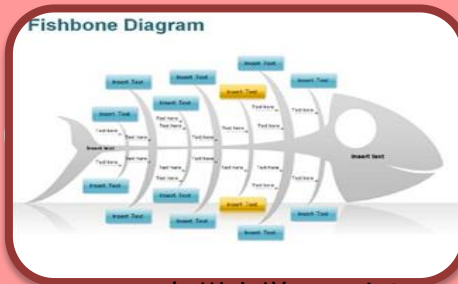
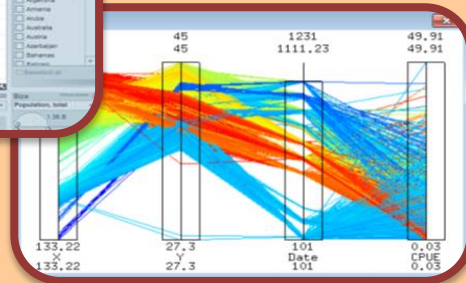
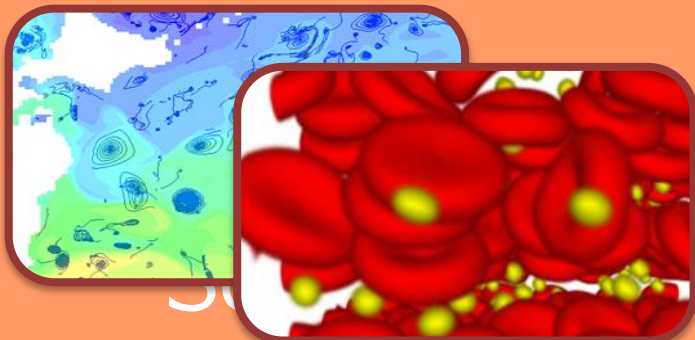
どれほどの気づきを得
たか、どんな行動変容
に結び付いたか？

Journal of Visualization

- History
 - Springer Vol.1(1998)-
- Editors-in-Chief:
 - K. Koyamada
 - K.C. Kim
- Scope
 - Visualization is an interdisciplinary imaging science devoted to making the invisible visible through the techniques of experimental visualization and computer-aided visualization.



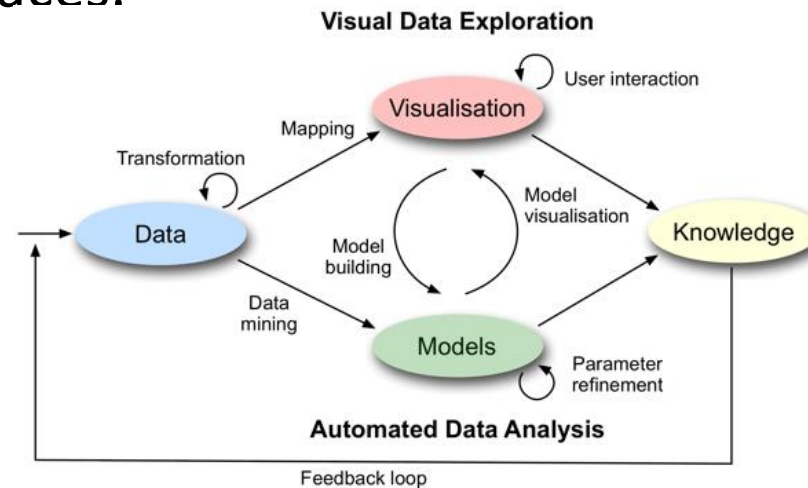
データ可視化技術



視覚的分析

J J Thomas and K A Cook "Illuminating the Path: The R&D Agenda for **Visual Analytics**" 2005

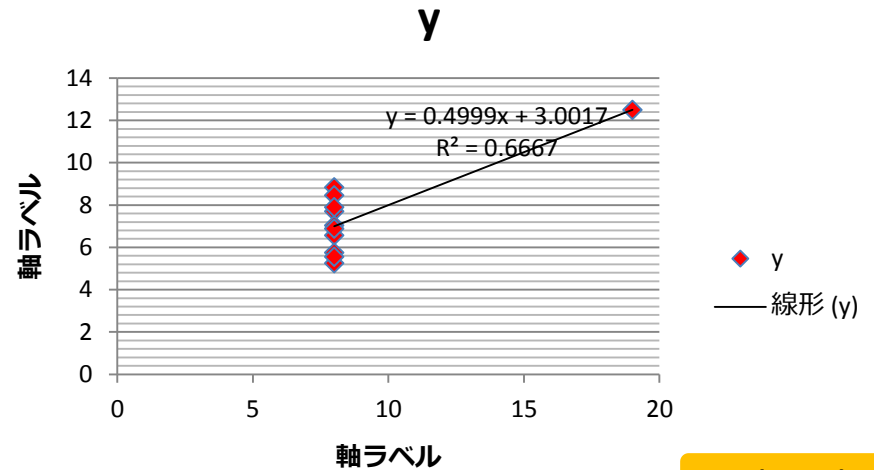
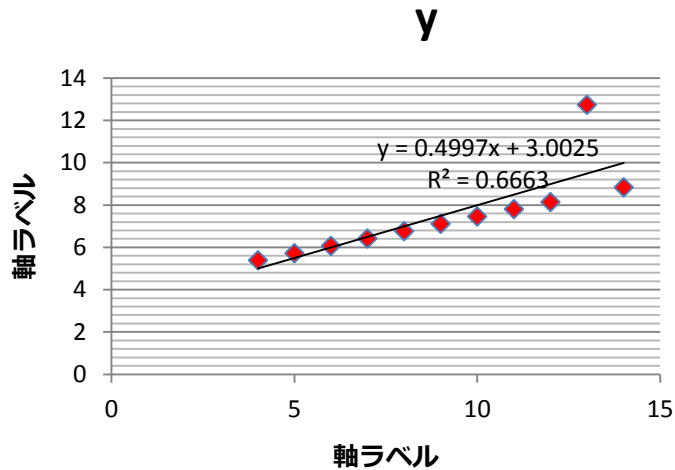
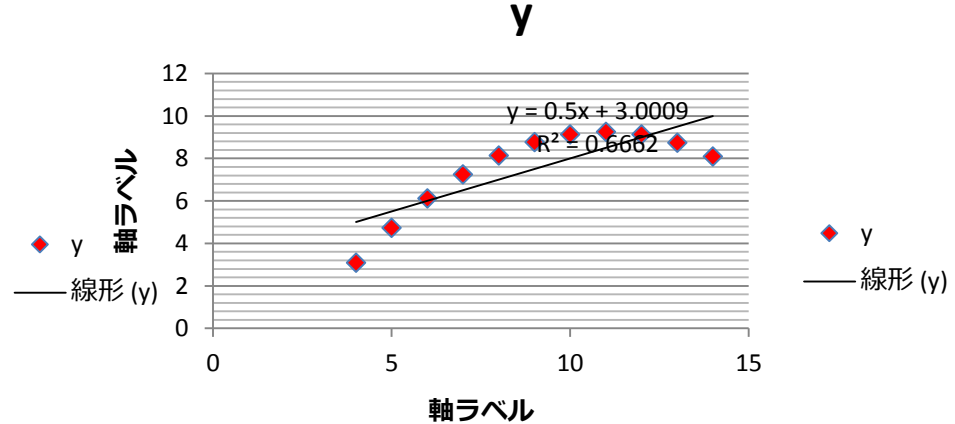
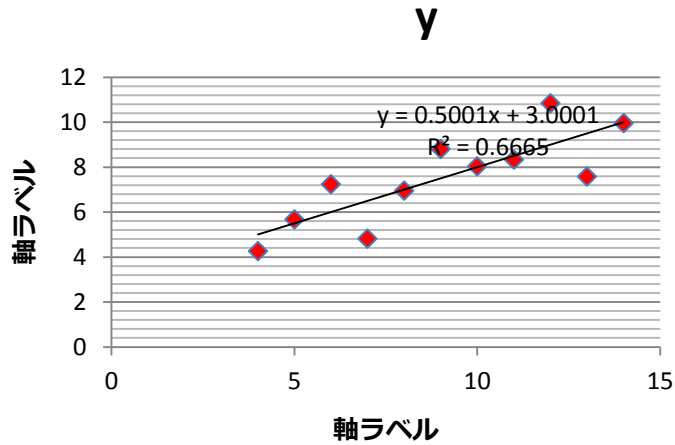
- The science of analytical reasoning supported by interactive visual interfaces.



"If we are to build a science of analytical reasoning,
we must apply **the scientific method**"

W. Ribarskya, B. Fisherb, W M. Pottengerc, "Science of Analytical Reasoning, " Information visualization, 8 (4), 2009

可視化は重要か？



Tedx video

F.J. Anscombe, "Graphs in Statistical Analysis," American Statistician, 27, 17-21(1973).

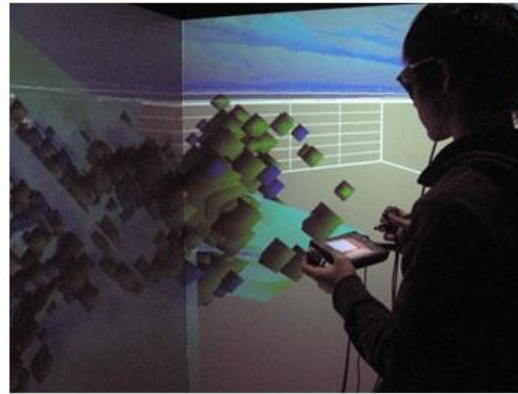
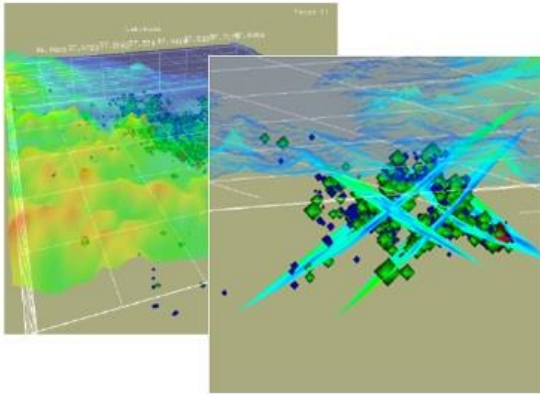
可視化マントラ

Ben Shneiderman

俯瞰

拡大&濾過

対話的詳細化



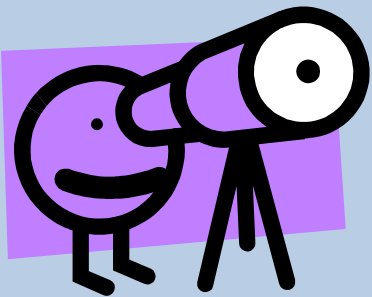
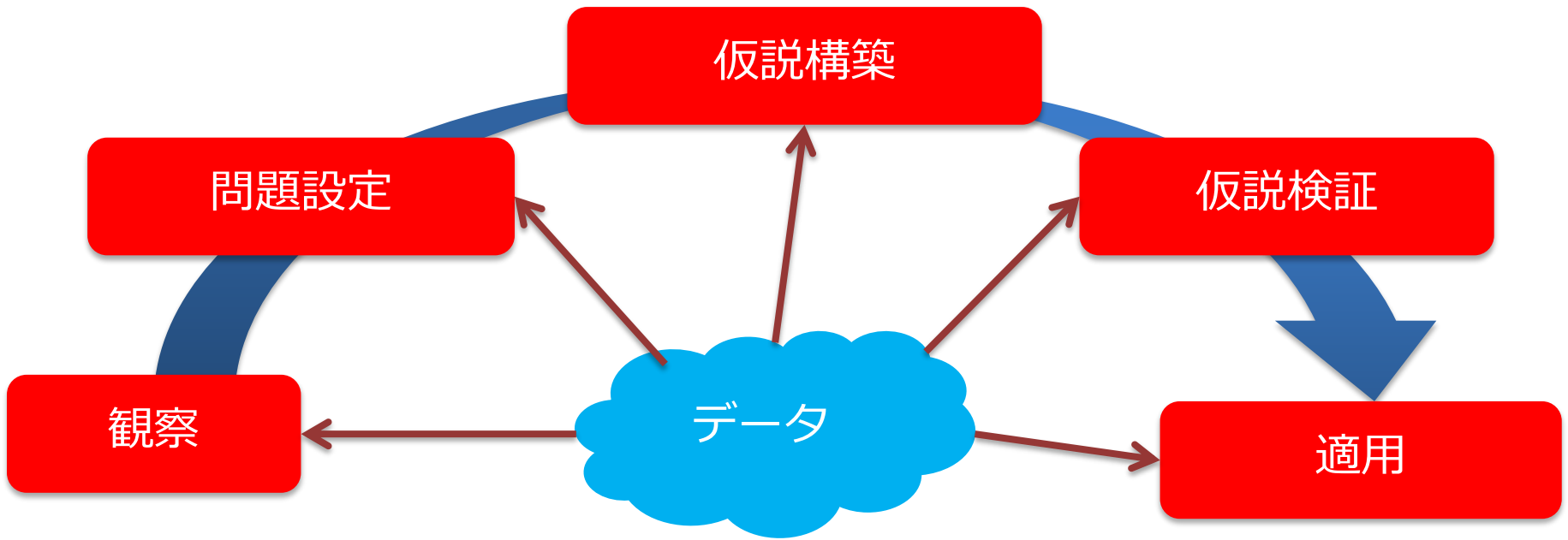
安原幸生, [坂本尚久](#), 江原康生, 片尾浩, [小山田耕二](#), "震源データからの断層面推定支援システムの開発", 日本シミュレーション学会誌, Vol.26, No.4, pp.212-218, 2007

データサイエンスを支える可視化

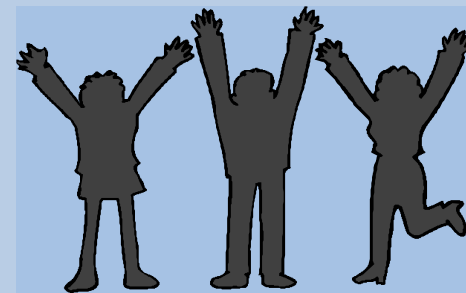
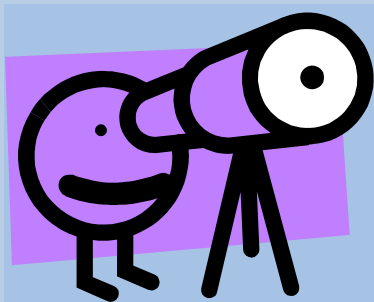
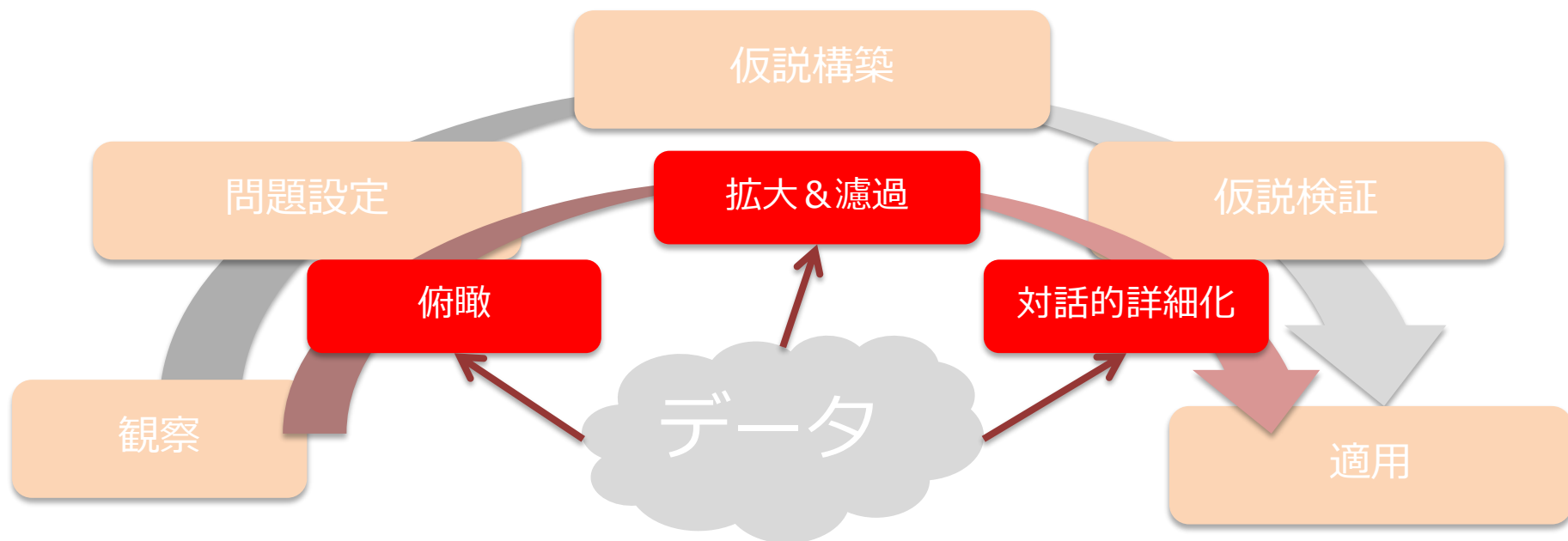
科学的可視化について

≠ サイエンスグラフィックス

科学的方法の流れ

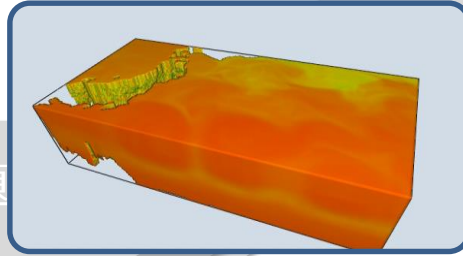


科学的可視化



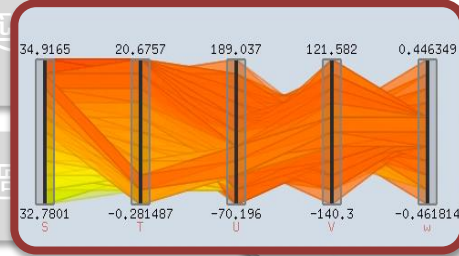
海洋科学における可視化

問題



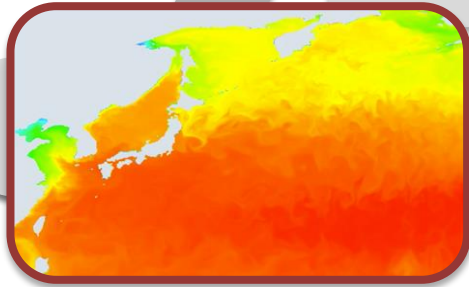
仮説構築

拡大&濾過

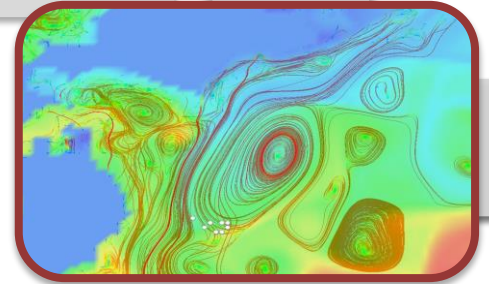


検証

俯瞰



対話的詳細化



データ



社会

九州大学ワークショップ

子供の科学

原田三夫, “この雑誌の役目”, 1924年10月号

・・・しかしこの雑誌の一番大切な目的は、ほんとうの**科学**というものが、どういうものであるかを、皆さんに知っていただくことであります。ちかごろは「**科学科学**」とやかましくいいますが、ほんとうに**科学**というものを知っている人はたくさんないようです。人は生まれながら美しいものを好む心を持っていますが、それと同じように自然の物事について詳しく知り、深くきわめようとする欲があります。昔からその欲の強い人が調べた結果自然の物事の間には、たくさんの定まった規則のあることがわかりました。**科学**ということは、この規則を明らかにすることです。多くの方が**科学**と言っているのは、たいていはその応用にすぎません。この規則を知ることによって人間は自然にしたがって無理のないように生き、楽しく暮らすことができ、これを応用して世が文明におもむくのです。

科学とは、物事の間因果関係を明らかにすること



科学する心

小林秀雄，“科学する心”，岩波書店講演集CD，1970年

・・・それは、まあこの環境を知る一種の**科学**論だなあ。**科学**論てものはねえ、とてももう面倒なものなんです。で、ただ、科学ってものはだなあ、ものがその、ものをこの本当にものをするのが**科学**ではない。あれはものの法則を知るんです。いいですか、そこはねえ、あのよおく考え貰わないといけないんだよ。つまりだなあ、えー、**科学**ってものはいつでも法則をめぐっているんです。ほんとに僕らの経験ってものを、ほんとの生きる経験ってものは**科学**はいらないんです。生きていなくてもいいんです。生きている人間ってものは、科学は、そんなものは認めないんです。いつでも、**科学**はねえ、その、規則をめぐらせるんです。ものともものはどういう関係にあるかってことを、因果関係だね、一口に言えば因果関係ってものは、自然はどういうふう動いているかって、いうその因果関係っていうものを目指しているんです・・・



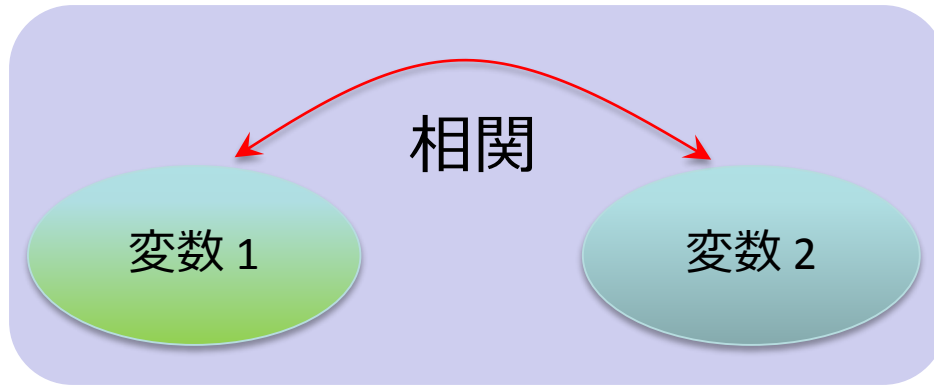
科学は因果関係を目指す

データサイエンスを支える可視化

因果関係の探索

九州大学ワークショップ

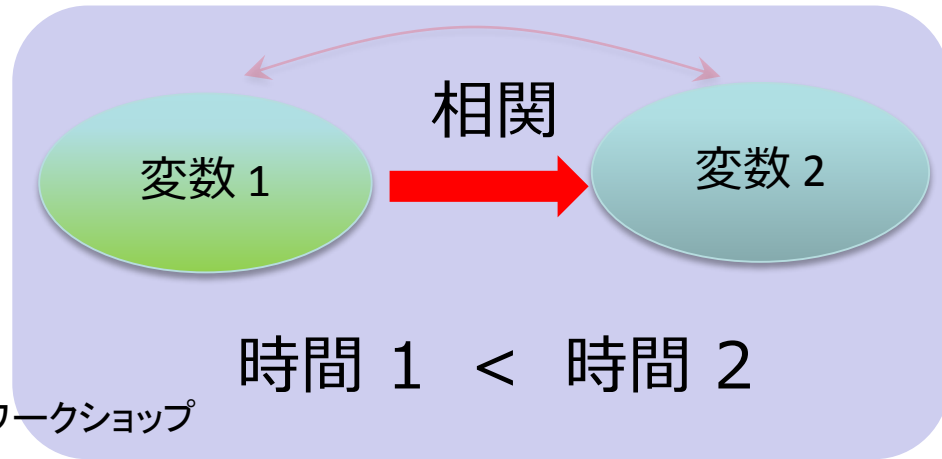
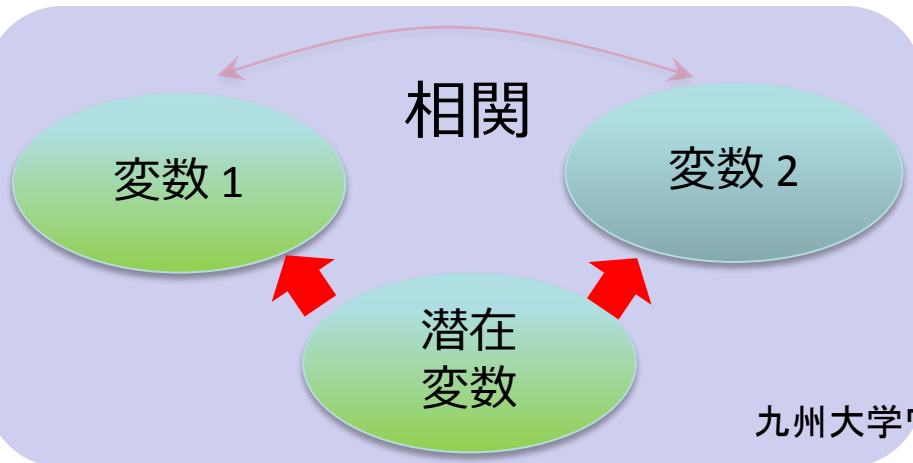
因果関係



相関は、必ずしも因果を意味しない

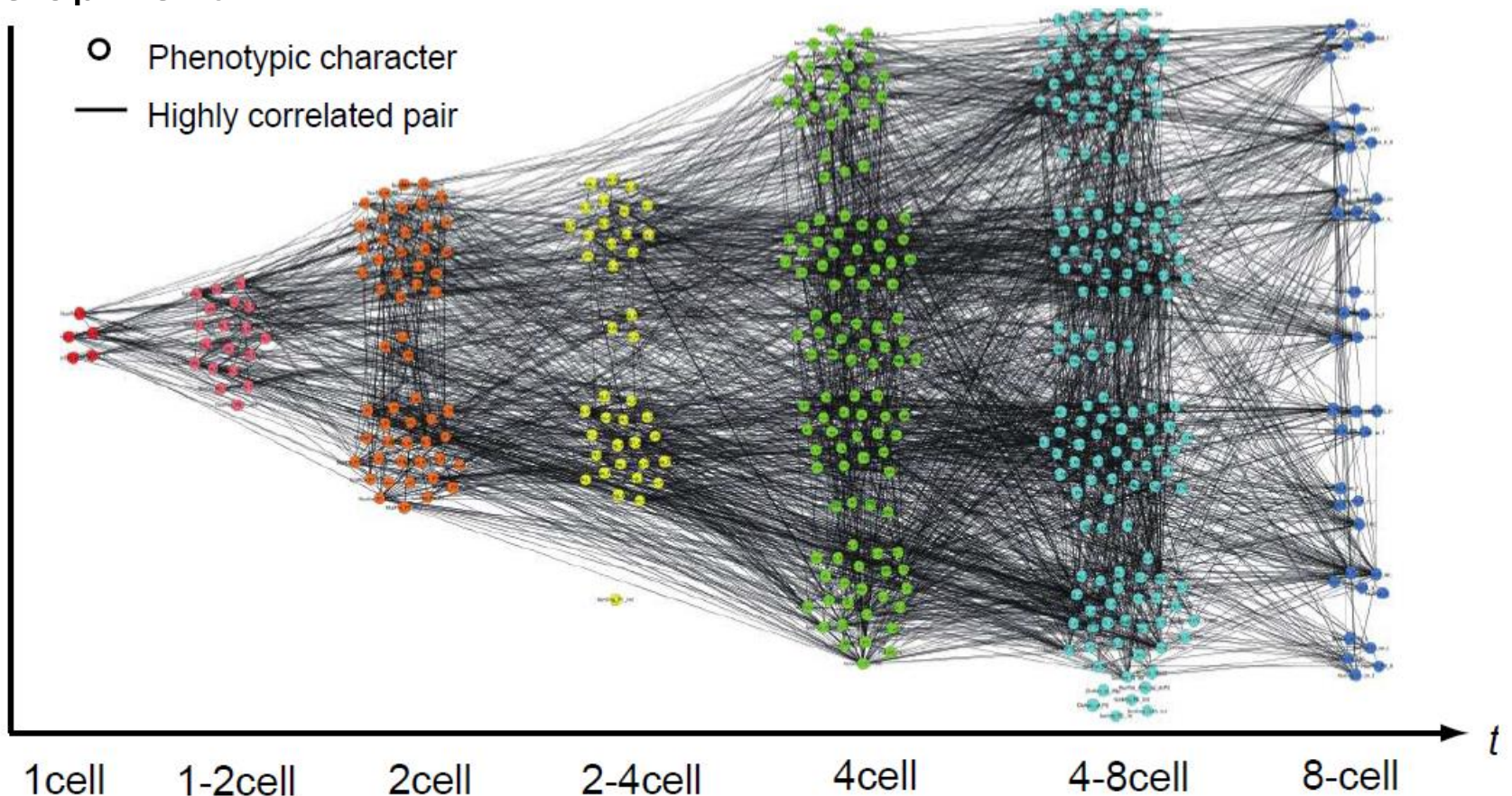
タイプ1 因果関係

タイプ2 因果関係



線虫の発生過程

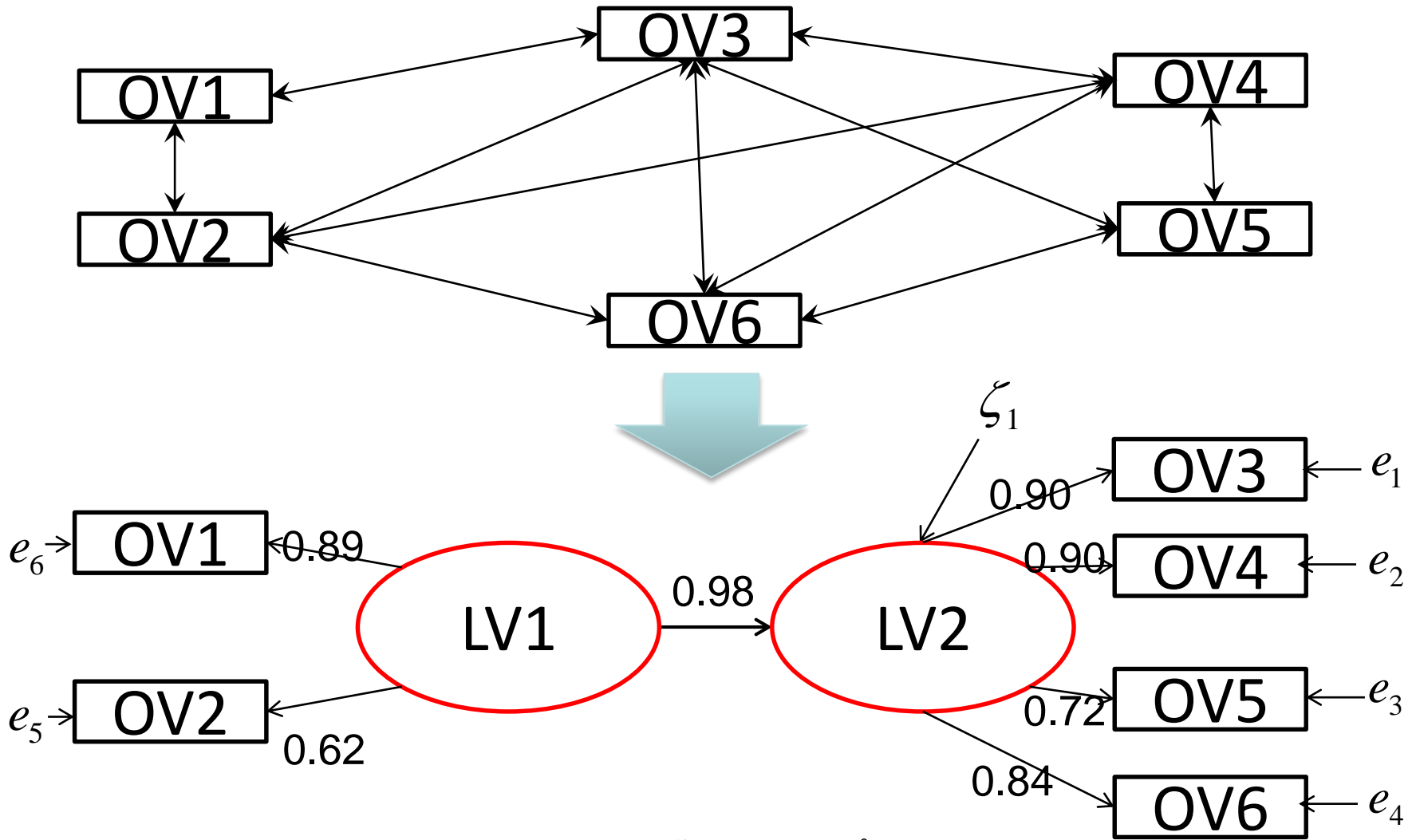
Correlations between phenotypic features in the c. elegans development



Data courtesy: Dr. Omani of RIKEN

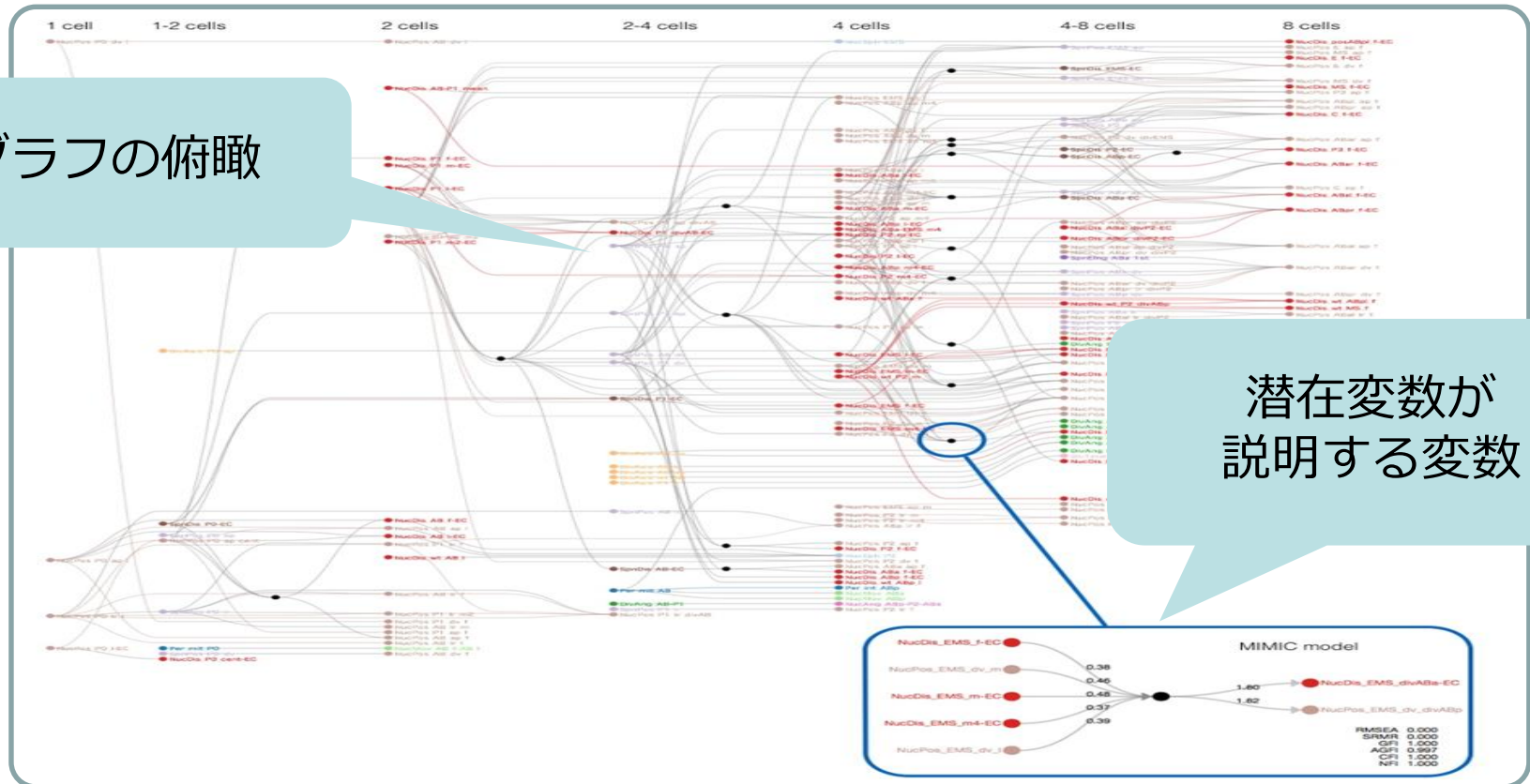
タイプ1 因果関係

構造方程式モデルによる潜在変数の導入



因果グラフにおける対話的可視化

グラフの俯瞰

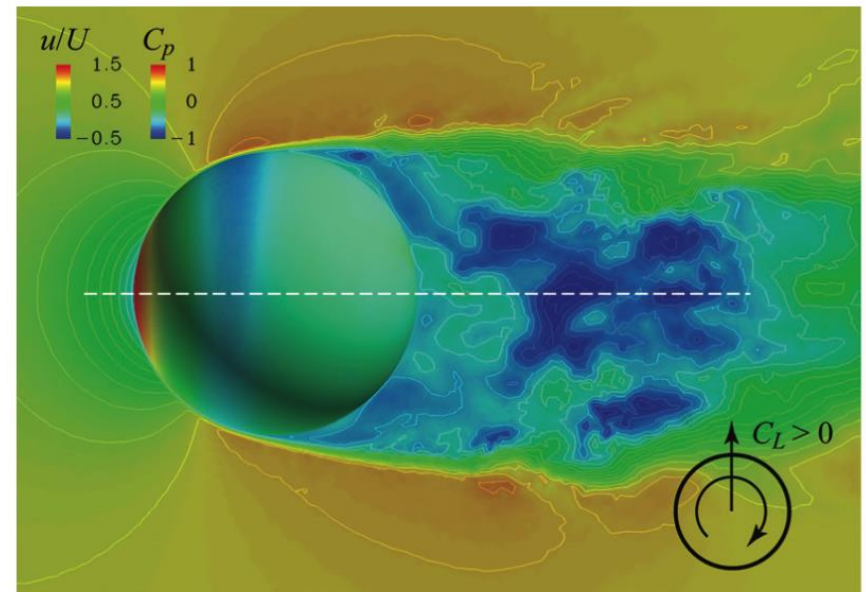
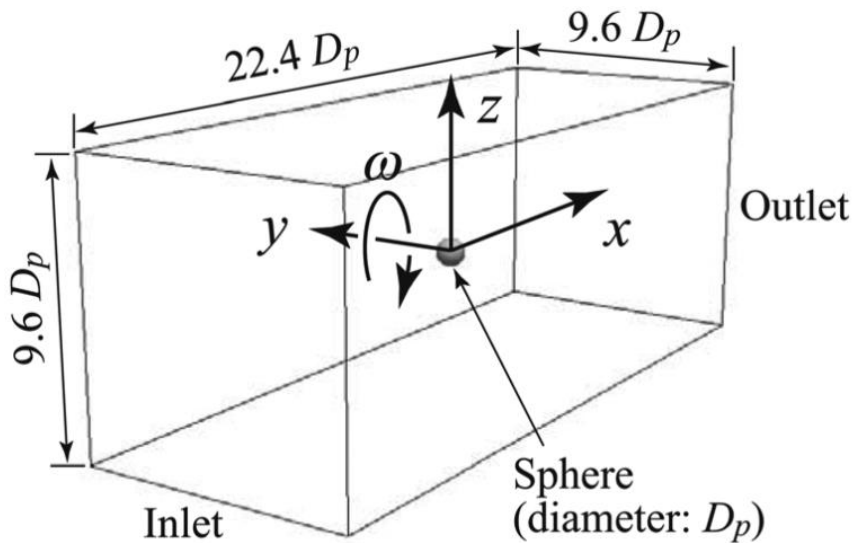


潜在変数が説明する変数

潜在変数の適切な設定のために対話的グラフ可視化技術が重要

熱流体解析モデル

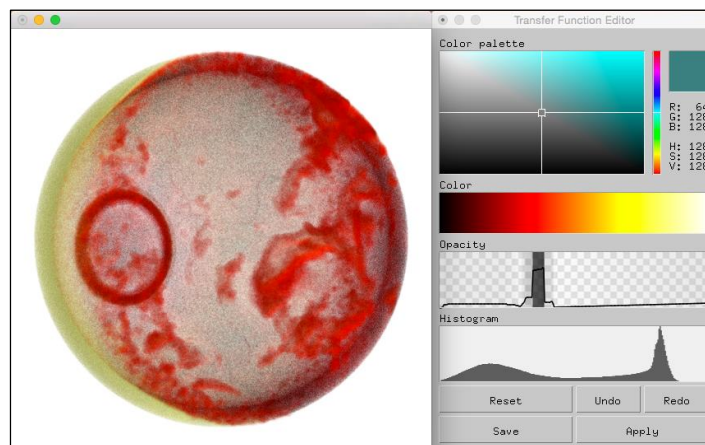
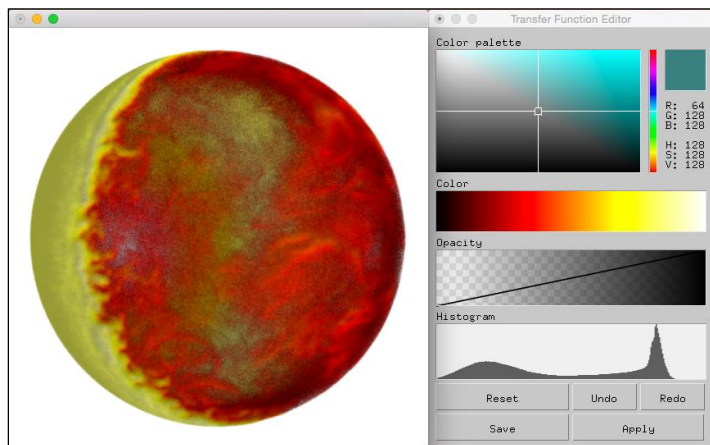
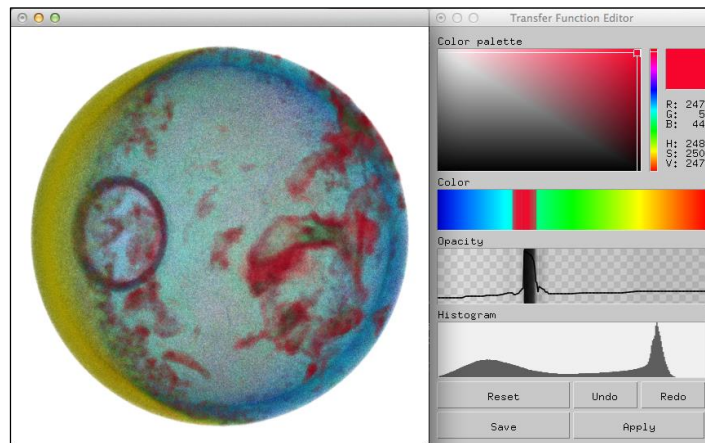
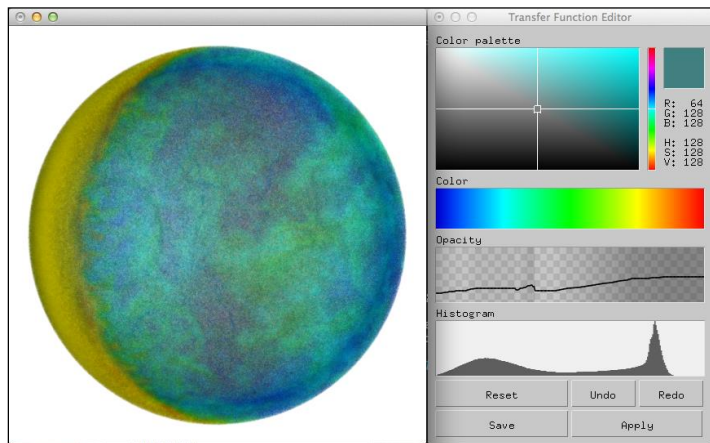
- 乱流場における加熱球体 まわりの熱流
解析計算結果データ



データ仕様

- 節点数 : 15,321,546
- 要素数 : 18,917,887 (prism)
27,545,304 (tet)
- タイムステップ数 : 198
- ファイルサイズ : 3.5GB per step

球体近傍における解析結果

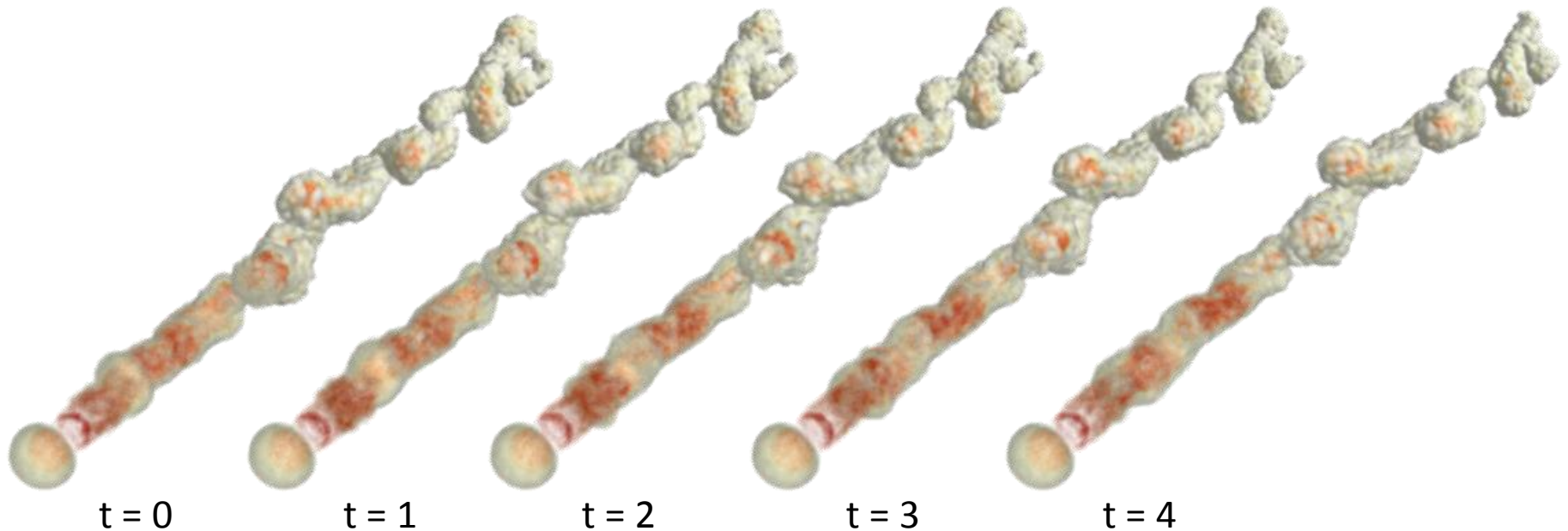


Average rendering speed: **8.77 fps**

Transfer function adjusting time: **163.9 milisecondes**

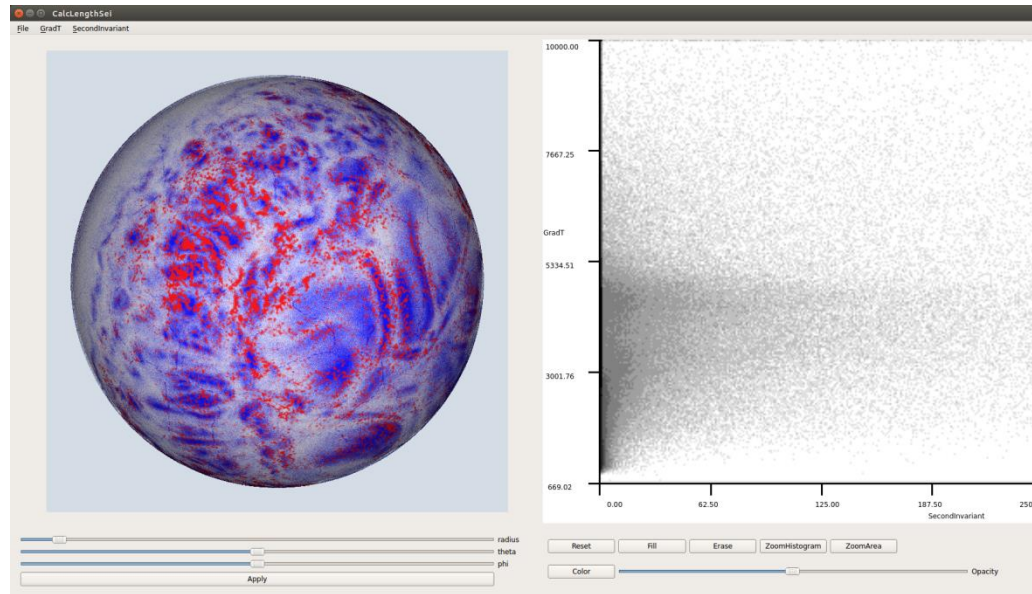
アニメーション表示

- 粒子レンダリング法の活用



研究対象としての因果関係

- 後流における熱輸送（熱伝達係数）に、渦がどのような影響を及ぼすか？

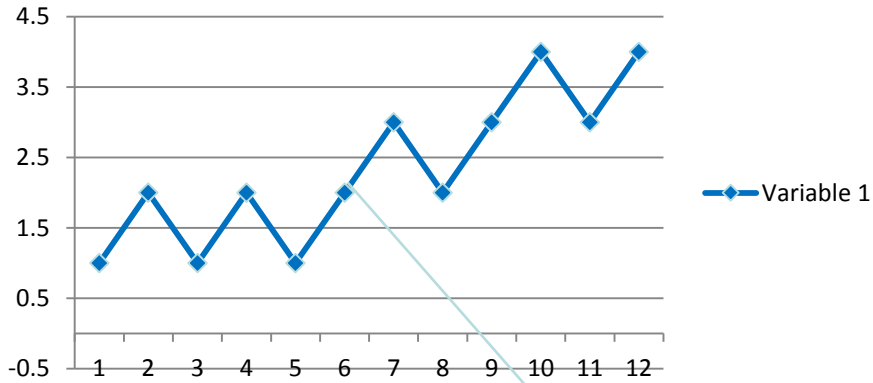


Naohisa Sakamoto, Kun Zhao and Koji Koyamada, Visual analytics system for finding a causal relationship between physical quantities from multivariate volume datasets, 1st Symposium on Visualization in Data Science (VDS), poster, 2015.

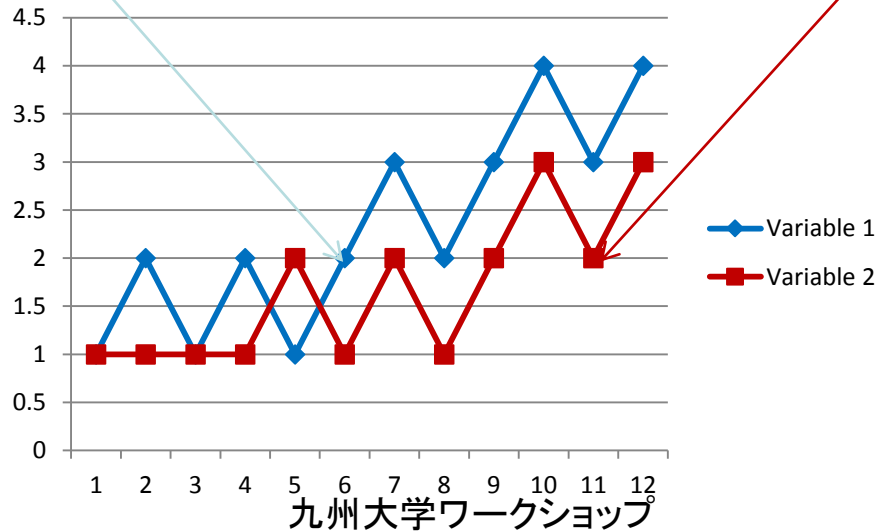
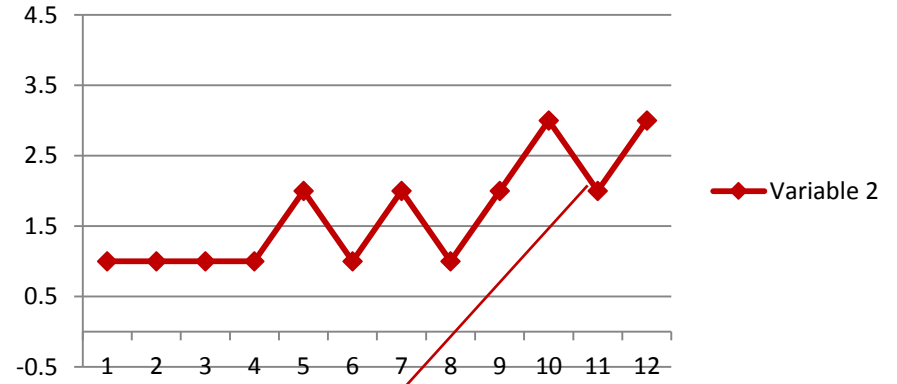
タイプ2 因果関係

時系列データ間の比較

変数 1



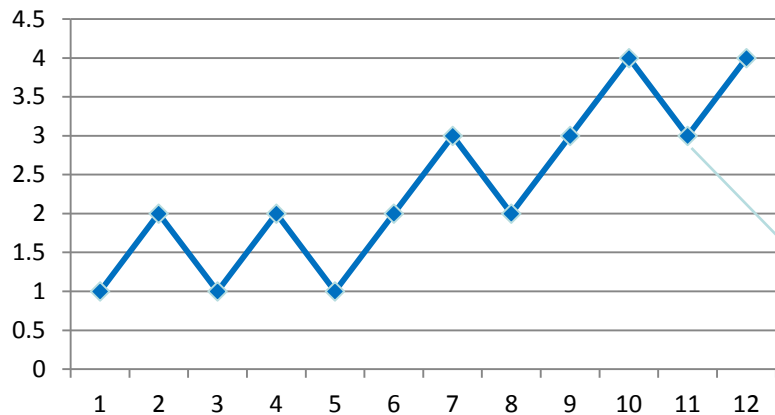
変数 2



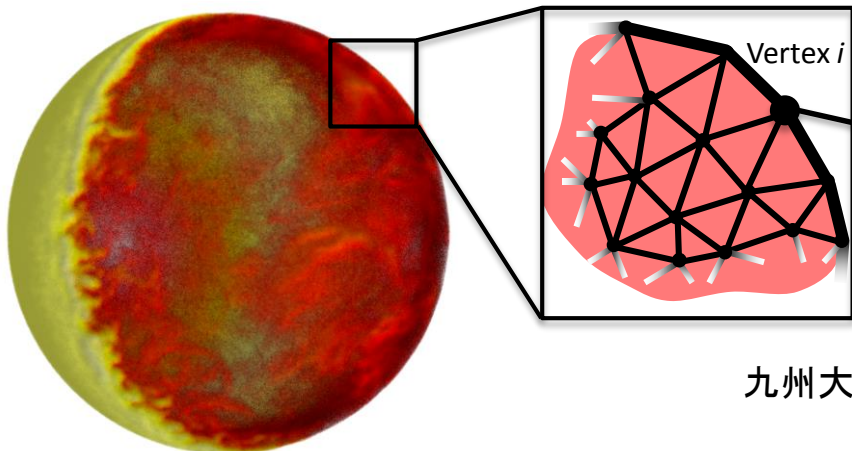
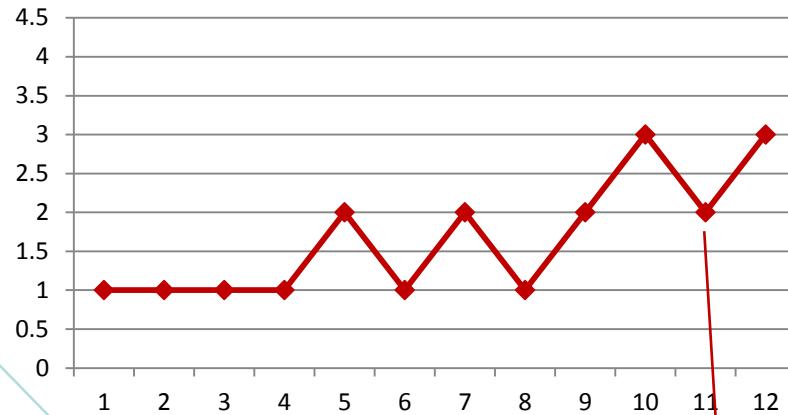
R=0.798023875

節点毎の時系列データ比較

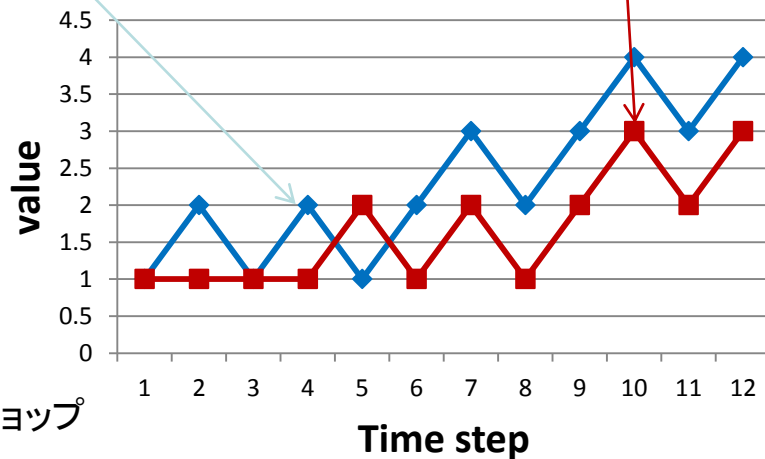
Temperature



Pressure

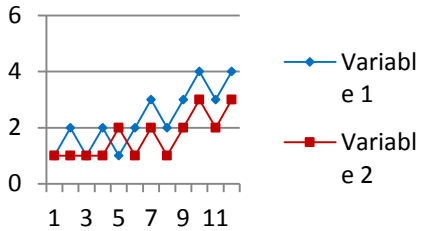


九州大学ワークショップ

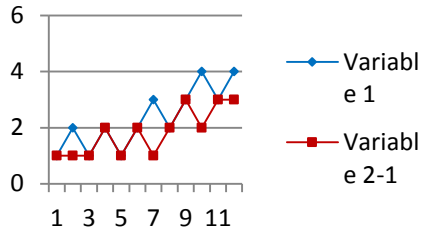


相互相関

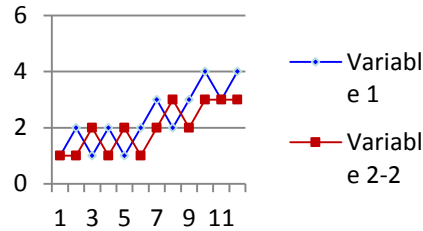
時間 shift = 0



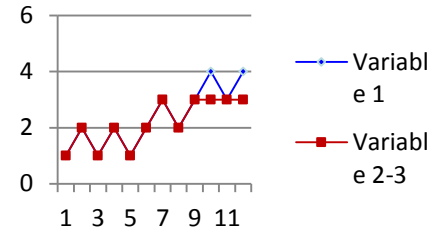
時間 shift = 1



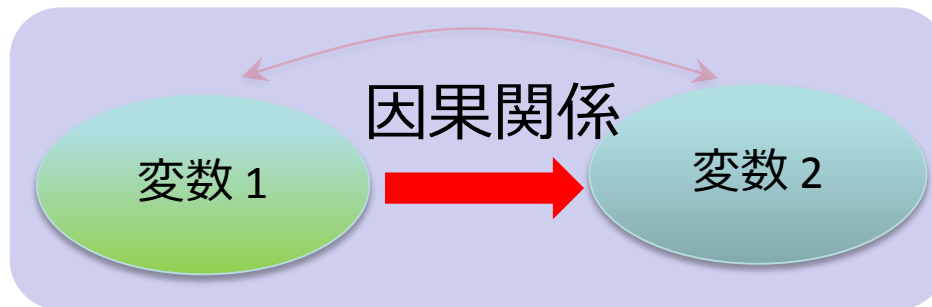
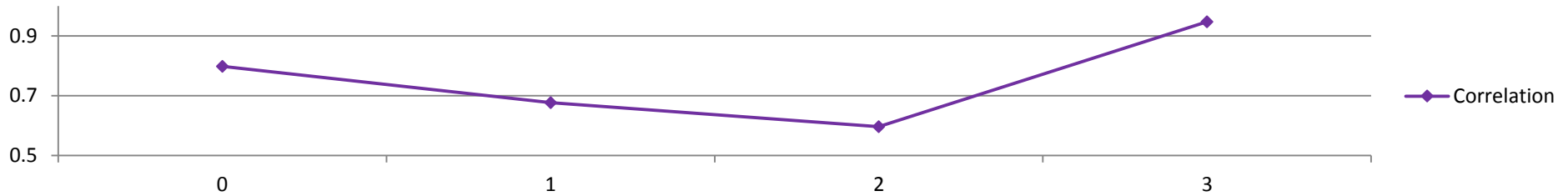
時間 shift = 2



時間 shift = 3

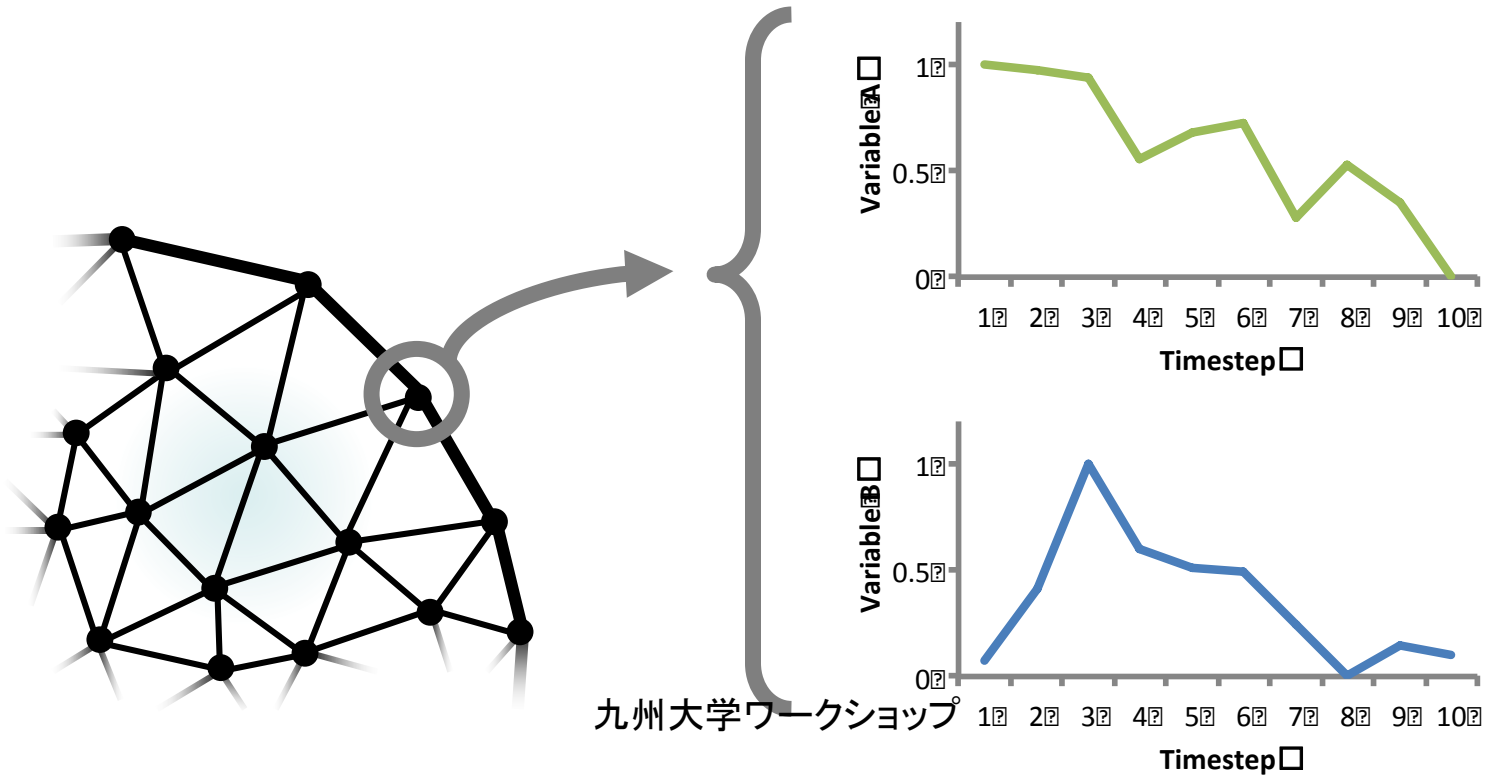


相関



動的時間伸縮法

- 二つの変数の時間変化（時系列データ）の類似度を分析する。



グレンジャー因果(1969, 1980)

二つの原理

1. 原因は、結果に先立って起こる
2. 原因は、結果にユニークな変化を引き起こす。言い換える
と原因となる時系列データには、結果となる時系列データ
について、他にはないユニークな情報が含まれる

以下のように情報を定める

- $I(t)$ – 時刻 t までの宇宙におけるすべての情報
- $I_{-x}(t) - I(t)$ のうち x に関する情報を取り除いたもの

ふたつの時系列 X と Y において以下の関係が成立するとき, X は、結果 Y の原因であると推論できる

$$P[Y(t+1) \in A | I(t)] \neq P[Y(t+1) \in A | I_{-x}(t)]$$

グレンジャー因果: 実用的な定義

We perform two vector auto-regressions as follows:

$$Y(t) = \sum_{l=1}^L a_l Y(t-l) + \epsilon_1 \quad (1)$$

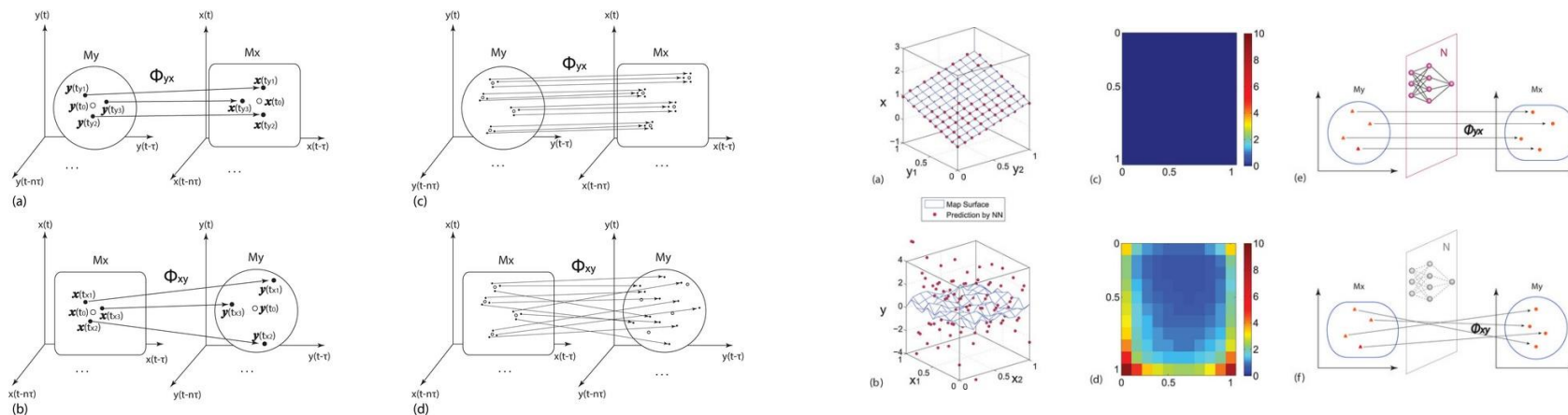
$$Y(t) = \sum_{l=1}^L a'_l Y(t-l) + \sum_{l=1}^L b'_l X(t-l) + \epsilon_2, \quad (2)$$

where L is the maximal time lag. We say X causes Y if eq (2) is statistically significantly better than eq (1).

短期時系列データを用い因果推論

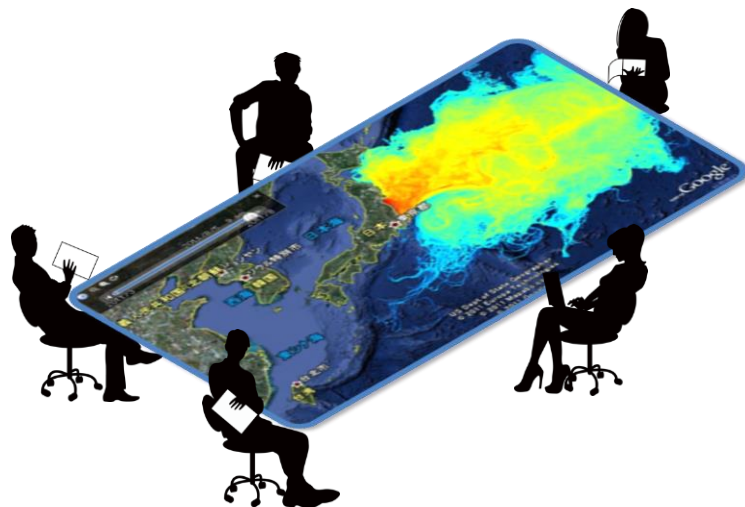
Huanfei Ma, Kazuyuki Aihara & Luonan Chen, Scientific Reports 4 (2014)

- 非線形力学系のアトラクターの埋め込み定理に基づく因果推論
 - 2つの観測された変数間の交差写像の滑らかさを数値化
 - ベンチマークからの数学的モデルと、生物系からの実際のデータの両方を使って有効性を実証



まとめ

- データサイエンスを支える可視化について説明した
 - 可視化について
 - 科学的可視化について
 - 因果関係の探索



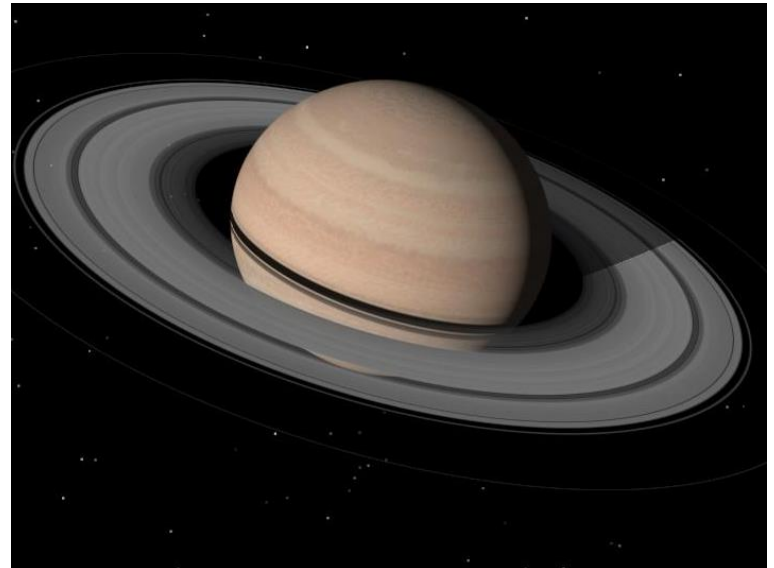
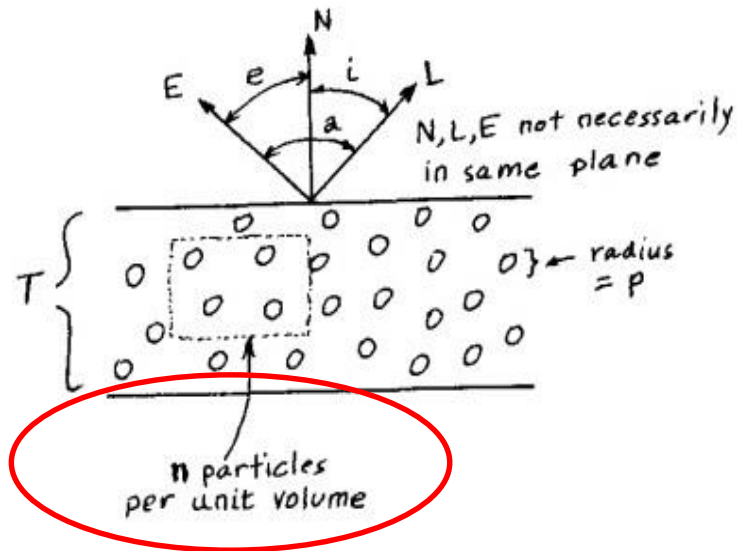
データサイエンスを支える可視化

可視化の歴史

粒子モデル

J. F. Blinn, "LIGHT REFLECTION FUNCTIONS FOR SIMULATION OF CLOUDS AND DUSTY SURFACES," 1982

- 粒子はポアソン分布に従う



All other particles are assumed to be of radius p also. For such a particle to be completely outside the cylinder its center must be outside this volume. Statistically, then, the attenuation of light traversing a cylinder of radius p and volume V is $P(0;V)$ = the probability of 0 particles in volume V . Now the expected number of particles in a given volume V is nV . If n is small this can be modelled as a Poisson process and

$$P(0;V) = \exp(-nV)$$

レンダリング方程式

James T. Kajiya. Brian P. Von Herzen. , “Ray tracing Volume Densities,” 1984

- 吸収と散乱を考慮した光学モデル

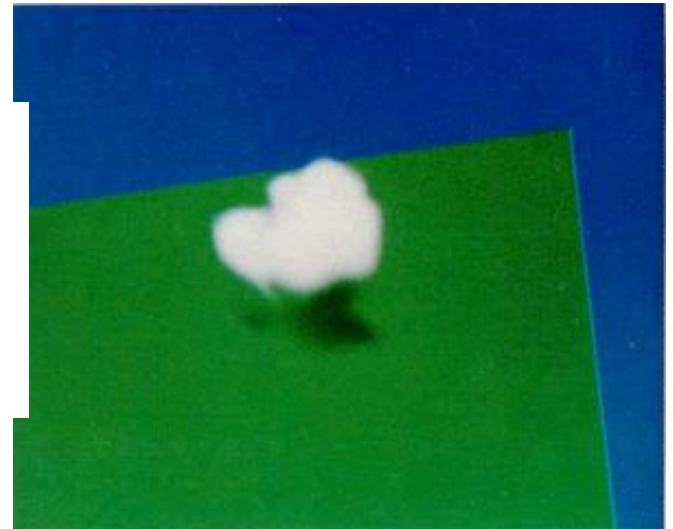
$$\begin{aligned}dI &= -\text{absorbed} + \text{emitted} \\ &= -\kappa\rho ds d\sigma d\omega + j\rho ds d\sigma d\omega\end{aligned}$$



$$\frac{dI}{ds} = \mathbf{s} \cdot \nabla_{\mathbf{x}} I.$$



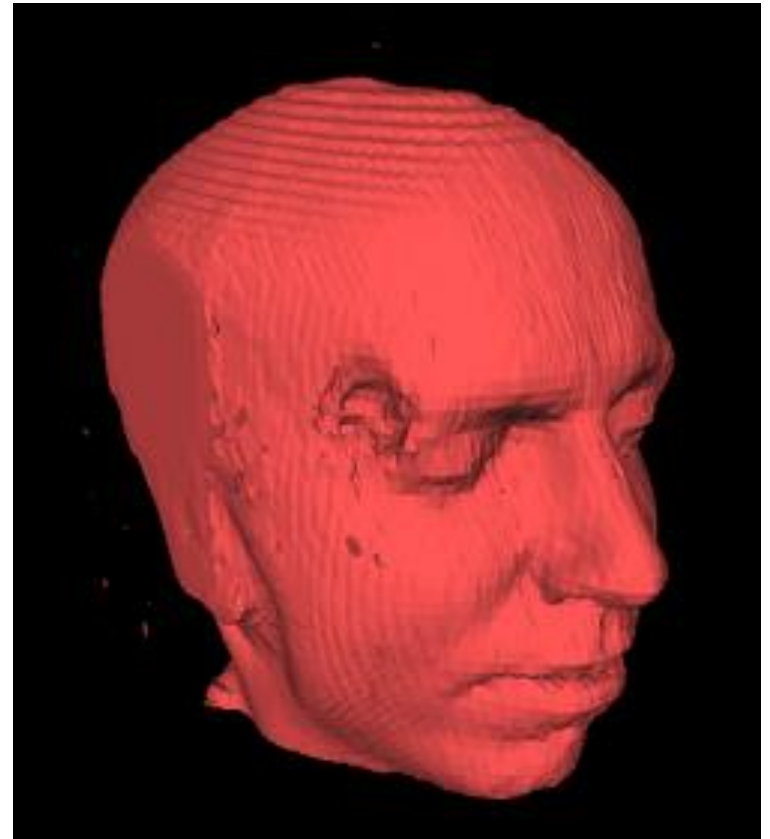
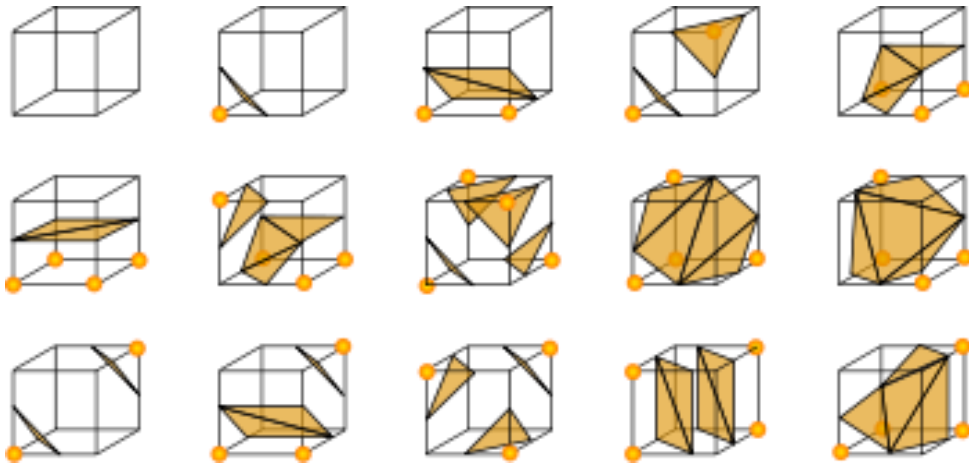
$$\begin{aligned}\frac{-1}{\kappa\rho} \mathbf{s} \cdot \nabla_{\mathbf{x}} I(\mathbf{x}, \mathbf{s}) - I(\mathbf{x}, \mathbf{s}) \\ + \frac{1}{4\pi} \int_{\|\bar{\mathbf{s}}\|=1} p(\mathbf{s}, \bar{\mathbf{s}}) I(\mathbf{x}, \bar{\mathbf{s}}) d\bar{\mathbf{s}} = 0.\end{aligned}$$



マーチングキューブ

WE Lorensen, "Marching cubes: A high resolution 3D surface construction algorithm," 1987

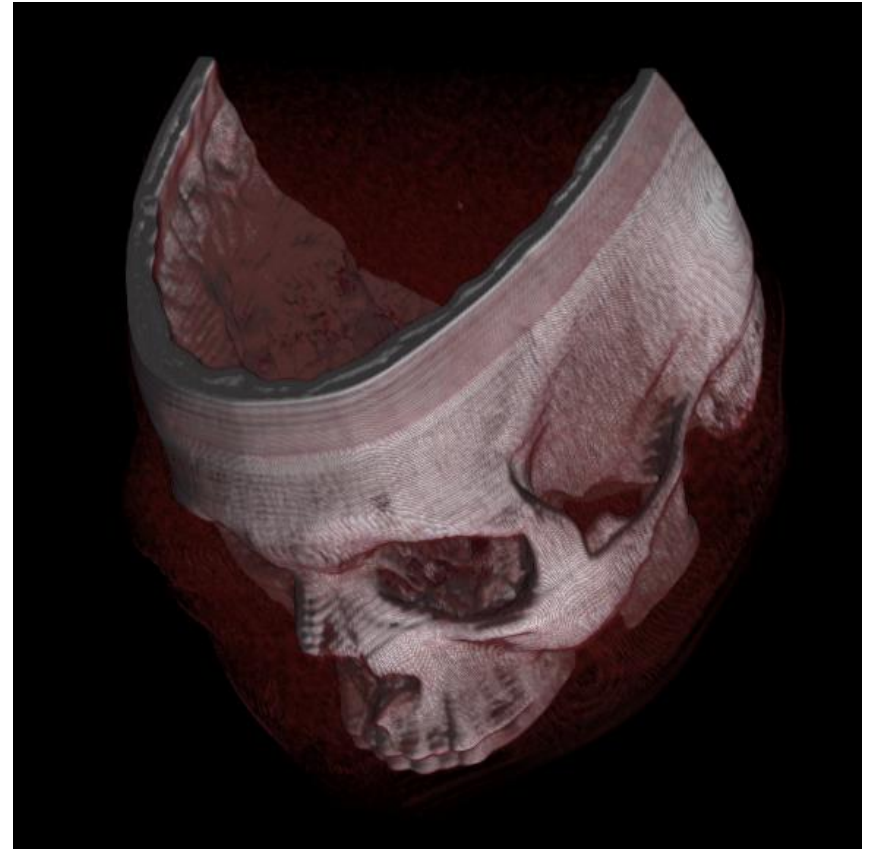
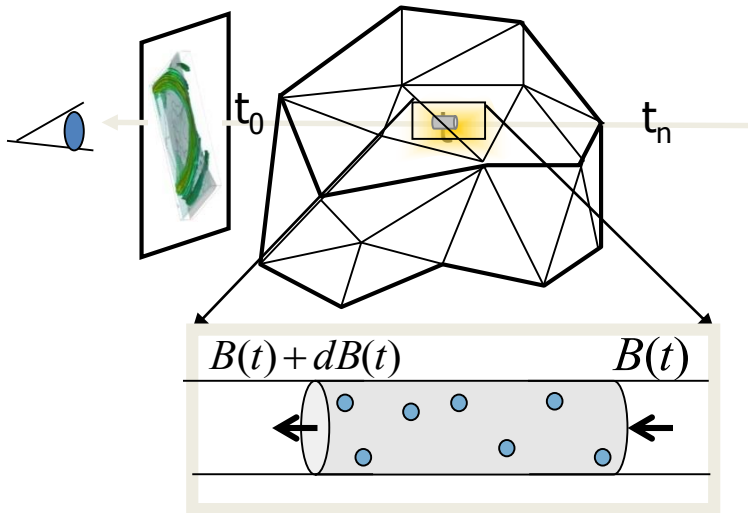
- 等値面のポリゴン表現



ボリウムレンダリング

Paolo Sabella, "A rendering algorithm for visualizing 3D scalar fields," 1988
Craig Upson, Michael Keeler, "V-buffer: visible volume rendering," 1988
Robert A. Drebin, Loren Carpenter, Pat Hanrahan, "Volume rendering," 1988

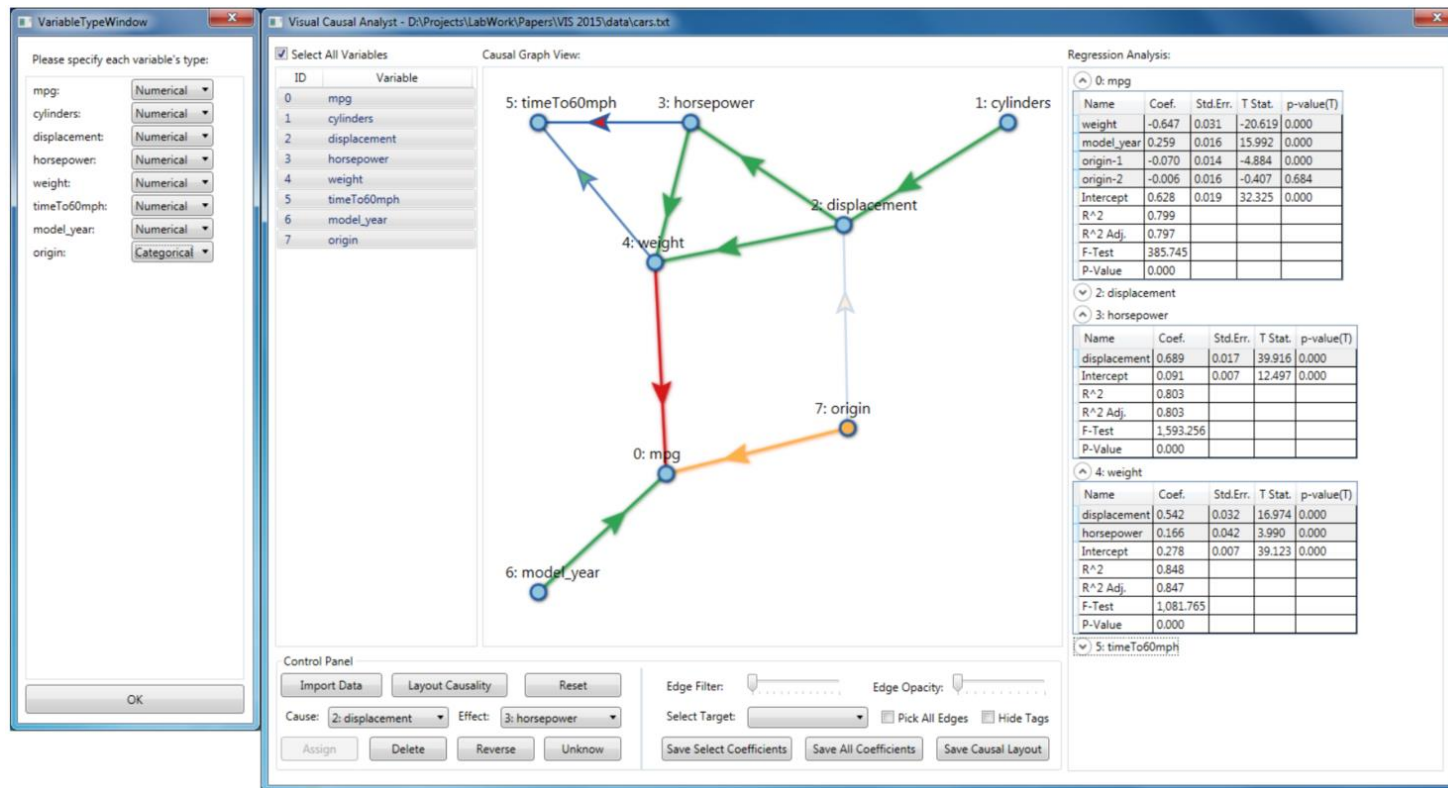
- スカラ場の雲状表現



The Visual Causality Analyst: An Interactive Interface for Causal Reasoning

Jun Wang and Klaus Mueller, IEEE Transactions on Visualization and Computer Graphics, 30(1)

- 専門家による因果関係のエラーチェックを支援



潜在変数や時系列データが扱えない