可視化の世界

計算科学が拓く世界

2016年10月19日（水）

学術情報メディアセンター

小山田耕二

内容

- 可視化について
- 可視化事例
- 因果推論技術
- 課題

本日のテーマと目的

- 可視化は、計算機や計測装置等から生成される膨大な数値データから気付きを得るための基盤技術として重要になっている。本講義では、計算科学と密接な関係にある可視化技術の基礎と応用について説明する。

可視化：ビッグデータ時代の科学を拓く

可視化について

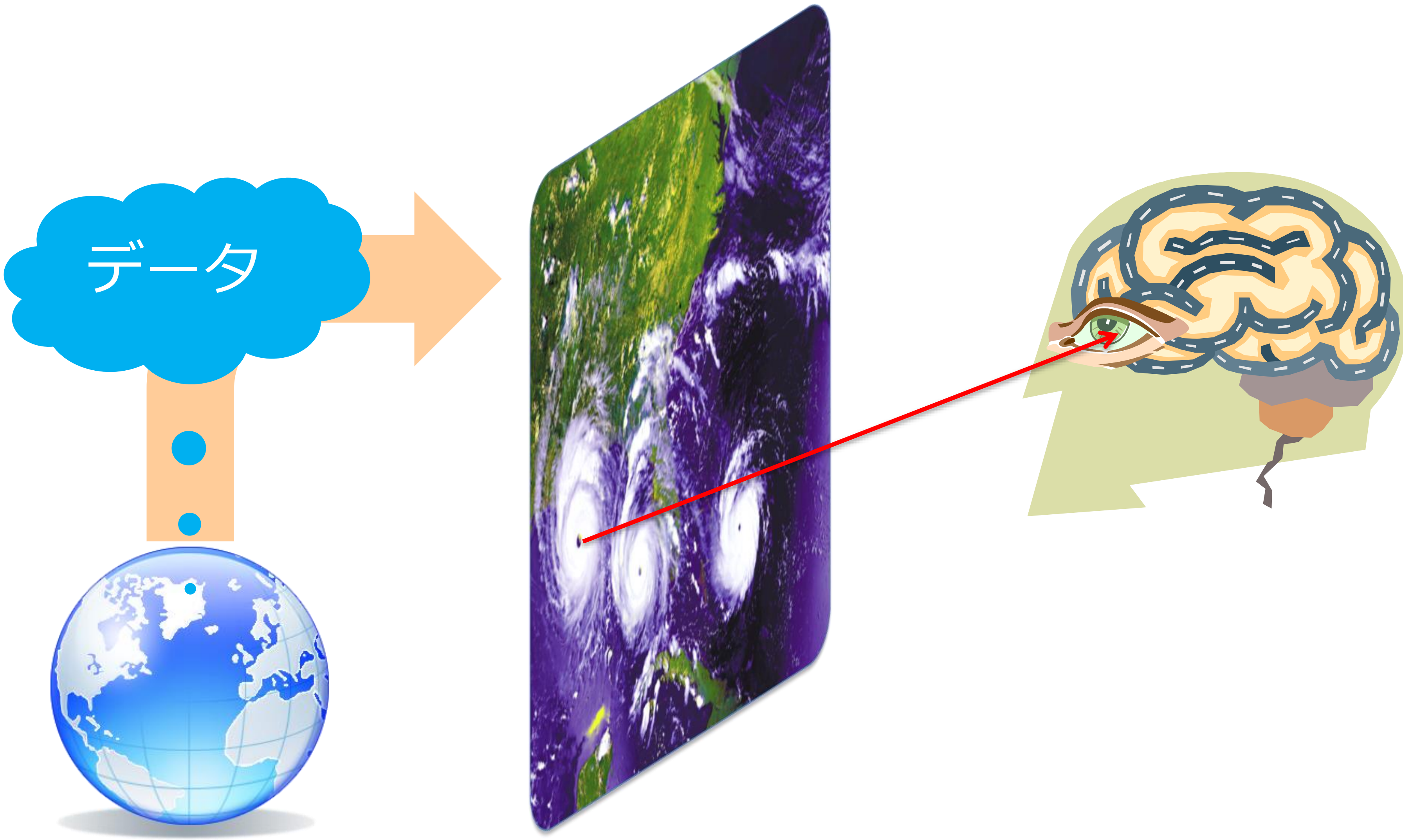
ビッグデータ時代の科学を支える素養

Source: IDC's Digital Universe Study, sponsored by EMC, June 2011

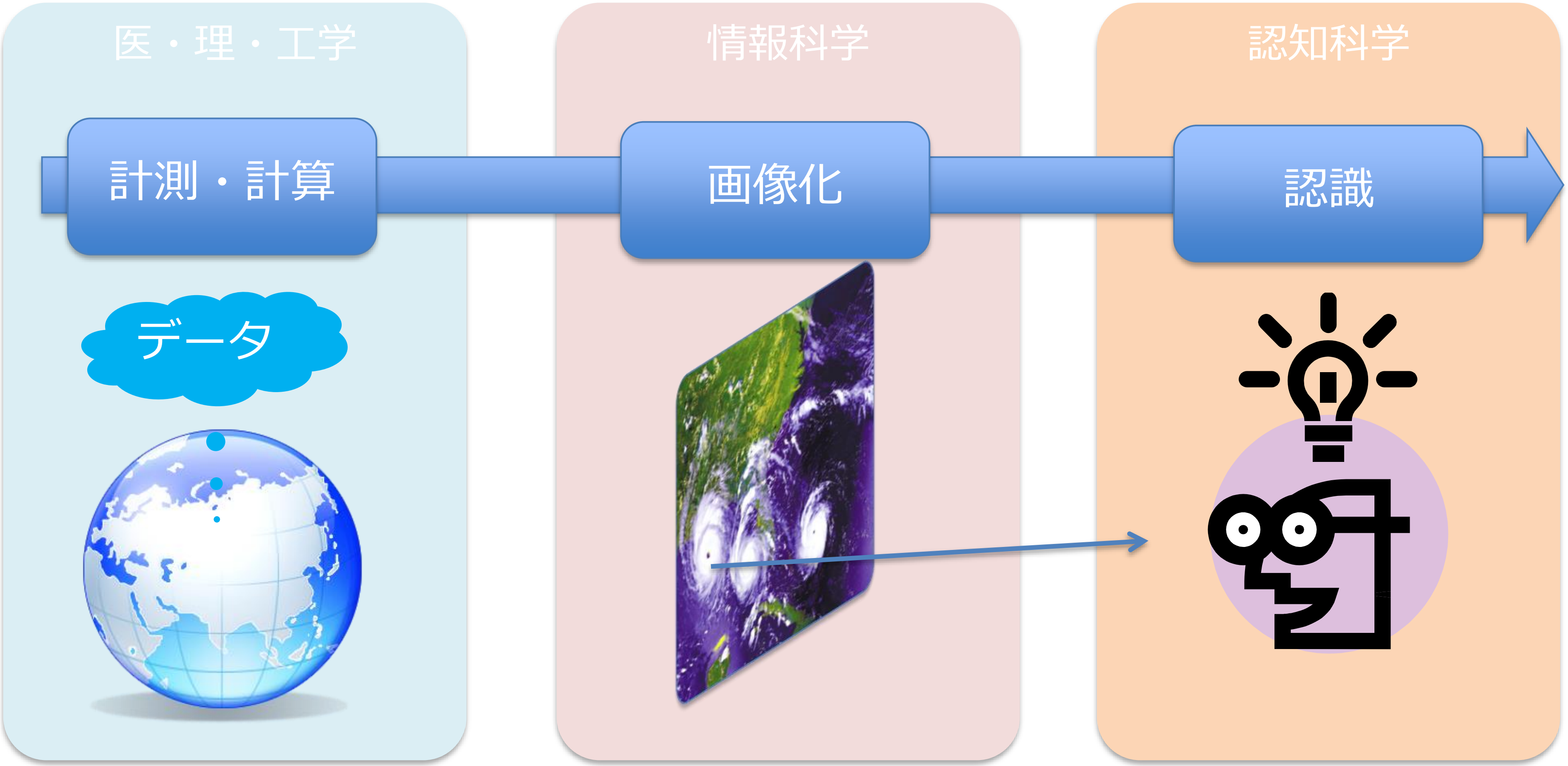
1.8 zeta bytes in 2011

データと情報

可視化：データと脳をつなぐ



可視化研究と性能評価



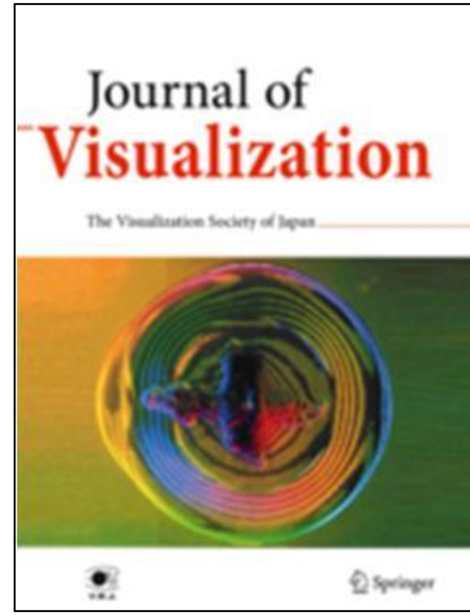
どんな現象をデータ化できたか？

どれほど効率よく画像化できたか？

どれほどの気付きを得たか、どんな行動変容に結び付いたか？

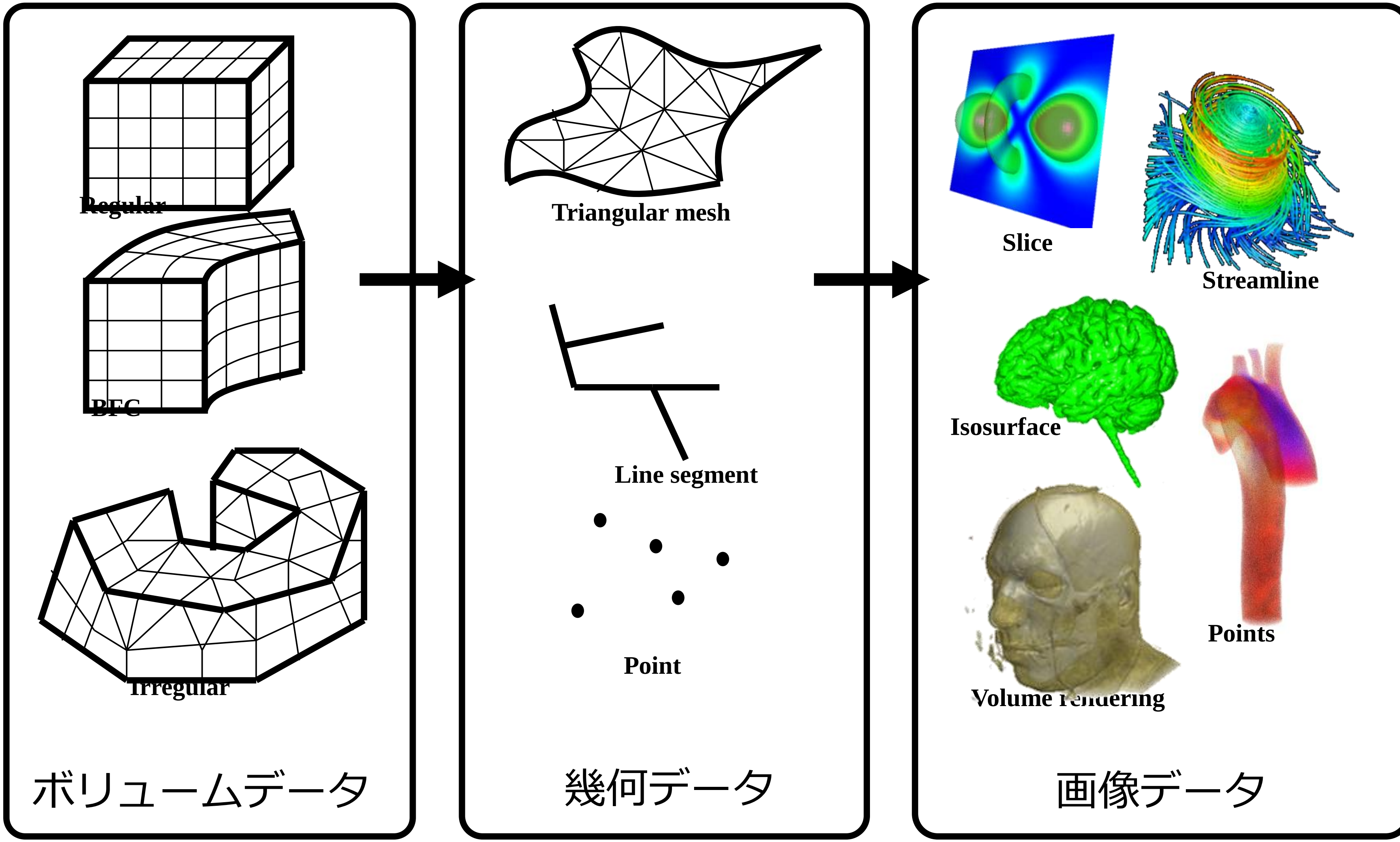
Journal of Visualization

- 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

データ可視化パイプライン



可視化：ビッグデータ時代の科学を拓く

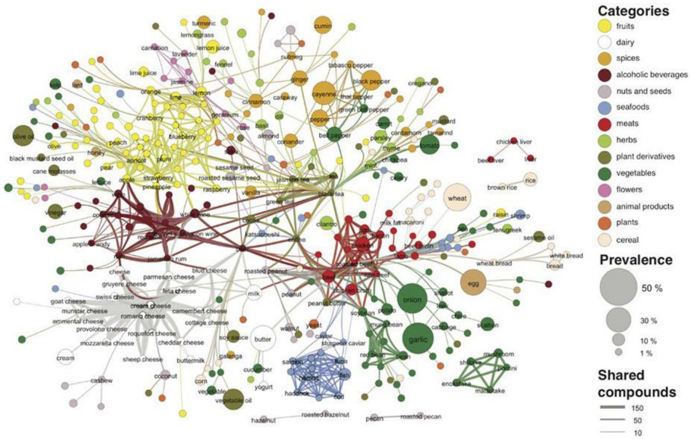
いろいろな可視化

計測データの可視化

流れの可視化
（データ提供：Gustavo R.S. Assi）

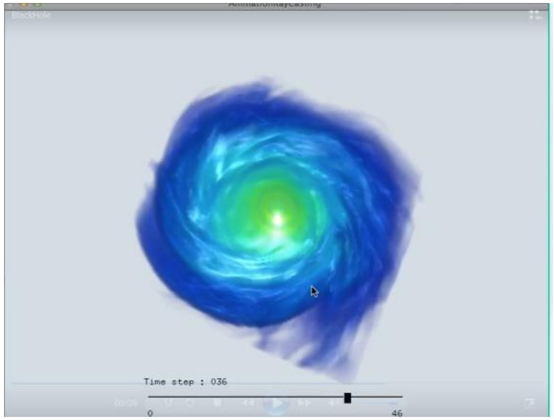
船鉾の可視化
（データ提供：田中先生@立命館大学）

医療データの可視化
（データ提供：www.oagpubco.com）

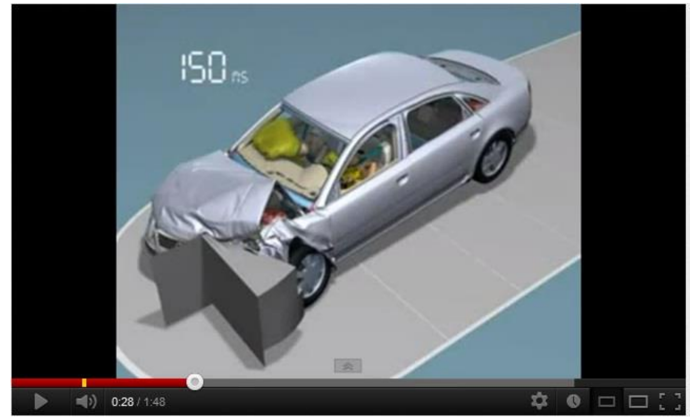


Yong-Yeol Ahn, et. al,
Flavor network and the principles of food pairing, Nature

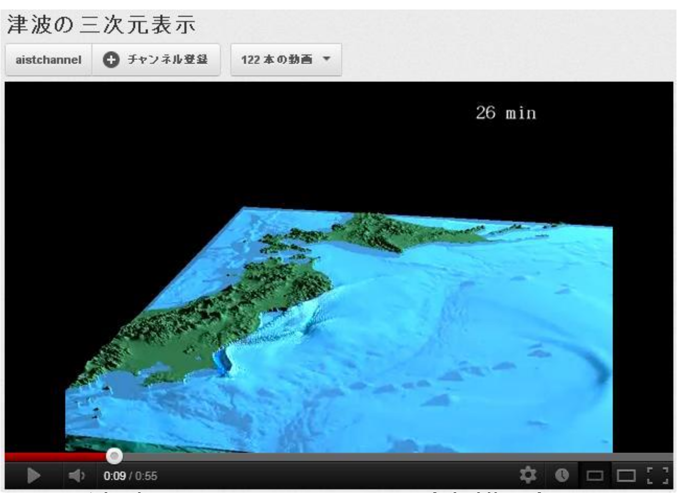
シミュレーション結果の可視化



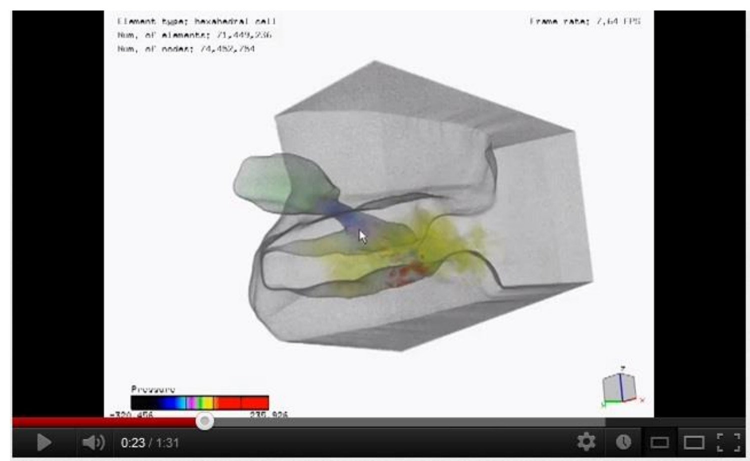
ブラックホールシミュレーション
（木内先生@京都大学）



自動車衝突解析(LOCTITE sitio Web www.loctite.com)

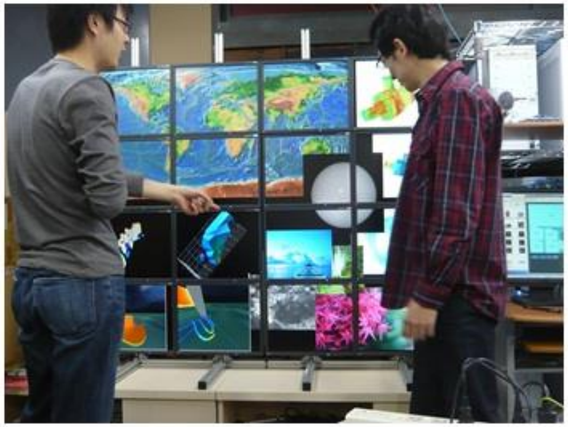


津波シミュレーション（産総研）



大規模口腔流体解析結果（野崎先生@大阪大学）

大型表示装置を使った可視化



複数アプリケーションの表示例

星間物質の乱流シミュレーション結果
（村主先生@白眉プロジェクト）

立体映像提示（アナグリフ方式）
（キッズサイエンススクール）

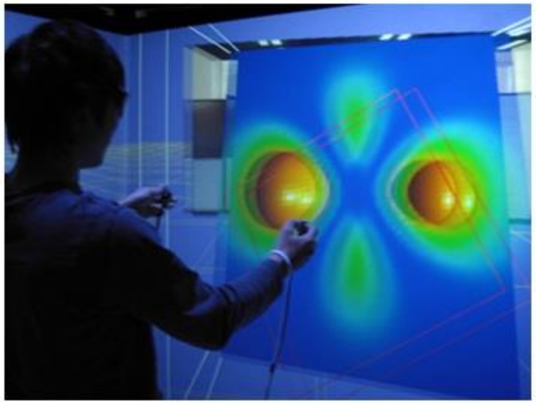
N体シミュレーション結果（暗黒物質）
（矢作先生@京大メディアセンター）

大規模ポンプの構造解析結果
（奥田先生@東京大学）

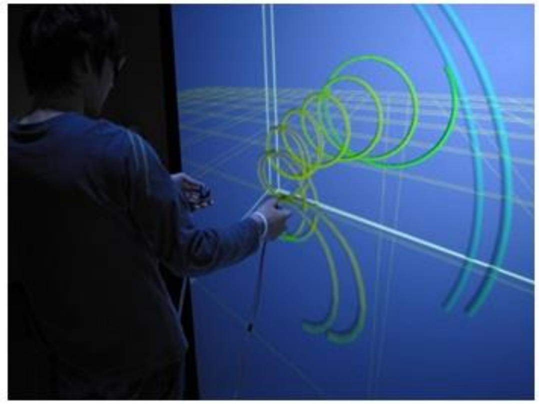
システム構成

PC x 11
NIC: 1Gbps
LCD: 40

没入表示装置を使った可視化

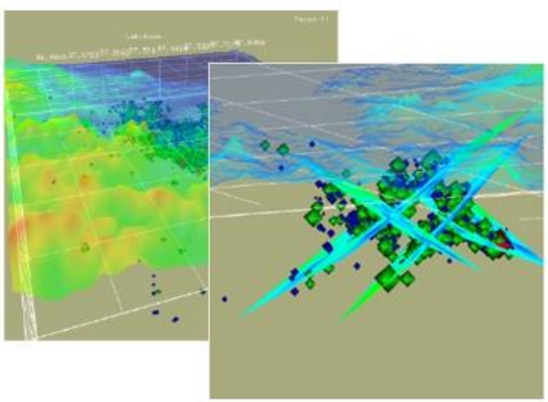


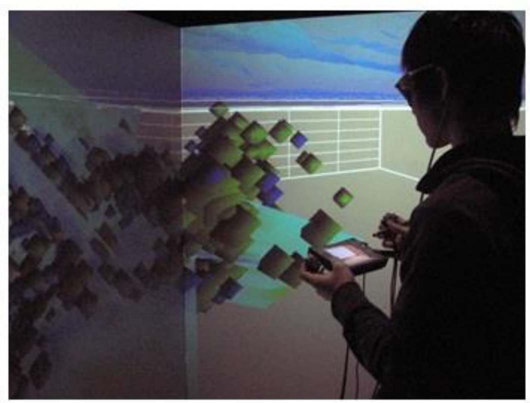
任意断面可視化



流線可視化

没入仮想空間提示





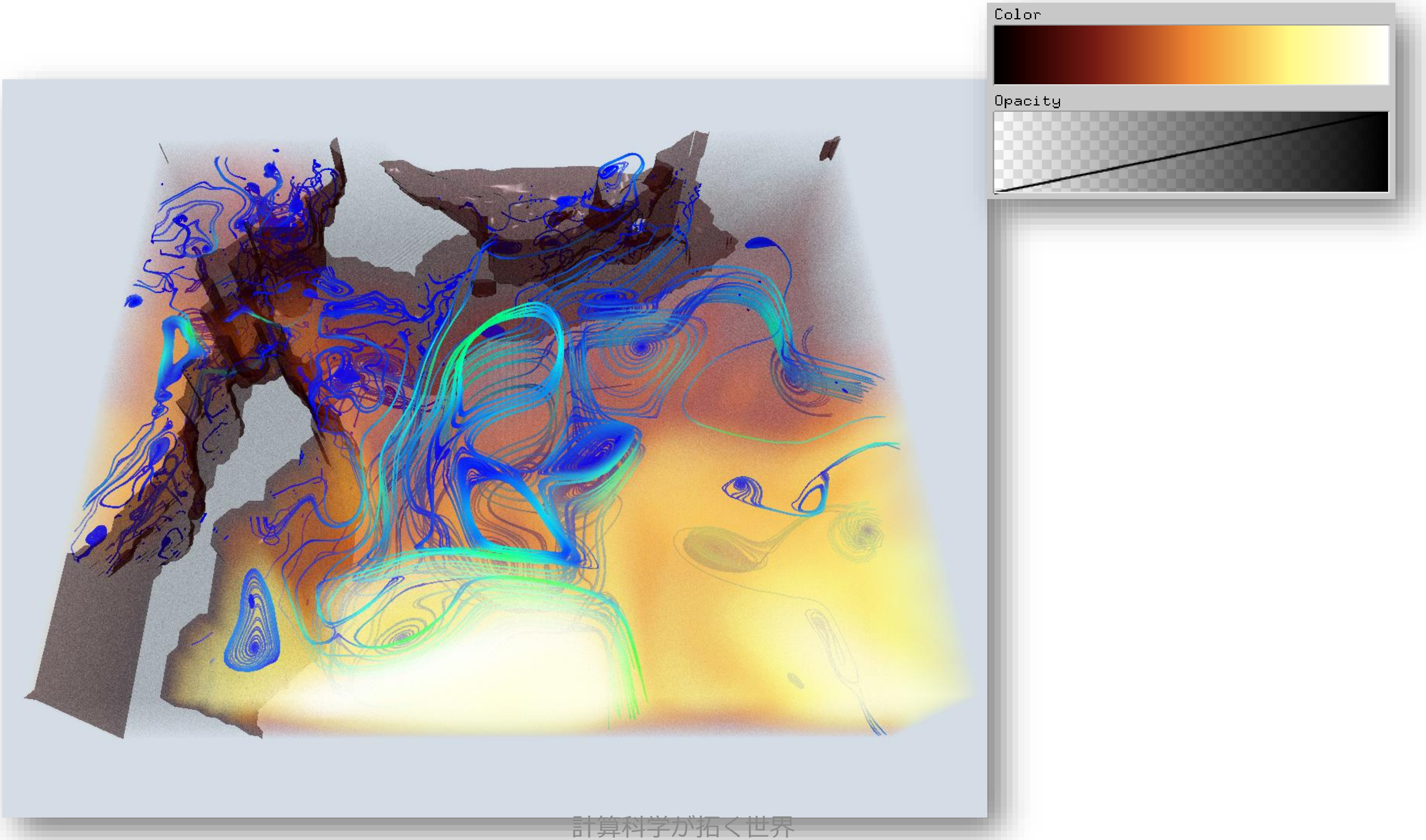
新潟中越地震の震源分布（断層面の推定）
（片尾先生@京大防災研）

- PC x 4
- 磁気装置
- 偏光メガネ
- 操作デバイス

システム構成

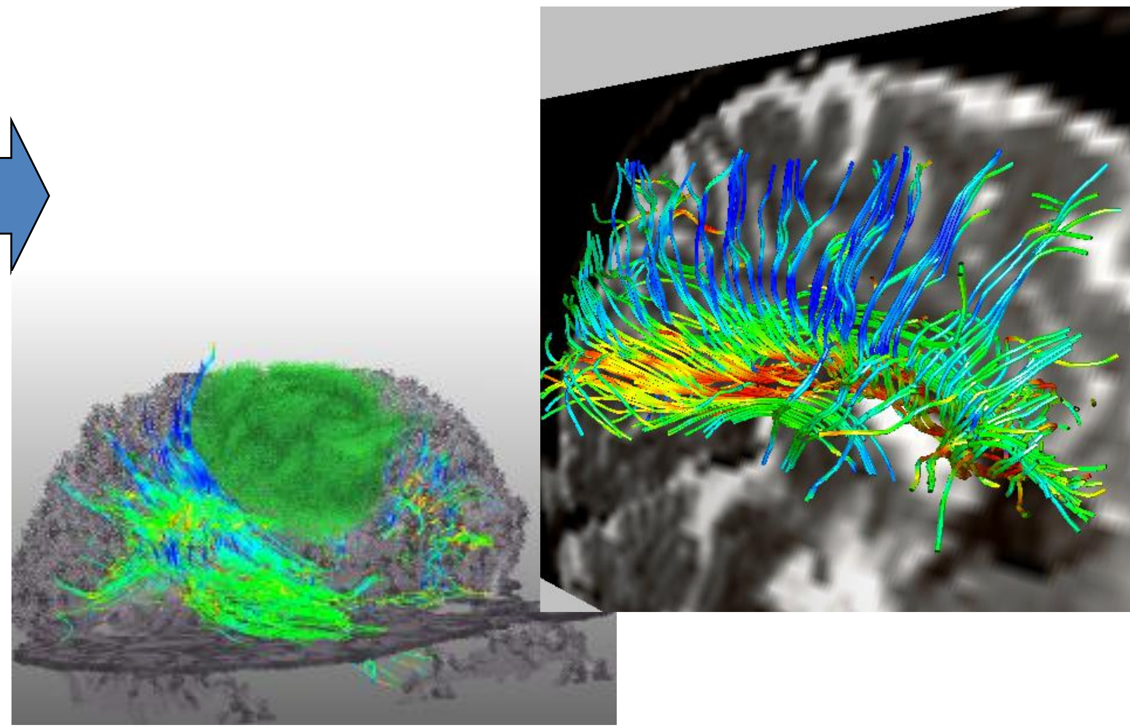
青森沖海流解析結果

（石川先生@京大）



脳白質神経線維可視化技術

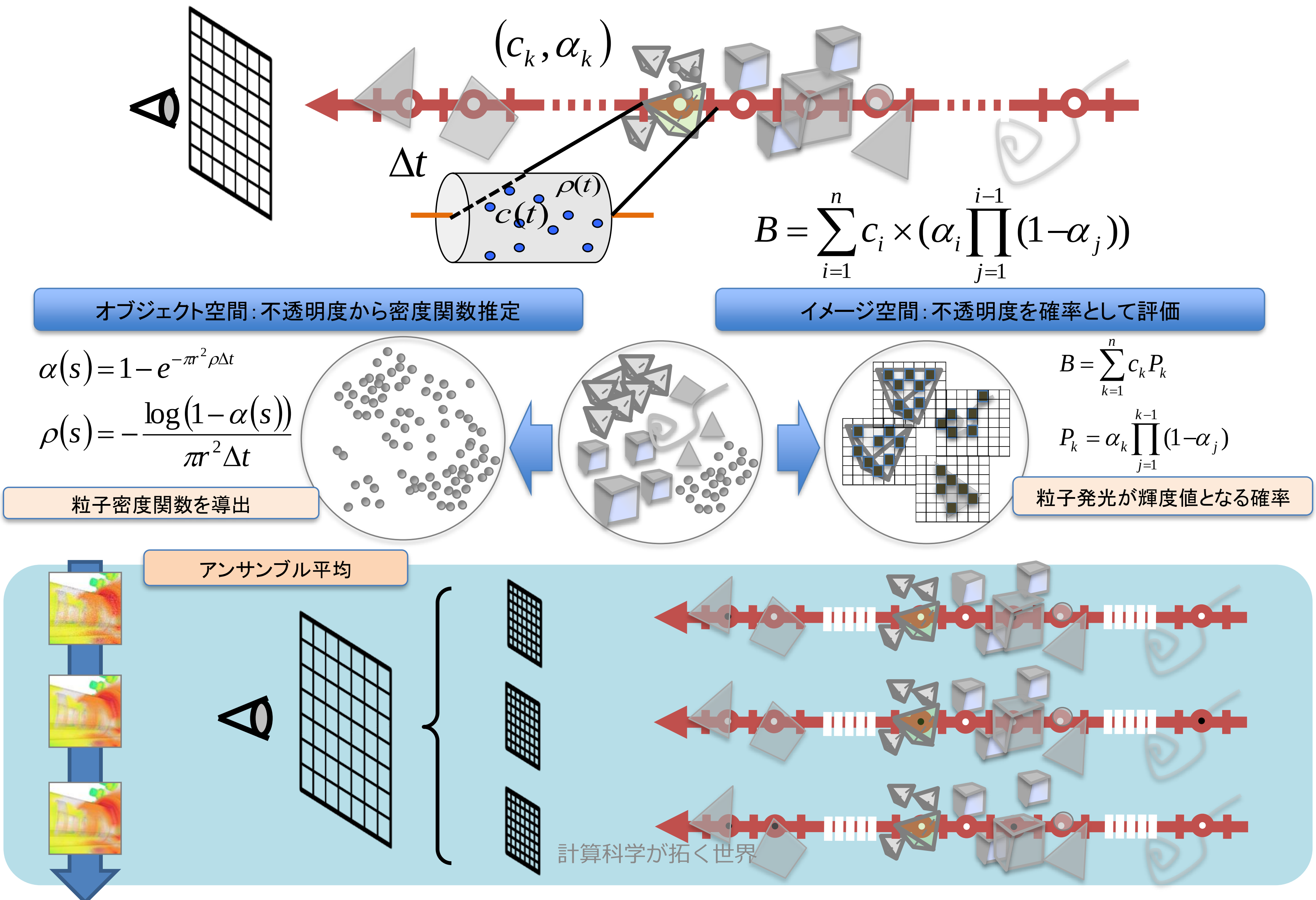
DT-MRIデータはテンソルデータの一種である．テンソルの最大固有値とその固有ベクトルから，神経線維の方向を計算し，流線可視化手法を適用することで，神経の流れを表現できる．



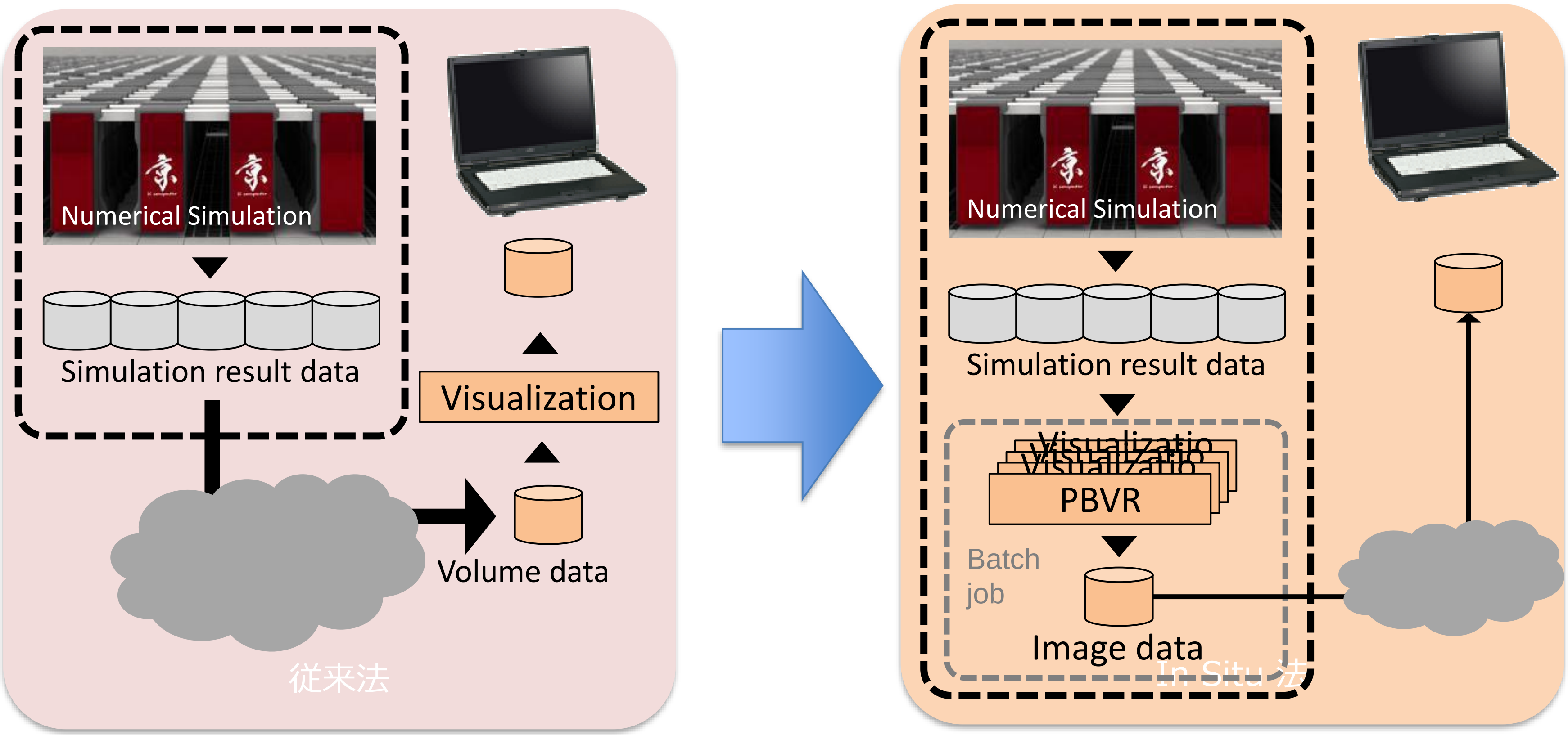
可視化：ビッグデータ時代の科学を拓く

可視化の拡張性

超拡張性の実現：粒子レンダリング



In-situ Visualization on K computer

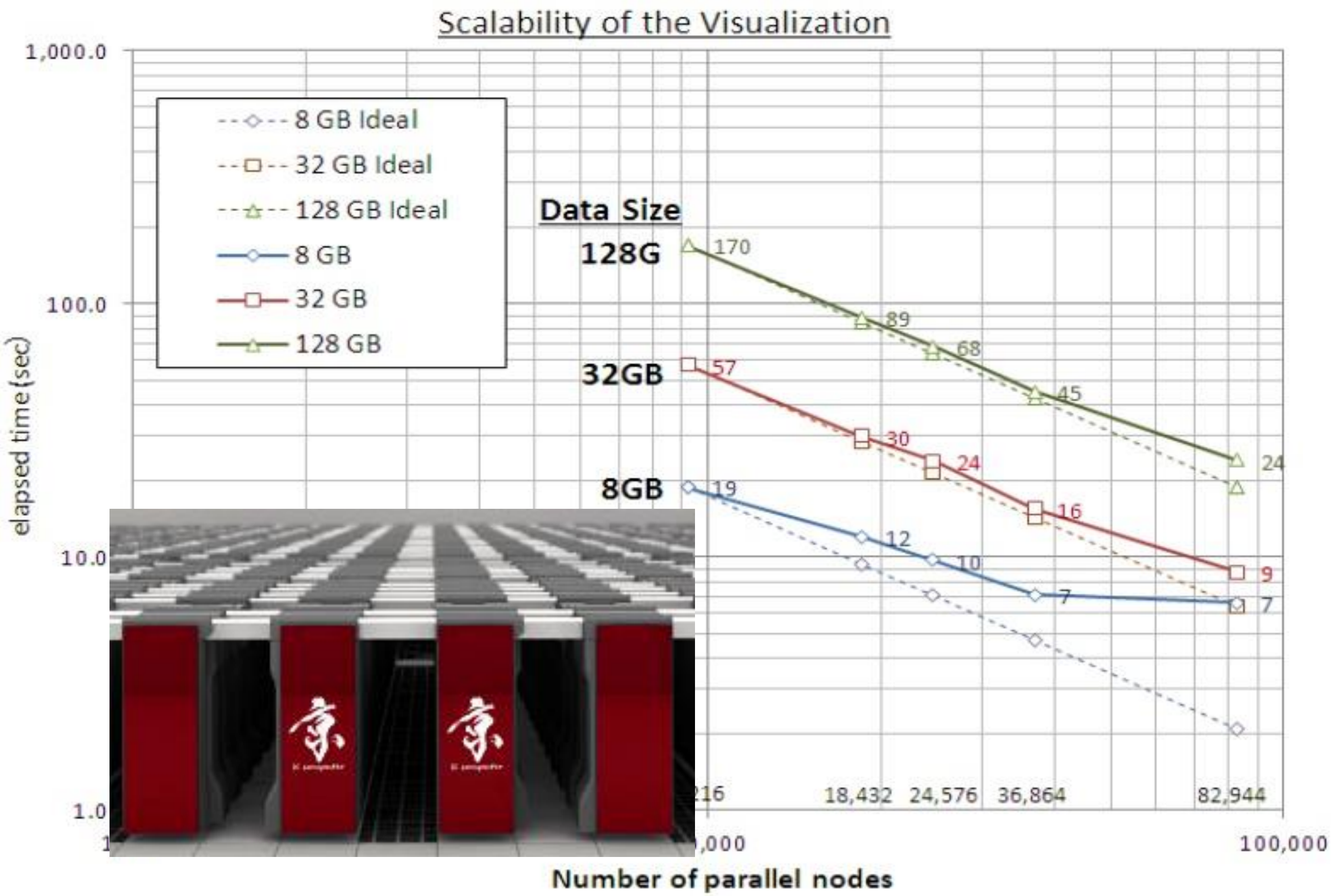


経過時間：1時間以上　＞　経過時間：10秒程度

Ogasa et al., "Visualization technology for the K computer", Fujitsu Scientific & Technical Journal, Vol.48, No.3, 2012.

Application

Visualization in K-computer system



Ogasa, A; Maesaka, H; Sakamoto, K; Otagiri, S, “Visualization Technology for the K computer,” Fujitsu Sci. Tech. Journal, Vol.38 No.4, 2012

タイルド表示装置
Tiled Display Wall (TDW)

可視化：ビッグデータ時代の科学を拓く

科学的可視化

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あります。多くの人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、物事の間の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、まぁこの環境を知る一種の科学論だなぁ。科学論てものはねぇ、とてももう面倒なものなんですよ。で、ただ、科学ってものはだなぁ、ものがその、ものをこの本当にものをし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、そこはねぇ、あのよおく考え貰わないといけないんだよ。つまりだなぁ、えー、科学ってものはいつでも法則をめがけているんです。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、科学はねぇ、その、規則をめがけるんです。ものと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に動いているかっと、いうその因果関係っていうものを目指しているんです・・・

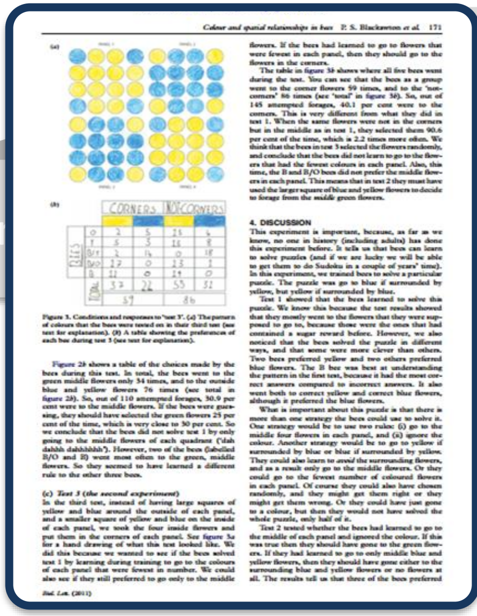
科学は因果関係を目指す

科学教育


biology letters

Animal behaviour

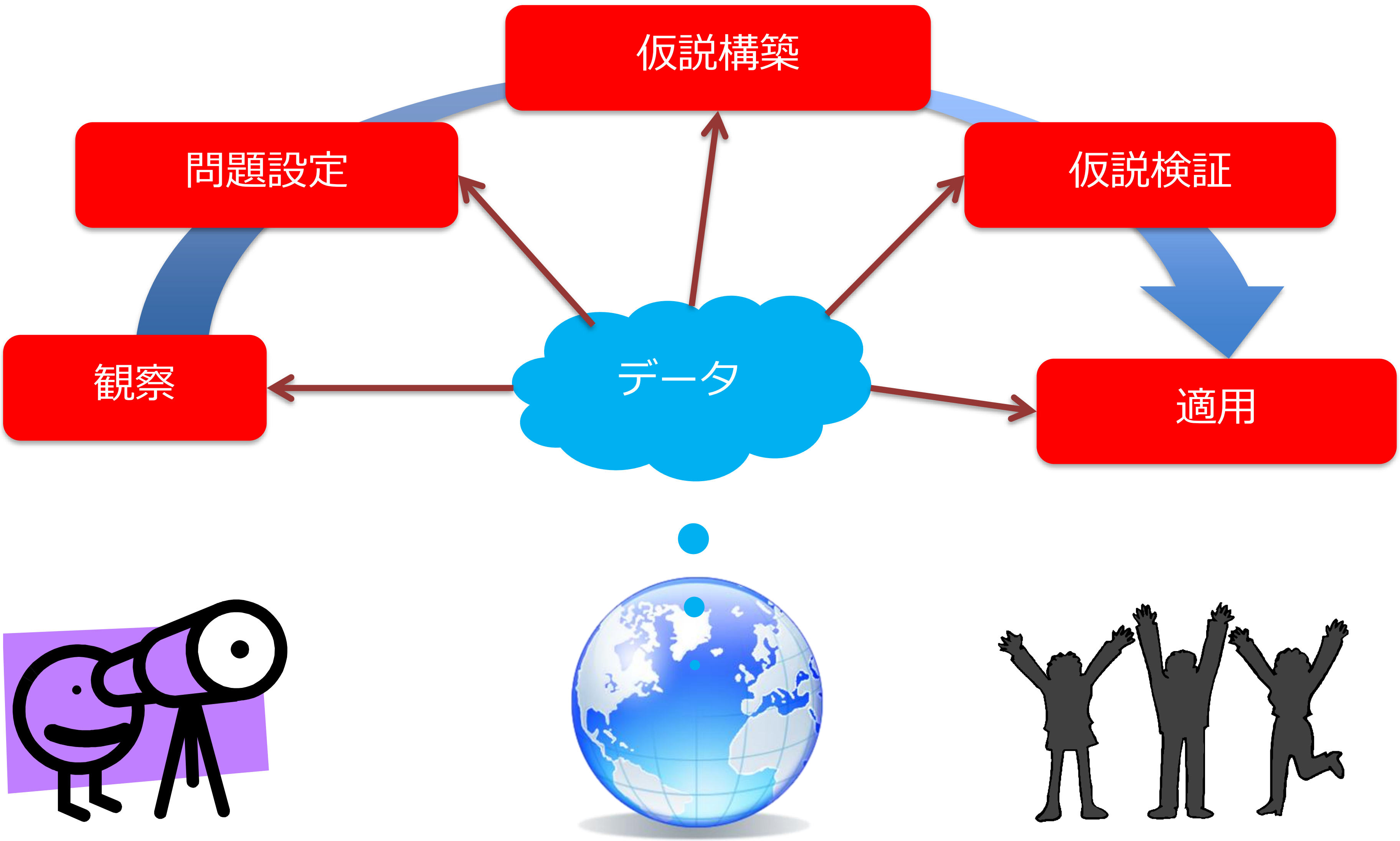
Blackawton bees



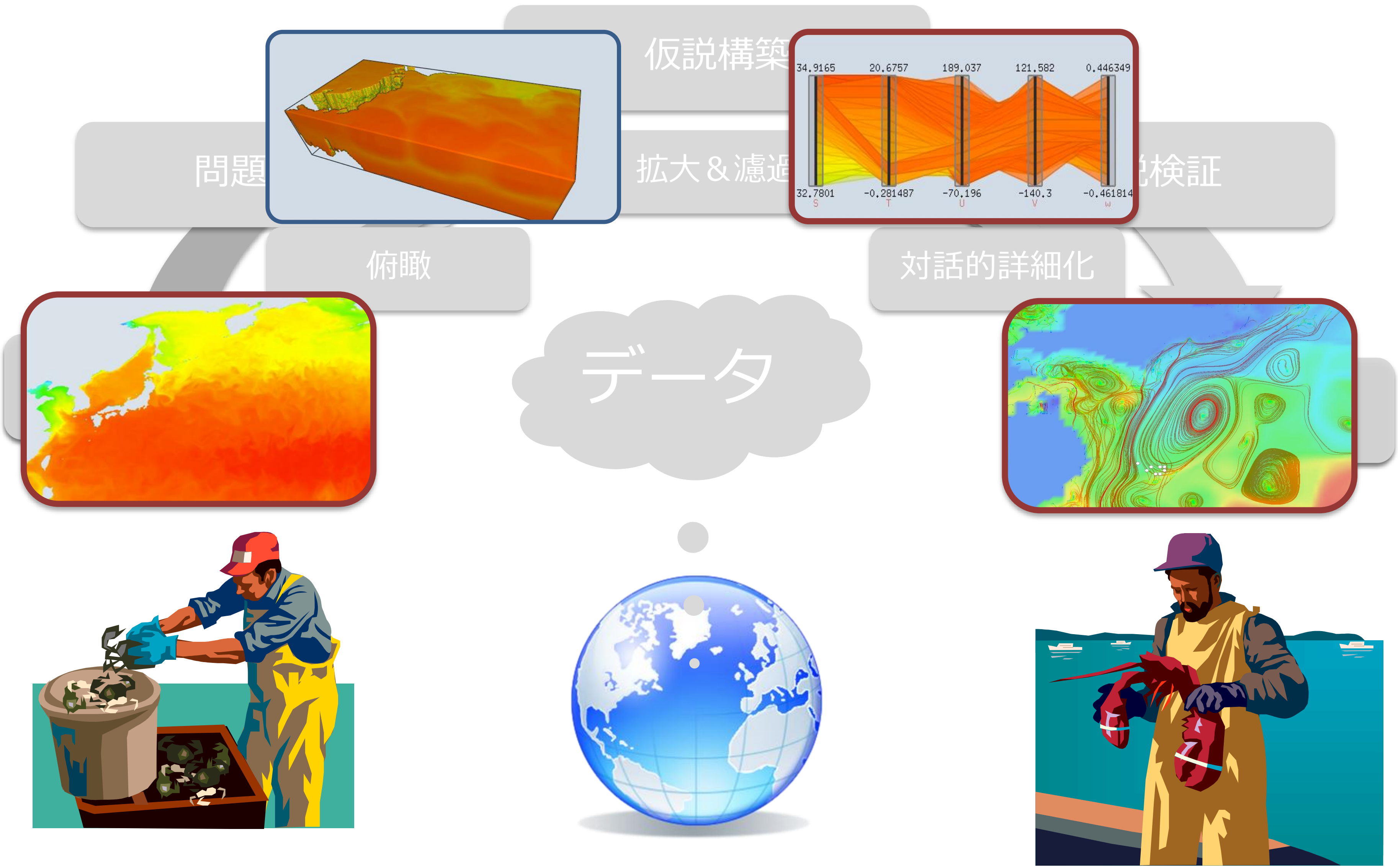
Elementary school students, "Blackawton bees," Biology Letters (2011) 7, 168–172

- USやEUでは, 小学校で科学的方法が教育されている
- 日本では, 大学や大学院で卒論や修論を通じて暗黙的に伝えられる
  - 日本では科目試験（正解がひとつ）に極端にこだわる
  - ほとんどの日本人学生は科学的方法をキチンと学ばない

科学的方法の流れ



海洋科学における可視化



可視化：ビッグデータ時代の科学を拓く

科学と因果発見技術

仮説

- ふたつの変数間のある方向をもった仮の関係
- その関係は、観測データによって検証される
- 「もしXならばY」のような因果関係によって記述される

因果関係発見は、科学における本質である

因果発見技術

M. Chen, et al., “From data analysis and visualization to causality discovery,” IEEE Computer , 44(10):84-87,

四つの挑戦

- 因果推論を促進するためのインフラ整備
- 可視化活用の能力の飛躍的向上
- 継続的なデータ洪水の対話的処理における負荷の軽減
- 推論過程における不確からしさの評価・可視化の対する要求への対応

INVISIBLE COMPUTING

From Data Analysis and Visualization to Causality Discovery

Min Chen, Anne Trefethen, René Bañares-Alcántara, Marina Jirotka, and Bob Coecke
University of Oxford, UK
Thomas Ertl and Albrecht Schmidt
University of Stuttgart, Germany

As data becomes invisible, emerging technologies can help human analysts and decision makers understand, model, and visualize causal relationships.

Nowadays the public, and even most decision makers, rarely see raw data. People are used to statistical results and visualized data presented by scientists, data analysts, or news readers. As computers disappear into the background, raw data is also becoming invisible.

The reliance on data analysis and visualization to improve the usability of information and communications technology echoes Mark Weiser's vision of ubiquitous computing: "... only when things disappear in this way are we freed to use them without thinking and so to focus beyond

CAUSALITY REASONING

Causality is the fabric of our dynamic world. We all frequently attempt to reason about the causes of everyday events—for example, why do I have a headache, or what has upset Alice?—in order to better manage them. We also seek to understand the origins of numerous complex phenomena such as social divisions, economic crises, global warming, and terrorism. Some of the greatest scientific discoveries, from Newton's laws of motion to Darwin's theory of natural selection, involve causality.

Causality has been studied in

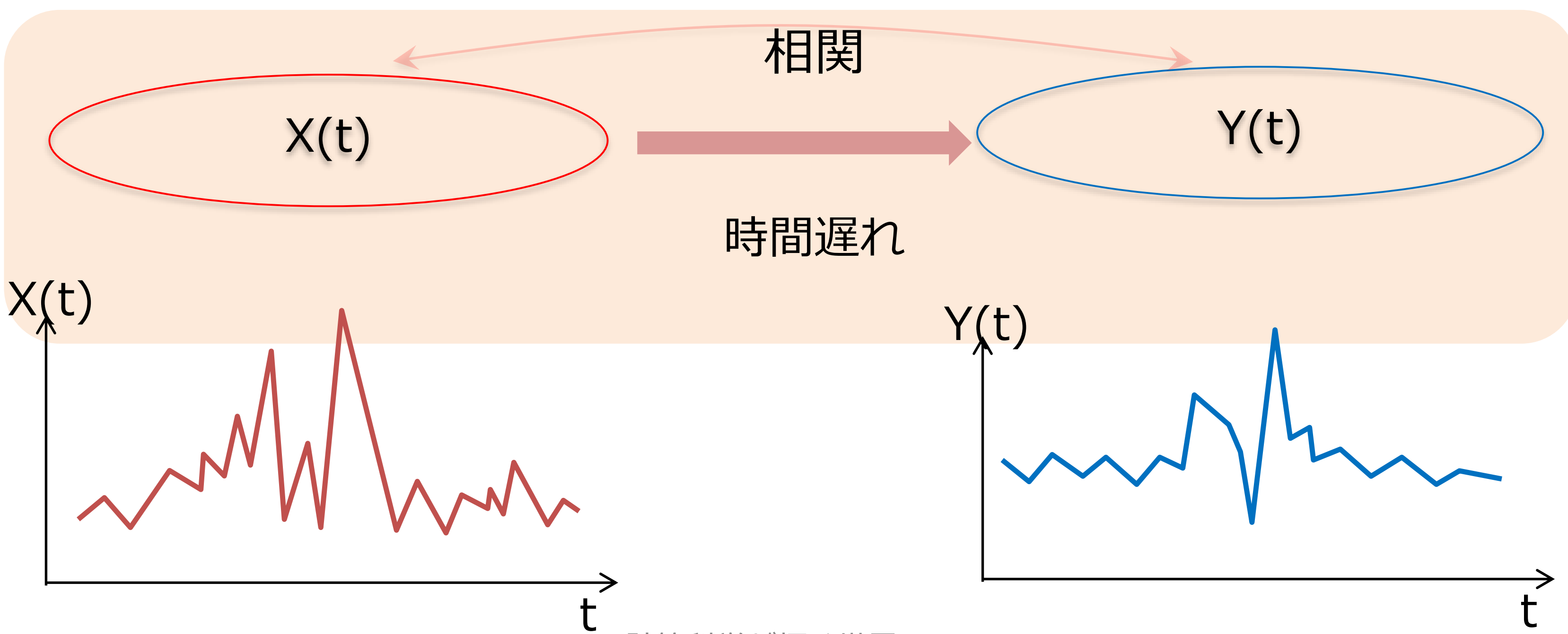
(*Natural Philosophy of Cause and Chance*, 1949).

Figure 1 illustrates Born's four levels of causality reasoning and the transitions between them. *Probabilistic causation* is a form of preliminary reasoning based on statistical correlation. "Overwhelming evidence" abstracts probabilistic causation to *logical causation*. When there is a scientific understanding of causality, *quantitative laws* of nature describe the functional relationship between measureable attributes of various events. According to Born, quantum mechanics offers the *fundamental understanding* of causation in physics.

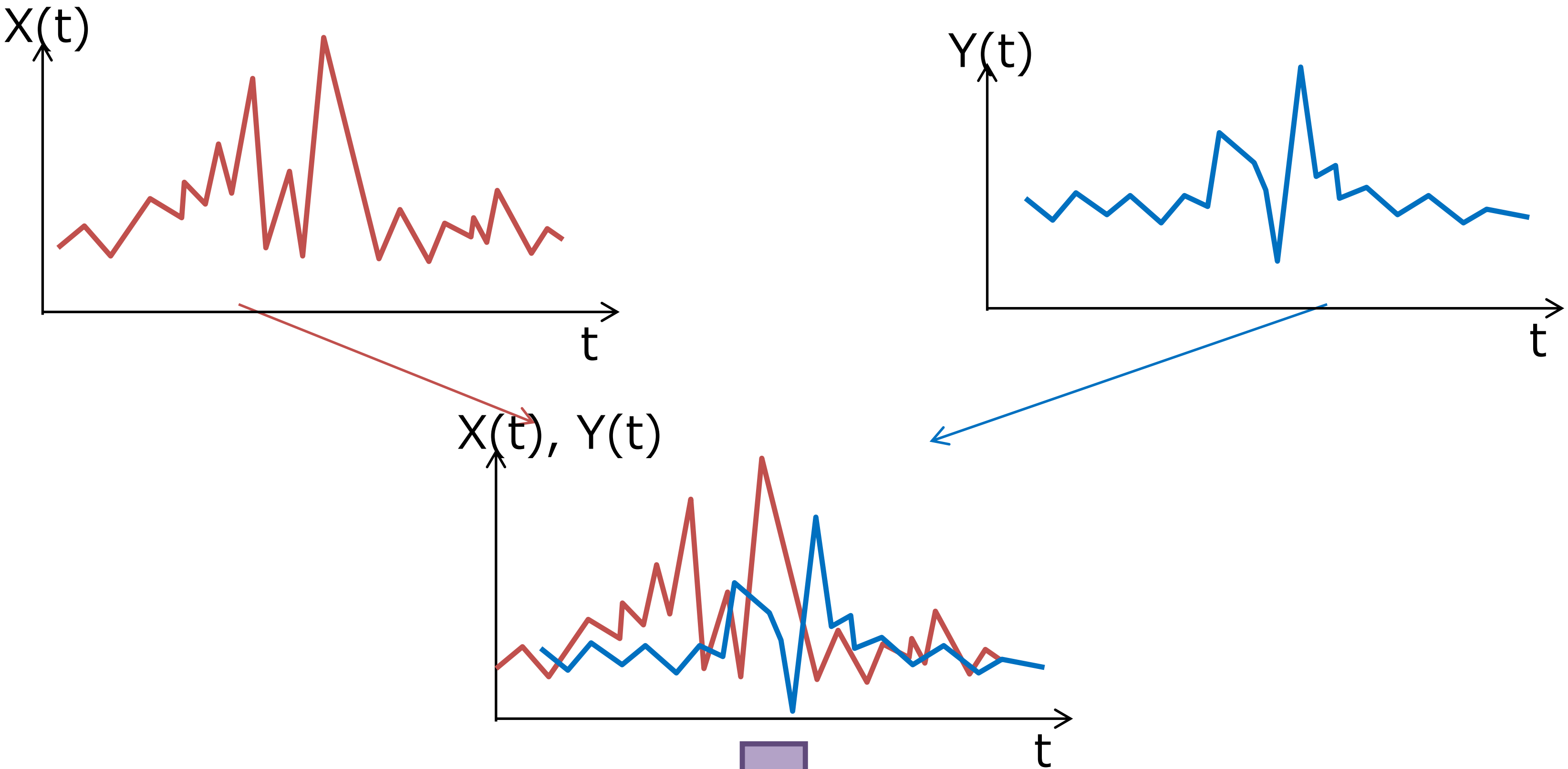
よりよい理解に向けて

時系列データ向け因果発見

- 因果関係は、あるプロセスともう一つのプロセスとの間の関係を表す
- 最初のプロセスは、二番目のプロセスに対して、部分的に関与する。
- 統一された定義はないものの、さまざまな因果発見技術が報告され、実際に活用されている。



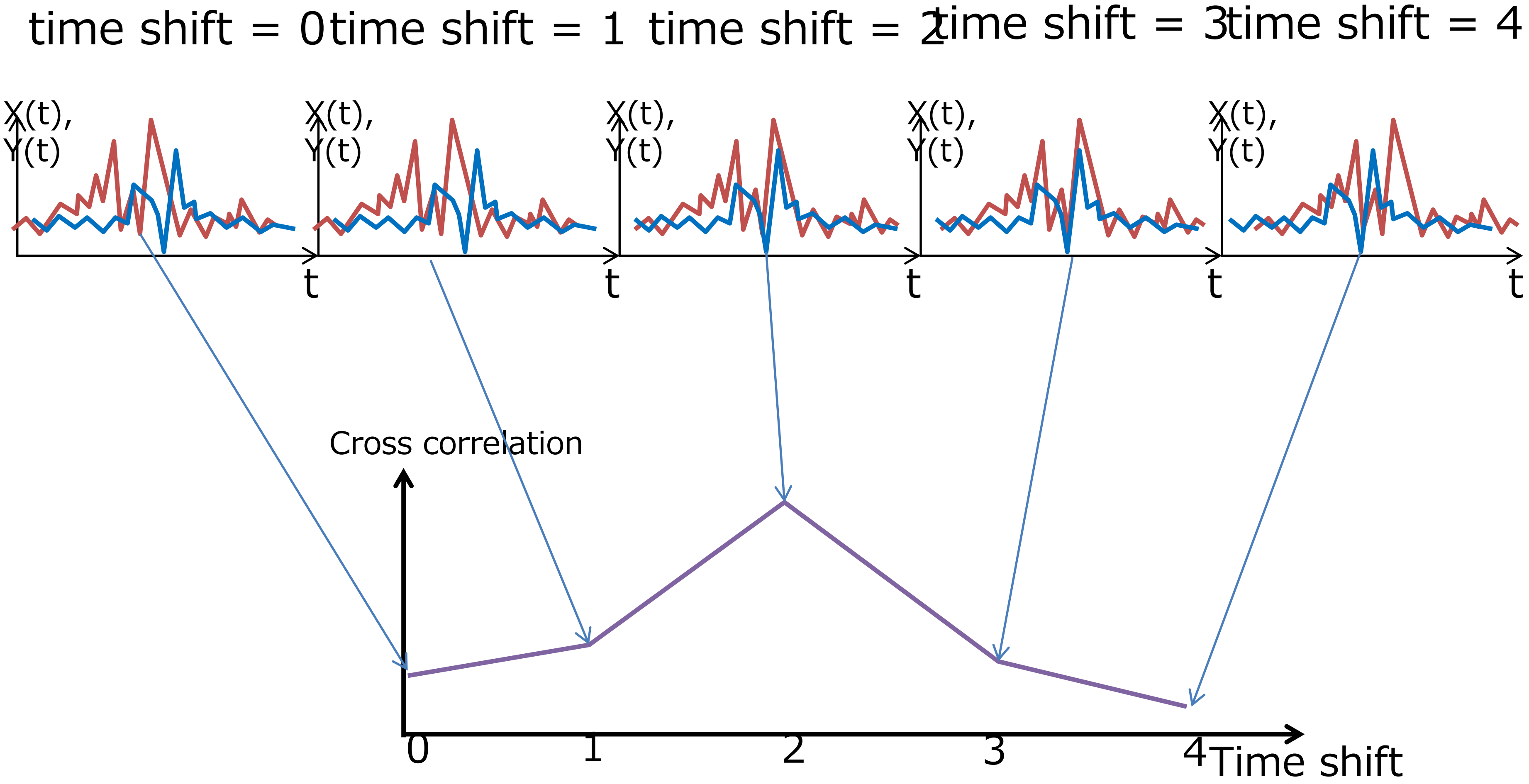
因果発見のための統計的検証法



- 相互相関
- グレンジャー因果
- 収束的交差写像

因果確率

相互相関


time shift = 0
time shift = 1
time shift = 2
time shift = 3
time shift = 4
X(t),
Y(t)
t
Cross correlation
0
1
2
3
4
Time shift

グレンジャー因果(GC： Granger Causality)

Granger, C. W. J. "Investigating Causal Relations by Econometric Models and Cross-spectral Methods". Econometrica 37 (3): 424–438 (1969)

It is a statistical test for a cause-and-effect relationship between two time series variables

We perform two vector auto-regressions as follows:

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

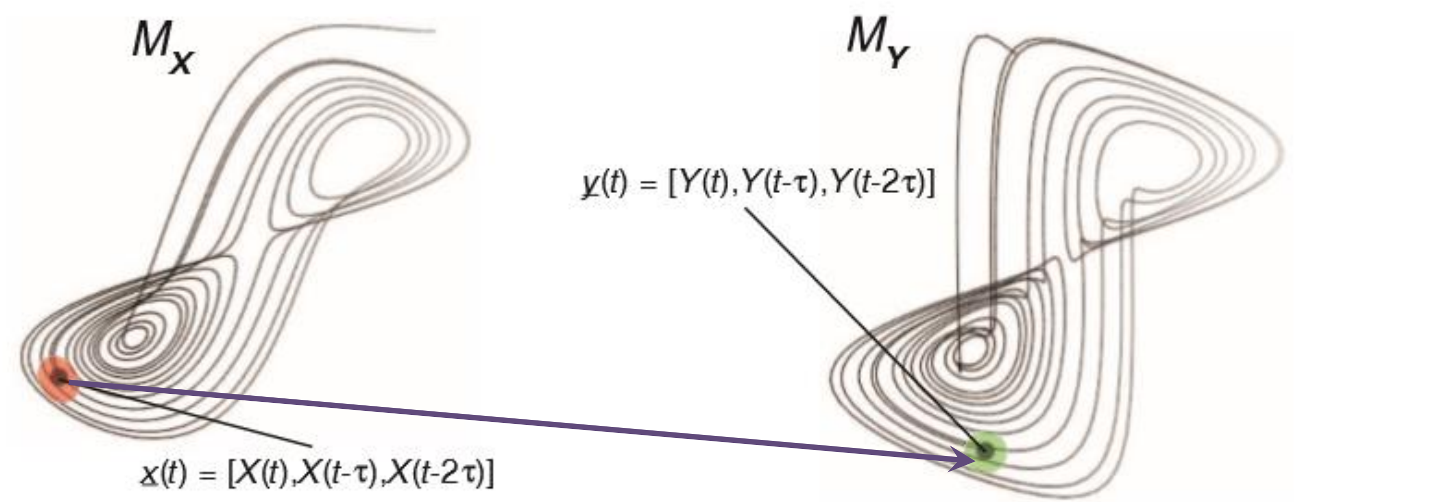
$$Y(t) = \sum_{l=1}^{L} a'_l Y(t-l) + \sum_{l=1}^{L} b'_l X(t-l) + \epsilon_2, \qquad (2)$$

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

収束的交差写像(CCM: Convergent Cross Mapping)

Sugihara, George; et al., "Detecting Causality in Complex Ecosystems," Science 338 (6106) pp. 496–500 (2012)

- It is a statistical test for a cause-and-effect relationship between two time series variables
- It constructs two shadow manifolds, M_X and M_Y using lagged-coordinate embedding X and Y, respectively
- If X forces Y uni-directionally, variable Y will contain information about X, but not vice versa.

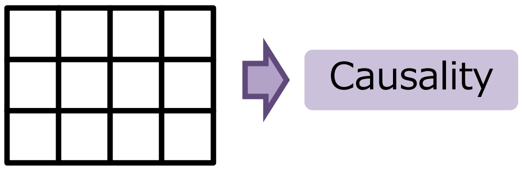


GC と CCM 使い分け

- どちらも相関は必ずしも因果を意味しないという問題を解決しようとする
- GCは、二つの変数の影響がお互いに独立で分離可能であるような確率的システムに適する
- CCMは、動的システム理論に基づき、二つの変数が相乗効果をもつようなシステムに適する

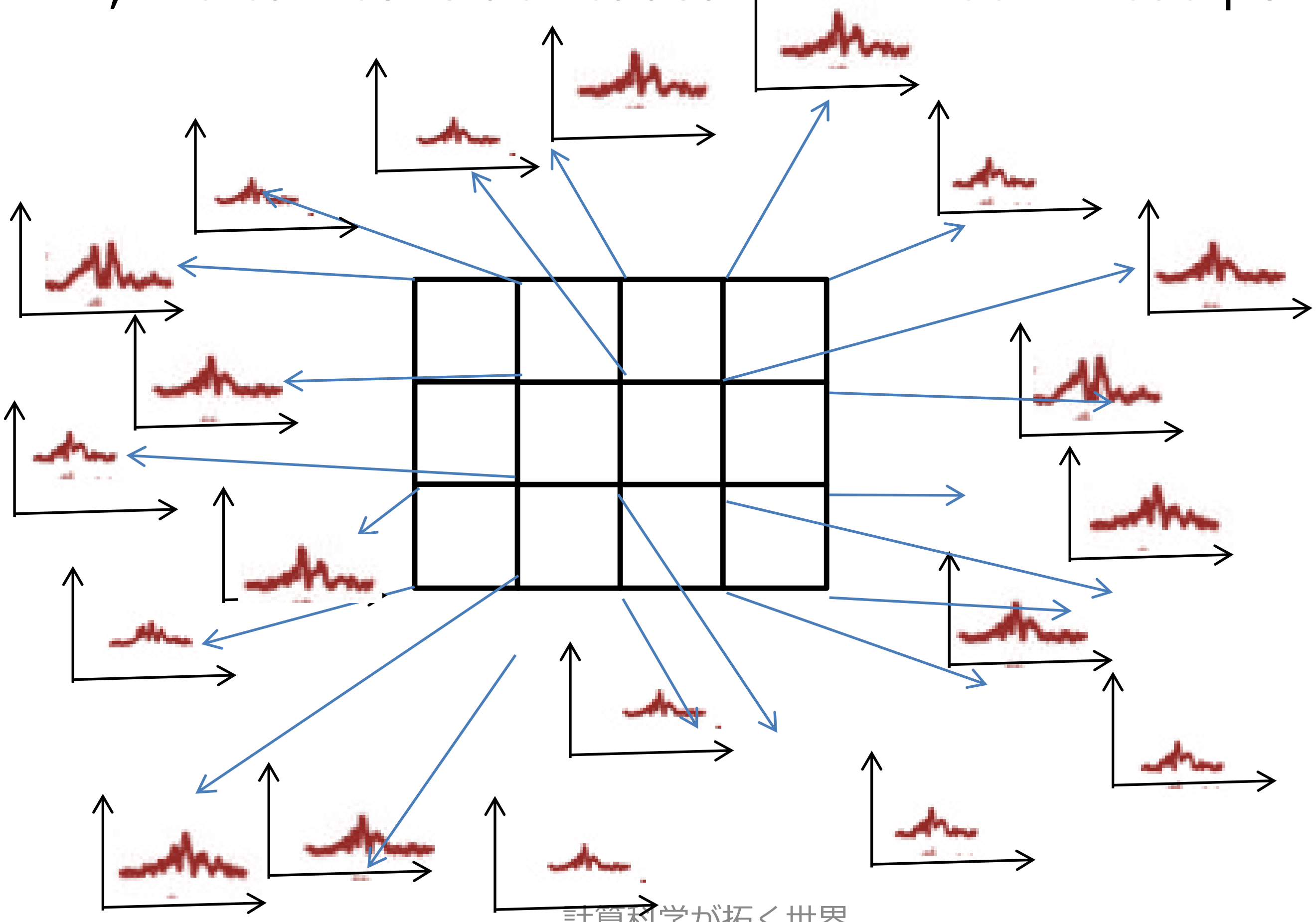
可視化：ビッグデータ時代の科学を拓く

時系列ボリュームデータにおける因果関係

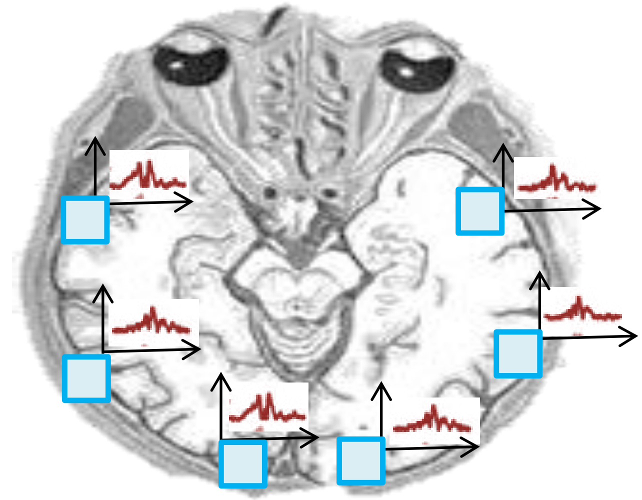


単一変数の時系列ボリュームデータ

- It defines univariate time series data defined at each grid point
- With it, we can consider causality between these points



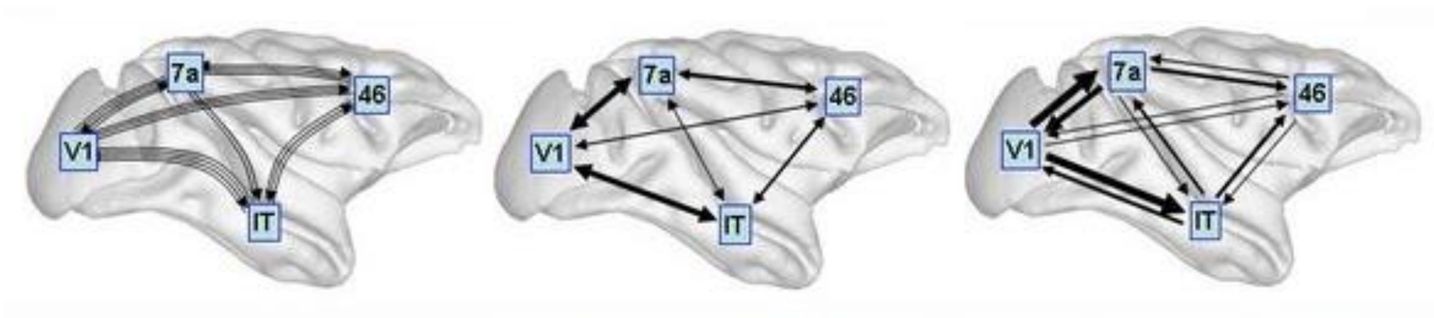
脳磁図(MEG: Magnetoencephalography)



- It records magnetic fields produced by electrical currents occurring naturally in the brain, using very sensitive magnetometers.
- Its applications include basic research into perceptual and cognitive brain processes, localizing regions affected by pathology before surgical removal, determining the function of various parts of the brain, and neurofeedback.

結合解析

structural connectivity　functional connectivity　effective connectivity

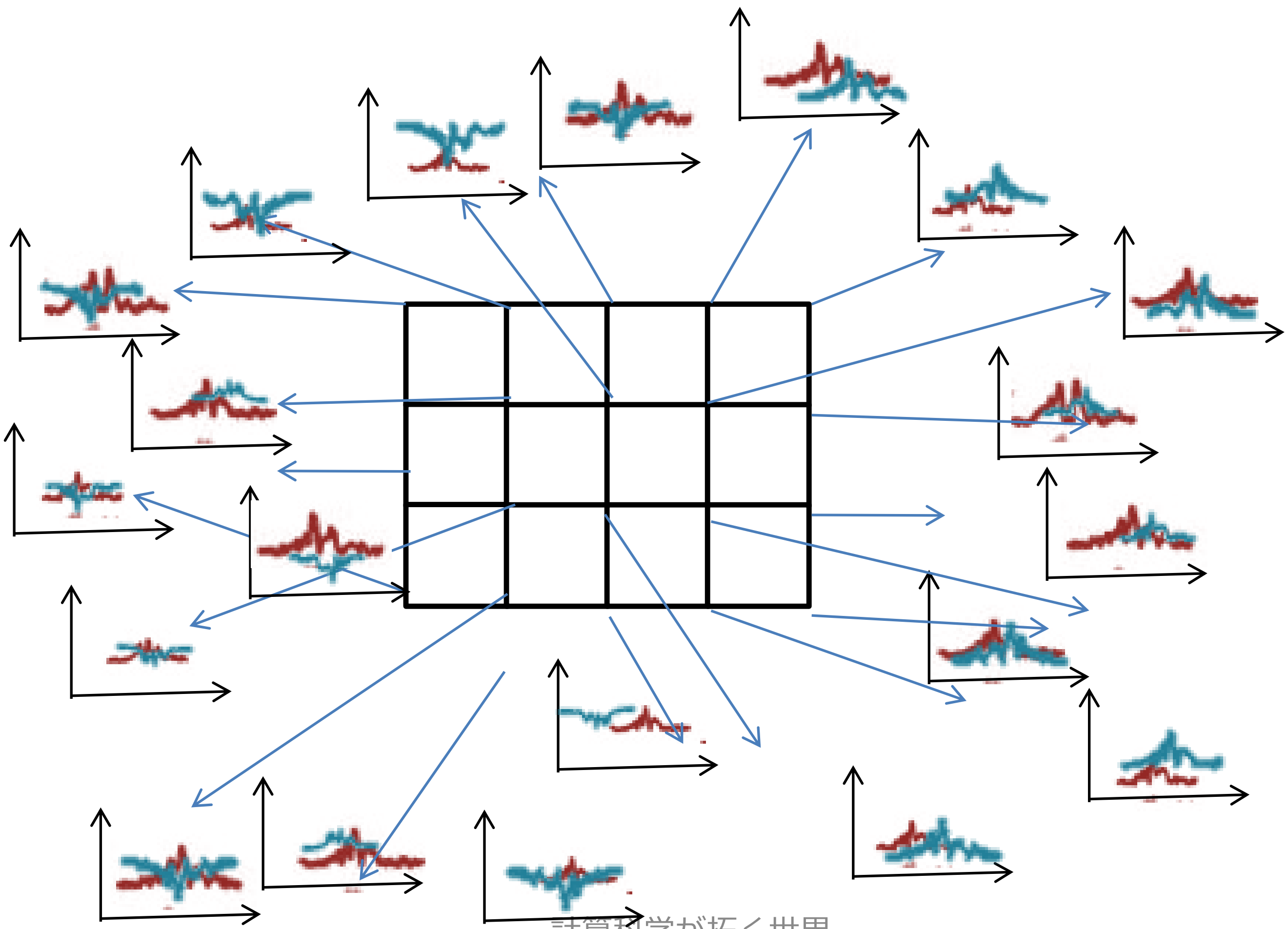


O. Sporns 2007, *Scholarpedia*

- Evaluation of structural connectivity (anatomical) using MR-DTI and Fiber-Tracking
- Confirmation of functional connectivity between activated areas using correlation
- Analysis of effective connectivity in neuronal groups using causation

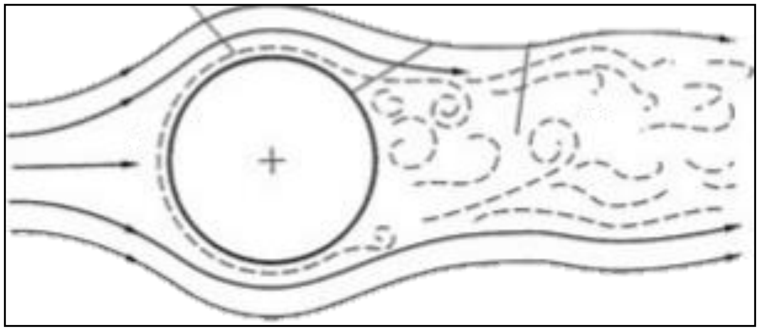
二変数の時系列ボリュームデータ

- It defines bivariate time series data defined at each grid point
- With it, we can consider causality at the point



洗濯効果

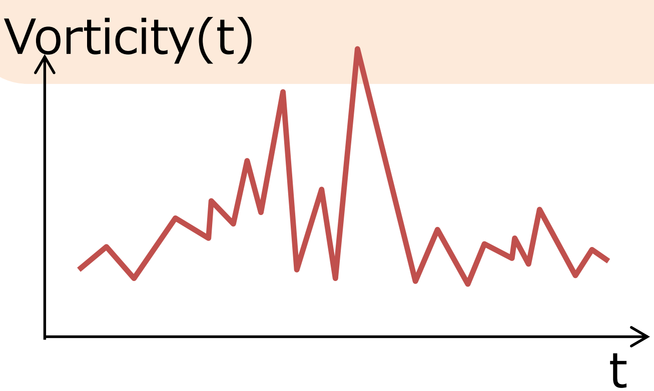
K. Suzuki, “UNSTEADY HEAT TRANSFER IN A CHANNEL OBSTRUCTED BY AN IMMERSED BODY”, 1994

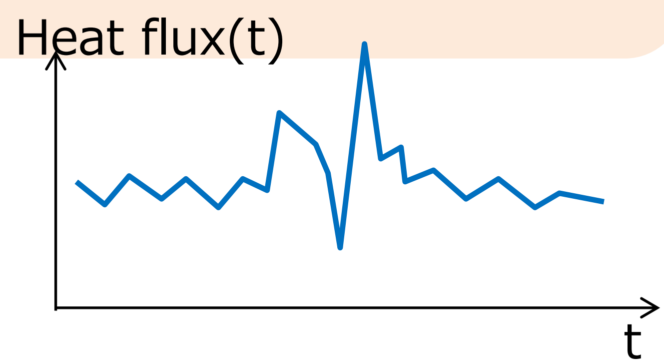


The vortex flow contributes to the enhancement of the heat transfer

Causality ?

Vorticity → Heat flux

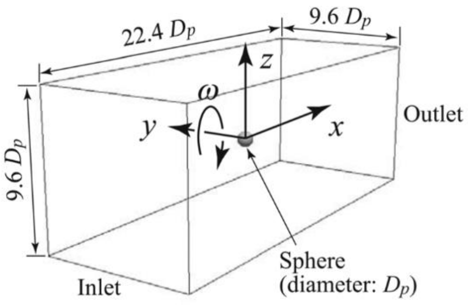


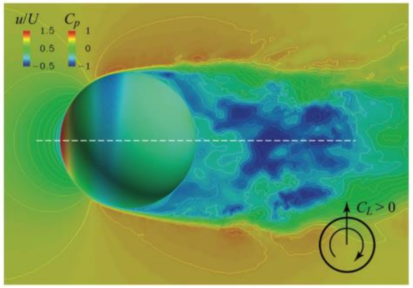


熱球周りの流体解析

Confirmation of the washing effect in the separated flow in the wake behind a sphere

- Model description
  - Number of nodes： 15,321,546
  - Number of cells： 18,917,887 (prism)
    27,545,304 (tet)
  - Number of time steps：198
  - File size：3.5GB per step

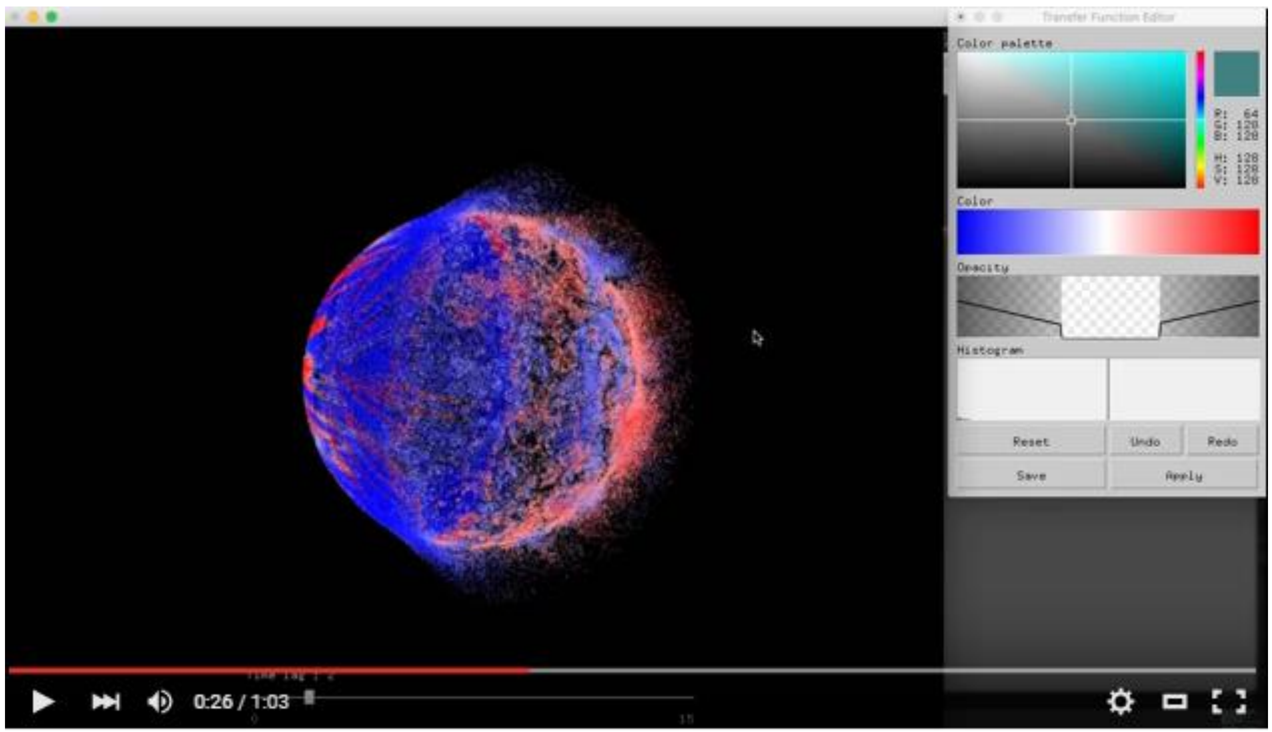




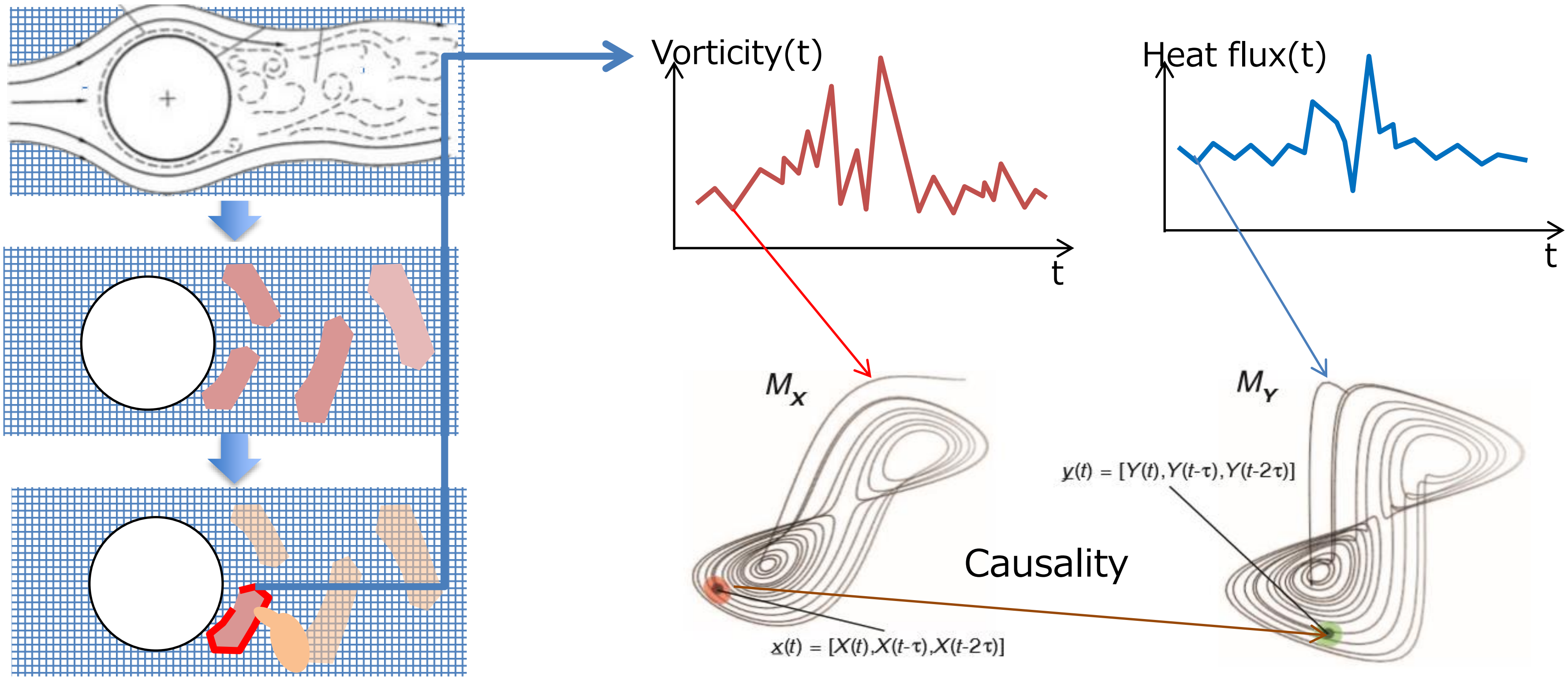
Negative Magnus lift on a rotating sphere at around the critical Reynolds number
Muto, Tsubokura, Oshima, Physics of Fluids, 24, 014102（2012）

因果関係ボリュームデータ

- Calculation of cross correlation at each grid point by changing the shift time
- Creation of a time-varying volume with 16 shift time steps



因果関係発見に向けて

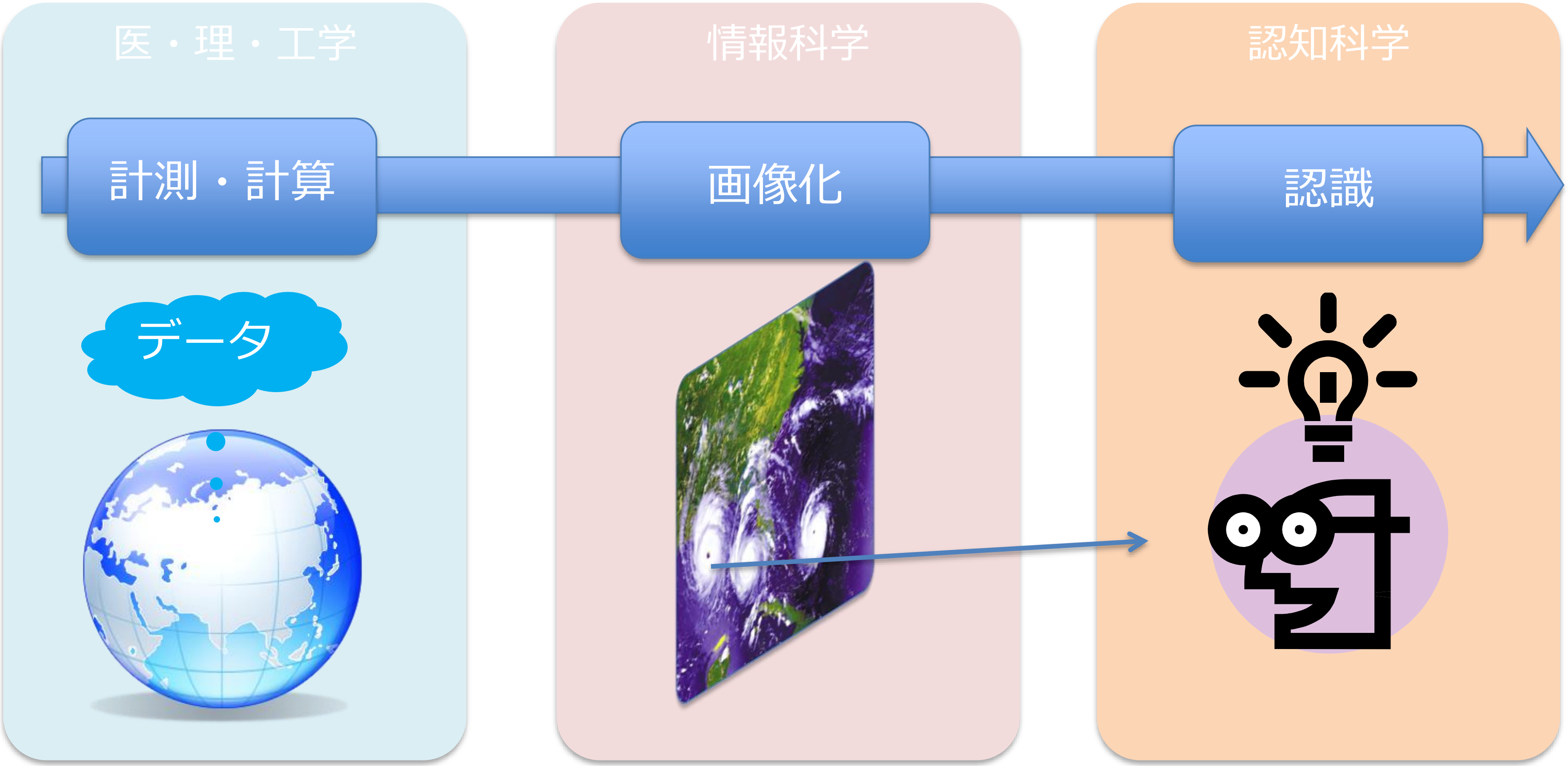


1. 融合可視化
2. 因果ボリュームデータの可視化
3. 興味領域の特定
4. 特定された領域での因果関係発見

可視化：ビッグデータ時代の科学を拓く

可視化の性能評価

可視化研究と性能評価



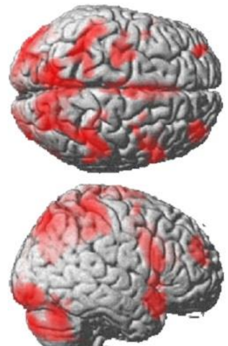
どんな現象をデータ化できたか？

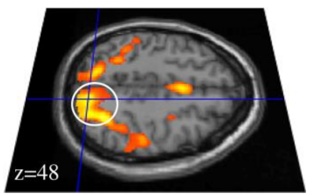
どれほど効率よく画像化できたか？

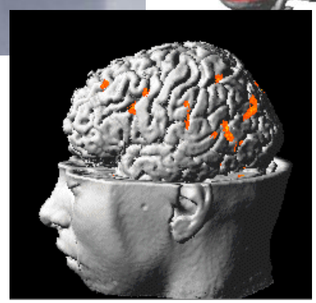
どれほどの気付きを得たか、どんな行動変容に結び付いたか？

脳機能計測装置による気付きの測定

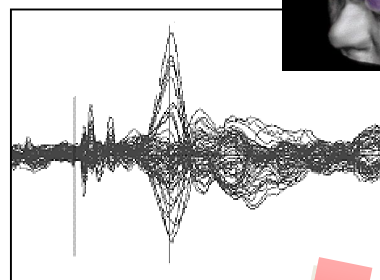
MRI

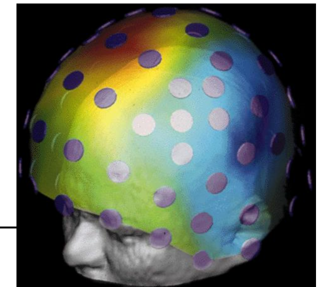






MEG

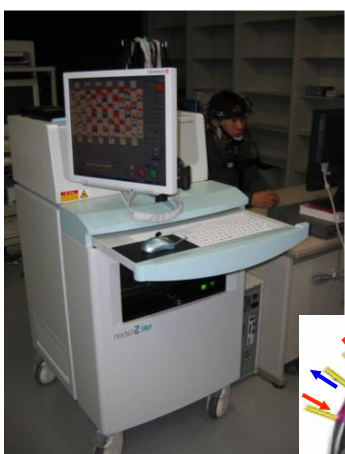


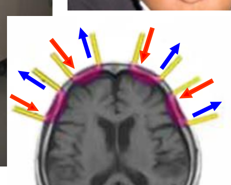


EEG

統合

時空間解像度



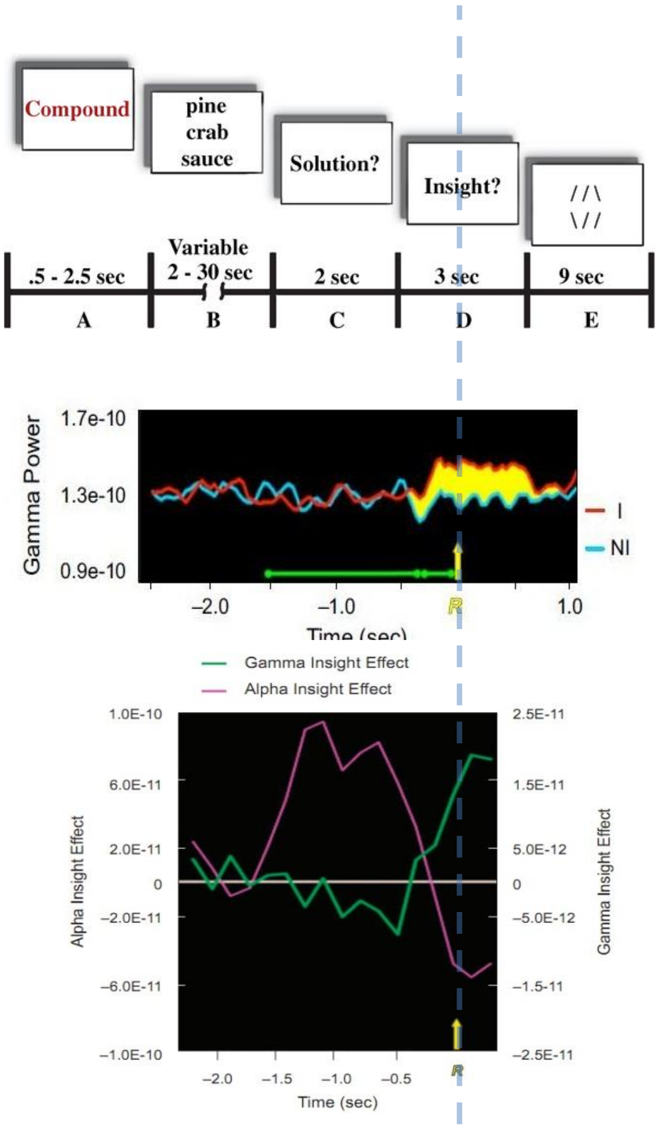


NIRS

視覚的気付きの探索

アハ体験についての研究

Kounios, J, Beeman, M, "The Aha! Moment-The Cognitive Neuroscience of Insight," 2009

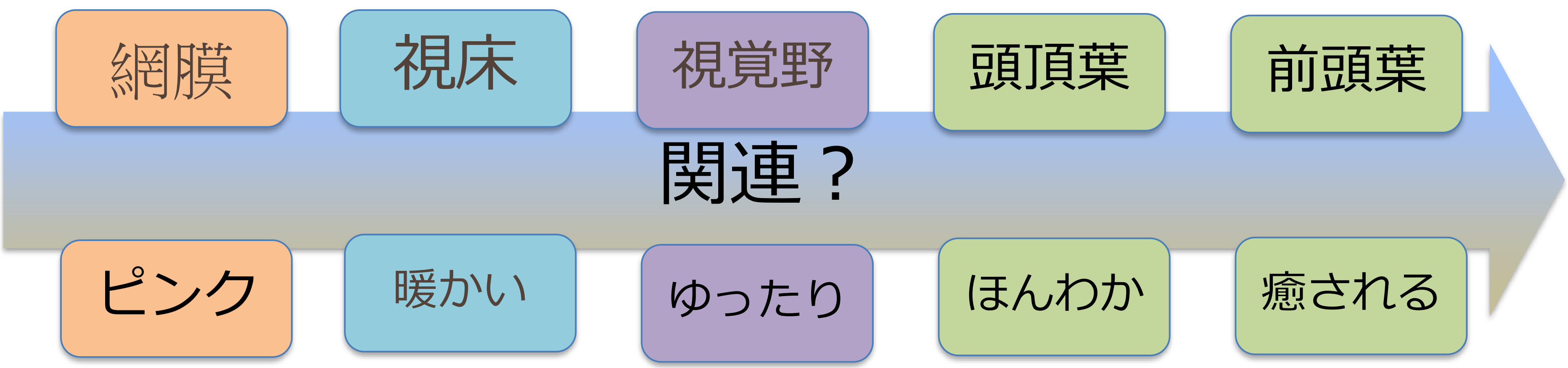


複合的遠隔性連想問題によるアハ体験

- ガンマ波 (約　40 ヘルツ) EEG 信号
- アルファ波 (約　10 ヘルツ) EEG 信号
- ひらめき時のガンマ波の組織的分布
- ひらめき時の機能的MRI分布

まとめ

- 科学的方法において重要な役割を果たす可視化研究
  - 粒子を用いた超拡張性可視化
  - 可視化による因果関係気付きの支援
  - 脳科学と連携する認知構造可視化



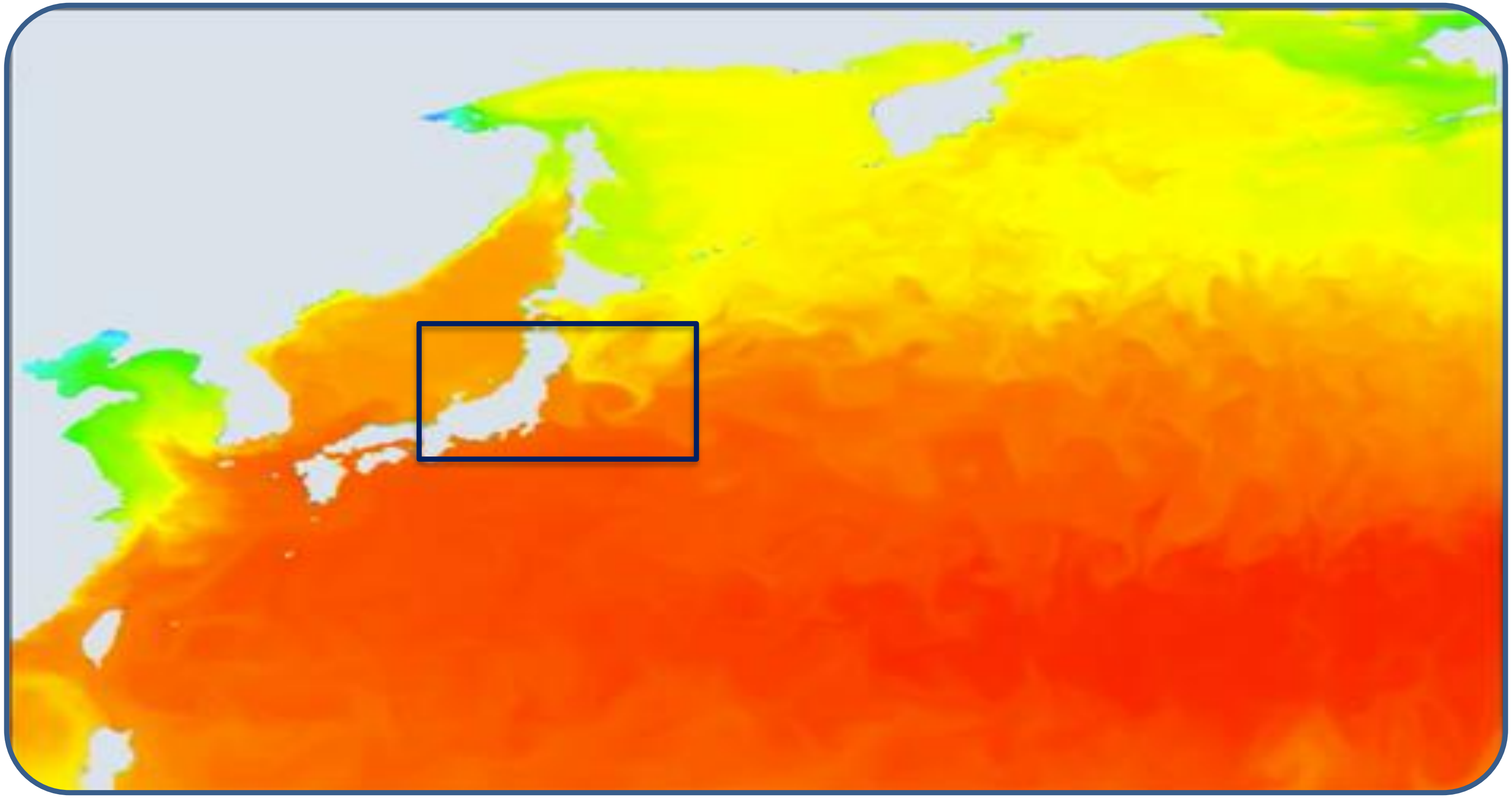
演習

- 指定された論文を読み、小学生の著者が考えた仮説とその検証方法を回答せよ。
  - 以下サイトにアクセスして、200字程度で記述せよ

https://goo.gl/6BgPxh

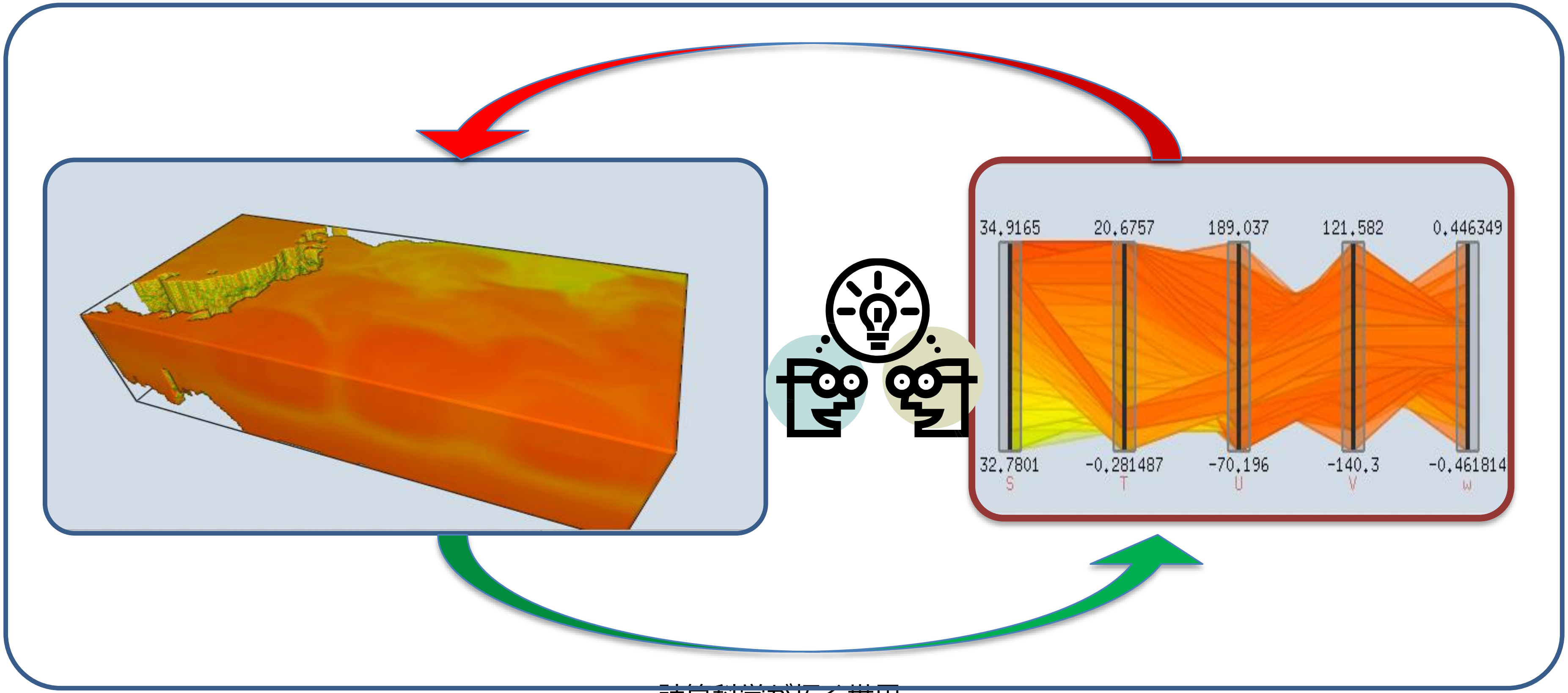
俯瞰可視化

- 問題を設定
  - 全体領域を俯瞰的に可視化
  - 関心の高い領域を選択



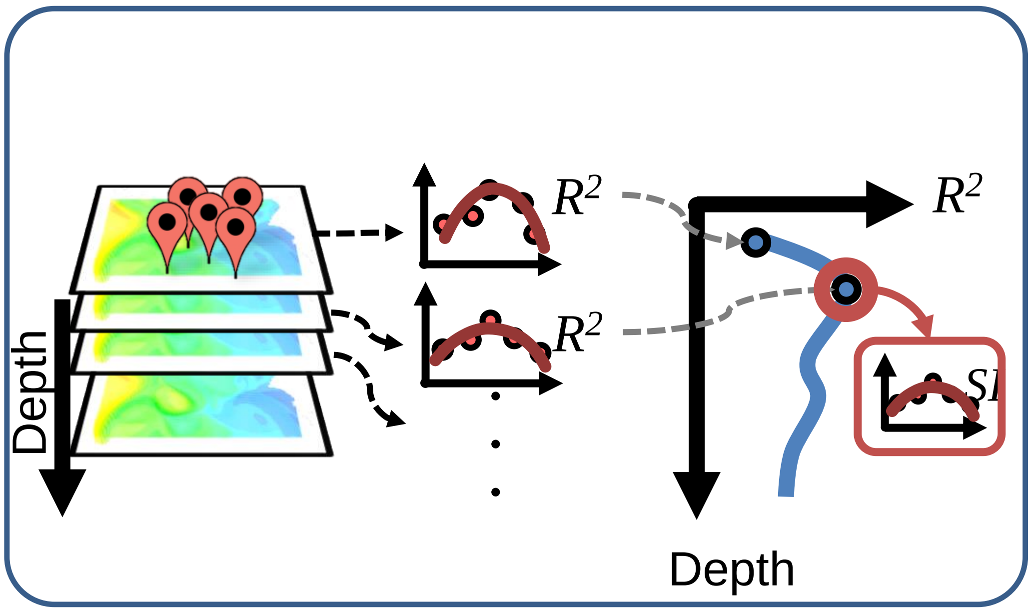
拡大可視化

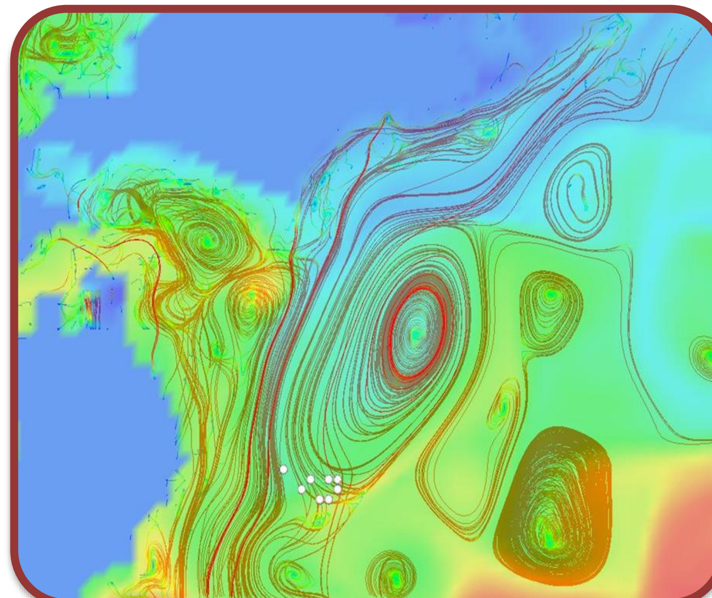
- 仮説を構築
  - 緯度や経度の範囲を変化させて領域を切り出し
  - 塩分濃度や海水温度の範囲を変化させて領域を切り出し



詳細可視化

- 仮説を検証
  - 漁獲高をうまく説明する深さを探索
  - 漁獲高予測マップを作成する

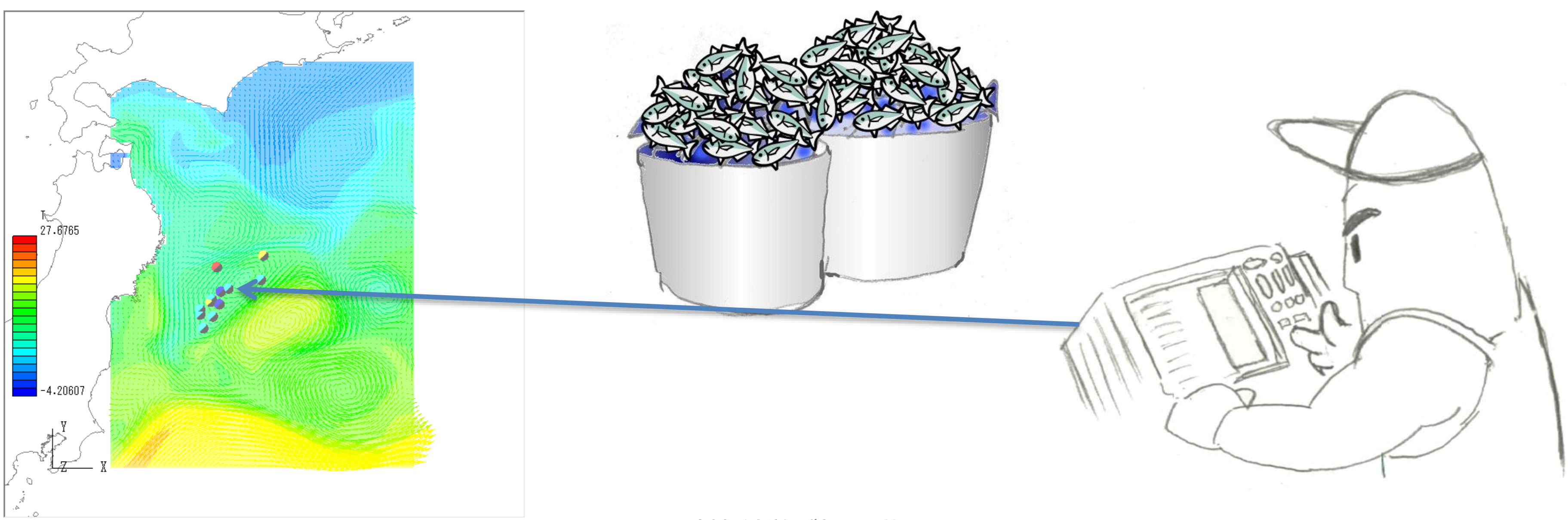




漁師による漁獲量計測

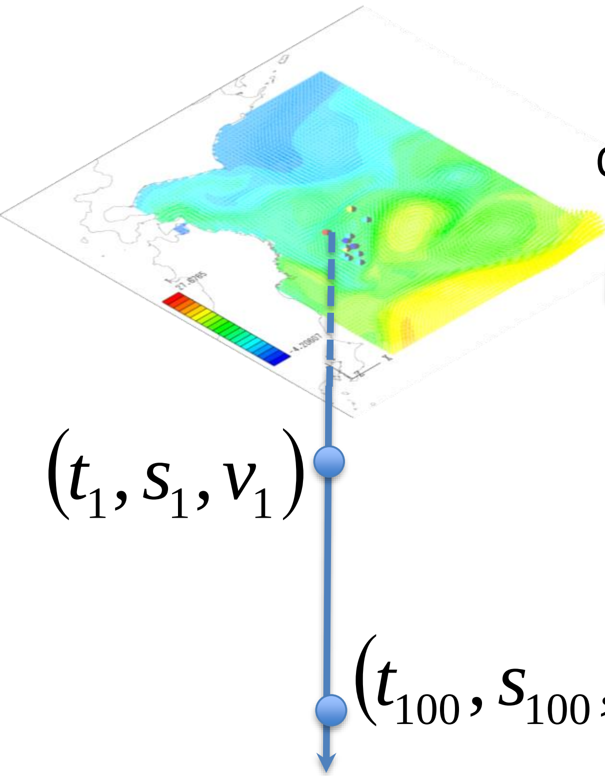
- 多くの漁師は以下の形式で漁獲量を計測し、記録
  - 漁船位置 (x,y)
  - 正規化漁獲量

$$CPUE = \frac{\sum catch}{\sum fishing\ \ days \sum fishing\ \ vessels}$$



CPUEをシミュレーション結果に含まれる変数で説明

- シミュレーション結果においてCPUEの定義される点を通る鉛直線上でサンプリング



(t_1, s_1, v_1)

$(t_{100}, s_{100}, v_{100})$

Original record　　　　Sampled variables

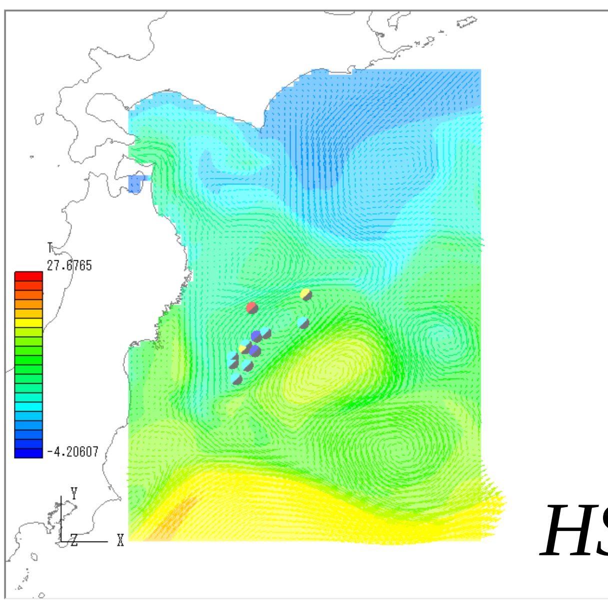
CPUE	x	y	t1	･･･	t100	s1	･･･	s100	v1	･･･	v100
0.642345	0.680518	0.755554	0.448537	･･･	0.920516	0.754839	･･･	0.113546	0.336512	･･･	0.214305
0.711656	0.283465	0.059334	0.099365	･･･	0.586174	0.890815	･･･	0.048899	0.536846	･･･	0.47709
0.553627	0.223513	0.395263	0.532903	･･･	0.985017	0.259018	･･･	0.884757	0.242159	･･･	0.699667
0.096016	0.086106	0.15024	0.131232	･･･	0.390091	0.958255	･･･	0.136579	0.506822	･･･	0.158313
0.927557	0.087157	0.576571	0.74707	･･･	0.311494	0.039833	･･･	0.265292	0.699515	･･･	0.433637
0.776234	0.66956	0.807943	0.680908	･･･	0.667134	0.189051	･･･	0.234158	0.83681	･･･	0.430659
0.339953	0.361467	0.427135	0.967647	･･･	0.251424	0.968415	･･･	0.437858	0.330161	･･･	0.401806
0.167646	0.512573	0.036514	0.060821	･･･	0.257233	0.103441	･･･	0.199014	0.339543	･･･	0.631877
0.525502	0.868313	0.127016	0.576511	･･･	0.665982	0.455188	･･･	0.379432	0.785053	･･･	0.58507
0.476963	0.512256	0.676056	0.301396	･･･	0.046038	0.187826	･･･	0.696615	0.44408	･･･	0.592365

$$CPUE = f(t_1, \cdots t_{100}, s_1, \cdots s_{100}, v_1, \cdots v_{100})$$

総合好適度(HSI) モデル

Tian, S., Chen, X., Chen, Y., Xu, L., Dai, X. 2009. Fisheries Research

- HSI は正規化されたCPUEを表す
- CPUEと高い相関を持つ変数を選択



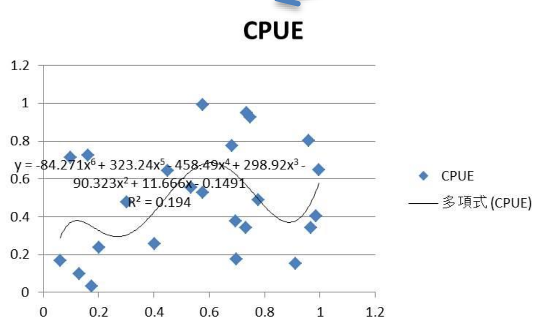
CPUE	x	y	t1	・・・	t100	s1	・・・	s100	v1	・・・	v100
0.642345	0.680518	0.755554	0.448537	・・・	0.920516	0.754839	・・・	0.113546	0.336512	・・・	0.214305
0.711656	0.283465	0.059334	0.099365	・・・	0.586174	0.890815	・・・	0.048899	0.536846	・・・	0.47709
0.553627	0.223513	0.395263	0.532903	・・・	0.985017	0.259018	・・・	0.884757	0.242159	・・・	0.699667
0.096016	0.086106	0.15024	0.131232	・・・	0.390091	0.958255	・・・	0.136579	0.506822	・・・	0.158313
0.927557	0.087157	0.576571	0.74707	・・・	0.311494	0.039833	・・・	0.265292	0.699515	・・・	0.433637
0.776234	0.66956	0.807943	0.680908	・・・	0.667134	0.189051	・・・	0.234158	0.83681	・・・	0.430659
0.339953	0.361467	0.427135	0.967647	・・・	0.251424	0.968415	・・・	0.437858	0.330161	・・・	0.401806
0.167646	0.512573	0.036514	0.060821	・・・	0.257233	0.103441	・・・	0.199014	0.339543	・・・	0.631877
0.525502	0.868313	0.127016	0.576511	・・・	0.665982	0.455188	・・・	0.379432	0.785053	・・・	0.58507
0.476963	0.512256	0.676056	0.301396	・・・	0.046038	0.187826	・・・	0.696615	0.44408	・・・	0.592365
Correlation			0.419613	・・・	0.34039	-0.33172	・・・	-0.03077	0.477163	・・・	0.08621

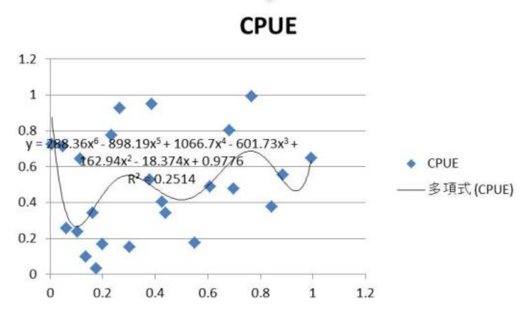
$$HSI = g\left(t_1, t_{100}, v_1\right) = \sqrt[3]{SI_1\left(t_1\right) SI_2\left(t_{100}\right) SI_3\left(v_1\right)}$$

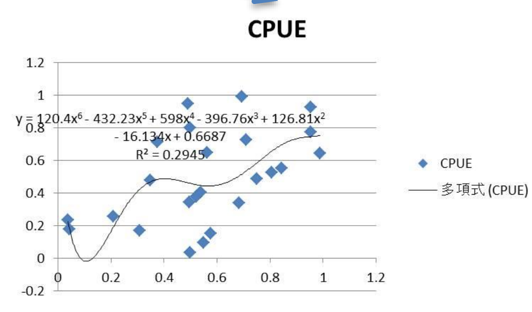
好適度(SI)

- SI 方程式はそれぞれの変数の回帰式で表現される。

$$HSI = \sqrt[3]{SI_1(t_1)SI_2(t_{100})SI_3(v_1)}$$







$$HSI = \sqrt[n]{\prod_{i=1}^{n} SI_i}$$

$$SI_i = \frac{CPUE_{fit\_i} - \min CPUE_{fit\_i}}{\max CPUE_{fit\_i} - \min CPUE_{fit\_i}}$$

HSI モデル開発の手順

以下の手順でモデルを開発する:

1. CPUEと高い相関を持つ変数を選択し、SI方程式を構築
2. SI方程式の相乗平均としてHISを構築
3. モデル妥当性検証のためにHIS分布を可視化