地球データ駆動科学の黎明期

東京大学・大学院新領域創成科学研究科
基盤科学系　複雑理工学専攻
岡田真人

日時:2025年3月25日(火) 13:05-13:40

要旨

- データ駆動科学を地球科学に導入した地球データ駆動科学の黎明期の研究の紹介
- 2010年の11月ごろに複雑理工学専攻の鳥海教授に依頼され、鳥海先生のPDであった桑谷氏の地殻媒質評価に関するセミナーに参加
- 計算論的視覚脳科学の運動知覚のモデルで用いたマルコフランダムフィールド(MRF)モデルで、地殻媒質評価できることに気付く。
- 本講演では、このような地球データ駆動科学の黎明期の経緯を紹介し、今後の地球データ駆動科学の展望を述べる。

内容

- 自己紹介
- 計算論的視覚的脳科学
  - David Marrの三つのレベル
  - 運動知覚の統合のモデル
- 知覚媒質評価
  - データ駆動学の三つのレベル
- 新学術領域「スパースモデリングの深化と高次元データ駆動科学の創成」
- 地球データ駆動科学の展望

自己紹介

- 大阪市立大学理学部物理学科 (1981 - 1985)
  - アモルファスシリンコンの成長と構造解析
- 大阪大学大学院理学研究科(金森研) (1985 – 1987)
  - 希土類元素の光励起スペクトルの理論
- 三菱電機 (1987 - 1989)
  - 化合物半導体(半導体レーザー)の結晶成長
- 大阪大学大学院基礎工学研究科生物工学(福島研) (1989 - 1996)
  - 畳み込み深層ニューラルネット
  - 情報統計力学(ベイズ推論と統計力学の数理的等価性)
- JST ERATO 川人学習動態脳プロジェクト (1996 - 2001)
  - 計算論的神経科学
- 理化学研究所 脳科学総合研究センタ (甘利T) (2001 - 04/06)
  - ベイズ推論，機械学習，データ駆動型科学
- 東京大学・大学院新領域創成科学研究科 複雑理工学専攻
  - データ駆動科学 (2004/07 –)

内容

- 自己紹介
- 計算論的視覚的脳科学
  - David Marrの三つのレベル
  - 運動知覚の統合のモデル
- 知覚媒質評価
  - データ駆動学の三つのレベル
- 新学術領域「スパースモデリングの深化と高次元データ駆動科学の創成」
- 地球データ駆動科学の展望

計算論的神経科学

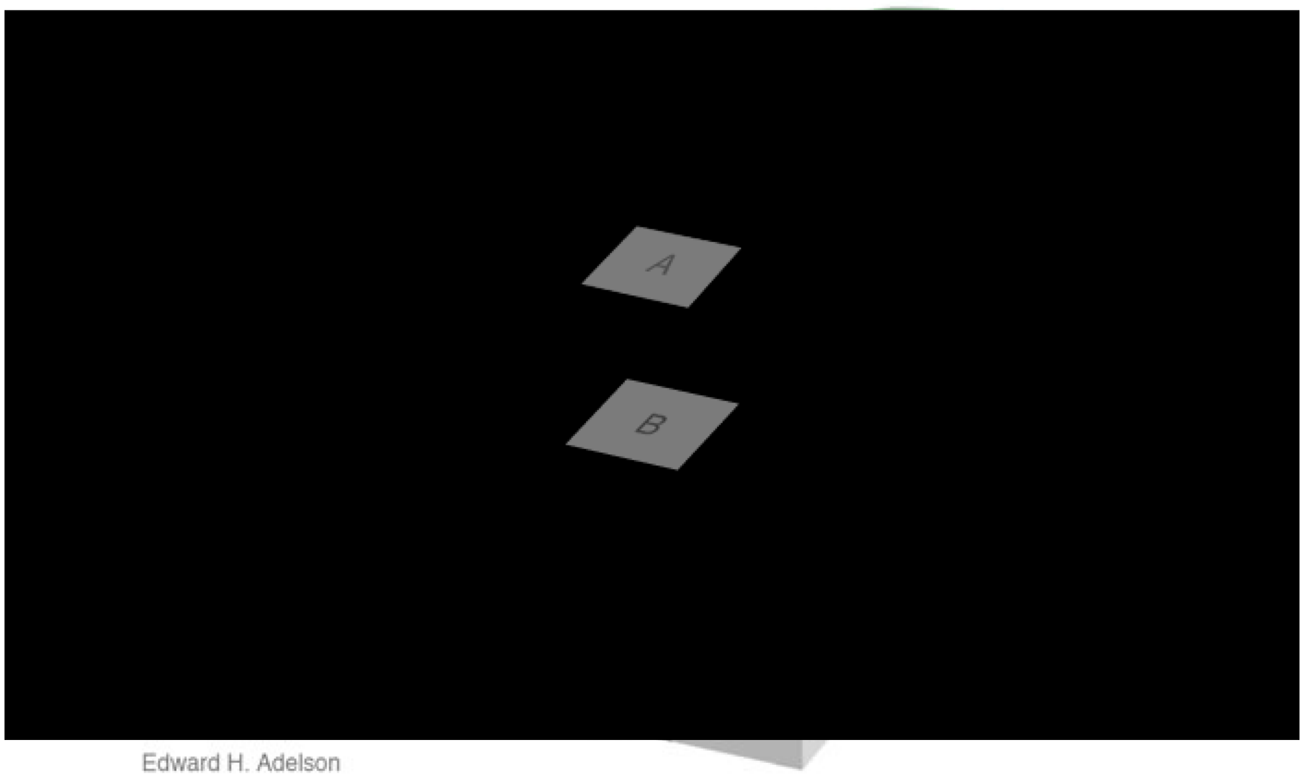
3次元視覚世界の再構成
(David Marr, 1982)

3次元視覚世界

2次元網膜

3次元再構成のキュー
- 両眼視差
- 陰影からの構造復元
- 運動からの構造復元
- 遮蔽 ....

岡田研究室　脳の物理学を創る

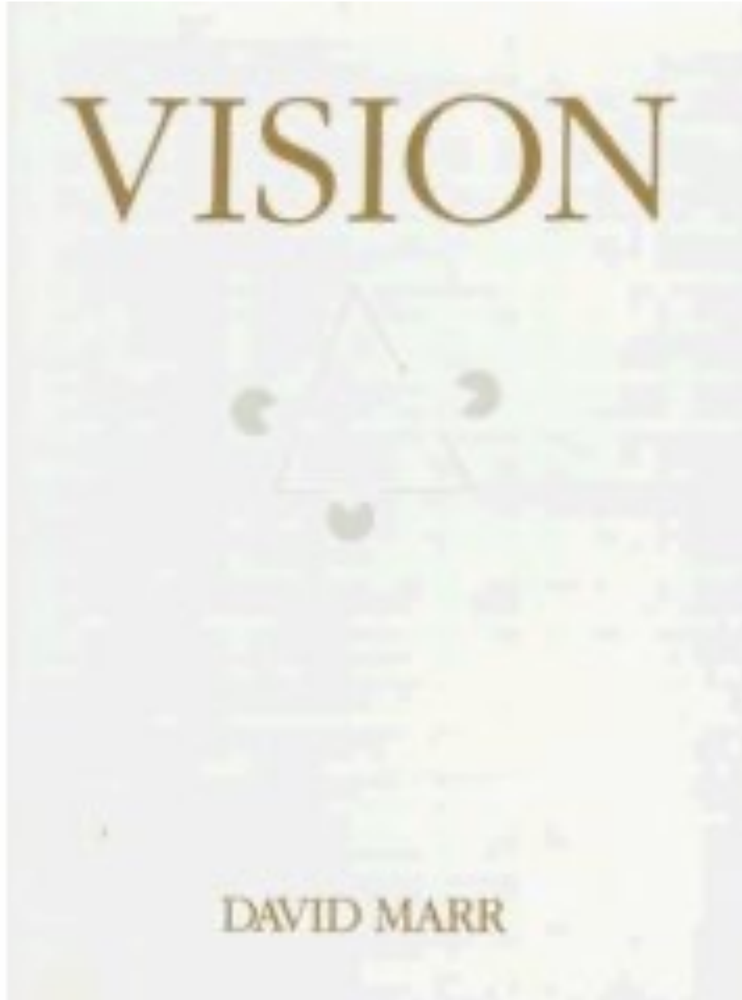


陰影からの構造復元

視覚現象を数学的に解明（３次元世界の再構成)し，
脳でどのような形で(生物学的に)実行されているかを探る

計算論的神経科学
Computational Neuroscience

David Marr (1945-1980)



Vision (1982)

David Marrの三つのレベル

計算理論

計算の目的とその適切性を議論し，実行可能な方法の論理を構築

表現・アルゴリズム

計算理論の実行方法．特にその入力と出力の表現と変換のためのアルゴリズム

ハードウェア実装

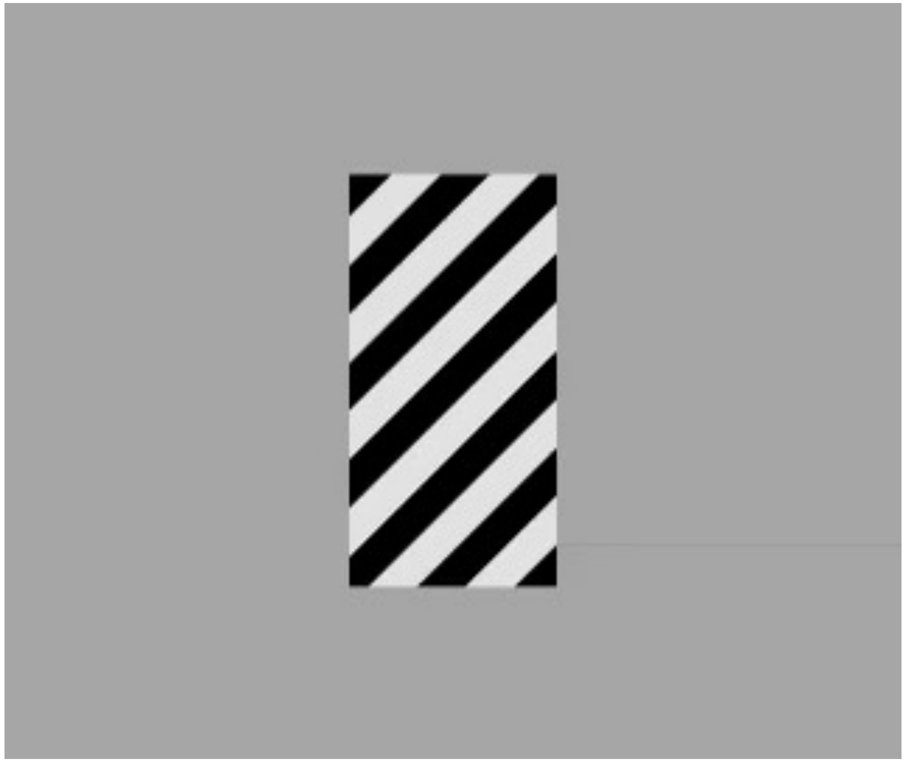
表現とアルゴリズムの物理的な実現
ニューラルネットワーク

Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, MIT Press, 1982.

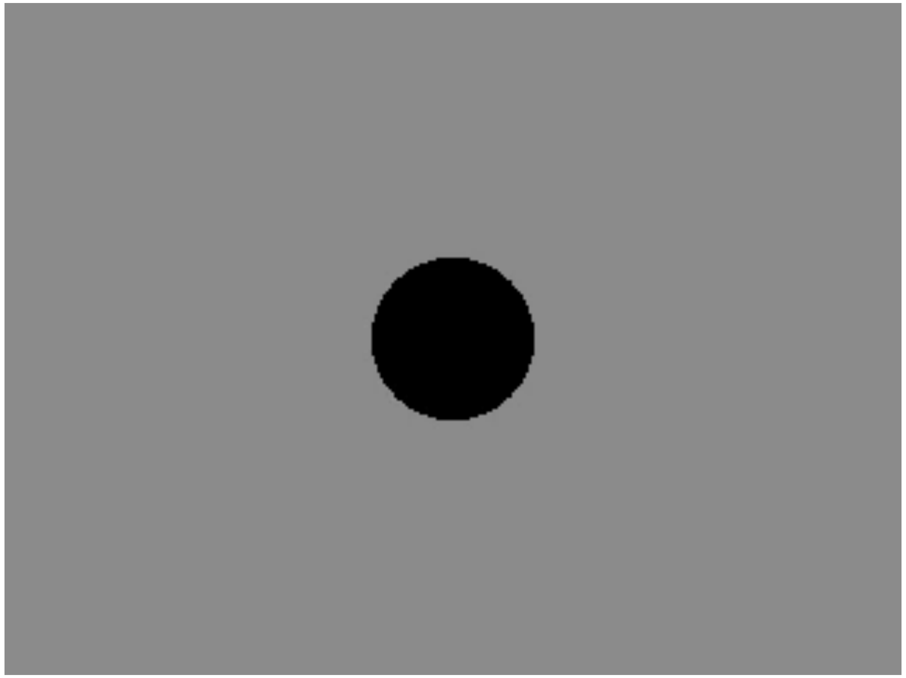
Okada, Nishina and Kawato
“The neural computation of the aperture problem: an iterative process”,
NeuroReport, 14(14), 1767-1771, (2003)

バーバーポール

真ん中の部分の動きが変わる

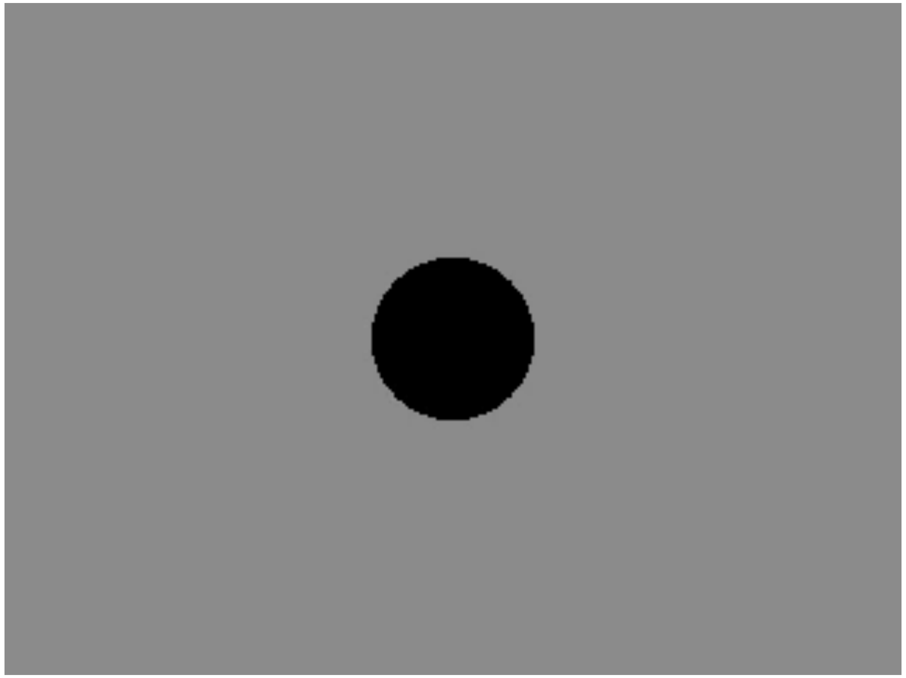


小窗問題

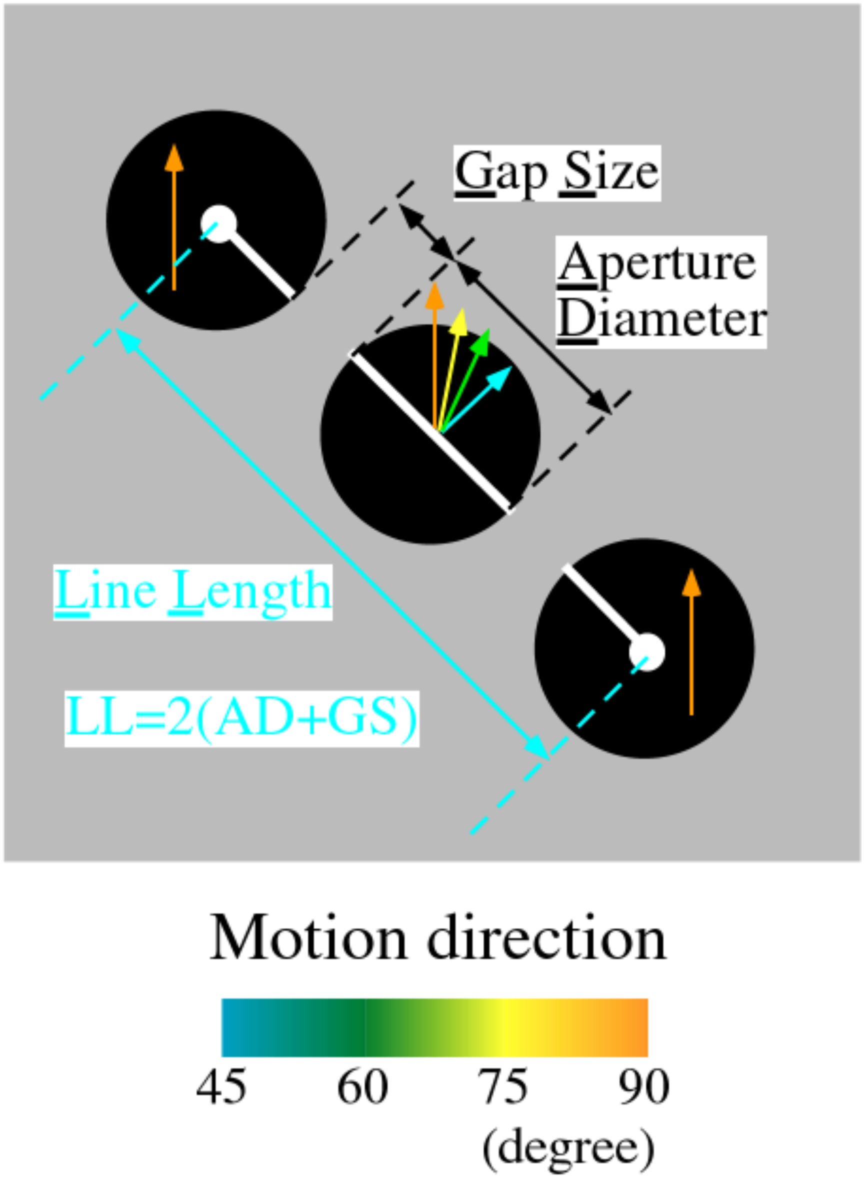


視覚計算の不良設定性(小窓問題)

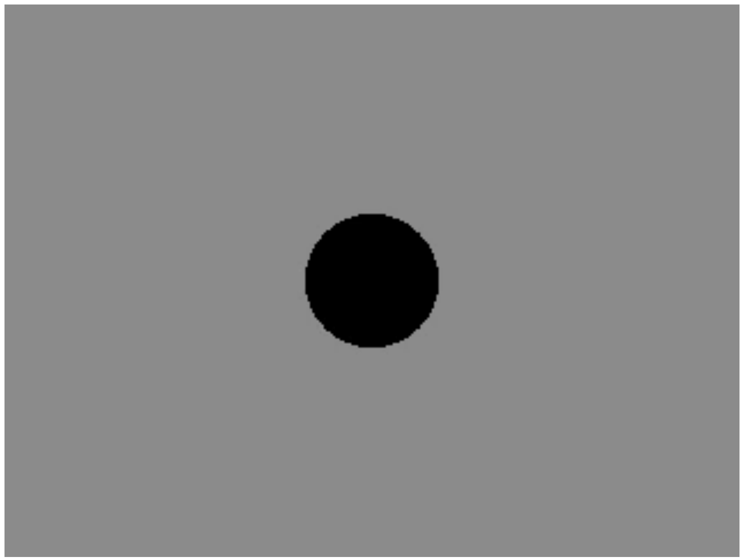
小窓問題



視覚刺激の曖昧性から情報統合の必要性



小窓から見える三本線分は,
まるで一本の線分の一部で
あるように見える
→離れた視覚情報の統合
離れていも近くの確実な情報
を利用する
→近い情報が神経場上でも近い
と統合に用いやすい



Hildreth model

$$f\left(\{\mathbf{V}(s)\}\right)=\int_{I_v} ds\left(\mathbf{V}(s)\cdot\mathbf{N}(s)-V^N(s)\right)^2$$

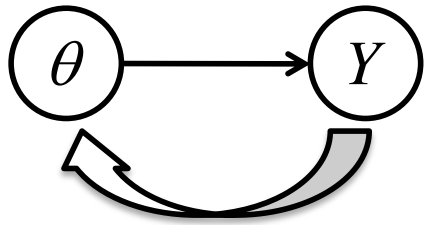
$$+\lambda_1\int ds\left(\frac{\partial}{\partial s}\mathbf{V}(s)\right)^2$$

- $\boldsymbol{N}(s)$: Normal vector at coordinate s
- $\boldsymbol{V}(s)$: Velocity vector estimate at coordinate s
- V^N(s): Observed value of the normal component of the velocity vector at coordinate s
- λ_1: Regularization parameter

Bayesian inference

$$p(Y,\theta) = \underline{p(Y \mid \theta)p(\theta) = p(\theta \mid Y)p(Y)}$$

Generative model
(causality)



Bayes' theorem

$$p(\theta \mid Y) = \frac{p(Y \mid \theta)\,p(\theta)}{p(Y)}$$

$p(\theta \mid Y)$: Posterior probability. The probability of the parameter θ given the data.

$p(\theta)$: Prior probability. It is necessary to set in advance. Scientific knowledge accumulated so far

Posterior probability

Prior probability

$$p\left(\{\mathbf{V}(s)\}\right) \sim \exp\left(-\frac{\int ds \left(\frac{\partial}{\partial s} \mathbf{V}(s) \right)^2}{2\sigma^2} \right)$$

Conditional probability

$$p\left(V^N(s) \mid \{\mathbf{V}(s)\}\right) \sim \exp\left(-\frac{\int ds \left(\mathbf{V}(s) \cdot \mathbf{N}(s) - V^N(s) \right)^2}{2\sigma'^2} \right)$$

Posterior probability

$$p\left(\{\mathbf{V}(s)\} \mid V^N(s)\right) \sim p\left(V^N(s) \mid \{\mathbf{V}(s)\}\right) p\left(\{\mathbf{V}(s)\}\right)$$

Cost function

Logarithm of posterior probability and cost function

$$\log p\left(\{\mathbf{V}(s)\} \mid V^N(s)\right) \sim -f\left(\{\mathbf{V}(s)\}\right)$$

$$f\left(\{\mathbf{V}(s)\}\right) = \int ds \left(\mathbf{V}(s)\cdot\mathbf{N}(s) - V^N(s)\right)^2$$

$$+\frac{\sigma^2}{\sigma'^2}\int ds \left(\frac{\partial}{\partial s}\mathbf{V}(s)\right)^2$$

Proposed model

$$f\left(\{\mathbf{V}(s), b(s)\}\right) = \int_{I_v} ds \left(\mathbf{V}(s)\cdot\mathbf{N}(s) - V^N(s)\right)^2$$

$$+\lambda_1 \int_{I_V+I_O} ds\, b(s)\left(\frac{\partial}{\partial s}\mathbf{V}(s)\right)^2 + \lambda_2 \int_{I_V+I_O} ds \left(\frac{\partial}{\partial s} b(s)\right)^2$$

- $\boldsymbol{N}(s)$: Normal vector at coordinate s
- $\boldsymbol{V}(s)$: Velocity vector estimate at coordinate s
- V^N(s): Observed value of the normal component of the velocity vector at coordinate s
- λ_V, λ_b : Regularization parameters
- $b(s)$: Binding signal
- I_V: Visible area, I_O : occluded area

Relaxation computation
Steepest descent method

$$\frac{\partial \mathbf{V}(s)}{\partial t} = -\frac{\partial}{\partial \mathbf{V}(s)} f\left(\{\mathbf{V}(s), b(s)\}\right)$$

$$= \lambda_1 b(s) \frac{\partial^2}{\partial s^2} \mathbf{V}(s) - \left(\mathbf{V}(s) \cdot \mathbf{N}(s) - V^N(s)\right) \quad s \in I_V$$

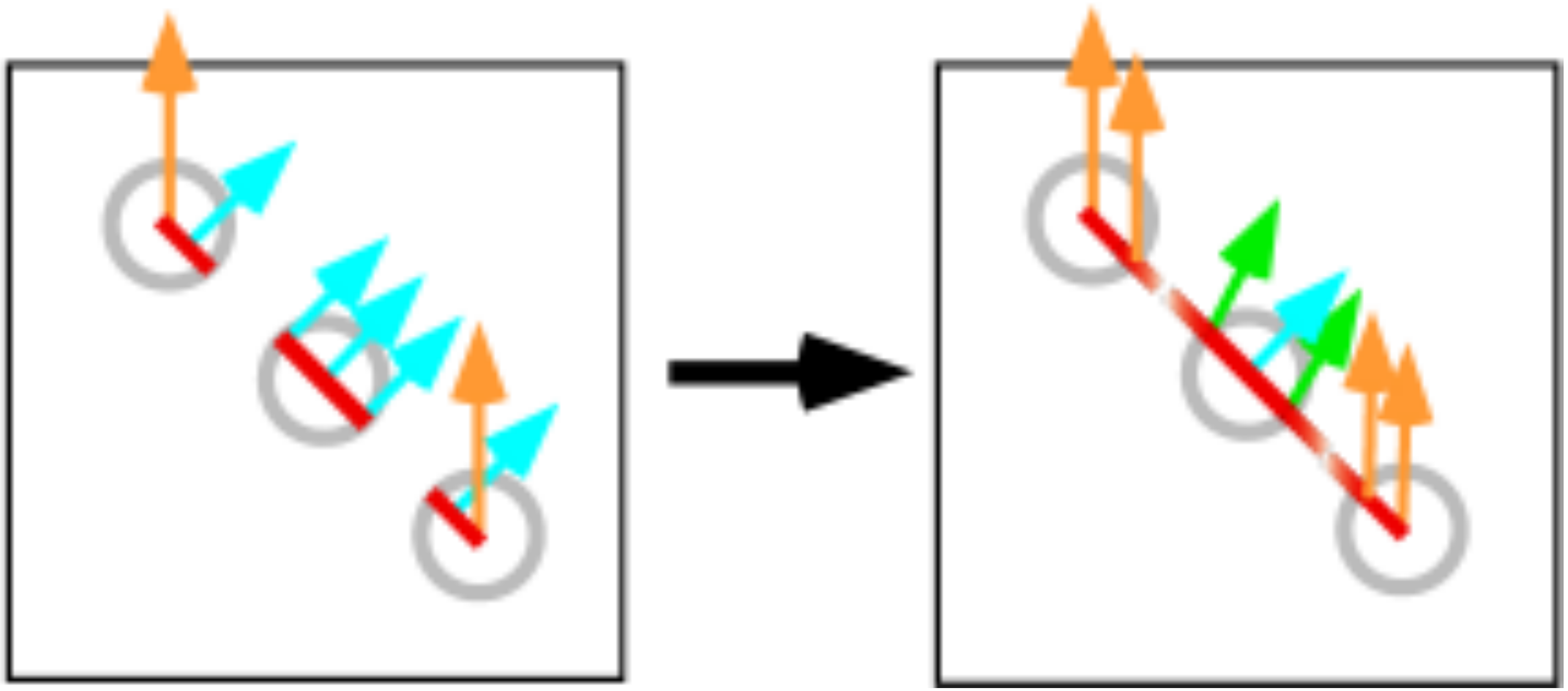
$$= \lambda_1 b(s) \frac{\partial^2}{\partial s^2} \mathbf{V}(s) \qquad s \in I_O$$

$$\frac{\partial b(s)}{\partial t} = -\frac{\partial}{\partial b(s)} f\left(\{\mathbf{V}(s), b(s)\}\right)$$

$$= \lambda_2 \frac{\partial^2}{\partial s^2} b(s) - \lambda_1 \left(\frac{\partial}{\partial s} \mathbf{V}(s)\right)^2$$

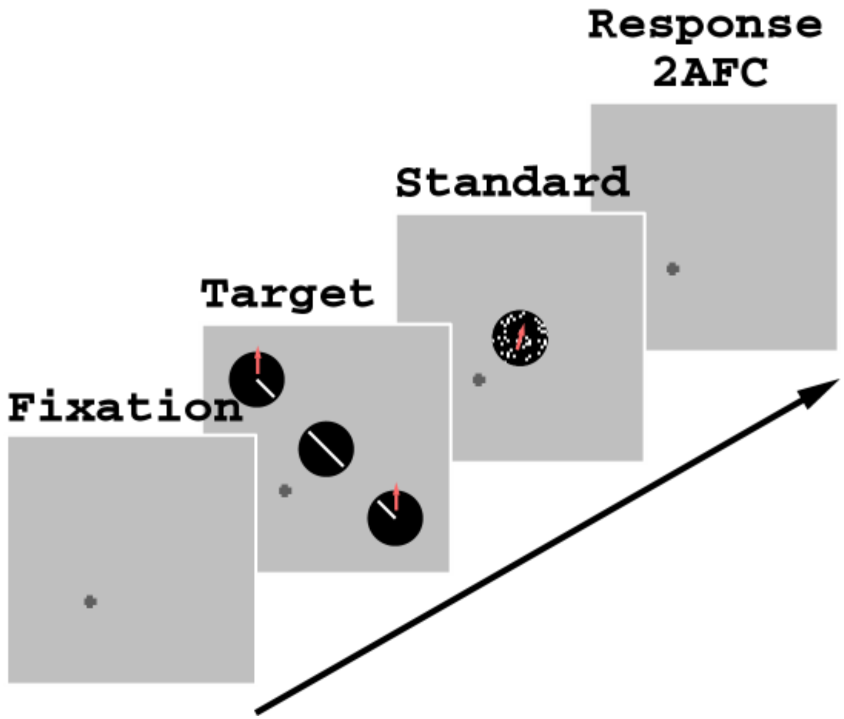
Model

Target stimulus



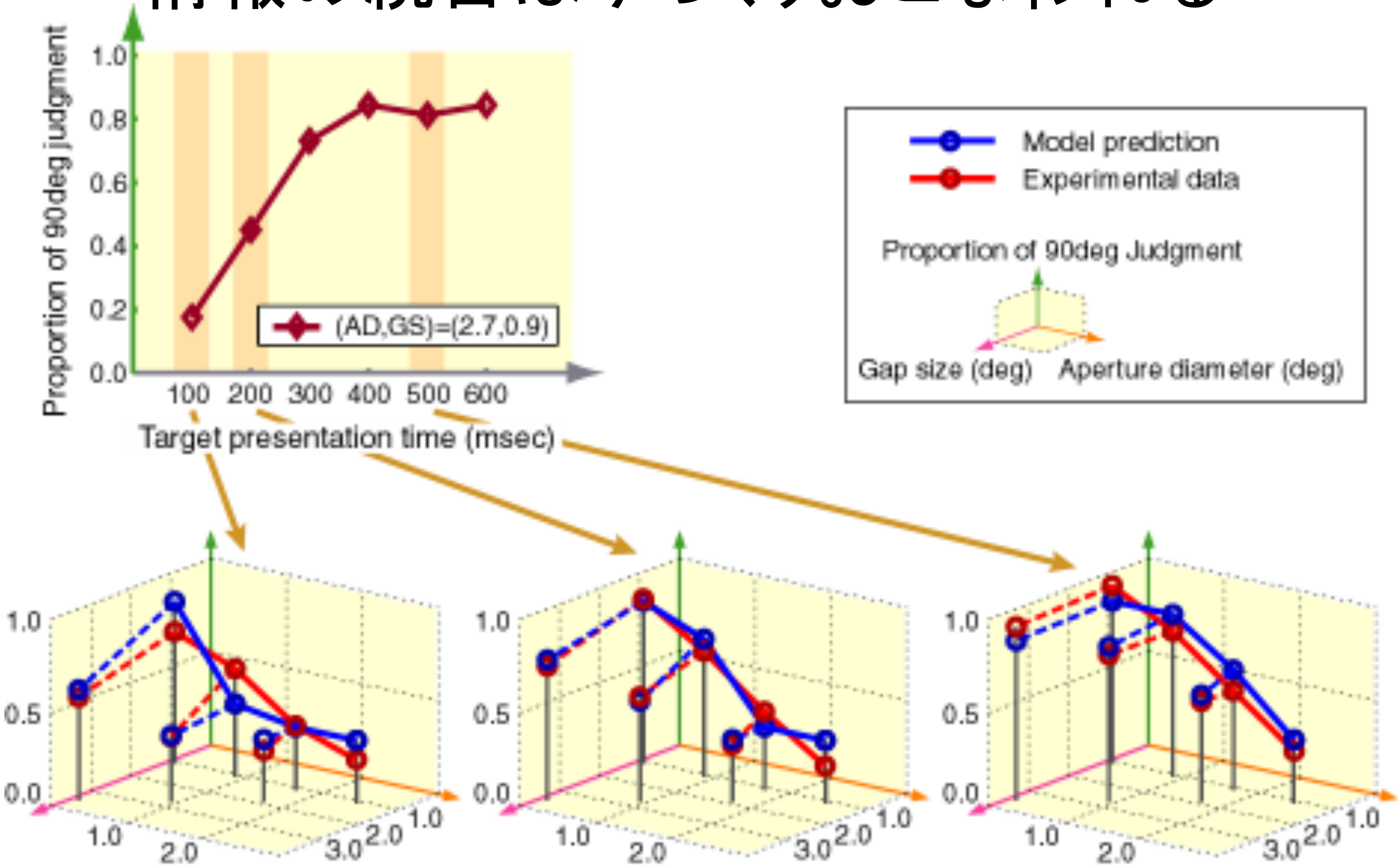
Initial state

実験1



- 呈示時間=100, 200, 300, 400, 500, 600msec
- 被験者はStandardのランダムドットの動きに比べて，90度に近いか遠いかを二者選択で答える(強制二者選択課題).

情報の統合はゆっくりおこなわれる



(Okada, Nishina and Kawato, 2003)

内容

- 自己紹介
- 計算論的視覚的脳科学
  - David Marrの三つのレベル
  - 運動知覚の統合のモデル
- 知覚媒質評価
  - データ駆動学の三つのレベル
- 新学術領域「スパースモデリングの深化と高次元データ駆動科学の創成」
- 地球データ駆動科学の展望

Kuwatani, Nagata, Okada and Toriumi

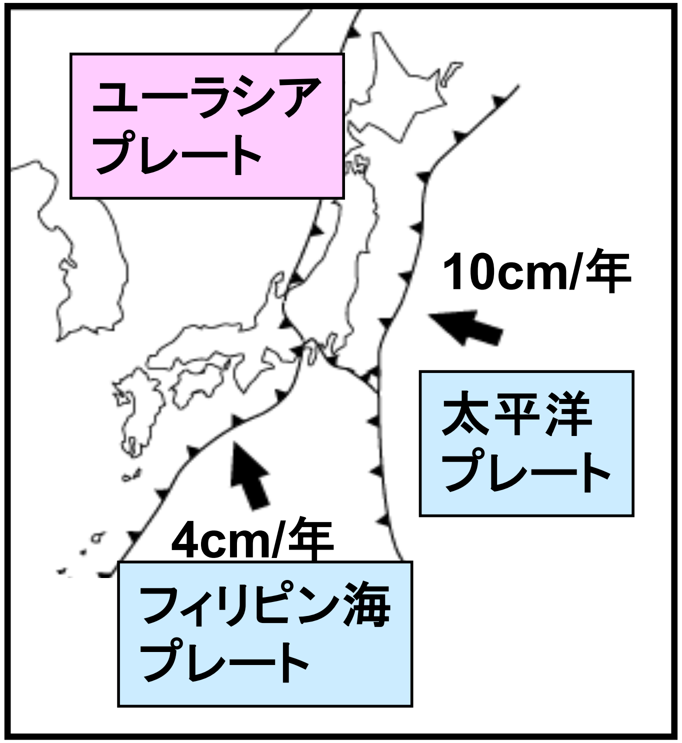
"Markov random field modeling for mapping geofluid distributions from seismic velocity structures"

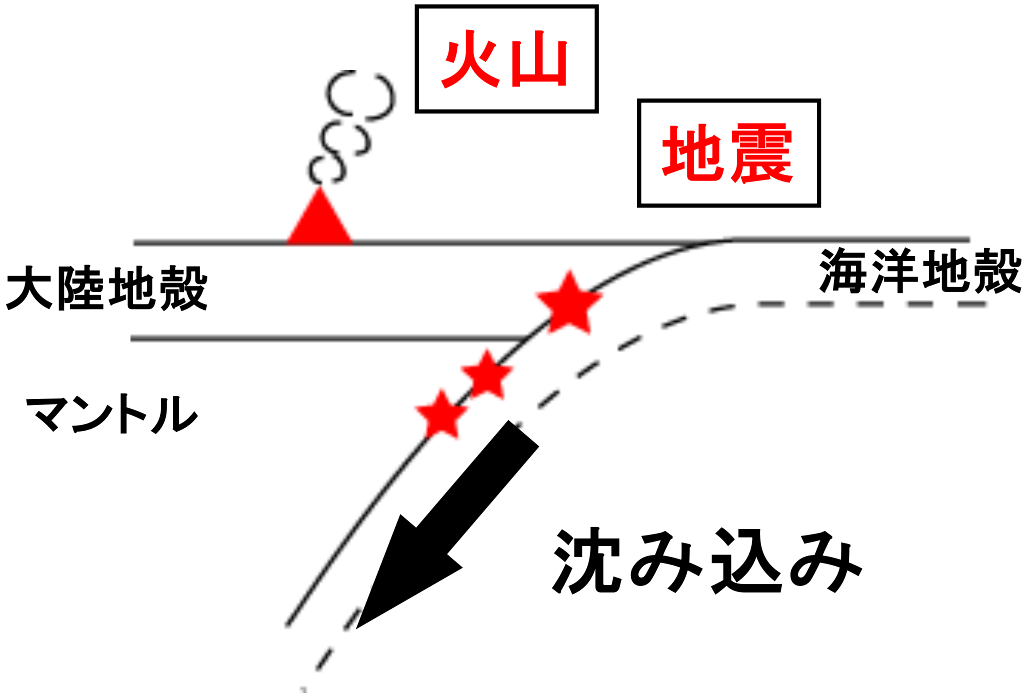
Earth, Planets and Space, 66(5), 1-9, (2014)

目次

- はじめに
  - 地震波速度と流体
- Markov random field (MRF) モデル
- 人工データテスト
- 自然データに適用
- 展望とまとめ

沈み込み帯

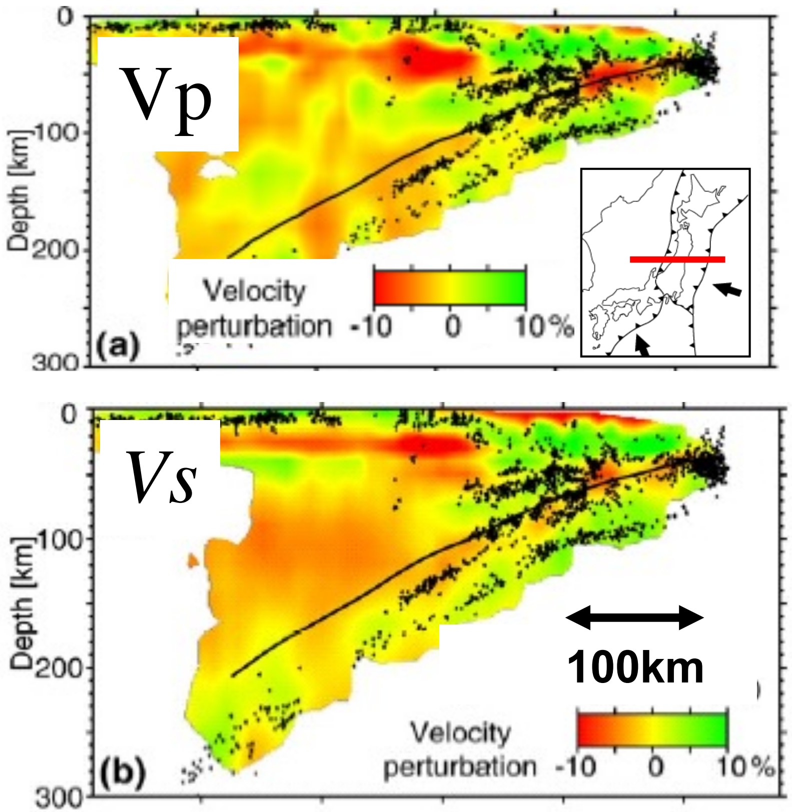




沈み込み帯のダイナミクスの理解には，
地下深部の物質科学的な理解が不可欠.

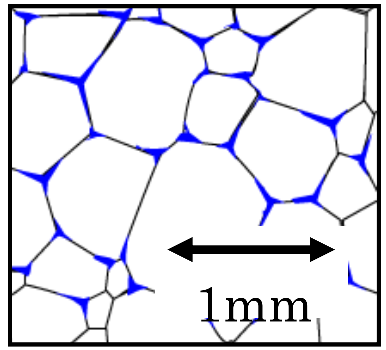
地殻媒質の評価

地球物理学観測データ



地震波速度(Vp,Vs)

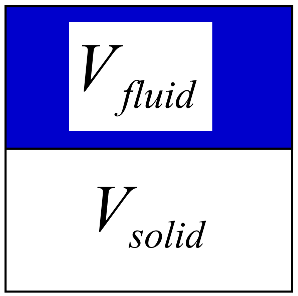
地質学的物質学的情報



岩石の種類 (z)
流体の種類 (w)
流体の量　(Φ)
流体の形状(α)

流体の量と形状

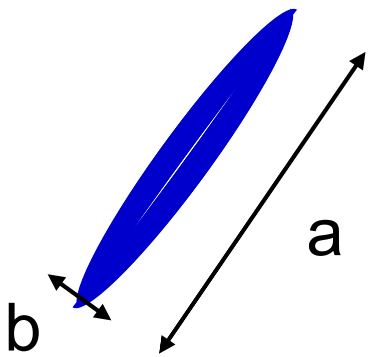
φ: 空隙率(ポロシティ)



$$\phi \equiv \frac{V_{fluid}}{V_{solid} + V_{fluid}}$$

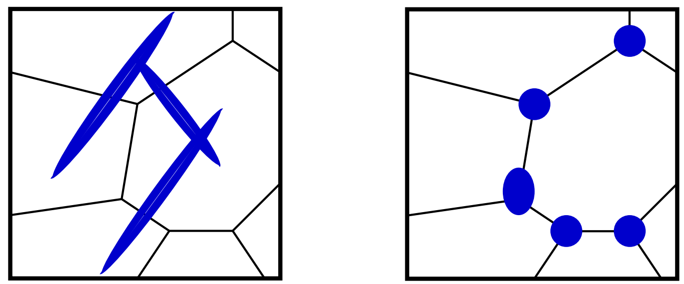
データ範囲: 0 - 1

α: 空隙の形状(アスペクト比)



$$\alpha \equiv \frac{b}{a}$$

データ範囲:
0(クラック状)- 1(球状)



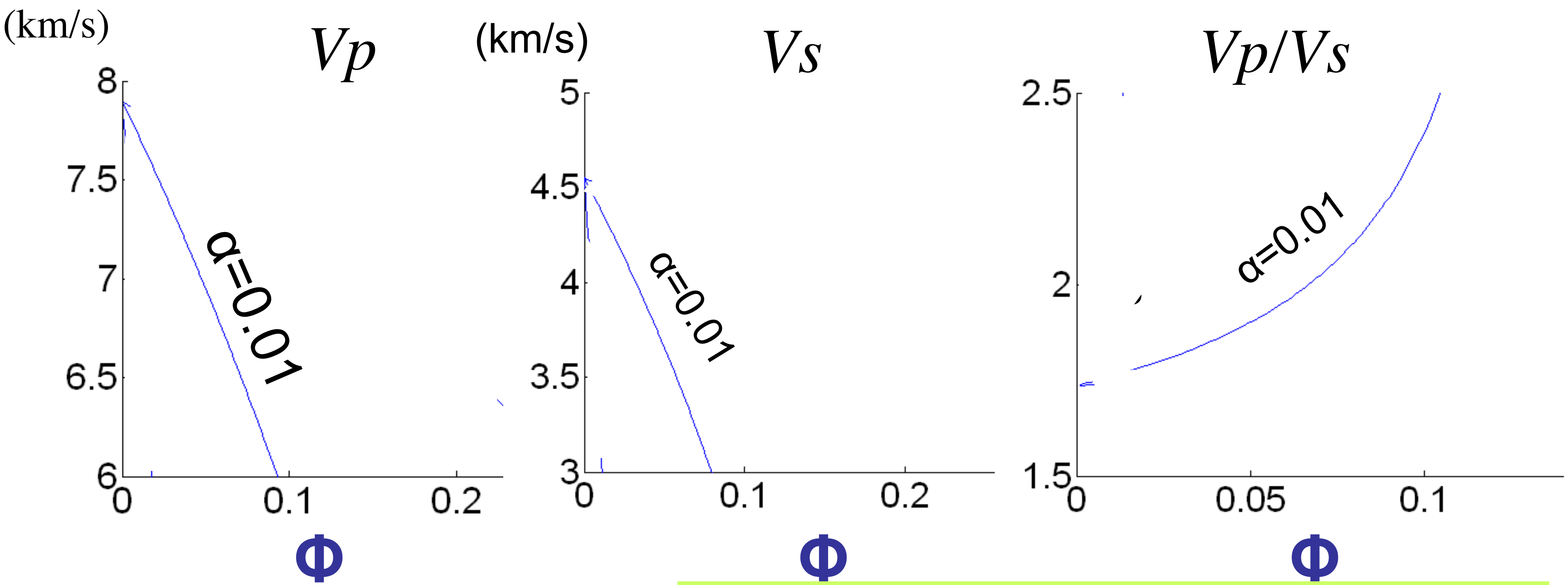
地震波速度と流体の関係

岩石の種類(z)，流体の種類(w)，温度(T)を一定とすると，

$$\boxed{\begin{aligned} Vp &= f_P(\phi,\alpha) \\ Vs &= f_S(\phi,\alpha) \end{aligned}}$$

φ: 空隙率(ポロシティ),
α: 空隙の形状（アスペクト比）

Takei 2002



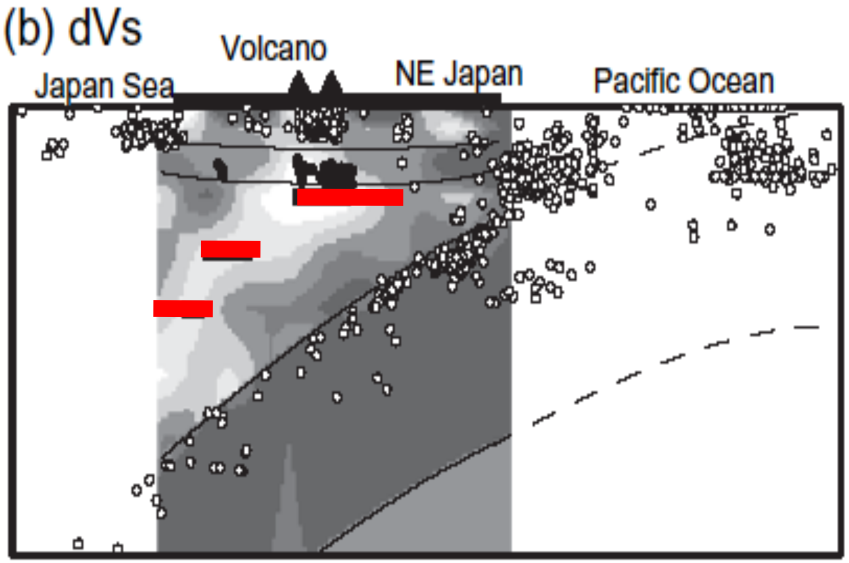
カンラン岩 $(Vp0,Vs0)=(7.9\text{km/s}, 4.5\text{km/s})$，メルト

先行研究

Nakajima, Takei, Hasegawa (2005)

地震波速度データ(Vp,Vs)から逆関数(Takei, 2002)を使って流体量・形状(Φ, α)を決定.

各深さの観測値の平均値を使用.



深さ	流体	φ(%)	α
40km	melt	1-2	～ 0.03
65km	melt	～0.05	～ 0.001
90km	melt	3-6	～ 0.1

本研究

問題点

1．観測データの誤差・不確定性

2．観測変数< 未知変数

目的

Markov Random Field (MRF)モデルを用いて、
地震波速度構造(Vp,Vs)から流体分布(Φ, α)を推定する。

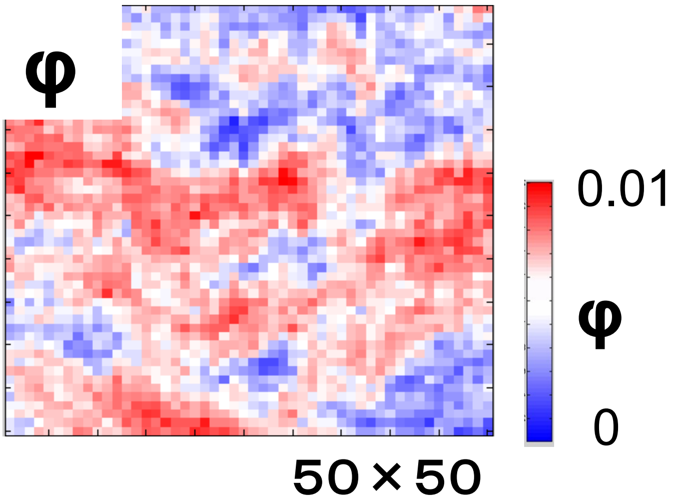
MRFモデル：画像解析によく用いられるベイズ統計的手法

ノイズを含んだ観測量
不定問題

物理量のある程度の連続性を仮定することで最も確からしい推定を行う。

MRFモデル

知りたい物理量（φ）の空間分布

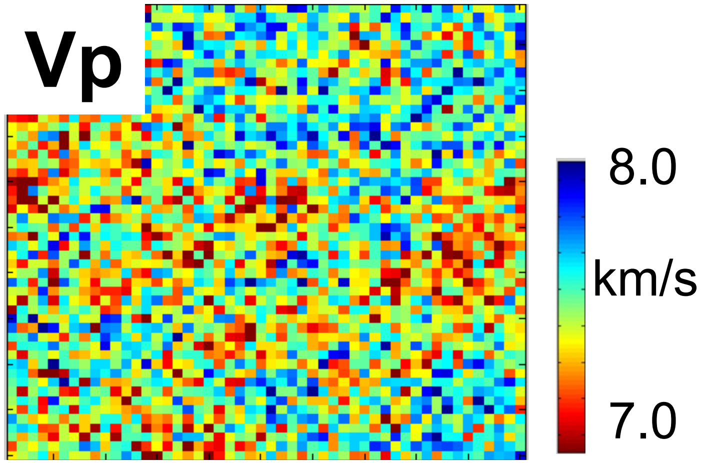


ノイズを含む観測

$$V_i = f(\phi_i) + \varepsilon_i$$

推定

観測量（V）の分布



$$E(\phi) = \sum_i \left(V_i - f(\phi_i)\right)^2$$

観測量の再現性

評価関数 $E(\phi)$ **を最小にするφが最も確からしいφ**

ベイズ推定

Φの事前分布を仮定

隣り合う物理量はだいたい似たような値をとる. ➡ $p(\phi) \propto \prod_i \exp\left(-\frac{\{\phi_i - \phi_{i-1}\}^2}{2\sigma_m{}^2} \right)$

尤度関数（生成モデル）

$V_i = f(\phi_i) + \varepsilon_i$ ➡ $p(V_i \mid \phi_i) = \frac{1}{\sqrt{2\pi\sigma_n}} \exp\left(-\frac{(V_i - f(\phi_i))^2}{2\sigma_n^2} \right)$

事後確率 $p(\phi \mid V)$ **はベイズの定理により,**

$$\boxed{p(\phi \mid V) = \frac{p(V \mid \phi)\, p(\phi)}{p(V)}}$$

$p(\phi \mid V)$ **を最大にするΦを推定値とする. (MAP推定)**

ベイズ推定:評価関数の導出

$p(\phi | V)$ の最大化は $-\log p(\phi | V)$ の最小化と等しいから,

$$\boxed{E(\phi) \equiv -\log p(\phi | V) = \sum \left(V_i - f(\phi_i)\right)^2 + \frac{\sigma_n^2}{\sigma_m^2} \sum \left(\phi_{i+1} - \phi_i\right)^2}$$

で定義される評価関数の最小化を行えばよい.

$$\boxed{\lambda = \sigma_n^2 / \sigma_m^2}$$

$\sigma_n^2 < \sigma_m^2 \rightarrow$ 観測値を重視する.

$\sigma_n^2 > \sigma_m^2 \rightarrow$ 物理量の連続性を重視する.

ハイパーパラメータ推定

事後確率 $p(\lambda \mid V)$ を最大化するλを決定する.

ベイズの定理を用いて,

$$\boxed{p(\lambda \mid V) = \frac{p(\lambda)\cdot p(V \mid \lambda)}{p(V)}}$$

事後確率 $p(\lambda \mid V)$ の最大化は

$$\boxed{F(\lambda) \equiv -\log p(\lambda \mid V) = -\log \int \exp\left(-E(\phi,\lambda)\right) d\phi}$$

で定義される自由エネルギーの最小化を行えばよい.

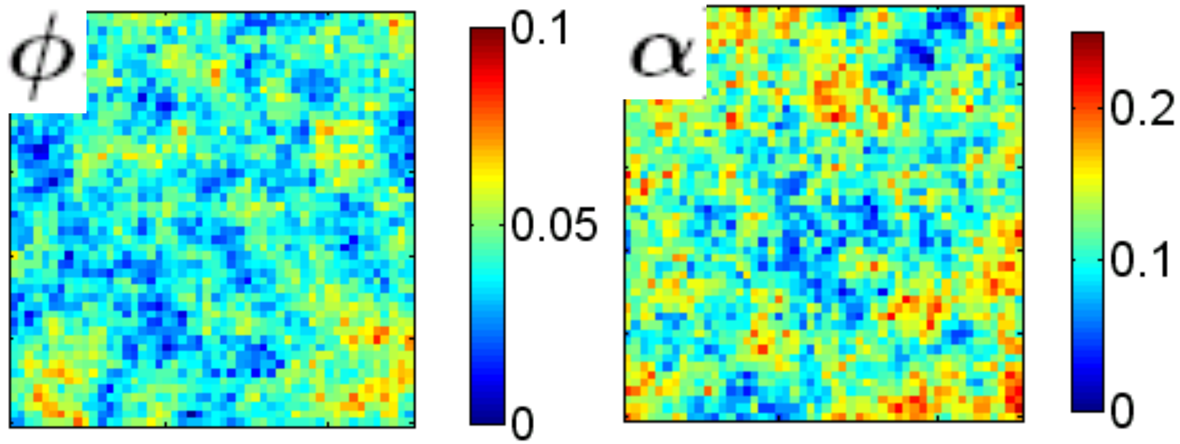
⇩

MCMC (Markov chain Monte Carlo)法を用いた最急降下法

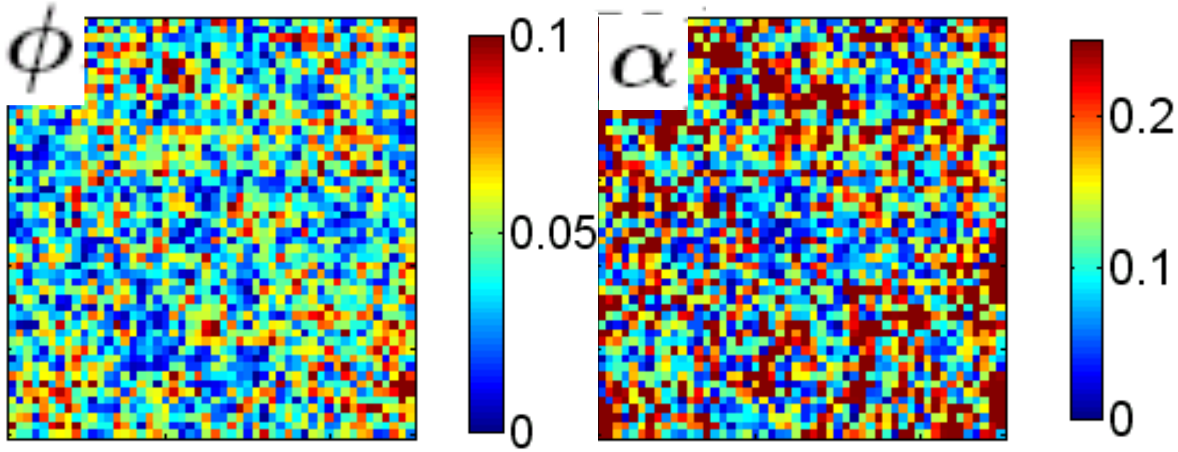
人工データの解析

人工データ

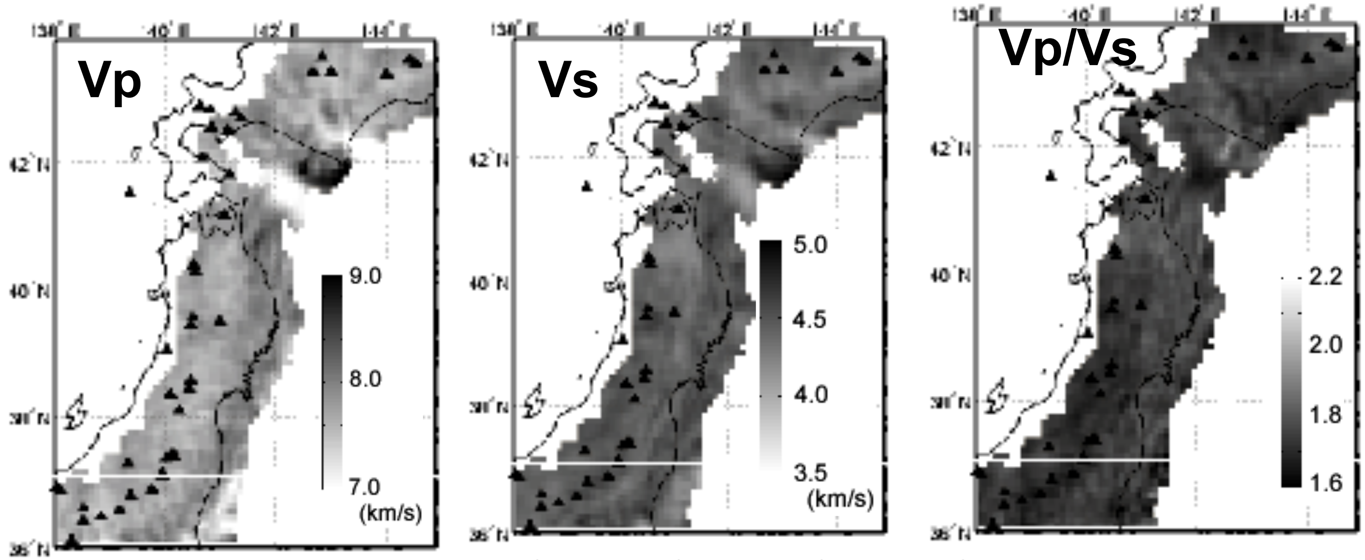
$$(\sigma_\phi^2, \sigma_\alpha^2) = (0.001, 0.0001)$$



従来手法（逆関数）



天然系への応用(Depth=50km)



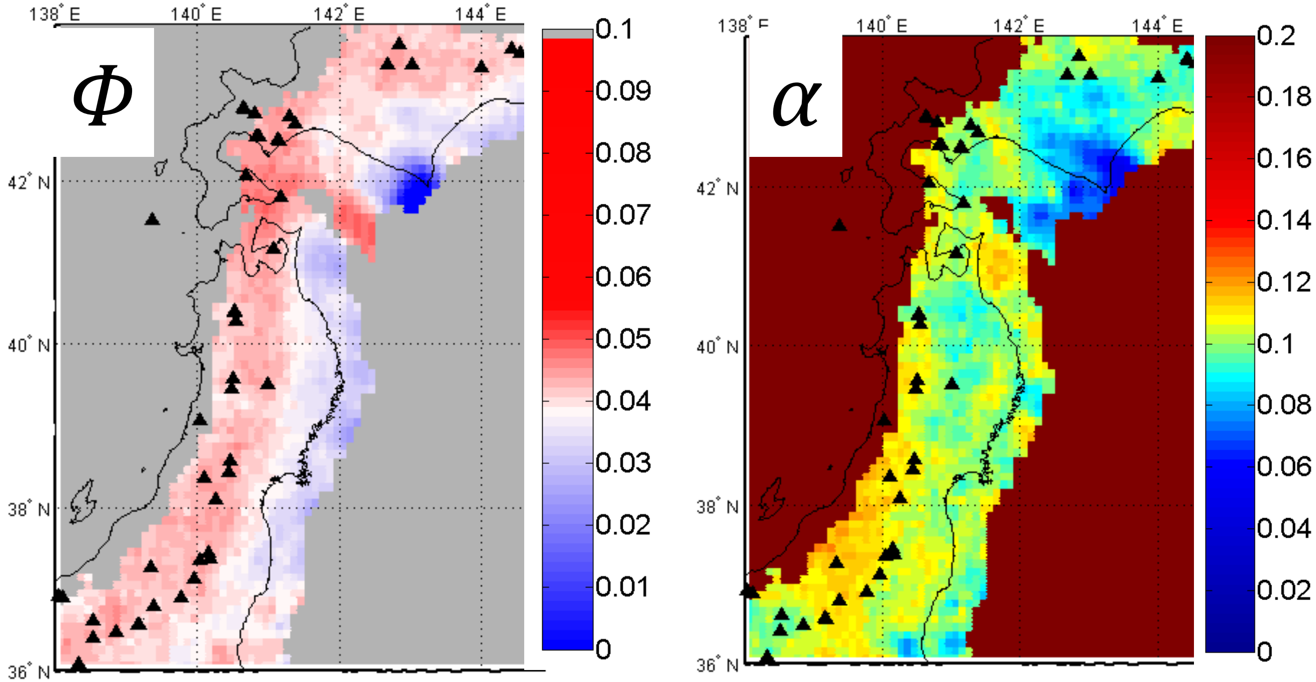
Matsubara et al. 2008

$(Vp0, Vs0)=$
$(8.6\mathrm{km/s}, 5.2\mathrm{km/s})$

流体＝melt

Hyperparameter推定

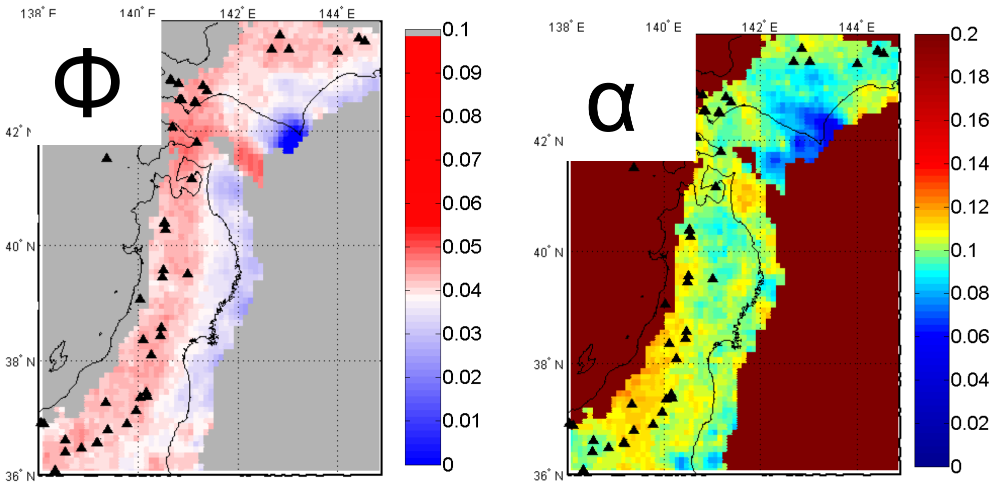
$$(\sigma_\phi^2, \sigma_\alpha^2) = (0.0001, 0.0018)$$
$$(\sigma_P^2, \sigma_S^2) = (0.052, 0.014)$$



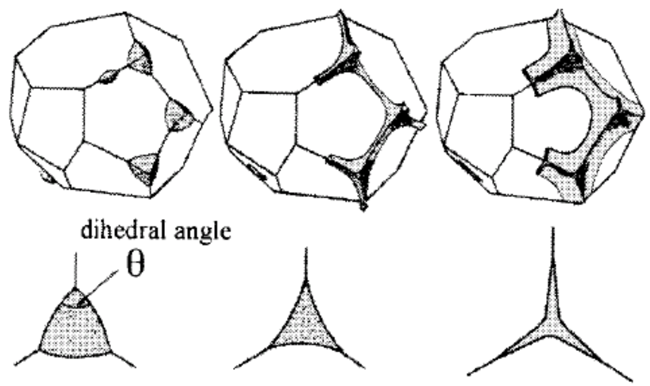
地球物理学的インプリケーション

- 火山直下において，
  Φ～0.05　→　岩石の部分融解によるマグマの生成.
  アスペクト比が0.1程度→　組織的平衡

Depth=50km



流体の存在形態



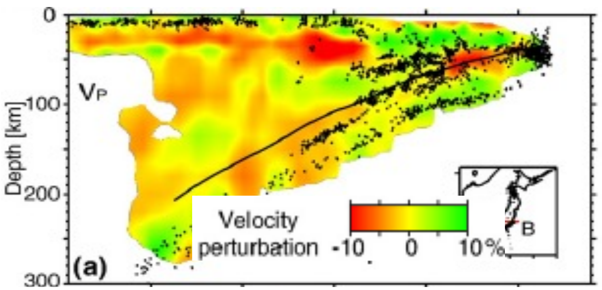
問題点と展望

- 推定した観測ノイズの分散が大きすぎる.

  → Φ, α以外の物理量が必要？

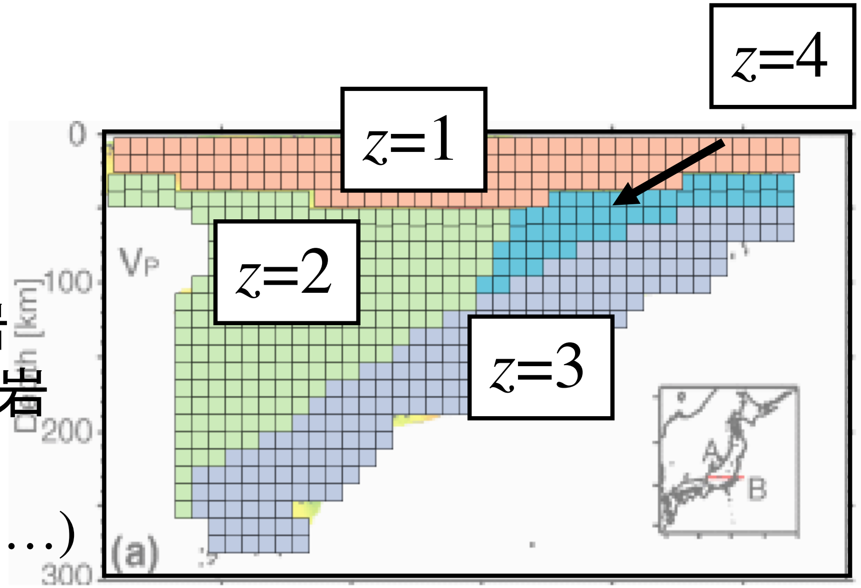
- 不連続に変化する岩石の種類(z) も求めたい. → ラベルプロセスの導入

地震波速度(Vp,Vs)



z = 岩相

(1: カコウ岩
2: カンラン岩
3: 玄武岩
4: 蛇紋岩 …)

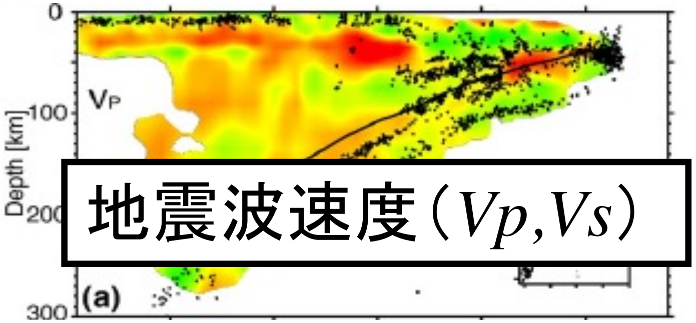


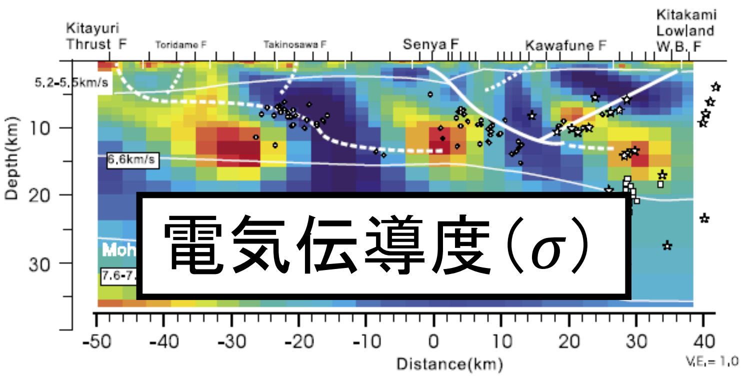
地球科学におけるベイズ推定の有効性

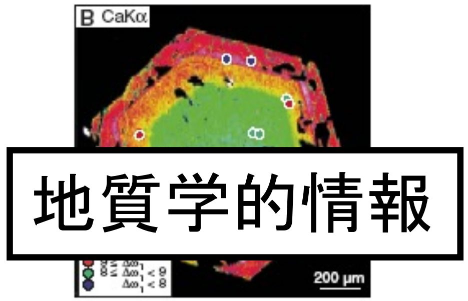
1．観測変数<未知変数のときも解析可能

変数：Z（岩石の種類），W（流体の種類）
T(温度), f(連結度)．．．

2．異なる観測データや先験的知識を容易に導入可能







3．不連続性の検出

内容

- 自己紹介
- 計算論的視覚的脳科学
  - David Marrの三つのレベル
  - 運動知覚の統合のモデル
- 知覚媒質評価
  - データ駆動学の三つのレベル
- 新学術領域「スパースモデリングの深化と高次元データ駆動科学の創成」
- 地球データ駆動科学の展望

データ駆動科学の三つのレベル (2016)

計算理論(対象の科学，計測科学)

データ解析の目的とその適切性を議論し，実行可能な方法の論理(方略)を構築

モデリング(統計学，理論物理学，数理科学)

計算理論のレベルの目的，適切さ，方略を元に，系をモデル化し，計算理論を数学的に表現する

表現・アルゴリズム(統計学，機械学習，計算科学)

モデリングの結果得られた計算問題を，実行するのためのアルゴリズムを議論する.

Igarashi, Nagata, Kuwatani, Omori, Nakanishi-Ohno and M. Okada, “Three Levels of Data-Driven Science”, *Journal of Physics: Conference Series*, 699, 012001, 2016.

内容

- 自己紹介
- 計算論的視覚的脳科学
  - David Marrの三つのレベル
  - 運動知覚の統合のモデル
- 知覚媒質評価
  - データ駆動学の三つのレベル
- 新学術領域「スパースモデリングの深化と高次元データ駆動科学の創成」
- 地球データ駆動科学の展望

内容

- 自己紹介
- 計算論的視覚的脳科学
  - David Marrの三つのレベル
  - 運動知覚の統合のモデル
- 知覚媒質評価
  - データ駆動学の三つのレベル
- 新学術領域「スパースモデリングの深化と高次元データ駆動科学の創成」
- 地球データ駆動科学の展望

新学術領域研究　平成25～29年度
スパースモデリングの深化と高次元データ駆動科学の創成

領域代表岡田真人の個人的な狙い

世界を系統的に記述したい
その方法論と枠組みを創りたい
ヒトが世界を認識するとは?

実験・計測

モデリング

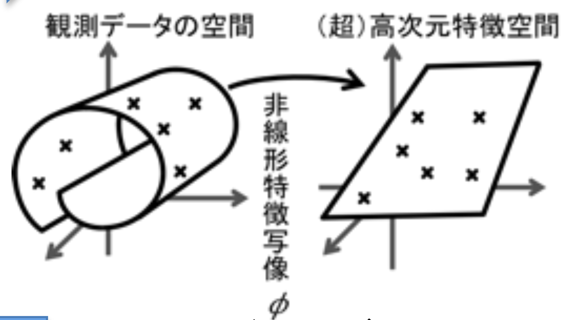
データ　　説明変数

$\boldsymbol{y} =$ 基底Φ $\boldsymbol{x}$

非0　0

数理・情報

性能評価



アルゴリズムの
高速・高精度化

新学術SpMの若手立役者

研究打ち合わせに同席することで、研究者の素養を見抜く力と、プロジェクト立案のノウハウをOJTで習得し、自らのキャリアップに繋げる。

大森 敏明
神戸大　准教授
当時: 東大　助教

永田 賢二
NIMS　主任研究員
当時:東大　特任研究員

桑谷 立
JAMSTEC　GL
当時:東大　特任研究員

地球データ駆動科学の今後展望

- MRFモデルを応用することで，地震波速度構造(Vp, Vs)から流体の分布(Φ, α)を推定する枠組みを構築した.
  - フィルタ幅をデータのみから決定できる.（ハイパーパラメータ推定）
- 東北火山弧直下では岩石が数%部分融解.

- 第一原理だけでは取り扱えない地球科学には、データ駆動的なアプローチが必須である。
- データ駆動科学(ベイズ推定)は地球科学にとって非常に強力である.