

地球データ駆動科学の現状とこれから

桑谷 立 (Tatsu KUWATANI)

kuwatani@jamstec.go.jp

<https://researchmap.jp/7000005501>

国立研究開発法人海洋研究開発機構 (JAMSTEC)・海域地震火山部門
火山・地球内部研究センター ・ 固体地球データ科学研究グループ

海洋研究開発機構 (JAMSTEC: Japan Agency for Marine-Earth Science and Technology)

- 海洋に関する基盤的研究開発・技術開発を目的とした国立研究開発法人. @横須賀・横浜・高知・青森
- 職員: 約1000人
- 海洋・地球・生命がメインターゲット

掘削船「ちきゅう」



しんかい6500



地球シミュレータ



経歴

くわたに たつ
(桑谷 立)

- 東京大学理学系研究科地球惑星科学専攻・博士課程(-2008)
 - 変成岩岩石学 (指導教官: 鳥海光弘)
- 東京大学新領域創成科学研究科・特任研究員(2009-2013)
 - 数理・情報科学の勉強を開始
- 東北大学環境科学研究科・助教(2013- 2015)
- 海洋研究開発機構(JAMSTEC)・研究員⇒主任研究員 (2015-現在)
 - 数理・情報科学の地球科学への応用
- JST「情報計測」 さきがけ研究者, CREST 主たる共同研究者 (2016-現在)
- 神戸大学大学院 システム情報学研究科・客員准教授 (2017-現在)

もともとの専門：岩石学

岩石：固体地球の歴史や地震・火山・資源形成
解明のための唯一の物質科学的情報源.

野外調査

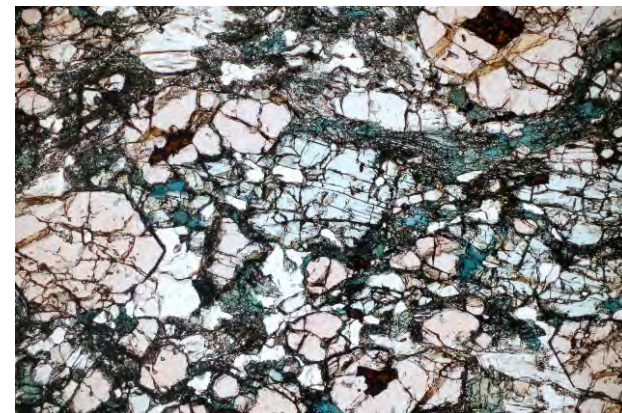


2008 @愛媛県法皇山脈

試料の
採取



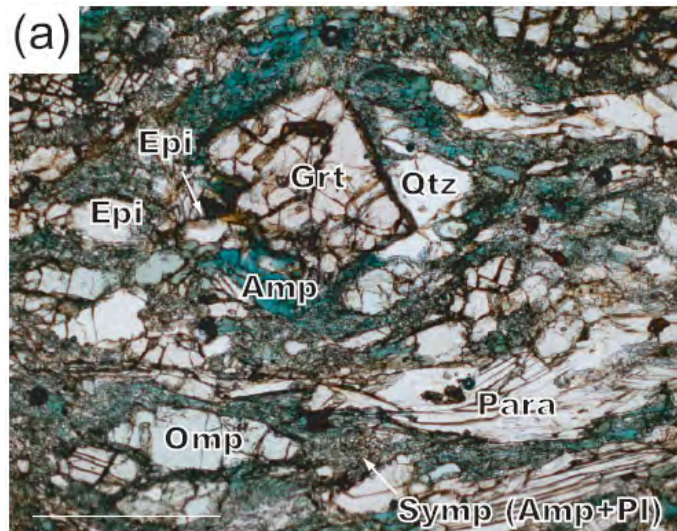
薄片
写真



数理・情報科学的な研究をはじめた動機

岩石学：岩石（結果）から，地球プロセス（原因）を推定する

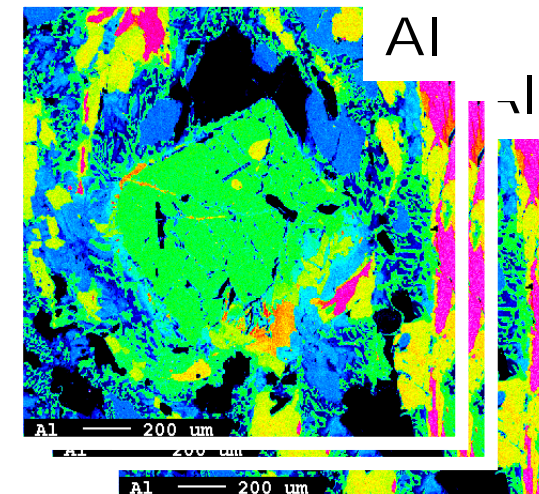
岩石薄片の顕微鏡写真



多数のプロセスが重合 &
最終結果のみから過去を推定

大量・多次元の化学組成データ

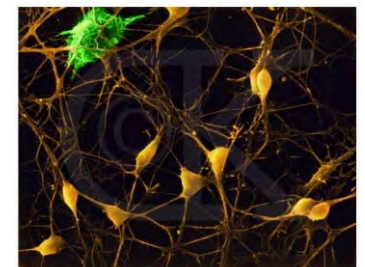
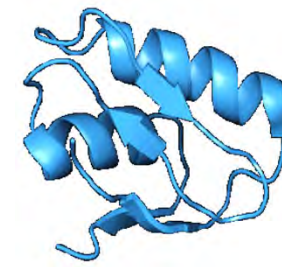
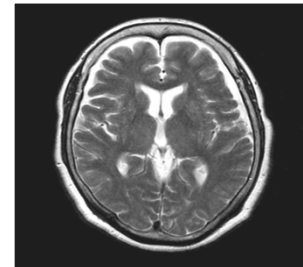
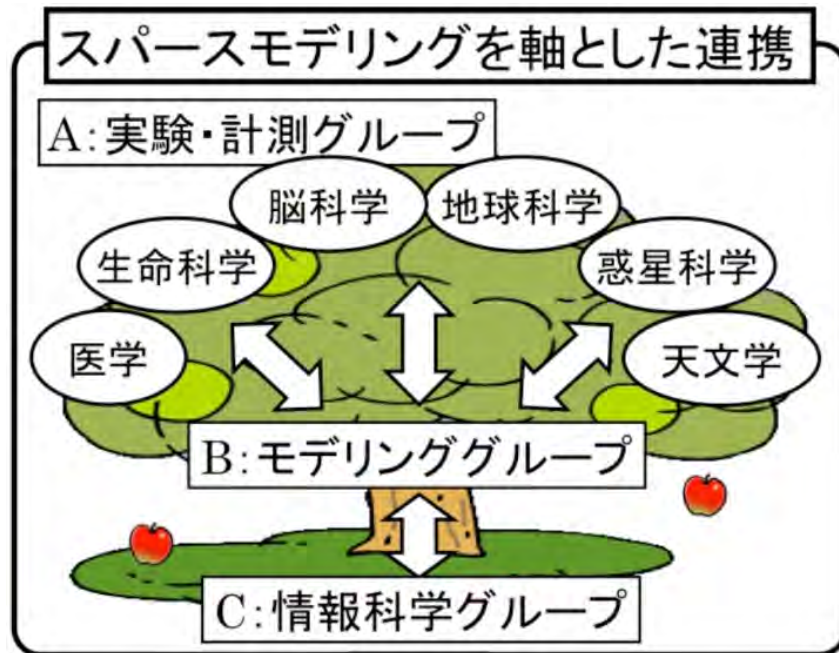
計測技術の進歩



データを最大限活用したい → 数理科学者と共同研究開始

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」(2013-2017)

- 領域代表: 岡田真人 (東大新領域)
- 様々な自然科学と情報科学の学融合型共同研究プロジェクト
 - 13の計画研究と第I, II期あわせて70以上の公募研究

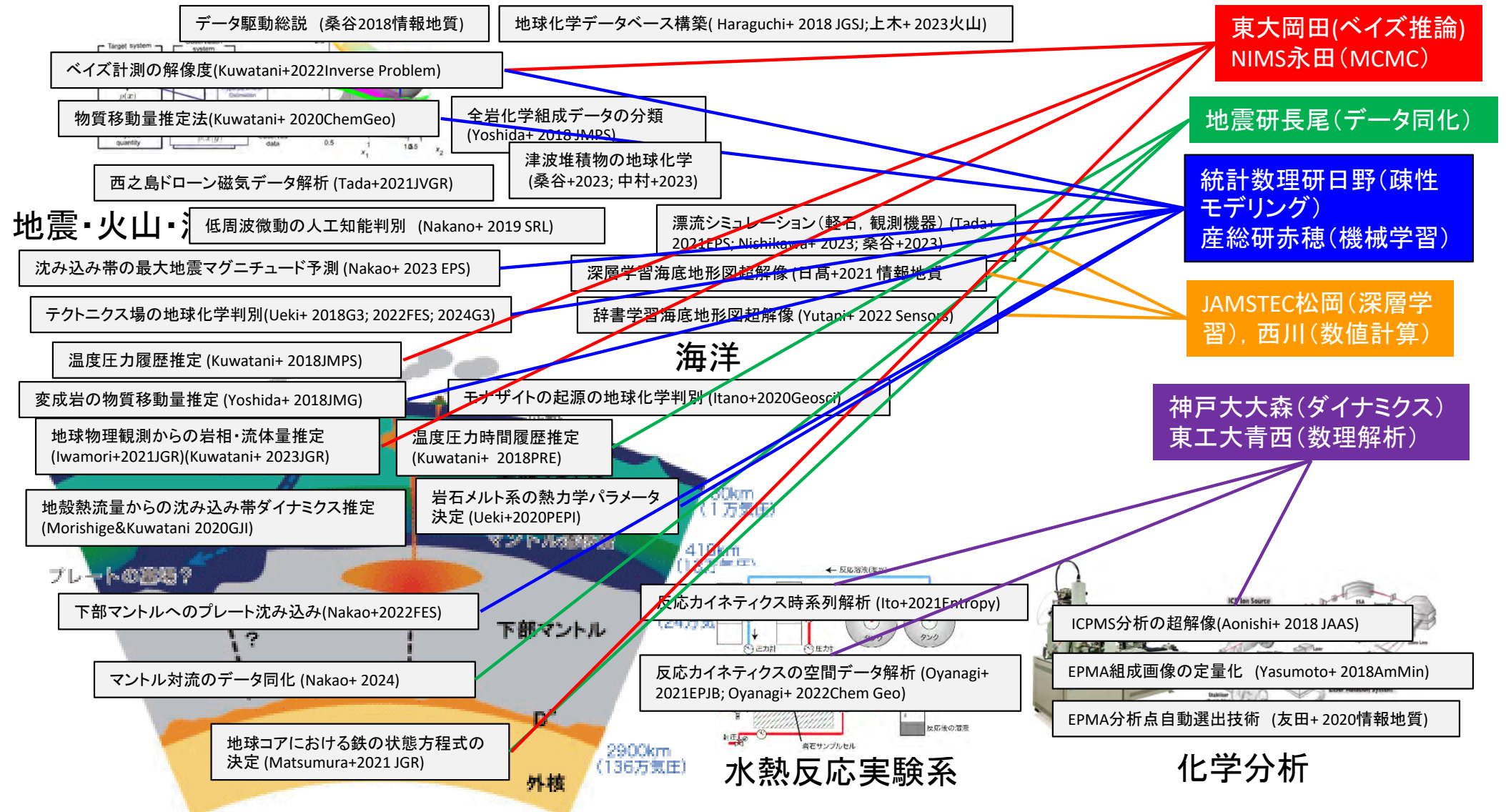


<http://sparse-modeling.jp/>

自然＋情報科学, 自然＋自然科学の学融合

コアから表層, μm から全球スケール, 多様なアプローチ(地球物理学・地球化学・地質学)を含む, 多様な対象に対してデータ科学研究を実施

数理・手法開発

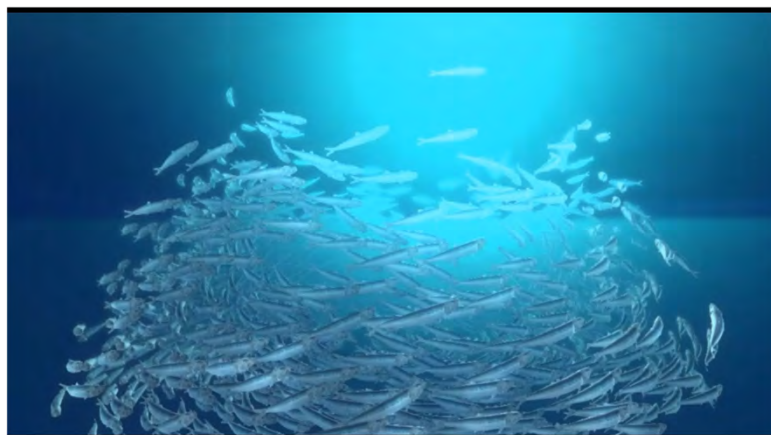


地球科学者と数理・情報科学者を橋渡しして同時多発的なイノベーション創出

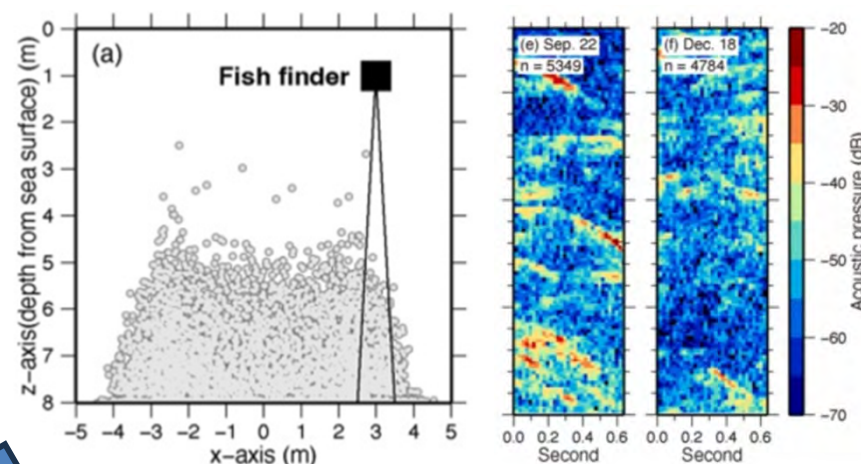
最近の研究例：魚探を用いた養殖魚計測

- 魚群探知機＋機械学習により，生け簀内の尾数を推定

生け簀網内の様子(simulated CG)

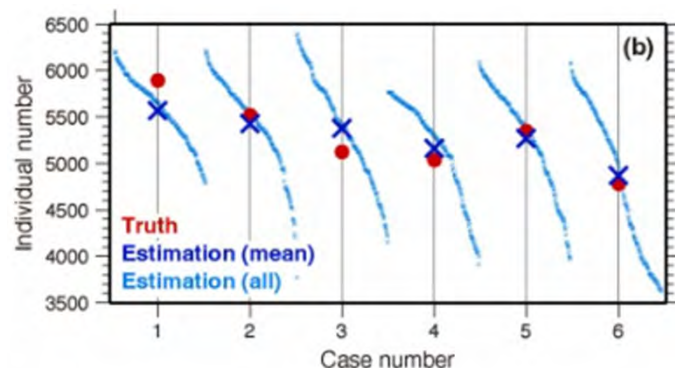


魚探による計測



古野電気様のご協力

養殖ブリの生け簀で性能検証



黒瀬水産様のご協力

CNNにより，エコー画像から尾数を推定
(学習データは魚群シミュレーションで作成)

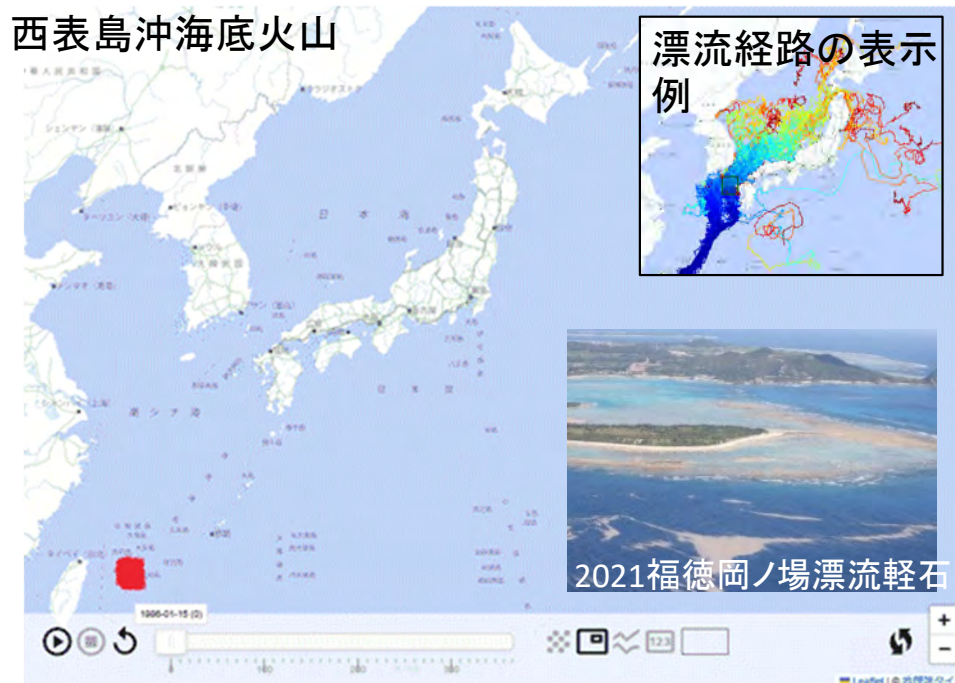
現状で7%以内の誤差⇒さらなる改善が必要

最近の研究例：海域火山由来の漂流軽石研究

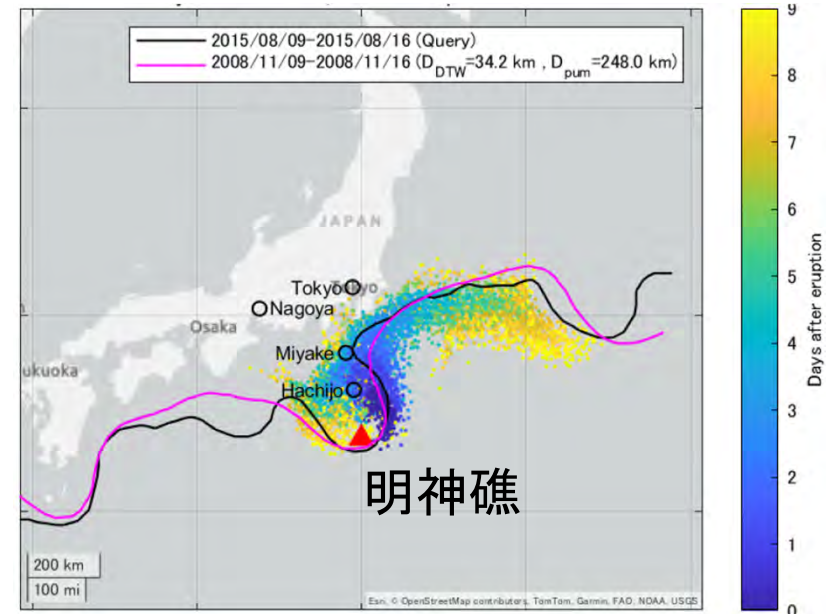
- きっかけ：2021年の福德岡の場海底火山噴火，南西諸島を中心に大量漂着し，海上交通や漁業，観光業等に大きな影響

事前のリスク評価のための漂流シミュレーション

プロアクティブデータ駆動漂流予測
(もし今日，海底火山噴火が起きたら...)



Nishikawa+. (2023, *PEPS*), 桑谷+(2023, 情報地質)



Kuwatani+. (under review in *npj Natural Hazards*.)

自治体や重要インフラ施設などのためのハザード評価システムを構築中

固体地球科学の特徴

- 知りたい物理量が観測・計測データとして直接的に獲れない
 - 内部が知りたいのに、計測データは表面のみ
 - 物質が知りたいのに、計測データは物性のみ
 - 過去が知りたいのに、計測データは現在のみ



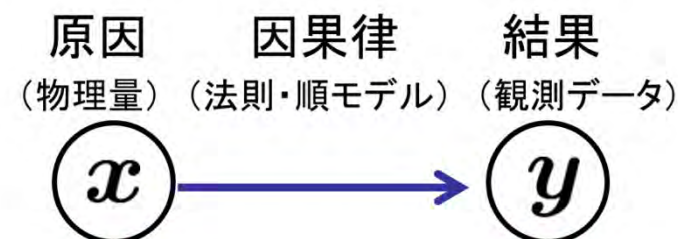
- 知りたい物理量に対してデータの量・質が十分でない
 - 連立方程式の (式の数) < (未知変数の数)

計測モデルと先験的知識を、系統的かつ確率的に
導入可能なベイズ計測が有効であると考えられる

ベイズ推論

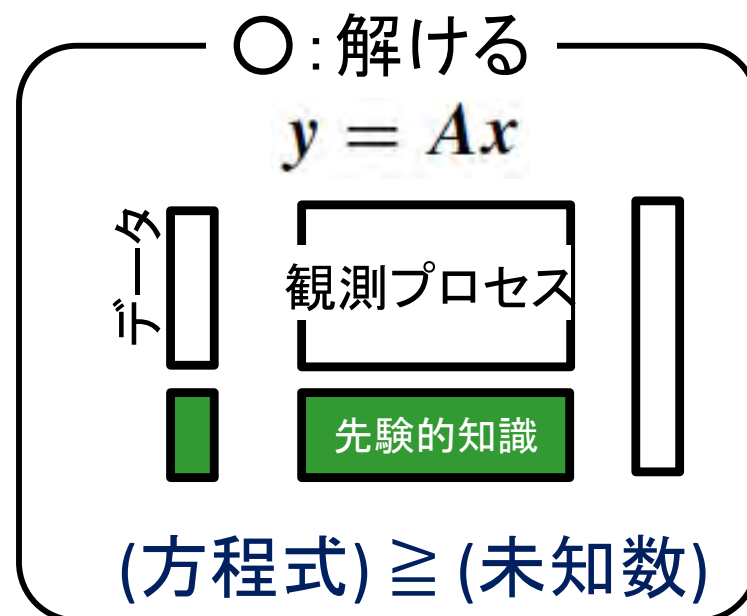
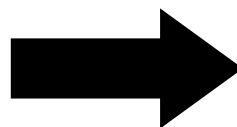
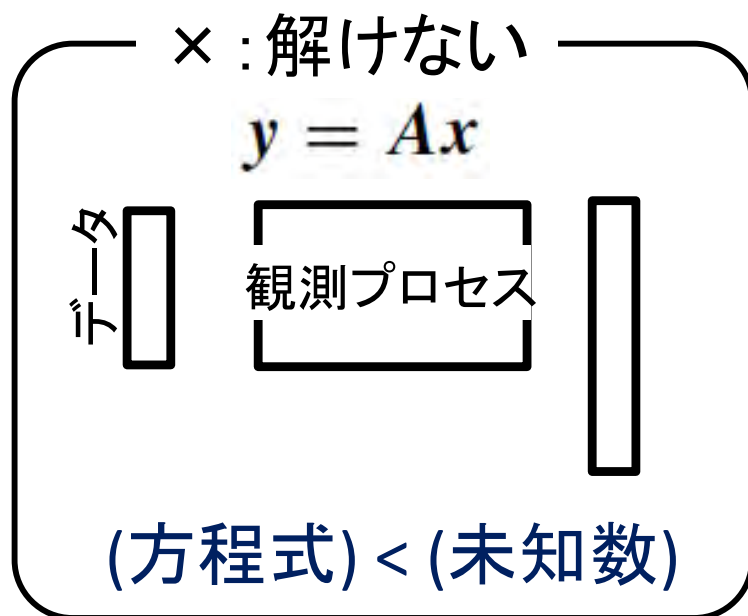
- ベイズの定理により因果律を遡り、結果から原因を確率論的に推定する逆解析の枠組み
- 観測プロセスと先験的知識を系統的に導入

空間的連続性・非負・
ゼロが多い(=スパースモデリング)



因果律を遡る

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$



ベイズ推論により、多様で難しい逆問題を系統的に扱うことが可能

ベイズ推論 の定式化

観測（データの生成過程）を確率で記述

$p(y|x)$: 尤度関数

観測プロセス（因果律）

輸送・分光スペクトル・ノイズ重畳

知りたい物理量
（原因）

x



y

観測生データ
（結果）

因果律を遡る

観測対象に関する先
験的知識：非負・連続
性・スパース性

$p(x)$: 事前確率

ベイズの定理で事後確率を書きくだす

$$\underline{p(x|y)} = \frac{\underline{p(y|x)} \underline{p(x)}}{p(y)}$$

$$E(x) \equiv -\log p(x|y)$$

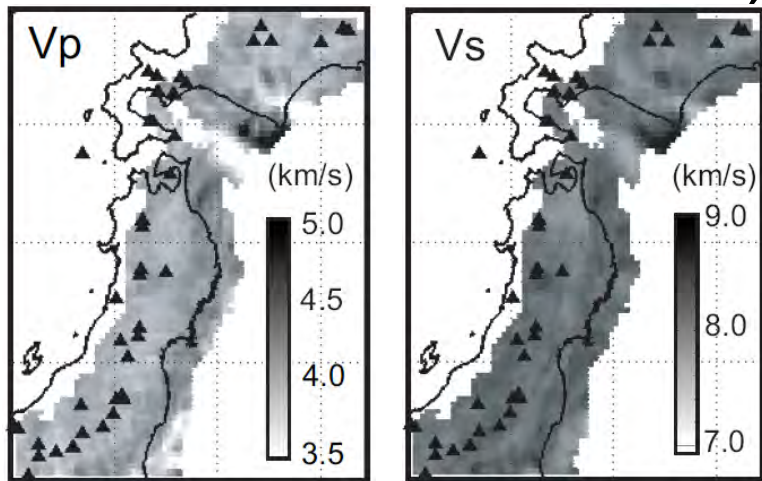
対象を深層学習（生成AI）

解析的もしくはサンプリングなどにより事後確率の分布を推定

東北日本地下深部の流体分布推定

$$E(\mathbf{x}) = \underbrace{\sum_{i=1}^n (y_i - f(x_i))^2}_{\text{観測量の再現性}} + \lambda \underbrace{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}_{\text{物理量の空間連続性}}$$

地震波速度 (V_p , V_s)



Depth=40km (Matsubara et al. 2008)

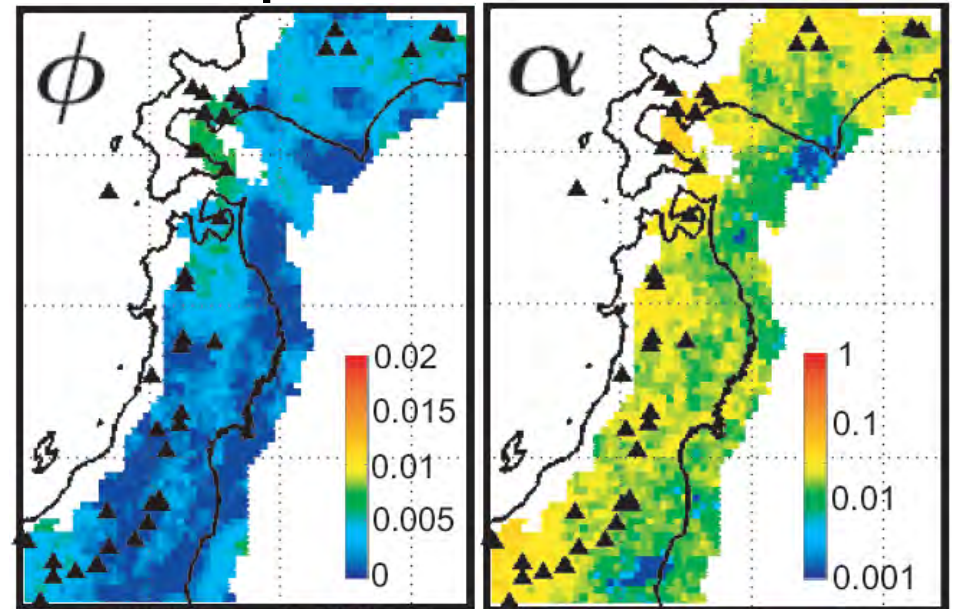
$$V_P = f_P(\phi, \alpha)$$

$$V_S = f_S(\phi, \alpha)$$

岩相 (カンラン岩)

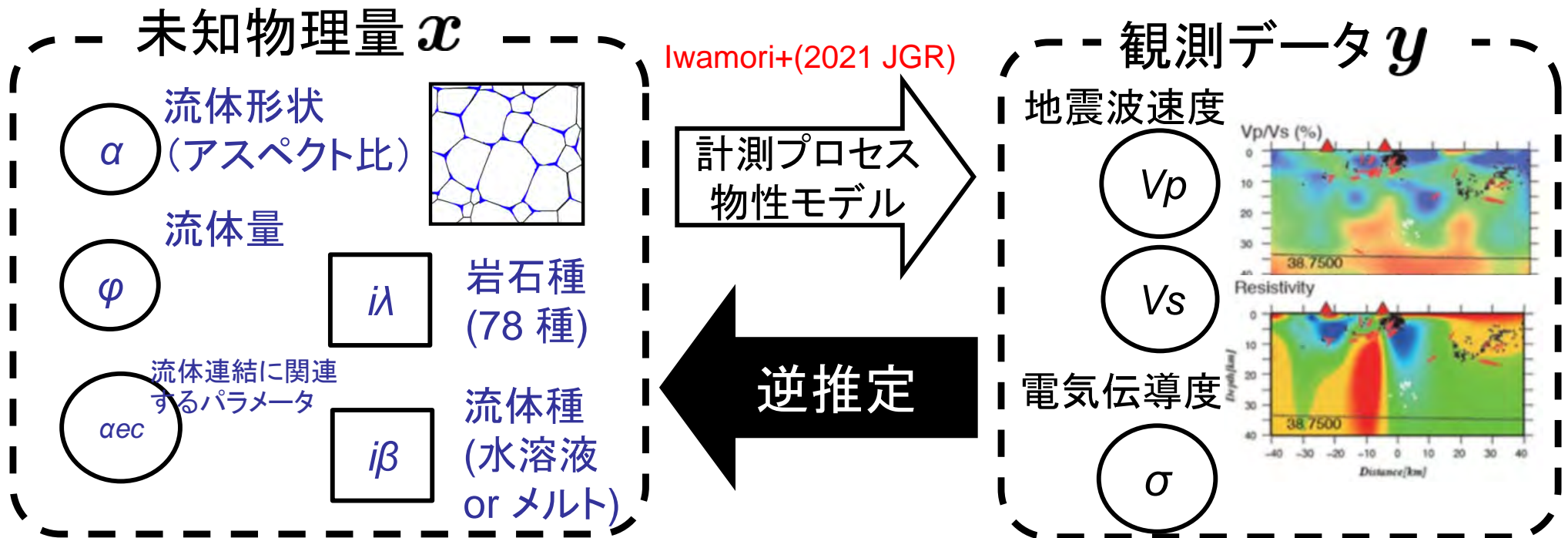
流体相 (メルト) を仮定

流体分布: Φ (porosity), α (aspect ratio)



地殻媒質評価

- 観測データから、計測プロセスを遡って、未知物理量を逆推定



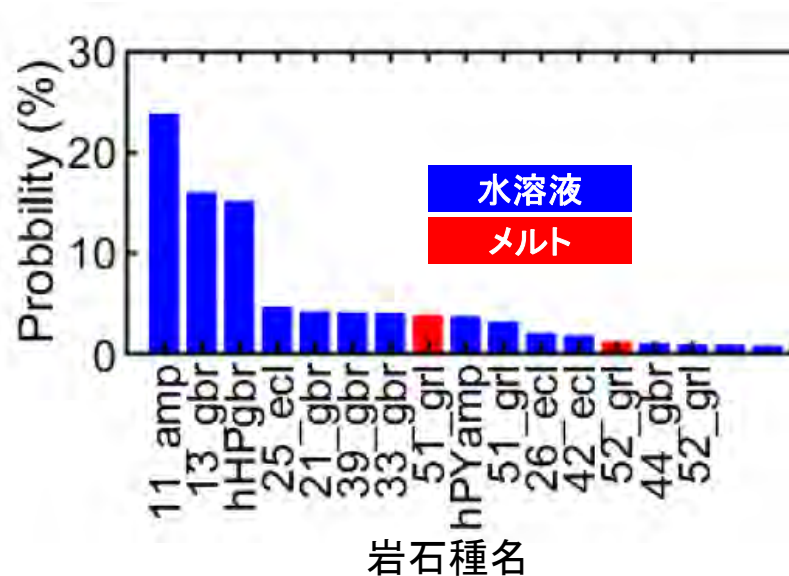
観測データよりも未知数が多い+ノイズ \Rightarrow 従来手法では困難
 \Rightarrow 確率論的アプローチが必要 \Rightarrow ベイズ計測

ベイズ計測の概念を用いた推定

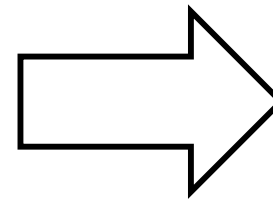
- 提案手法: 周辺化により, 岩石 & 流体種をモデル選択

$$p(\Delta|y) = \int_{-\infty}^{\infty} p(x, \Delta|y) dx$$

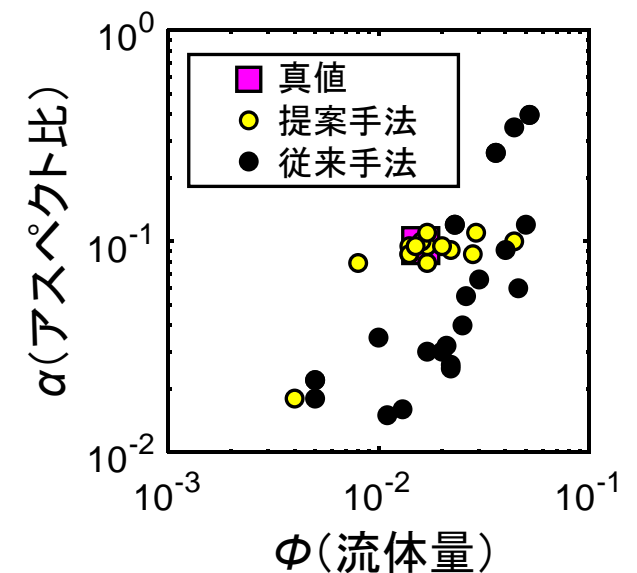
156種類の中から最適モデルを選択



モデル選択後,
流体量等を推定



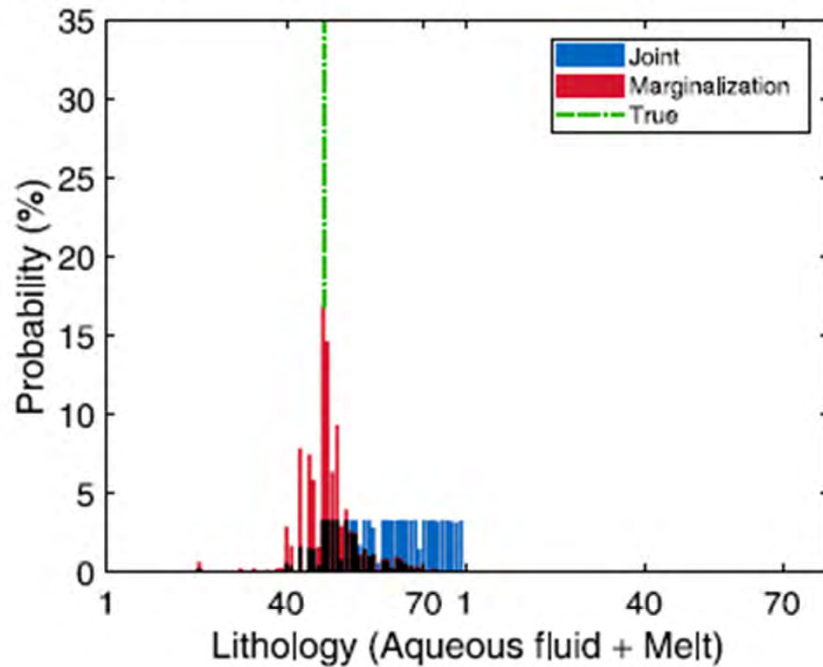
観測ノイズを変えた
20ケースの推定結果



条件の異なる数十ケースの人工データ解析テストで有効性を検証

人工データ解析例

(a) Case I



(b) Case I

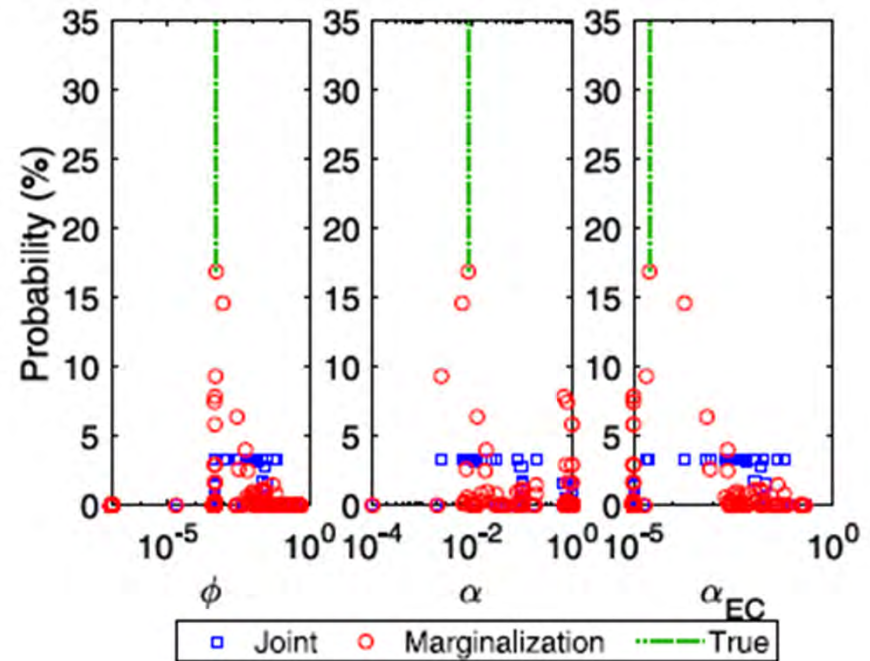


Figure 4. Estimated results for Case (i). (a) The probability of the discrete variables (lithology and liquid type). The blue and red bars indicate the joint-probability and marginalization estimation, respectively. The black ticks indicate the overlapping part between two bars. The numbers of the horizontal axis indicate the lithology and liquid type (Table 1). The lithology is listed in descending order of the bulk composition SiO_2 . Numbers 1–78 in the left part correspond to the liquid phase being the aqueous fluid, and numbers 1–78 in the right part correspond to the melt case. (b) The scatter diagrams plot the probability of the discrete variables on the vertical axis and the value of the liquid-related geofluid parameter (ϕ , α , and α_{EC}) on the horizontal axis for each combination of lithology and liquid type. The blue square and red circle indicate the joint-probability and marginalization estimation, respectively.

実データ解析例

- 鳴子火山地域データに適用
 - 地震波: Okada+(2014), 電気比抵抗: Ogawa+(2014)
 - 低比抵抗アノマリー(C1)の中心域 ($D=20\text{km}$)
 - V_p, V_s : 6.30 km/s, 3.55 km/s, σ : 0.396 S/m
- 仮定: 0.545GPa, 828.518C, 5 wt% NaCl水溶液or玄武岩質マグマ.
- 結果: porosity: 0.0437, aspect ratio: 0.871, α_{EC} : 0.0115

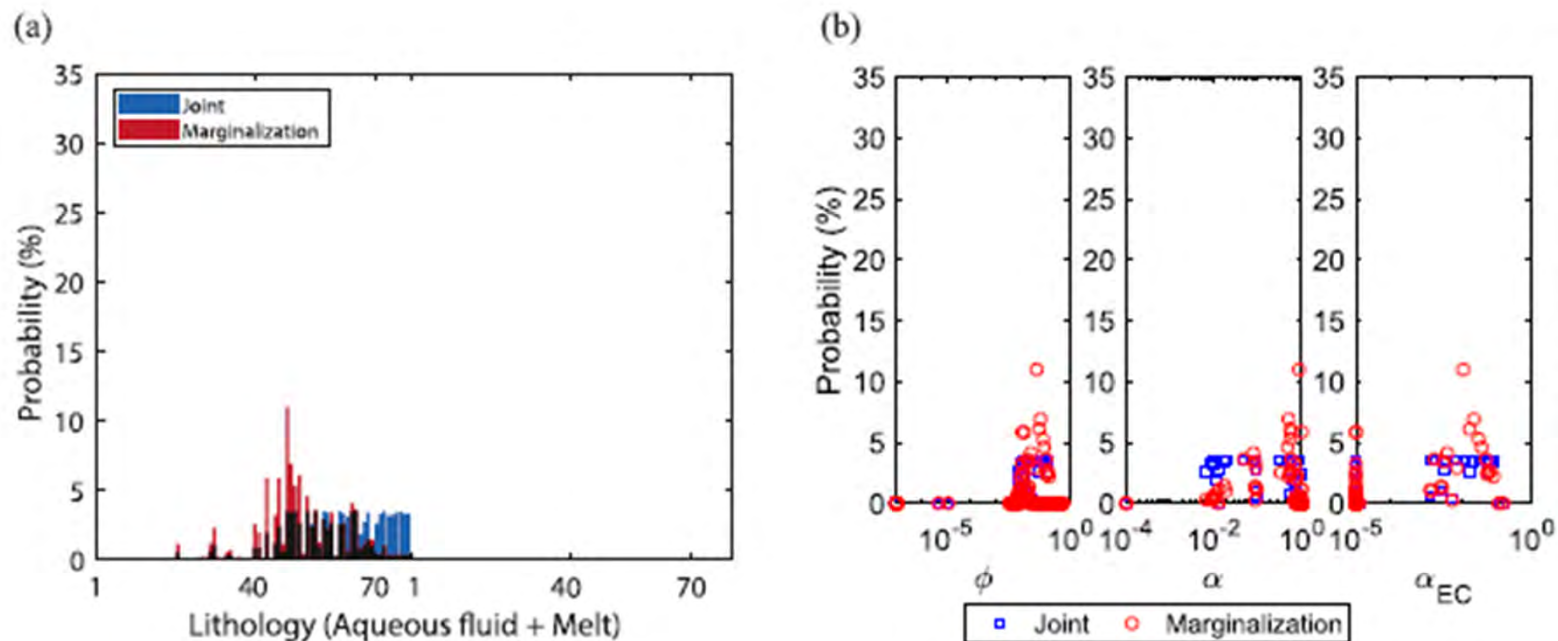


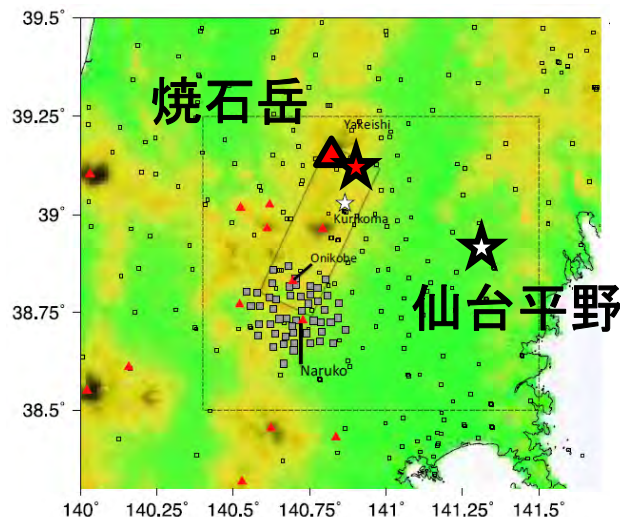
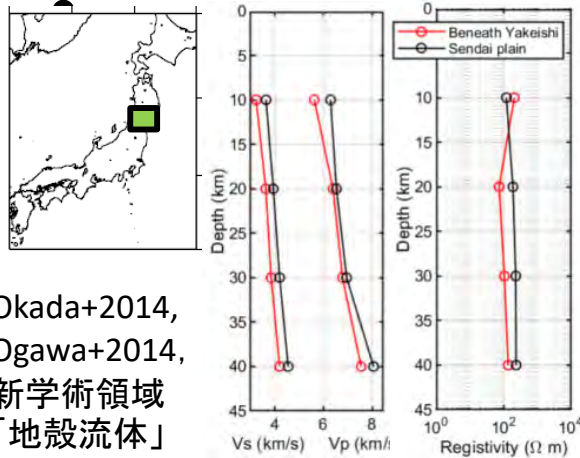
Figure A1. Estimated results for the real-world data analysis. (a) The probability of the discrete variables (lithology and liquid type). (b) The scatter diagram, which plots the probability to the vertical axis and the value of the liquid-related geofluid parameter (ϕ , α , and α_{EC}) for each combination of lithology and liquid type. Please also see the caption in Figure 4.

実データ解析結果(プレリミナリ)

未公表データを含む

- 東北地方鳴子地域の高精度・稠密観測データに適用

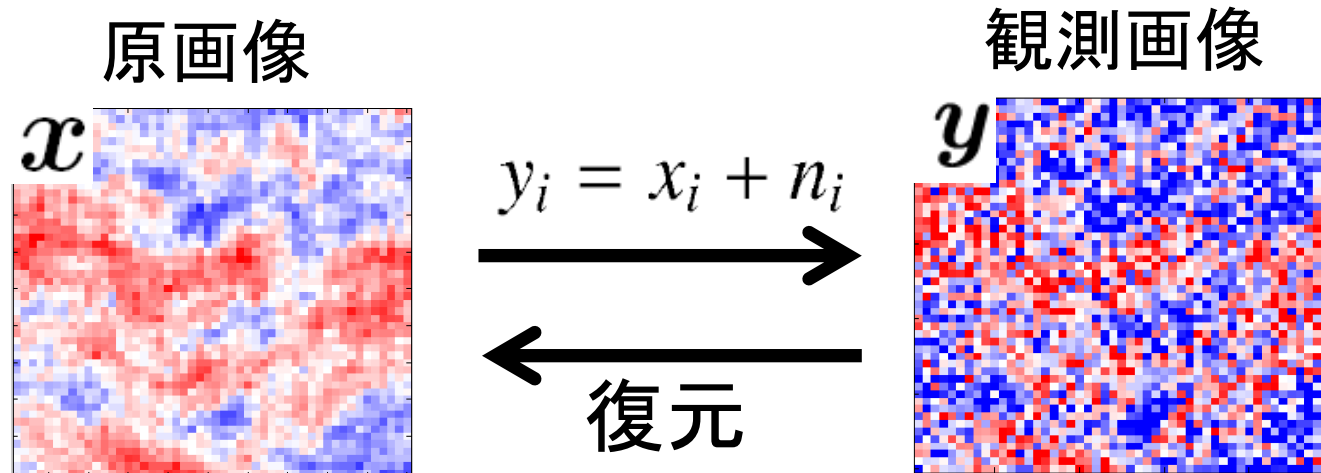
⇒ 各場所の岩石・流体種・量・形状を推定



地下深部の物質科学的な詳細を可視化
⇒沈み込み帯ダイナミクスの理解に貢献

Markov Random Field (MRF) モデル

- 画像解析等によく用いられるベイズ統計手法
- ノイズがのった観測画像から原画像を復元



評価関数

$$H(x) \equiv \sum_{i=1}^N (y_i - x_i)^2$$

観測量の再現性

地球データ駆動科学の「現状」

- 地球データ駆動科学は, SpM新学術(2013-2018)とともに創成され, それ以降も, 情報計測CREST/さきがけ(2016-)などとともに順調に発展
- 多数の研究成果
 - 毎年平均6,7編, 2025年は既に7編
- 多数の若手データ駆動科学人材が成長
 - 中尾, 上木, 吉田, 赤松, 松村, 大柳, 板野, 三木,...
- 一方, いろいろな課題もあり



地球データ駆動科学の「これから」

- 研究現場でのニーズは高い ⇒ 若手人材育成が必須
⇒ 研究コミュニティ(エコシステム)が必要
⇒ データ駆動関係の学会や国際専門誌などが理想
- 数理的本質を考えずに「ダウンロードサイエンス」化している現状も否めない.
- 原点に戻って, SpM新学術のような学融合領域が必要ではないか.