総 説

地球科学プロセス解明のためのデータ駆動型解析 —地質学分野における応用例—

桑谷 立^{1), 2)}

Data-driven Analysis towards Understanding of Geoscience Processes: Application Examples in Geology

Tatsu KUWATANI^{1), 2)}

Abstract : Recent developments in computer technology and progress in information science continue to provide powerful information processing technologies for extracting useful information at the maximum extent from an available dataset. They are called data-driven analyses and applied in various natural sciences, engineering fields and industry. In this paper, I briefly explain the concepts of Bayesian estimation and sparse modeling, which are considered as fundamental mathematical schemes of data-driven analysis, and overview the application examples which have been studied by the author's group in geology. For data analysis on rock textures, Bayesian estimation is effective, because it can introduce a priori knowledge about spatiotemporal structure and physicochemical model objectively in inversion analysis. For analysis of geochemical data, sparse modeling is particularly effective, because it can automatically extract essential combination of elements and/or governing earth-scientific processes from a high-dimensional dataset. Collaboration between information scientists and geologists leads to the development of new powerful methods for specific and general problems in geology, and the importance of data-driven analysis will increase in the future.

Key words : Data-driven science, Inverse problem, Machine learning, Bayesian estimation, Sparse modeling

1. はじめに

地質学において、様々な観察や計測・分析データをもとに地球の過去および現在の物理・化学的過程（以下、プロセスと呼ぶ）を明らかにすることは根本命題の一つである。計測・分析技術の絶え間ない発展は、これまで得られなかったような大量もしくは新たな種類のデータ取得を可能にし、地球システム・現象の理解に著しく貢献している。しかし、得られるデータは知りたい物理量やプロセスに対して不十分なことが多く、さらに精度・解像度・計測範囲などの限界の存在のため、地質学における問題解決は容易ではない。一方、近年の計算機技術の発達と数理・情報科学の進展は、取得データから最大限に情報を抽出するための強力なデータ科学手法群を生み出し続けており、機械学習や人工知能は産業界や実社会などで既に広く活用

されている（Bishop, 2006）。これらはデータ駆動型解析として、自然科学や工学分野においても応用が進んでおり、それぞれの分野に関する課題解決に貢献しつつある。

このような状況を背景に生まれた学問分野がデータ駆動科学である。データ駆動科学は、自然科学者と数理・情報科学者が協働することで、データの背後に潜む構造やプロセスの抽出に関して、特定の対象や分野に依存しない普遍的な数理・情報科学的方法論の構築を目指している（Igarashi *et al.*, 2016）。2013年に、文部科学省科学研究費補助金新学術領域研究「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」（略称：疎性モデリング、領域代表者：岡田真人、助成期間：2013年度～2017年度、<http://sparse-modeling.jp/>）と題した大型研究プロジェクトが開始し、地球科学・惑星科学を含む生物学・地学を中心

2018年4月26日受付 2018年5月9日受理

日本情報地質学会・地質情報整備活用機構 共催シンポジウム2017（2017年11月、東京）で一部講演。

1) 国立研究開発法人海洋研究開発機構地球内部物質循環研究分野 Department of Solid Earth Geochemistry, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka 237-0061, Japan. E-Mail: kuwatani@jamstec.go.jp

2) 国立研究開発法人科学技術振興機構戦略的創造研究推進事業さきがけ

PRESTO, Japan Science and Technology Agency (JST), Kawaguchi 332-0012, Japan.

第1図 新学術領域「疎性モデリング」の組織構成

とする幅広い自然科学分野において、最先端のデータ解析手法を応用する研究が活発に行われている(第1図)。異なる分野が多数協働することで、ある分野で用いられているデータ解析手法が、別分野にスムーズに伝搬され、従来、個別分野内において解決が困難だった問題に対しても、新たな数理・情報科学的解決方法が見つかるようなケースも生まれてきている。

冒頭にも述べたように、地質学分野では複雑なシステムに関する挙動と構造を常に不十分な観測・計測データから解明するという難解な逆問題に取り組まなければならない。そのため、地質学はデータを最大限に活用して情報抽出を行う数理統計解析や機械学習、つまり、データ駆動型解析の恰好の適用対象といえよう。実際に、地質学分野では古くから数理統計解析手法が応用されているほか(e.g. Davis, 2002)、地質学と情報科学の融合分野である「情報地質学」が発展してきた(e.g. 弘原海, 1990; 正路, 1998)。著者らの研究グループにおいても、新学術領域「疎性モデリング」の発足前後から現在に至るまで、地質学・地球化学・地球物理学・環境科学を含む幅広い地球科学関連分野の様々な対象に対して、データ駆動型解析を適用してきた。

本総説では、はじめに、データ駆動型解析のキーテクノロジーであるベイズ推論(逆解析に順モデルと先験的知識を有効に導入する確率論的アプローチ)とスパースモデリング(高次元データから少数の本質的な説明変数を自動抽出する数理的方法論)についてその概念を簡単に解説する。その後、著者らのグループがデータ駆動型解析を地質学分野、主に岩石および表層堆積物の化学組成データへ応用した研究例を紹介する。これらの研究の多くは、地質学者と数理・情報科学者との密接な連携により生まれた成果である。最後に、本稿の内容をまとめるとともに、現状の課題と今後の展望について議論する。

2. データ駆動科学のキーテクノロジー

データ駆動科学において最も重要な数理的方法論はベイズ推論とスパースモデリングであろう。前者は、系や対象に関するモデル・先験的知識が既知のときに、それらを最

大限に活用して、対象にアプローチすることが可能である。後者は、系や対象に関するモデルが未知の場合にモデル自体を抽出する問題や、推論対象の構造に関してゼロ要素が多いこと、つまりスパース性(疎性)があらかじめ予想される問題などに対して、その威力を発揮する。

2-1. ベイズ推論

ベイズ推論は、計測・観測データ y から知りたい物理量 x を、ベイズの定理により推定する確率論的な逆解析法である(第2図)。ベイズの定理は、観測量 y が与えられた下での物理量 x が得られる条件つき確率を意味する事後確率 $p(x|y)$ を記述する恒等式である：

$$p(x|y) = \frac{p(y|x) \cdot p(x)}{p(y)} \quad (1)$$

ここで、 $p(x|y)$ は物理量 x が与えられた下での観測量 y の生成・出現確率を示し、尤度関数と呼ばれる。 $p(x)$ は物理量 x が計測・観測に関わらずもともと満たす確率分布であり、事前確率と呼ばれる。 $p(y)$ は規格化定数としての意味しか持たないため、実際の解析では省略されることが多い。ベイズ推論では、取得した観測データと先験的知識の下で、事後確率 $p(x|y)$ の分布を決めることで、もっとも確からしい物理量 x を推定する。

第2図 ベイズ推論の枠組み。因果律を遡り、結果である観測データ y から原因である物理量 x を推定する。

尤度関数 $p(y|x)$ は物理法則や観測プロセスなどの順モデルを確率関数の形式で記述したものであり、事前確率 $p(x)$ は注目する物理量に関する先験的知識を確率分布により明示的に表現したものである。一方、事後確率 $p(x|y)$ は、我々の論理的・物理的思考の基盤である因果律と逆方向であるため、直接的に定式化することは難しい。式(1)で記述されるベイズの定理は、定式化の難しい事後確率を尤度関数と事前確率の積として表現することで、その解析を可能にする。つまり、ベイズの定理は、因果律を遡ることで結果から原因の推定を実現するという意味を持つ。事後確率最大化の問題は数学的単純化のため、次に定義する評価関数の最小化に関する最適化に帰着させることが多い。

$$E(\mathbf{x}) \equiv -\ln p(\mathbf{x}|\mathbf{y}) \quad (2)$$

ベイズ推論は、順モデルやさまざまな先験知識を逆解析に組込むことが可能であるため、最適化手法の発展と計算機性能の向上を背景に数多くの自然科学分野のデータ解析に導入されつつある (e.g. Gregory, 2005; Bishop, 2006)。

ベイズ推論は、その柔軟な定式化のため、様々な地質学的情報や観測機器の情報などを先見的知識として事前確率に反映させることで、解析精度の向上も期待できる。さらに、ベイズ推論を用いることで、求めたい物理量や観測量に関する確率分布を支配する重要な変数 (ハイパーパラメータ) や順モデル自体をデータのみから決定することができる。これは、データのみから、より深い階層に存在する物理過程や潜在構造などを抽出することを意味している。また、ハイパーパラメータ推定の枠組みは、地球物理学においてジョイントインバージョンと呼ばれるような、様々な観測データを統合する逆解析手法の基盤技術になるものと考えられる。従来型の地球科学データプロセッシングでは、既存の地球科学的モデルを先験的に仮定した上で解析を進めることが多かった。ベイズ推論は、大量のデータを利用することで、従来は解析者が恣意的に決めていたモデル自体を客観的に選択することができる枠組みであり、データ駆動で未知の地球プロセスや構造を系統的に導くためのキーテクノロジーとなることが期待される。

2.2. スパースモデリング

スパースモデリングとは、高次元データの持つスパース性 (疎性: ゼロが多いこと) を利用し、本質的な低次元空間を抽出する数理解析手法の総称である。抽出された低次元空間は、データの持つ構造やその生成プロセスを反映していると考えられることから、対象とする系に成立する隠れた物理法則やモデルの推定に役立つものと期待される (第3図)。

スパースモデリングの最も代表的な数理解析技術が、評

第3図 スパースモデリングの概念図。

高次元データから本質的な低次元空間を抽出する。

価関数にL1正則化項を導入し最適化を実施する圧縮センシング (Tibshirani *et al.*, 2005; Donoho, 2006) である。圧縮センシングに関しては、田中 (2010) の解説論文などを参照いただきたい。未知物理量 $\mathbf{x}$ に関する既知の線形観測プロセスが $\mathbf{A}$ で表現され、観測値 $\mathbf{y}$ が得られるような観測方

程式を考えた場合、圧縮センシングの評価関数は、たとえば下記のように表現できる。

$$E(\mathbf{x}) = (\mathbf{y} - \mathbf{A}\mathbf{x})^2 + \lambda |\mathbf{x}| \quad (3)$$

ここで、 λ は正則化パラメータ、 $|\mathbf{x}|$ はベクトル $\mathbf{x}$ のL1 (絶対値) ノルムを示す。右辺第一項は観測データの再現性を促し、第二項は解の疎性を促す項である。よって、評価関数を最小化する推定解は、観測量の再現性と推定解の疎性を同時に満たす。この評価関数は線形計画法を利用した高速な最適化が可能であるため、現実的な計算時間で疎な解 $\mathbf{x}$ を求めることができる。また、評価関数の右辺第一項は尤度関数に対応し、L1正則化項は事前知識としての解のスパース性に対応することから、ベイズ推論の枠組みの中で数理的に解釈することも可能である。

近年、L1最適化はその数理解析の進展を背景に、医療画像解析をはじめとして天文学など様々な分野に適用され、イメージングの超解像や計測時間短縮化を実現している (e.g. Lustig *et al.*, 2007; Honma *et al.*, 2014)。地球科学分野においても、例えば全球測位衛星システム (GNSS) により得られた測地学的観測データからの地球内部断層変位の逆解析に用いられている (e.g. Evans and Meade 2011; Nakata *et al.*, 2016; Nakata *et al.*, 2017)。そもそも、L1最適化は、実用的な逆問題の解析法として、反射法地震波探査などを中心に1970年代から利用されてきた (e.g. Claerbout and Muir, 1973; Taylor *et al.*, 1979)。L1最適化の地球物理学への応用の詳細は、Tarantola (2005) や桑谷ほか (2016) を参照されたい。

一方、実際に解くべき問題によっては、圧縮センシングがうまく機能しない場合も存在する。圧縮センシングは現実的な計算時間内で解を得ることができる有効な方法であるが、常にL0最適化 (L0ノルムはベクトル要素中の非ゼロの個数を意味する) と同じ解を出すとは限らない。むしろ、実データ解析においてはこのようなケースは特殊である。このため、自然科学分野の研究対象に圧縮センシングを適用する際には常に注意を払う必要がある。自然科学分野では、コストや計算時間がかかったとしてもグローバルな最適解およびその分布を推定したいことが多い。そのような場合には、全ての変数に関して「利用する/利用しない」の2通りの変数組み合わせについて、モデルの評価を行う「組み合わせ全探索法」が有効である (Igarashi *et al.*, 2016)。モデルの評価基準としては、交差検証法により計算した汎化性能を比較することが多い。汎化性能は、未知データへの予測性能を意味し、交差検証法とは、手持ちのデータを訓練データとテストデータに分割することで汎化性能を推定する方法のことである (e.g. Bishop, 2006)。

組み合わせ全探索法は、次節で説明する津波堆積物の地球化学判別 (Kuwatani *et al.*, 2014c) の他、神経生理学実験によるサルの個体識別の問題 (Nagata *et al.*, 2015)、脳活動計測による発達障害症状の判別 (Ichikawa *et al.*, 2014)、Ia

第4図 鉱物粒子からの温度圧力履歴推定。(a)ザクロ石粒子の鉱物組成累帯構造を示すFe元素のX線強度画像。(b)分析された化学組成データ。データは常にノイズを含む。(c)推定される温度圧力履歴。従来手法ではノイズの影響により、推定値が大きく振動してしまう。桑谷ほか(2014)を改変。

型超新星の極大等級説明変数抽出(Igarashi *et al.*, 2018)など幅広い分野に適用されている。変数が多い場合は、組み合わせの個数が指数関数的に増えるため、現実的な計算時間で解を求めることはもちろん不可能であるが、最近では、最適解の全体的な特徴を捉えるための近似的な全探索法なども開発されている(Nagata *et al.*, 2015)。

圧縮センシングと組み合わせ全探索法が最も代表的なスパースモデリングの数理解析技術であるが、一般的な多変量解析手法に関しても次元圧縮に利用できるものは全てスパースモデリングと呼ぶことができる。主成分分析(PCA: Principal Component Analysis)は、多次元データの直交変換により、互いに無相関の合成変数(主成分基底)を抽出する代表的な多変量解析手法である。固有値の小さい主成分基底を無視することで、情報量の損失をできるだけ少なくしたうえでの次元圧縮が可能である。数学的な原理が単純でアルゴリズムも頑健である主成分分析は最も汎用的なスパースモデリングであるといつてよい。

3. 地質学分野における応用例

本節では、著者らのグループが行ってきた地質学分野に関連する研究例を中心に、解析対象により便宜的に3つに分けて紹介する。

3-1. 岩石組織データ等への応用

岩石学は、現在取得できる岩石という限られた情報源から、岩石の経験した過去の歴史や地球内部の物理化学過程を明らかにする学問分野である(岡本・桑谷, 2017)。岩石はその組織と化学組成に過去の情報を保存しており、岩石試料中を計測・分析した各ポイントの化学組成データをもとに、岩石形成時の物理化学条件を推定することが最も一般的な研究アプローチである。データ駆動型解析を岩石組織の解析に適用した先駆的な研究として、ベイズ推論により変成岩の温度圧力履歴を推定したKuwatani *et al.* (2012; 2018)を紹介する。

変成岩中には、しばしば組成累帯構造を持つ鉱物粒子が存在する。組成累帯構造とは、鉱物粒子の中心部から縁辺部に向かって、元素の化学組成が同心円状に変化している

ことを示す(第4図a)。化学組成の変化は岩石の経験した温度 T と圧力 P の変化(温度圧力履歴)の様子を反映しており、熱力学的な逆解析により温度圧力履歴を推定することが可能である。しかし、実際には化学組成データ y はノイズなどを含んでいるため、従来手法ではおおまかな変化の傾向しか捉えられないことが多かった(第4図b, c)。そこで、ベイズ推論的画像解析手法の一種であるマルコフ確率場(MRF: Markov Random Field)モデル(e.g. Geman and Geman, 1984; 桑谷ほか, 2014a)を適用した。尤度関数 $p(y|P, T)$ は化学平衡論に基づく熱力学関数で表現し、事前確率 $p(P, T)$ は温度・圧力の時間に対する連続性を基に数学的に仮定した。このように推定したい物理量の連続性の仮定を導入したベイズ推論的な画像解析手法をMRFモデルと呼ぶ。最大化したい事後確率 $p(P, T|y)$ はベイズの定理に基づき式(1)により計算することが可能である。実際の計算では、式(2)で定義される評価関数を最小化することで温度と圧力の連続的な推移を推定することができる。Kuwatani *et al.* (2012)は、岩石中に存在する全ての鉱物が平衡であるという仮定が成立している場合の温度圧力履歴推定法を確立している。Kuwatani *et al.* (2018)では、同様の手法をより汎用的な鉱物包有物温度圧力計に適用するように拡張することで、手法の適用範囲を大幅に拡大している。

上記の温度圧力履歴推定の例では、電子プローブマイクロアナライザ(EPMA: Electron Probe Micro Analyzer)によって、1次元プロファイル上において離散的に定量分析された化学組成データに応用している(第4図b)。最近になって、EPMAによって得られる2次元定性面分析データ(第4図a)を定量化学組成図に変換する試みが行われている(e.g. Lanari *et al.*, 2014)。Yasumoto *et al.* (in press)では、定性面分析データから定量組成データへの変換する関数を理論に基づいて改善するとともに、鉱物の自動同定の過程に密度ベースクラスタリング(Titterton, 1985)を適用することで、微細組織を含む広範囲の2次元定量化学組成図の高速作成を可能にしている。このような試みは、従来の1次元の化学組成データと比較して、圧倒的な情報の増加をもたらすため、岩石組織の定量的な理解を各段に進めるきっかけになるであろう。

第5図 多変量解析の例。 y および a はそれぞれ観測データと抽出された基底ベクトルを示す。(a)主成分分析 (PCA), (b)独立成分分析 (ICA), (c)非負値行列分解 (NMF), (d)クラスタ分析, (e)判別分析。

岩石の形成プロセスは多数の要素が様々な時間スケールで複雑に絡み合うため、その解析は非常に困難である。さらに、推定対象が過去の地球内部のプロセスであるため、得られた解の検証自体も難しい。そこで、岩石-水相互作用に関するいくつかの問題について、検証可能および時間的情報を得られる水熱実験との対応を通して、物理現象の本質を残しつつ最大限の単純化を行い、順モデルの構築およびデータ駆動型の逆解析を進めている。

順モデル構築の例として、地球内部の水理学的性質や力学的性質に決定的な影響を与えるシリカ (SiO_2) の流体内部沈殿現象の核形成プロセスの解析を実施した (Okamoto *et al.*, 2015)。本研究では、シリカの沈殿の際に出現する複数の準安定相の表面エネルギーの効果を統計力学のポツツピンで表現し、その時間発展を交換モンテカルロ法 (Hukushima and Nemoto, 1996) で数値計算することで、オストワルドの段階則 (物質析出時に準安定相が段階的に出現し最終安定相まで変化する現象) の再現に成功している。将来的には、開発した順モデルを岩石組織データの逆解析に導入することで、岩石形成時の岩石-水相互作用の復元を試みる予定である。

データ駆動型解析の例としては、実験系の時系列化学分析データから非線形反応ダイナミクスを抽出する基礎的な数理的方法論の構築を進めている。Omori *et al.* (2016) では、固相の表面積に反応速度が依存するような不均質反応について、化学組成の時系列データから状態空間モデルを用いて、反応定数を推定するベイズ逆解析法を開発した。現在は、対象を時空間データに拡張し、分析データから反応係数と拡散係数を同時に推定する手法を開発中である。さらに、天然の岩石に近い複雑な化学反応系への適用や、反応速度を支配する表面積幾何モデルの自動的選択手法の開発にも取り組んでいる。

ここでは、化学反応に関連する岩石組織を中心に取上げたが、変形プロセスに関連する組織にもデータ駆動型解析は有効である。最近では、ブーディン構造と呼ばれる岩石中の鉱物粒子の破断構造組織についての数理モデル化や情報量基準に基づいた物理モデルの選択などがなされている (Matsumura *et al.*, 2017a, b; in press)。

3-2. 岩石の全岩組成データへの応用

岩石試料群の主要・微量元素濃度や同位体組成からなる全岩化学組成データセットは、起源物質やその試料群が被った様々な地球科学プロセスを反映している。これらの高次元データから本質的な物理化学プロセスや構造を抽出し、地球内部の現象と構造を理解することが固体地球化学の本質である。これまで、固体地球化学では、理論的もしくは経験的に重要とされる特定の元素および元素比からなる2, 3次元空間上において元素組成をプロットし定性的にプロセスを推定する方法や、スパイダーグラムと呼ばれる横軸に元素、縦軸に対数スケールで濃度をプロットしたグラフ群の形状から議論することが一般的であった (e.g. Rollinson, 1993)。その一方で、1970年代から多変量解析を用いて、多数試料、多元素組成のデータセットから岩石学的な知見を得ようとするデータ駆動型の研究が主流ではないものの進められてきた (e.g. Le Maitre, 1982)。

近年、大量に生産・蓄積されつつある地球化学データを統合した汎用のデータベースの整備が進められており (例えば、GEOROCやPetDB)、大量の地球化学データからの定量的な情報抽出の重要性がますます増してきている (e.g. Keller *et al.*, 2015; 桑谷・岩森, 2018)。これらを背景に、従来の多変量解析手法および現代的な洗練された機械学習の手法が岩石全岩組成データに適用されつつある (第5図)。岩森らのグループは、独立成分分析 (ICA: Independent Component Analysis) を地球化学分野に初めて導入し、同位体比データからマントル内の化学的な不均質構造を生むプロセスの抽出に成功している (e.g. Iwamori and Albarede, 2008)。Ueki and Iwamori (2017) では、東北地方仙岩地域の火山岩の微量元素データにスパースモデリングとしてPCAを適用し、化学組成のトレンドを作り出した鉱物分化やマグマ混合などのプロセスを特定した。Iwamori *et al.* (2017) では、k-means クラスタ分析 (KCA: K-means Cluster Analysis) の解析結果がデータの preprocessing に依存してPCAもしくはICAの解析結果と対応することを示すとともに、それぞれの分析手法の相違点と有用性について議論している。

Yoshida *et al.* (2018) は、非負値行列分解 (NMF: Nonnegative Matrix Factorization) およびKCAを用いて、主

要・微量元素全岩組成データから変成作用中の物質移動プロセスの定量化を行った。組成データは、総和が常に100%となる定数制約の性質から、一般的な多変量解析を用いた場合、得られる統計的変量にはバイアスがかかってしまう(Aitchison, 1986)。本研究では、トピックモデルと呼ばれるニュース記事内容の解析にも用いられるNMF手法を用いることで、組成データを構成する複数の端成分の推定を試みている。三波川変成帯四国中央部に分布する泥質変成岩の試料群に対して解析を行った結果、原岩のものと化学組成バリエーション由来と考えられる成分と、変成度の上昇に伴って系統的に増加する成分を分離できた。変成作用の熱力学的なモデリングと比較して考察を行った結果、脱水反応の進行に伴い、選択的に元素が物質移動する新たなメカニズムの存在が示唆された。

Ueki *et al.* (2018) では、全地球の火成岩化学組成データセットを用いたテクトニクス場判別の問題に複数の機械学習手法を適用し比較している。テクトニクス場の地球化学判別は、全岩化学組成データの最も一般的な使用法とされる。現在までに様々な地球化学判別図が提案されてきた(e.g. Pearce and Cann, 1973)。近年になって線形判別関数などを用いた客観的かつ定量的な判別がなされてきた。特に最近では、Petrelli and Perugini (2016) によって、サポートベクトルマシン(SVM: Support Vector Machine)と呼ばれる機械学習の代表的な判別手法(Cortes and Vapnik, 1995)の適用がなされている。Ueki *et al.* (2018) では、SVMの他に乱数決定木(RF: Random Forest) (Breiman, 2001) およびスパース多項式ロジスティック回帰(SMR: Sparse Multinomial Regression) (Krishnapuram *et al.*, 2005) と呼ばれる判別手法を導入した。RFやSMRを利用することで、単に予測性能の高い判別を行うのみではなく、テクトニクス場判別に重要な元素やその組み合わせを知ることができ、地球科学プロセスの解釈にも貢献できる。Ueki *et al.* (2018) では、それぞれのテクトニクス場を最も代表するような玄武岩マグマ化学組成を決定したほか、テクトニクス場とマグマ生成プロセスの関連性について詳細に議論している。

以上のように、データ駆動型解析は、地球化学分野において岩石形成に関連する物理化学プロセスの抽出に効果的である。一方、これまで示した研究例は試料の全岩組成データのみを解析対象にしているが、地球化学データセットは本来、試料採取の空間座標を有している。Yoshida *et al.* (in press) では、Yasumoto *et al.* (in press) が鉱物の自動分類の過程で利用した密度ベースクラスタリングを用いることで、空間的近接性の情報を事前情報として導入し地質岩体を分類する実用的な方法を提案している。この研究では、Haraguchi *et al.* (in press) によりコンパイルされた国内文献中の化学組成データセットを利用し、手法の妥当性を検証している。Haraguchi *et al.* (in press) は、試料採取地点の空間座標特定の高精度化を行うとともに、扱いやすい統一的

なフォーマットのデータベースとして公開し国際的に広く活用を促している。今後、地球化学分野において、このようなオープンデータベースを活用して、高次元・大量のデータに対しデータ駆動型解析を適用するというビッグデータ駆動の研究スタイルが定着するものと考えられる(Till, 2015)。

3-3. 地球表層物質の地球化学データへの応用

著者らのグループでは、津波堆積物や陸域土壌・河川堆積物の地球化学データを対象とした表層環境試料の地球化学的判別・分類や、データ生成の原因となった地球科学プロセスの抽出を目的としたデータ駆動型解析を実施している。ここでは主に2011年東北沖津波堆積物を対象にした研究例を中心に紹介する。津波堆積物研究を通して、今回および過去の津波の実態や発生メカニズム、環境影響などを明らかにすることは、地質学・環境科学の面した喫緊の課題といえよう(駒井ほか, 2012)。解析に利用したデータセットは、東北大学大学院環境科学研究科および産業総合技術研究所が中心となったグループによって試料採取および化学分析されたものである(土屋ほか, 2012; 川辺ほか, 2012)。

Kuwatani *et al.* (2014c) は、津波堆積物の元素含有量を用いた地球化学的判別手法に注目し、サポートベクトルマシン(SVM)と組み合わせ全探索法を用いて、2011年東北沖津波堆積物と東北地方に分布する海成堆積物の地球化学判別の高精度化を試みている(第6図)。地球化学判別においては、識別にどのような元素の組み合わせを用いるかが重要である。従来は、先述のように、既存のモデル依存もしくは経験的に少数の元素を選択したうえで低次元空間において2次元図上で地球化学判別を行うことが一般的であった。本研究では、スパースモデリングの一種である組み合わせ全探索法を用いて(Igarashi *et al.*, 2016)、識別率の高い元素の組み合わせ群を決定している。すなわち、主要元素および重金属元素からなる全18元素に対して、「利用する/利用しない」の全 2^{18} 通りの元素組み合わせに対して、交差検定法により汎化性能を比較する。その結果、測定に使った全ての元素を用いるのではなく、最適な11元素(Al, Ca, Ti, Mn, Cr, Sb, Cu, Zn, As, Cd, Pb)を選択することで、使用した試料セットに対して100%の汎化性能で判別が可能であることがわかった。

元素組み合わせの探索により、地球化学判別を高精度化すると共に、津波堆積物の起源や浸水プロセスに関する情報を取得することも可能である。第6図(b)は、判別率の高いトップ50位について、使用する元素の組み合わせと判別率の傾きを示した図である。ヒ素(As)や鉛(Pb)などの重金属類元素が津波堆積物に多く含有することが見てとれる。これらの元素組み合わせの情報などから、津波堆積物の起源物質が、鉱山起源の重金属類を多く含んだ沿岸域の海底堆積物由来である可能性が示唆されている。

第6図 2011年東北沖津波堆積物のサポートベクトルマシン (SVM) による判別。(a)元素組み合わせとして[Mg, Si, Ca]を用いた場合の津波堆積物(赤丸)と非津波堆積物(青丸)の判別面。本研究では線形の判別関数($w \cdot y + b = 0$)を仮定しており、判別面は平面で表される。[Mg, Si, Ca]は主要7元素の組み合わせの中では最高の判別率91.2%を与える。(b)元素組み合わせの全探索により得られた判別率上位50組の元素組み合わせを示した図。縦軸は元素を示し、横軸は判別率の高いものから順番に並べた元素組み合わせを示す。1~50位は100%~98.5%の判別率に対応する。判別に利用しない元素については黒色で示し、判別に利用した元素についてはその判別面の傾きをカラーマップで示した。傾きが正の元素は津波堆積物に多く含まれる傾向にある。Kuwatani *et al.* (2014) をもとに一部を改変。

桑谷ほか(2014b)では、2011年東北沖津波堆積物を対象に、主成分分析(PCA)を利用した化学特性の抽出も行っている。津波堆積物の元素含有量は、津波堆積物の起源物質や津波浸水・堆積プロセス、その後の組成改変プロセスの結果できたものである。そのプロセスを知ることは、津波堆積物の判別や津波浸水メカニズムの解明に貢献する。また、そればかりでなく、津波堆積物中に含まれる重金属類が及ぼす環境汚染リスクの定量評価や津波堆積物の土木資材としての有効活用などのための基礎的な情報にもなりうる。本研究では、化学組成のバリエーションを生み出した元素挙動プロセスを特定するために、PCAを用いた。得られた主成分について、その主成分ベクトルの示す元素の相関関係をもとに、プロセスを類推した結果、津波堆積物の化学組成バリエーションを支配する要因として、「ケイ酸塩鉱物の含有量比」、「重金属類の付加」、「海水成分の影響」の3つが重要であることが明らかになった。また、得られたサンプルごとの主成分得点を基に、試料のグループ分類や試料地点上にプロットすることで、地理的特徴などとあわせて、津波堆積物の化学特性のバリエーションについて議論を行っている。

津波堆積物に対しては、重金属類元素及びその環境影響に焦点を絞った研究も進めている(Nakamura *et al.*, 2016)。この研究では、はじめに、環境影響評価に利用される純水および塩酸に対する津波堆積物の重金属類の溶出量データを取得した。これらのデータにPCAやクラスタ分析を実施することで、地域ごとの重金属類の溶出傾向を明らかにするとともに、重金属類元素の分類を行った。その結果、毒性の特に強いヒ素・鉛・カドミウムについて、純水と塩酸で異なった溶出傾向を示すことがわかった。これは、津波堆積物がおかれる酸化還元状態に依存して、各元素の吸

着メカニズムが変化していることに起因しているものと考えられる。つまり、津波堆積物の環境リスク評価には重金属類の吸着特性を理解することが重要であることが明らかになった。今後は、堆積学的情報や地理空間の情報なども導入することで津波堆積物の判別手法を高精度化すると共に、仙台平野などにおける地層コア連続データ等に適用し、歴史大津波到達範囲の推定や現在まで見逃されてきた未知の津波イベントの発見などを目指している(駒井ほか, 2016)。

これまで述べてきた津波堆積物研究以外にも、様々な表層環境試料にデータ駆動型解析を適用している。研究例として、仙台平野の河川堆積物由来の土壤中重金属類について、因子分析を用いた環境影響評価が行われているほか(中村ほか, 2016)、原らが採取した宮城県全域および茨城県全域の土壤試料群(原ほか 2008; 産業技術総合研究所地質調査総合センター, 2014)の主要・重金属元素含有量データにPCAを適用し、地球化学的特性の検討を行っている(中村ほか, 2018)。一方、Nakamura *et al.* (2017)では、重金属類の土壌への吸着特性に対する物理化学パラメータ(pH, 陽イオン交換容量, 比表面積など)の影響の定量化を試みている。吸着試験により得られた各元素の吸着係数と各土壌の物理化学パラメータのデータセットに対して、組み合わせ全探索法を用いた線形重回帰分析により、各重金属元素の吸着係数に対する物理化学パラメータの依存性を明らかにした。データ駆動型解析により、地球表層物質の化学的特徴の詳細を明らかにし、重金属元素類の濃集・移動メカニズムなどを定量化することで、重金属類を含んだ表層物質の長期にわたる溶出量を推定し周辺環境への影響評価を行うことが可能であると考えられる。

4. まとめ

本稿では、データ駆動型解析の数理科学的方法論のコンセプトと、地球科学プロセスの解明を目的とした地質学分野への応用について紹介してきた。ここでは著者らのグループの研究成果を中心とした紹介にとどまっているが、データ駆動型解析は地質学に関連する様々な分野で急速に浸透しつつある。特に資源探査工学を含む工学的な分野では、古くからデータ駆動型解析を実践する地球統計学を発達させてきた伝統もあり (e.g. Goovaerts, 1997), 機械学習や人工知能などの最新の解析技術が既に広く応用されている (e.g. Zuo, 2017)。

冒頭にも述べたように、データ駆動型解析はデータを扱う学問分野に普遍的であるため、ある分野で有効な手法がそのまま別の分野の異なった対象にも適用できることが多い。一方、分野や対象に関して固有の問題がある場合は、その解決を動機として、数理科学的に新たな方法論の開発につながることも期待できる。これは様々な自然科学分野と数理・情報科学分野が連携することの重要性を示しており、多数の学問分野の融合領域である新学術領域「疎性モデリング」発足の核心的なコンセプトの一つである。

データ駆動型解析の基盤数理スキームとして、ベイズ推論とスパースモデリングについて紹介した。地球科学で扱う対象は時間的・空間的構造を有していることが多い。よって、時間・空間に対する物理量の連続性を先験的知識として事前確率に客観的に導入できるベイズ推論は強力である。第3節において岩石組織のデータ解析の例として紹介した温度圧力履歴推定のモデリング方法 (Kuwatani *et al.*, 2012; 2018) は、そのまま地球物理学の地震波トモグラフィ (Kuwatani *et al.*, 2014a) や地殻媒質評価 (Kuwatani *et al.*, 2014b) の問題にも適用可能である。また、熱力学モデルや反応速度論のモデルなどの物理化学モデルを尤度関数として柔軟に逆解析に導入できる点も地質学分野におけるベイズ推論の強みの一つである。

一方、地質学分野で扱われるデータの特徴として高次元データがある。特に多数の主要元素・微量元素・同位体組成からなる地球化学データは、データそのものの構造を把握することすら難しい。スパースモデリングにより、本質的な低次元空間を自動抽出することで、対象の有する隠れた真の構造や支配物理化学プロセスに関して知見を得ることが期待できる。複雑な現象が絡み合う地球科学現象において、本質的ではない次元(変数)が刈り込まれ、システムに対する解釈性が増すことは特に重要である。

地質学分野において地球科学プロセスを抽出する際の課題点をここでは二つだけ指摘したい。一つは検証不可能性である。地質学は過去の事象を解明する問題設定が多く、さらに、扱う現象の時間・空間のどちらか、もしくは両方ともにスケールが大きいため、実験などによる検証がほとんど不可能な場合も多い。このような場合に対して、データ解析の結果の検証をどのように行うべきか、常に難しい

問題である。もう一つは組成データの取扱いである。知りたいシステム・対象に関して組成データを多数集めたとしても、組成データの持つ定数と制約の存在により、示量変数に関する情報は常に欠落したままである。このことが原因となり、組成データから物理化学プロセスを定量的に推定することは困難である場合が多い。ここで挙げた以外にも地質学に特有の多数の課題が存在するが、これらの解決のためにも、それぞれの課題に応じた新しいデータ駆動型解析手法の開発が必要であり、数理・情報科学者と地質学者の密な協働が今後ますます重要となるであろう。

以上みてきたように、ベイズ推論とスパースモデリングという二つの強力な情報科学的技術を用いることで、データ駆動科学を実践する学融合研究が地質学において進みつつある。その適用範囲は、現状では個別の問題や特定の対象などに限られているが、将来的には手法の洗練により、広い時空間レンジで相互作用する複数のシステムにも拡大され、ダイナミックな地学現象の理解や地球環境変動の予測などにも貢献することが期待される。

謝辞: 本論文の投稿の機会をいただきました古宇田亮一氏に深く感謝いたします。また、編集委員長の小池克明教授および匿名査読者の方には貴重なご指摘をいただきました。ここに厚く御礼を申し上げます。本研究は、文部科学省科学研究費補助金・新学術領域研究・「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」の計画研究 A02-1 地球科学班 (研究課題番号 JP25120005) の研究成果の一部をまとめたものです。領域代表者の岡田真人教授、地球科学班研究代表者の駒井武教授ならびに班員の岡本敦博士、中村謙吾博士、吉田健太博士、土屋範芳教授に深謝申し上げます。本研究は、地球科学および数理・情報科学分野の多くの共同研究者の方々との密接な連携によって得られたものです。この場を借りて深く御礼申し上げます。また、以下の研究費からの支援も受けました：JST 戦略的創造研究推進事業 さきがけ (JPMJPR1676), JST 戦略的創造研究推進事業 CREST (JPMJCR1761), 地震研特定共同研究 (B) (2015-B-04・2018-B-01), JSPS 科研費 (JP15K20864・JP18H03820)。

文 献

- Aitchison, J. ed. (1986) *The Statistical Analysis of Compositional Data*. Chapman and Hall, London, 416p.
- Bishop, C. (2006) *Pattern Recognition and Machine Learning*. Springer, 738p.
- Breiman, L. (2001) Random forests. *Machine Learning*, vol.45, no.1, pp.5-32. <https://doi.org/10.1023/A:101093340>
- Claerbout, J. F. and Muir, F. (1973) Robust modeling with erratic data. *Geophysics*, vol.38, no.5, pp.826-844.
- Cortes, C. and Vapnik, V. (1995) Support-vector networks. *Machine Learning*, vol.20, no.3, pp.273-297.

- Davis, J. C. (2002). *Statistics and Data Analysis in Geology*. New York: John Wiley & Sons Inc., third edition, 638pp.
- Donoho, D. L. (2006) Compressed sensing. *IEEE Transactions on Information Theory*, vol.52, no.4, pp.1289-1306.
<https://doi.org/10.1109/TIT.2006.871582>
- Evans, E. L. and Meade, B. J. (2012) Geodetic imaging of coseismic slip and postseismic afterslip: Sparsity promoting methods applied to the great Tohoku earthquake, *Geophysical Research Letters*, vol.39, pp.L11314:1-7.
<https://doi.org/10.1029/2012GL051990>
- Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.6, no.6, pp.721-741.
<https://doi.org/10.1109/TPAMI.1984.4767596>
- Goovaerts, P. (1997) *Geostatistics for natural resources evaluation*. Oxford Univ. Press, New York, 512 pp.
- Gregory, P. (2005) *Bayesian Logical Data Analysis for the Physical Sciences*. Cambridge Univ. Press, 488pp.
- 原 淳子・川辺能成・駒井 武・井本由香利・杉田 創 (2008) 表層土壌評価基本図～宮城県地域～. 産総研地質調査総合センター, CD-ROM.
- Haraguchi, S., Ueki, K., Yoshida, K., Kuwatani, T., Mohamed, M., Horiuchi, S. and Iwamori, H. (in press) Geochemical database of Japanese islands for basement rocks: compilation of domestic article. 地質学雑誌.
- Honma, M., Akiyama, K., Uemura, M. and Ikeda, S. (2014) Super-resolution imaging with radio interferometer using sparse modeling. *Publ. Astron. Soc. Japan*, vol.66, no.5.
<https://doi.org/10.1093/pasj/psu070>
- Hukushima, K. and Nemoto, K. (1996) Exchange Monte Carlo method and application to spin glass simulations. *Journal of Physics Society of Japan*, vol.65, no.6, pp.1604-1608.
<https://doi.org/10.1143/JPSJ.65.1604>
- Ichikawa, H., Kitazono, J., Nagata, K., Manda, A., Shimamura, K., Sakuta, R., Okada, M., Yamaguchi, M.K., Kanazawa, S. and Kakigi, R. (2014) Novel method to classify hemodynamic response obtained using multi-channel fNIRS measurements into two groups: exploring the combinations of channels. *Front. Hum. Neurosci.*, vol.8, article 480.
<https://doi.org/10.3389/fnhum.2014.00480>
- Igarashi, Y., Nagata, K., Kuwatani, T., Omori, T., Nakanishi-Ohno, Y. and Okada, M. (2016) Three levels of data-driven science. *Journal of Physics: Conference Series*, vol.699, pp.012001:1-13.
<https://doi.org/10.1088/1742-6596/699/1/012001>
- Igarashi, Y., Takenaka, H., Nakanishi-Ohno, Y., Uemura, M., Ikeda, S. and Okada, M. (2018) Exhaustive search for sparse variable selection in linear regression. *Journal of Physics Society of Japan*, vol.87, no.4, pp.044802:1-8.
<https://doi.org/10.7566/JPSJ.87.044802>
- Iwamori, H. and Albarède, F. (2008) Decoupled isotopic record of ridge and subduction zone processes in oceanic basalts by independent component analysis. *Geochemistry, Geophysics, Geosystems*, vol.9, no.4, pp.Q04033:1-15.
<https://doi.org/10.1029/2007GC001753>
- Iwamori, H., Yoshida, K., Nakamura, H., Kuwatani, T., Hamada, M., Haraguchi, S. and Ueki, K. (2017) Classification of geochemical data based on multivariate statistical analyses: Complementary roles of cluster, principal component, and independent component analyses. *Geochemistry, Geophysics, Geosystems*, vol.18, no.3, pp.994-1012.
<https://doi.org/10.1002/2016GC006663>
- 川辺能成・原 淳子・保高徹生・坂本靖英・張 銘・駒井 武 (2012) 東日本大震災における津波堆積物中の重金属類とそのリスク. 土木学会論文集 G (環境), vol.68, no.3, pp.195-202. <https://doi.org/10.2208/jscej.68.195>
- Keller, C. B., Schoene, B., Barboni, M., Samperton, K. M. and Husson, J. M. (2015) Volcanic-plutonic parity and the differentiation of the continental crust. *Nature*, vol.523, pp.301-307. <https://doi.org/10.1038/nature14584>
- 駒井 武・川辺能成・原 淳子・坂本靖英・張 銘 (2012) 東日本沿岸における津波堆積物の性状に関する緊急調査—津波堆積物に起因する土壤汚染リスク—. GSJ 地質ニュース, vol.1, no.6, p.182-184.
- 駒井 武・桑谷 立・中村謙吾・土屋範芳 (2016) スパースモデリングによる津波堆積物の判別—津波到達推定への応用. 電子情報通信学会会誌, vol.99, no.5, pp.418-423.
- Krishnapuram, B., Carin, L., Figueiredo, M. A. T. and Hartemink, A. J. (2005) Sparse multinomial logistic regression: Fast algorithms and generalization bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.27, no.6, pp.957-968.
<https://doi.org/10.1109/TPAMI.2005.127>
- 桑谷 立・岩森 光 (2018) ビッグデータ解析. 「図説 地球科学の事典」朝倉書店.
- Kuwatani, T., Nagata, K., Okada, M. and Toriumi, M. (2012) Precise estimation of pressure—temperature paths from zoned minerals using Markov random field modeling: theory and synthetic inversion. *Contributions to Mineralogy and Petrology*, vol.163, no.3, pp.547-562.
<https://doi.org/10.1007/s00410-011-0687-3>
- Kuwatani, T., Nagata, K., Okada, M. and Toriumi, M. (2014a) Markov-random-field modeling for linear seismic tomography. *Physical Review E*, vol.90, pp.042137:1-7.
<https://doi.org/10.1103/PhysRevE.90.042137>
- Kuwatani, T., Nagata, K., Okada, M. and Toriumi, M. (2014b) Markov random field modeling for mapping geofluid distributions from seismic velocity structures. *Earth, Planets*

- and Space*, vol.66, pp.5:1-9.
<https://doi.org/10.1186/1880-5981-66-5>
- Kuwatani, T., Nagata, K., Okada, M., Watanabe, T., Ogawa, Y., Komai, T. and Tsuchiya, N. (2014c) Machine-learning techniques for geochemical discrimination of 2011 Tohoku tsunami deposits. *Scientific Reports*, vol.4, pp.7077:1-6.
<https://doi.org/10.1038/srep07077>
- Kuwatani, T., Nagata, K., Yoshida, K., Okada, M. and Toriumi, M. (2018) Bayesian probabilistic reconstruction of metamorphic P-T paths using inclusion geothermobarometry. *Journal of Mineralogical and Petrological Science*. vol.113, no.2, pp.82-95. <https://doi.org/10.2465/jmps.170923>
- 桑谷 立・永田賢二・岡田真人 (2014a) マルコフランダムフィールドモデルによる計測画像データからの潜在構造抽出. 映像情報メディア学会誌, vol.68, no.12, pp.892-896.
<https://doi.org/10.3169/itej.68.892>
- 桑谷 立・中村謙吾・渡邊隆弘・小川泰正・駒井 武 (2014b) 主成分分析を用いた次元圧縮に基づく東北地方太平洋沖地震による津波堆積物の地球化学的特性評価. 地学雑誌, vol.123, no.6, pp. 923-935.
<https://doi.org/10.5026/jgeography.123.923>
- 桑谷 立・中田令子・岡田真人・堀 高峰 (2016) スパースモデリングの地球物理学への応用. 電子情報通信学会会誌, vol.99, no.5, pp.406-410.
- Lanari, P., Vidal, O., De Andrade, V., Dubacq, B., Lewin, E., Grosch, E. G., and Schwartz, S. (2014) XMapTools: A MATLAB©-based program for electron microprobe X-ray image processing and geothermobarometry. *Computers & Geosciences*, vol.62, pp.227-240.
<https://doi.org/10.1016/j.cageo.2013.08.010>
- Le Maitre, R.W. (1982) *Numerical Petrology: Statistical Interpretation of Geochemical Data*. Elsevier, Amsterdam, 282p.
- Lustig, M., Donoho, D. and Pauly, J. M. (2007) Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, vol.58, pp.1182-1195.
<https://doi.org/10.1002/mrm.21391>
- Matsumura, T., Kuwatani, T., Ando, Y. and Masuda, T. (in press) Application of the inverse Batschelet distribution to measuring the preferred orientation of tourmaline grains. *Journal of Structural Geology*.
- Matsumura, T., Kuwatani, T. and Masuda, T. (2017a) The relationship between the proportion of microboudinaged columnar grains and far-field differential stress: A numerical model for analyzing palaeodifferential stress. *Journal of Mineralogical and Petrological Science*, vol.112, no.1, pp.25-30. <https://doi.org/10.2465/jmps.160711>
- Matsumura, T., Kuwatani, T. and Masuda, T. (2017b) Statistical model selection between elastic and Newtonian viscous matrix models for the microboudin palaeopiezometer. *Earth, Planets and Space*, vol.69, pp.83:1-12.
<https://doi.org/10.1186/s40623-017-0669-4>
- Nagata, K., Kitazono, J., Nakajima, S., Eifuku, S., Tamura, R. and Okada, M. (2015) An exhaustive search and stability of sparse estimation for feature selection problem. *IPJSJ Online Trans.* vol.8, pp.25-32. <https://doi.org/10.2197/ipsjtrans.8.25>
- Nakamura, K., Kuwatani, T., Kawabe, Y. and Komai, T. (2016) Extraction of heavy metals characteristics of the 2011 Tohoku tsunami deposits using multiple classification analysis. *Chemosphere*, vol.144, pp.1241-1248.
<https://doi.org/10.1016/j.chemosphere.2015.09.078>
- 中村謙吾・桑谷 立・駒井 武・山崎真一 (2018) 主成分分析を用いた表層土壌の元素濃度の地球科学的特徴の抽出. *Journal of MMIJ*, vol.134, no.2, pp.13-21.
<https://doi.org/10.2473/journalofmmij.134.13>
- 中村謙吾・佐藤海里・川辺能成・桑谷 立・駒井 武 (2016) 名取川水系の重金属類の環境評価に関する調査及び数値統計解析. *Journal of MMIJ*, vol.132, no.1, pp.22-30.
<https://doi.org/10.2473/journalofmmij.132.22>
- Nakamura, K., Yasutaka, T., Kuwatani, T. and Komai, T. (2017) Development of a predictive model for lead, cadmium and fluorine soil-water partition coefficients using sparse multiple linear regression analysis. *Chemosphere*, vol.186, pp.501-509.
<https://doi.org/10.1016/j.chemosphere.2017.07.131>
- Nakata, R., Kuwatani, T., Okada, M. and Hori, T. (2016) Geodetic inversion for spatial distribution of slip under smoothness, discontinuity, and sparsity constraints. *Earth, Planets and Space*, vol.68, pp.20:1-10.
<https://doi.org/10.1186/s40623-016-0396-2>
- Nakata, R., Hino, H., Kuwatani, T., Yoshioka, S., Okada, M. and Hori, T. (2017) Discontinuous boundaries of slow slip events beneath the Bungo Channel, southwest Japan. *Scientific Reports*, vol.7, pp.6129:1-7.
<https://doi.org/10.1038/s41598-017-06185-0>
- Ohno, Y., Nagata, K., Kuwatani, T., Shouno, H. and Okada, M. (2012) Deterministic algorithm for nonlinear Markov random field model. *Journal of Physical Society of Japan*, vol.81, pp.064006:1-6.
<https://doi.org/10.1143/JPSJ.81.064006>
- Okamoto, A., Kuwatani, T., Omori, T. and Hukushima, K. (2015) Free-energy landscape and nucleation pathway of polymorphic minerals from solution in a Potts lattice-gas model. *Physical Review E*, vol.92, pp.042130:1-9.
<https://doi.org/10.1103/PhysRevE.92.042130>
- 岡本 敦・桑谷 立 (2017) 変成岩組織と鉱物組成累帯構造からの情報抽出：フォーワードモデルと逆解析. 地質学雑誌, vol. 123, no.9, pp.733-745.
<https://doi.org/10.5575/geosoc.2017.0034>

- Omori, T., Kuwatani, T., Okamoto, A. and Hukushima, K. (2016) Bayesian inversion analysis of nonlinear dynamics in heterogeneous reactions. *Physical Review E*, vol.94, pp.033305:1-11. <https://doi.org/10.1103/PhysRevE.94.033305>
- Pearce, J. A. and Cann, J. (1973) Tectonic setting of basic volcanic rocks determined using trace element analyses. *Earth and Planetary Science Letters*, vol.19, no.2, pp.290-300. [https://doi.org/10.1016/0012-821X\(73\)90129-5](https://doi.org/10.1016/0012-821X(73)90129-5)
- Petrelli, M. and Perugini, D. (2016) Solving petrological problems through machine learning: The study case of tectonic discrimination using geochemical and isotopic data. *Contributions to Mineralogy and Petrology*, vol.171, no.10, pp.81:1-15. <https://doi.org/10.1007/s00410-016-1292-2>
- Rollinson H. R. (1993) *Using Geochemical Data: Evaluation, Presentation, Interpretation*. Longman, London, 384p.
- 産業技術総合研究所地質調査総合センター (2014) 表層土壌評価基本図～茨城県地域～, CD-ROM.
- 正路徹也 (1998) 地球情報学: その目指すもの. 情報地質, vol.9, no.3, pp.113-119. https://doi.org/10.6010/geoinformatics1990.9.3_113
- 田中利幸 (2010) 圧縮センシングの数理. *IEICE Fundamentals Review*, vol.4, no.1, pp.39-47. <https://doi.org/10.1587/essfr.4.39>
- Tarantola, A. (2005) *Inverse problem theory and methods for model parameter estimation*. SIAM, Philadelphia. <https://doi.org/10.1137/1.9780898717921>
- Taylor, H. L., Banks, S. C. and J.F. McCoy, (1979) Deconvolution with the l1 norm. *Geophysics*, vol.44, no.1, pp.39-52.
- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. and Knight, K. (2005) Sparsity and smoothness via the fused lasso. *J. Royal. Statist. Soc. B*, vol.67, pp.91-108.
- Till, C. (2015) Earth science: Big geochemistry. *Nature*, vol.523, pp.293-294. <https://doi.org/10.1038/523293a>
- Titterton, D. M., Smith, A. F., and Makov, U. E. (1985) *Statistical Analysis of Finite Mixture Distributions*. Wiley, New York, 243p.
- 土屋範芳・井上千弘・山田亮一・山崎慎一・平野伸夫・岡本 敦・小川泰正・渡邊隆広・奈良郁子・渡邊則昭・東北地方津波堆積物研究グループ (2012) 東北地方太平洋沖地震による岩手, 宮城, 福島県沿岸域の津波堆積物のヒ素に関するリスク評価. 地質学雑誌, vol.118, no.7, p.419-430. <https://doi.org/10.5575/geosoc.2012.0043>
- Ueki, K., Hino, H. and Kuwatani, T. (2018) Geochemical discrimination and characteristics of magmatic tectonic settings; a machine learning-based approach. *Geochemistry, Geophysics, Geosystems*. <https://doi.org/10.1029/2017GC007401>
- Ueki, K. and Iwamori, H. (2017) Geochemical differentiation processes for arc magma of the Sengan volcanic cluster, Northeastern Japan, constrained from principal component analysis. *Lithos*, vol.290, pp.60-75. <https://doi.org/10.1016/j.lithos.2017.08.001>
- Yasumoto, A., Yoshida, K., Kuwatani, T., Nakamura, D., Svojtka, M. and Hirajima, T. (in press) Fast and precise quantitative electron probe chemical mapping technique and its application to ultrahigh-pressure eclogite from the Moldanubian Zone of the Bohemian Massif (Nové Dvory, Czech Republic). *American Mineralogist*.
- Yoshida, K., Kuwatani, T., Hirajima, T., Iwamori, H. and Akaho, S. (2018) Progressive evolution of whole-rock composition during metamorphism revealed by multivariate statistical analyses. *Journal of Metamorphic Geology*, vol.36, no.1, pp.41-54. <https://doi.org/10.1111/jmg.12282>
- Yoshida, K., Kuwatani, T., Yasumoto, A., Haraguchi, S., Ueki, K. and Iwamori, H. (in press) GEOFCM: a new method for statistical classification of geochemical data using spatial contextual information. *Journal of Mineralogical and Petrological Science*. <https://doi.org/10.2465/jmps.171127>
- 弘原海 清 (1990) 情報地質学の歩みと展望. 情報地質, vol.1, no.1, pp.25-29. https://doi.org/10.6010/geoinformatics1990.1.1_25
- Zuo, R. (2017) Machine learning of mineralization-related geochemical anomalies: A review of potential methods. *Natural Resources Research*, vol.26, no.4, pp.457-464. <https://doi.org/10.1007/s11053-017-9345-4>

要 旨

地球科学プロセス解明のためのデータ駆動型解析
—地質学分野における応用例—

桑谷 立

近年の計算機技術の発達と数理・情報科学の進展は、データから最大限に情報を抽出するための強力な情報処理技術を生み出し続けている。これらはデータ駆動型解析として様々な自然科学や工学分野、産業界において応用が進みつつある。本稿では、データ駆動型解析の基盤数理解析スキームであるベイズ推論とスパースモデリングについて、簡単に解説するとともに、著者らによる地質学分野へのデータ駆動型解析の適用事例について概観する。岩石組織のデータ解析には、時空間構造に関する先験的知識と物理化学モデルを客観的に逆解析に導入できるベイズ推論が特に有効であり、多元素からなる地球化学データの解析には、高次元データから本質的な次元を自動抽出するスパースモデリングが有効である。情報科学者と地質学者の密接な協働は、地質学的諸問題に有用な新たな解析手法群の開発につながり、データ駆動型解析の重要性は今後より一層増していくものと考えられる。

キーワード：データ駆動科学，逆問題，機械学習，ベイズ推論，スパースモデリング