

データ駆動科学による異分野共創

東京大学・大学院新領域創成科学研究科

複雑理工学専攻

岡田真人

日時:2025年2月17日(月) 13:05-13:40

九大伊都キャンパスウェスト1号館D棟D414

講演のスライドは岡田研HPにて公開予定

<https://mns.k.u-tokyo.ac.jp/lab.html#overview>

内容

- 自己紹介
- 異分野共創の目的と実践
 - データ駆動科学の三つのレベル
 - ベイズ的スペクトル分解を例にして
- 可解線形回帰モデル
- NMRの緩和モード分解
- SPring-8全ビームラインベイズ化計画
- まとめと今後の展開

自己紹介

- ・ 大阪市立大学理学部物理学科
 - アモルファスシリ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)
 - 希土類元素の光励起スペクトルの理論
- ・ 三菱電機
 - 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)
 - 置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所 複雑理工学専攻
 - データ駆動科学 (2004/07 -)

本講演の目的

- ・異分野共創の目的: ある分野の標準を刷新するために、異分野と共に創する
- ・主張: その数理・情報科学的基盤がデータ駆動科学である。
- ・異分野共創の数理・情報科学的基盤のデータ駆動科学の一翼を担うベイズ推論による、実践を以下の事例で紹介する。
- ・可解線形回帰とNMRの緩和モード分解でベイズ計測の基礎から展開を紹介
- ・SPring-8全ビームライン計画の紹介

内容

- ・自己紹介
- ・異分野共創の目的と実践
 - ・データ駆動科学の三つのレベル
 - ・ベイズ的スペクトル分解を例にして
- ・可解線形回帰モデル
- ・NMRの緩和モード分解
- ・SPring-8全ビームラインベイズ化計画
- ・まとめと今後の展開

異分野共創の目的と実践

- ・異分野共創の目的: ある分野で、**その分野だけでは行えないブレークスルーを実現**することである。
- ・その分野に協力する分野の研究も進めるこ**とができると良い。**
- ・この目的の実践のためには、**その分野のデータを中心に共創を考えるしかない。**
- ・つまり、データで科学を駆動するデータ駆動科学の枠組みを用いる必要がある。

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.

David Marrの三つのレベル (1982)

David Marrは複雑な情報処理装置を理解するには以下の三つのレベルが必要であると説いた

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

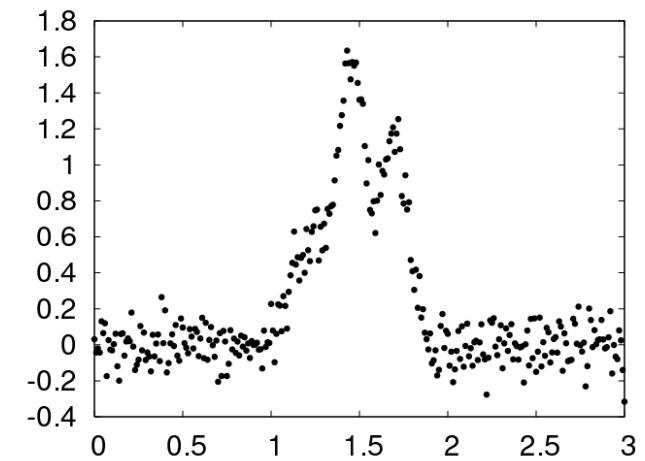
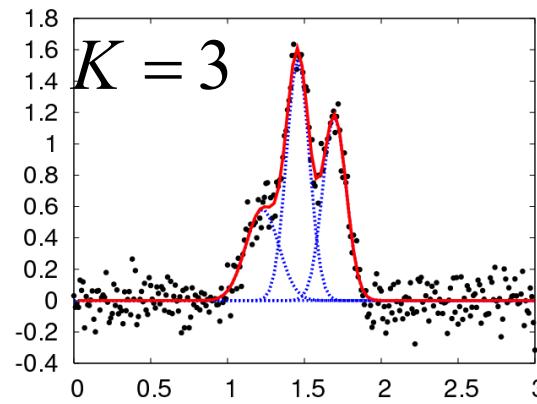
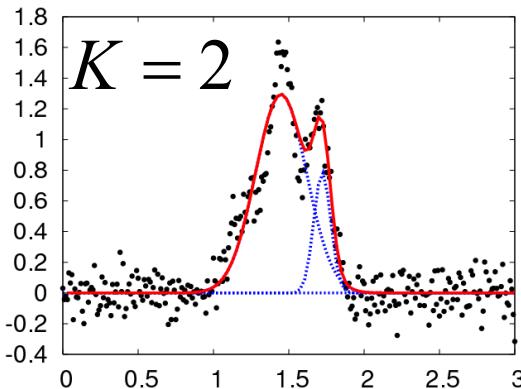
David Marr Vision: A Computational Investigation into the Human Representation and Processing of Visual Information (1982)

スペクトル分解

永田賢二, 杉田誠司, 岡田真人
東大新領域

Kenji Nagata, Seiji Sugita and Masato Okada,
"Bayesian spectral deconvolution with the
exchange Monte Carlo method", *Neural
Networks*, 28, 82-89 (2012)

ベイズ的スペクトル分解



Nagata, Sugita and Okada, Bayesian spectral deconvolution
with the exchange Monte Carlo method, *Neural Networks* 2012

スペクトル分解の三つのレベル (1/2)

スペクトル分解の計算理論

データ解析の目的: 多峰性スペクトルから背後にある離散電子のエネルギー準位を推定する

データ解析の適切さ: 多峰性スペクトルを単峰性関数の線形和で表し、その単峰性関数の個数を推定し、その単峰性関数の中心位置を電子のエネルギー準位とする
統計学の交差検証誤差やベイズ的モデル選択で単峰性関数の数を決める。

スペクトル分解のモデリング

多峰性スペクトルを単峰性関数の線形和に観測ノイズが付加されて生成されるとモデリングする

スペクトル分解の定式化

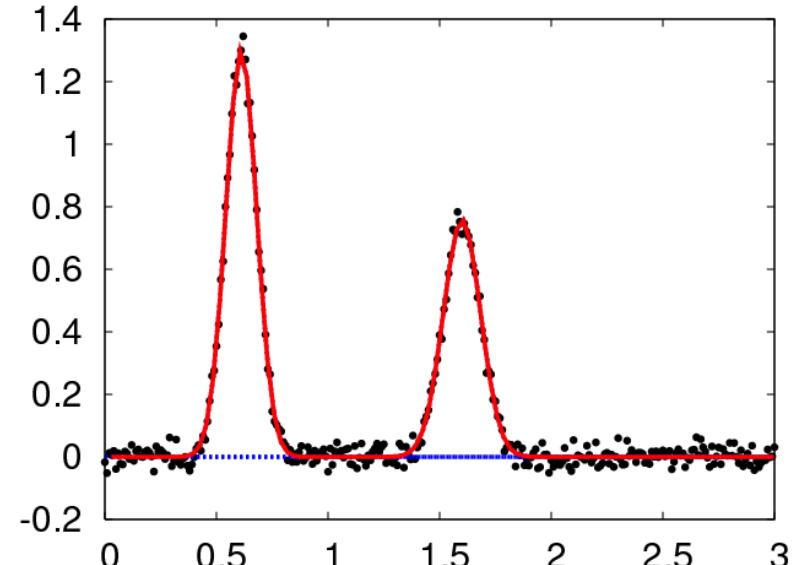
ガウス関数(基底関数)の足し合わせにより、スペクトルデータを近似

観測データ: $D = \{x_i, y_i\}_{i=1}^n$

x_i : 入力 y_i : 出力

$$f(x; \theta) = \sum_{k=1}^K a_k \exp\left(-\frac{b_k(x - \mu_k)^2}{2}\right)$$

$$\theta = \{a_k, b_k, \mu_k\} \quad k = 1, \dots, K$$



二乗誤差を最小にするようにパラメータをフィット(最小二乗法)

$$E(\theta) = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i; \theta))^2$$

スペクトル分解の三つのレベル (2/2)

スペクトル分解の表現・アルゴリズム

多峰性スペクトルを単峰性関数の線形和に観測ノイズが付加されて生成するとモデリングし、ベイズ推論を適用することで、 K 個の単峰性関数の大きさ、位置、幅。大きさの事後確率を求める。各 K に対して、ベイズ的自由エネルギーを求め、ベイズ的自由エネルギーを最小にする K を求める。その **K 個の单峰性関数の位置を、電子のエネルギー準位**とする。

Igarashi, Nagata, Kuwatani, Omori, Nakanishi-Ohno and M. Okada, “Three Levels of Data-Driven Science”, *Journal of Physics: Conference Series*, 699, 012001, 2016.
Nagata, Sugita and M. Okada, “Bayesian spectral deconvolution with the exchange Monte Carlo method”, *Neural Networks*, 28, 82-89 2012.

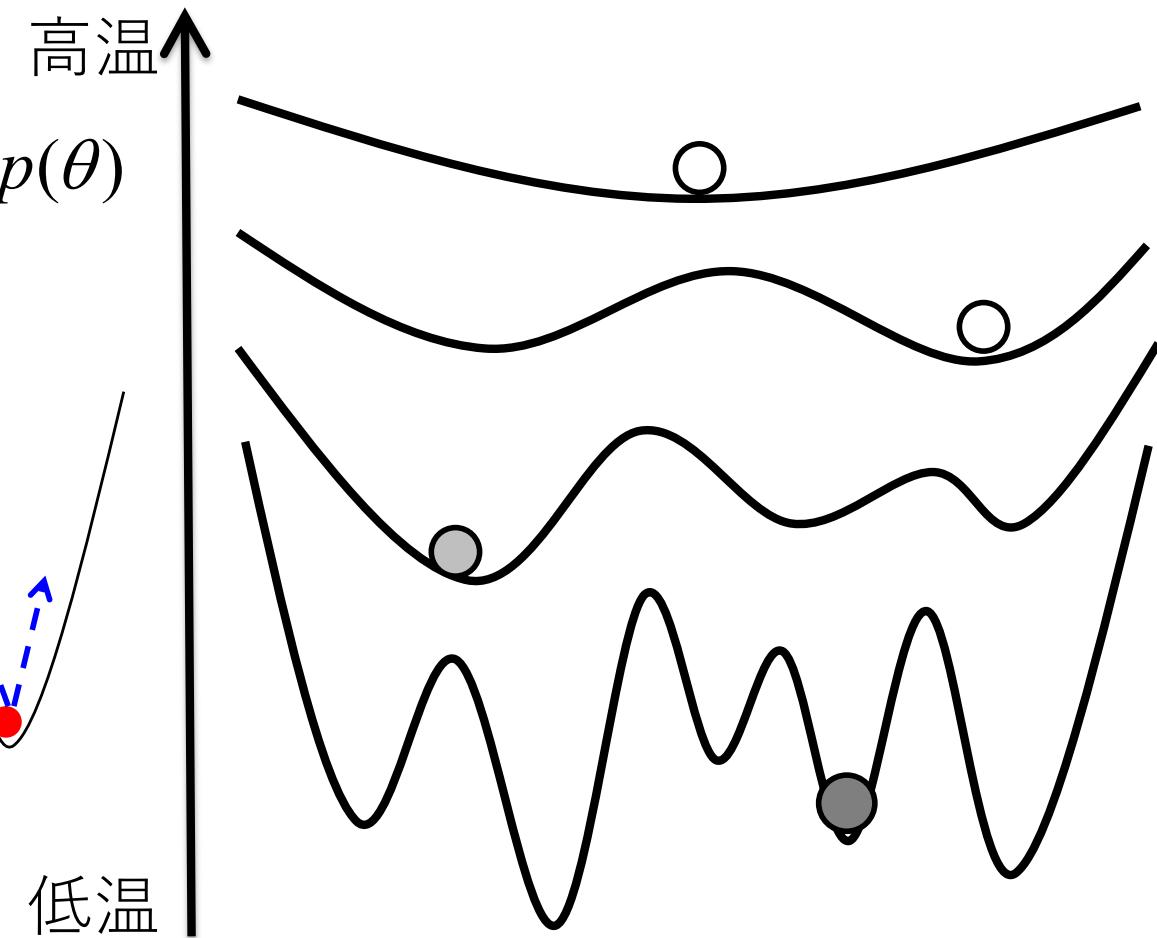
モンテカルロ法の適用 レプリカ交換モンテカルロ法

メトロポリス法

$$p_\beta(\theta) \propto \exp\left(-\frac{n}{\sigma^2} \beta E(\theta)\right) p(\theta)$$

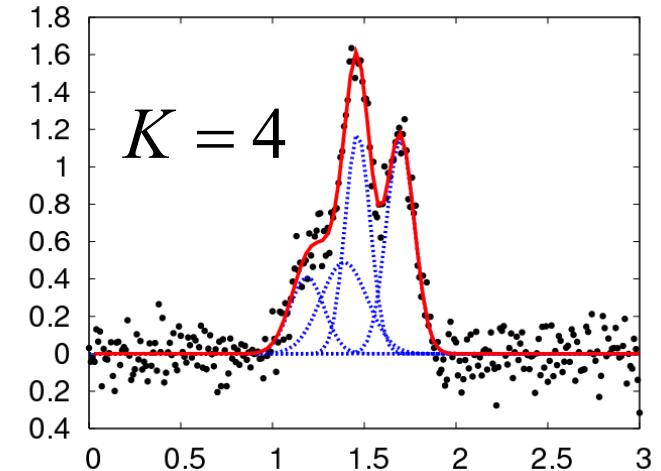
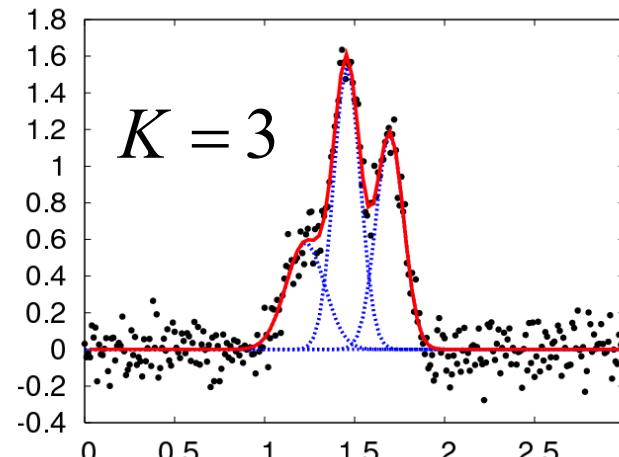
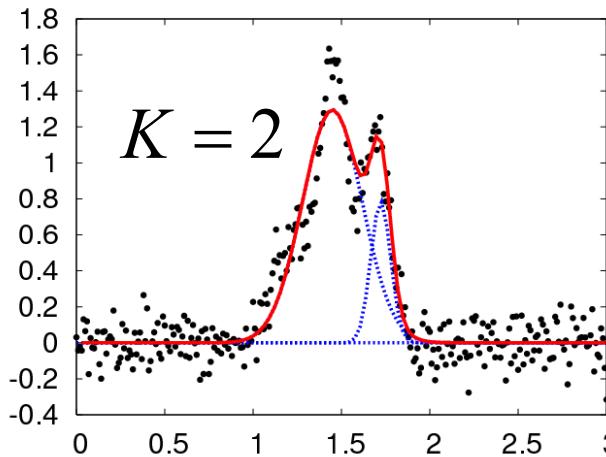
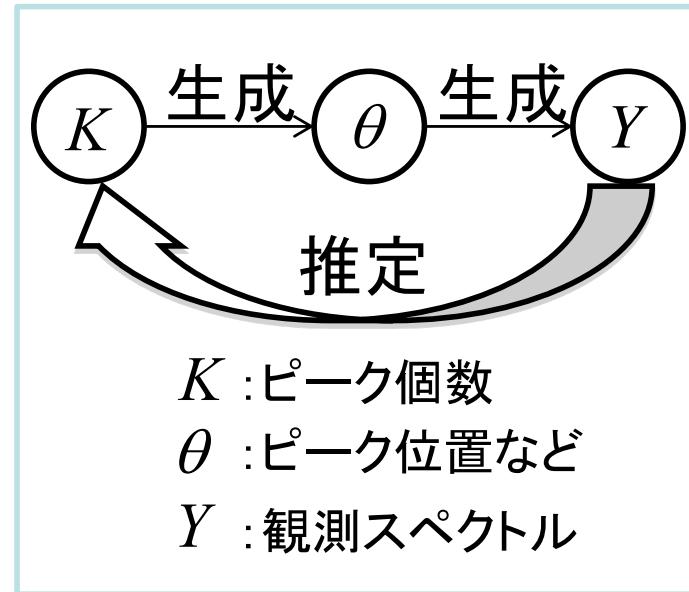
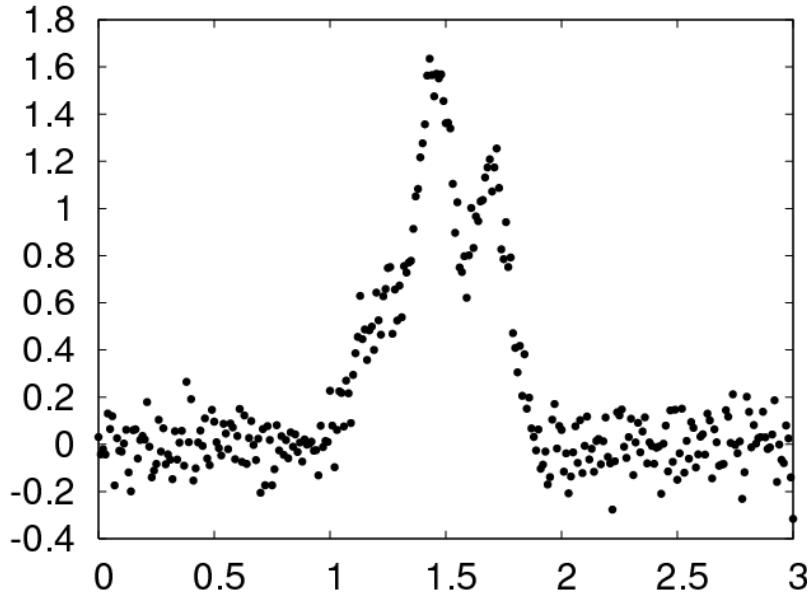


レプリカ交換モンテカルロ法



K. Hukushima, K. Nemoto, *J. Phys. Soc. Jpn.* **65** (1996).

モデル選択: K をどう選ぶか



Nagata, Sugita and Okada, Bayesian spectral deconvolution with the exchange Monte Carlo method, *Neural Networks* 2012

モデル選択: 自由エネルギーの導入

1. 欲しいのは $p(K|Y)$

2. θ がないぞ

3. $p(K, \theta, Y)$ の存在を仮定

$$p(K, \theta, Y) = p(Y|\theta, K)p(K)$$

$$p(Y|\theta, K) = \prod_{i=1}^n p(y_i|\theta) \propto \exp(-nE(\theta))$$

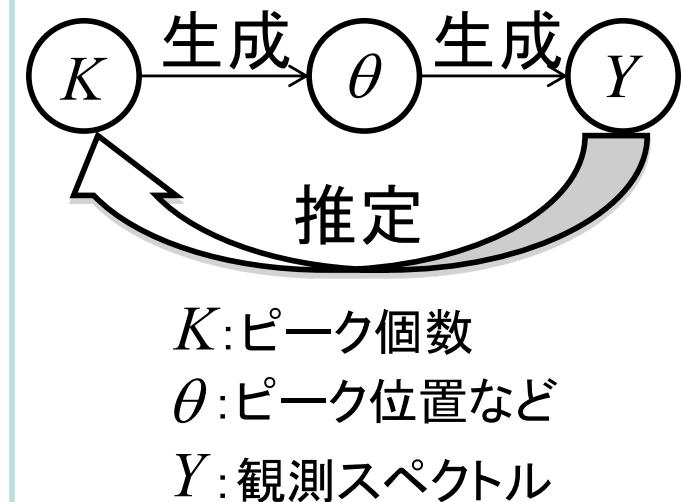
4. 無駄な自由度の系統的消去: 周辺化, 分配関数

$$p(K, Y) = \int p(K, \theta, Y) d\theta$$

$$p(K|Y) = \frac{p(Y|K)p(K)}{p(Y)} \propto p(K) \int \exp(-nE(\theta)) p(\theta) d\theta$$

$$F(K) = -\log \int \exp(-nE(\theta)) p(\theta) d\theta = E - TS$$

自由エネルギーを最小にする個数 K を求める.



スペクトル分解のハードウェア実装

Table 1. Execution times [seconds] of different algorithms and computing implementations.

Algorithm \ Computing	CPU	GPU
MCMC method	10371.5 sec	1326.9 sec
SVI metohod	260.8 sec	7.2 sec

$260.9/7.2=36.1$ times faster

Rapid, Comprehensive Search of Crystalline Phases
from X-ray Diffraction in Seconds
via GPU-Accelerated Sparse Inference

17

Murakami, Nagata, Matsushita and Demura¹
National Institute for Materials Science: NIMS

David Marrの三つのレベル (1982)

David Marrは複雑な情報処理装置を理解するには以下の三つのレベルが必要であると説いた

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

David Marr Vision: A Computational Investigation into the Human Representation and Processing of Visual Information (1982)

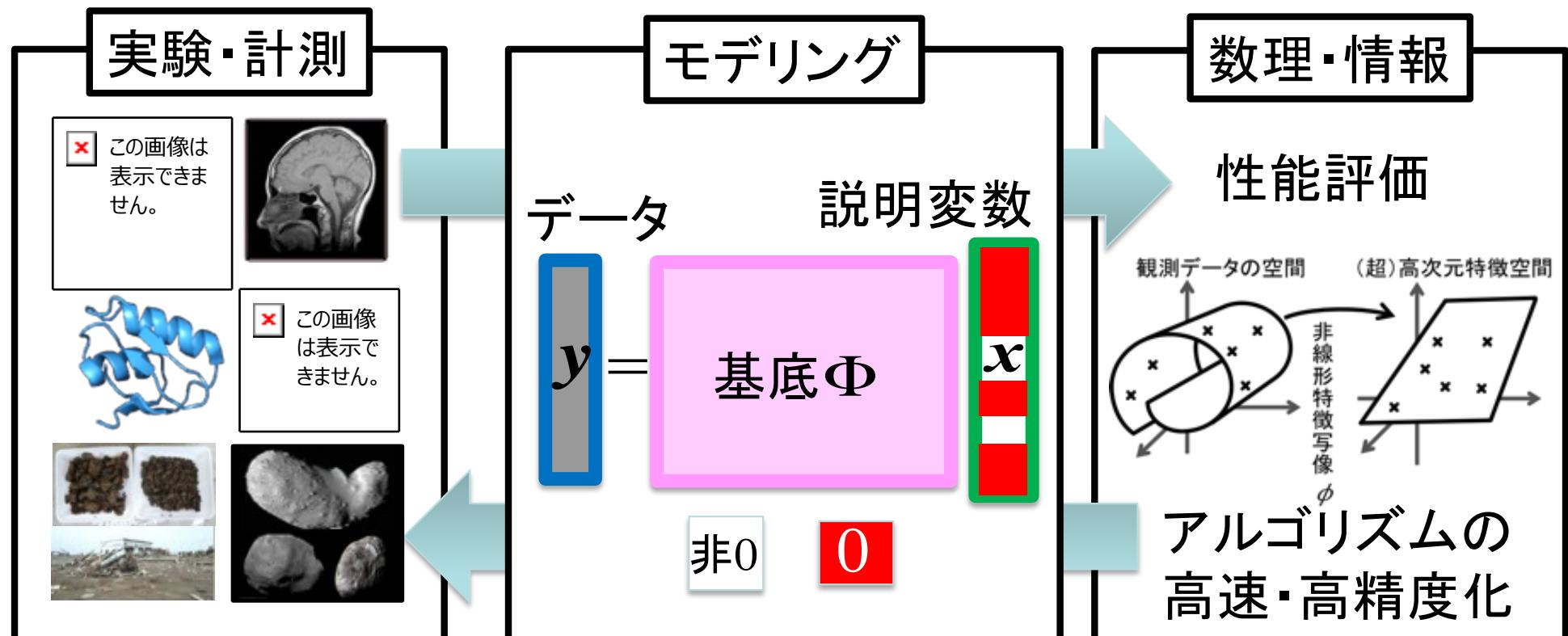
異分野共創でどのような 分野のブレークスルーを期待できるか？

- ・ 計算理論
 - データの生成元の分野、計測科学
 - ・ その分野の限界の突破と計測科学の刷新
- ・ モデリング/アルゴリズム
 - 物理学、数理科学、統計学
 - ・ モデリング手法の深化
 - 深層ネットの数理(特異統計学)と計測科学
- ・ ハードウェア実装
 - 計算機科学: 新規アルゴリズムの新たな実装

新学術領域研究 平成25～29年度 スペ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？



内容

- ・自己紹介
- ・異分野共創の目的と実践
 - データ駆動科学の三つのレベル
 - ベイズ的スペクトル分解を例にして
- ・可解線形回帰モデル
 - ・NMRの緩和モード分解
 - ・SPring-8全ビームラインベイズ化計画
 - ・まとめと今後の展開

線形回帰の可解ベイズ計測

片上舜^A, 柏村周平^A, 永田賢二^c, 水牧仁一朗^c,
岡田真人^A

A 東大新領域, B NIMS, C 熊大

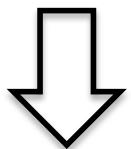
Shun Katakami, Shuhei Kashiwamura, Kenji Nagata,
Mizumaki Masaichiro and Masato Okada, "
Mesoscopic Bayesian Inference by Solvable Models
", <https://arxiv.org/abs/2406.02869>

ベイズ計測とは？

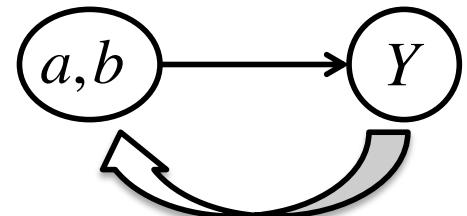
ベイズ推論

$$p(Y, a, b) = p(Y | a, b)p(a, b) = p(a, b | Y)p(Y)$$

↓
<ベイズの定理>



生成(因果律)



$$p(a, b | Y) = \frac{p(Y | a, b)p(a, b)}{p(Y)} \propto \exp(-nE(a, b))p(a, b)$$

$p(a, b | Y)$: 事後確率。データが与えられたもとでの物理パラメータの確率。

$p(a, b)$: 事前確率。あらかじめ設定しておく必要がある。これまで蓄積されてきた科学的知見

ベイズ計測三種の神器

1. 物理パラメータの事後確率分布定
2. モデル選択
3. データ統合

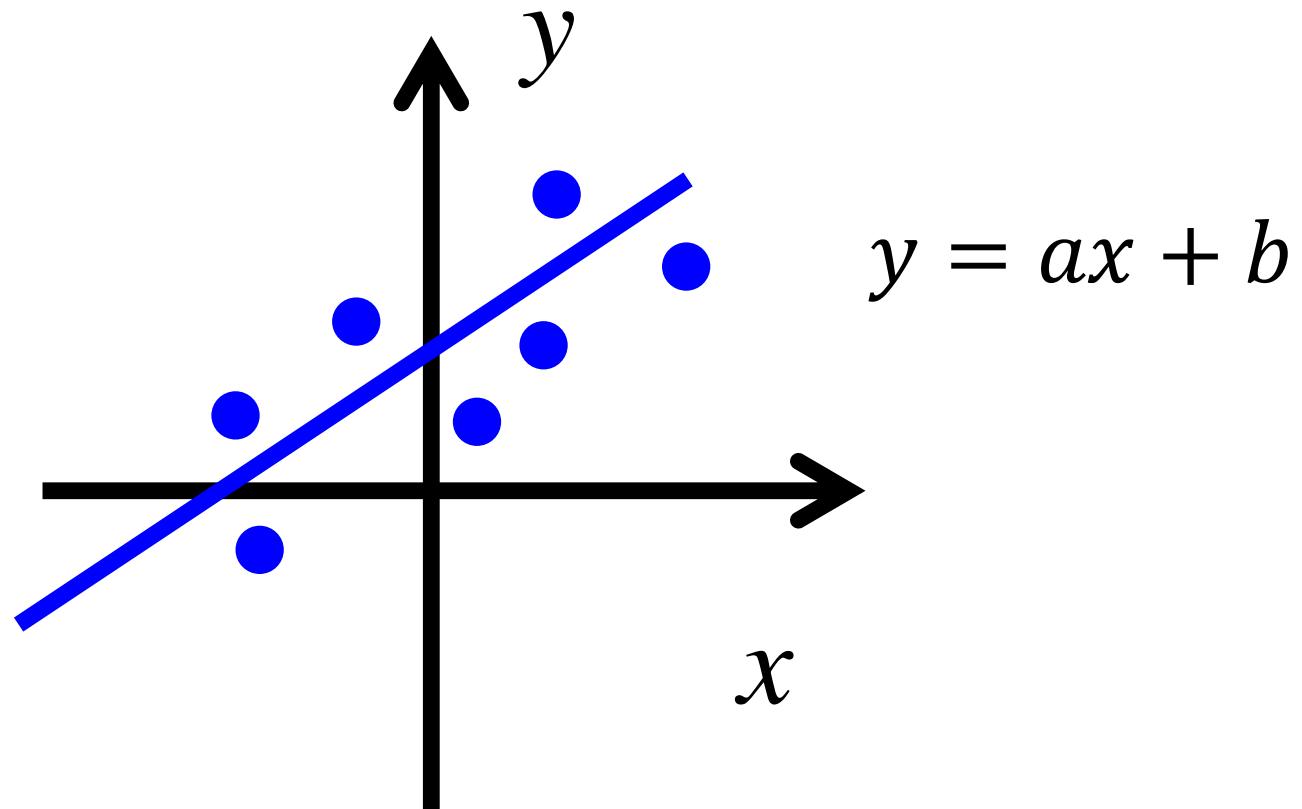
ベイズ計測の三種の神器

- ・ 神器1: パラメータの事後確率推定
 - 計測限界
- ・ 神器2: ベイズ的モデル選択
 - データを説明する複数のモデル候補から、データのみで、一つのモデルを選択
- ・ 神器3: ベイズ統合
 - 一つに物質に対する複数計測(マルチーモーダル計測)からパラメータを一組み決定

ベイズ計測の利点

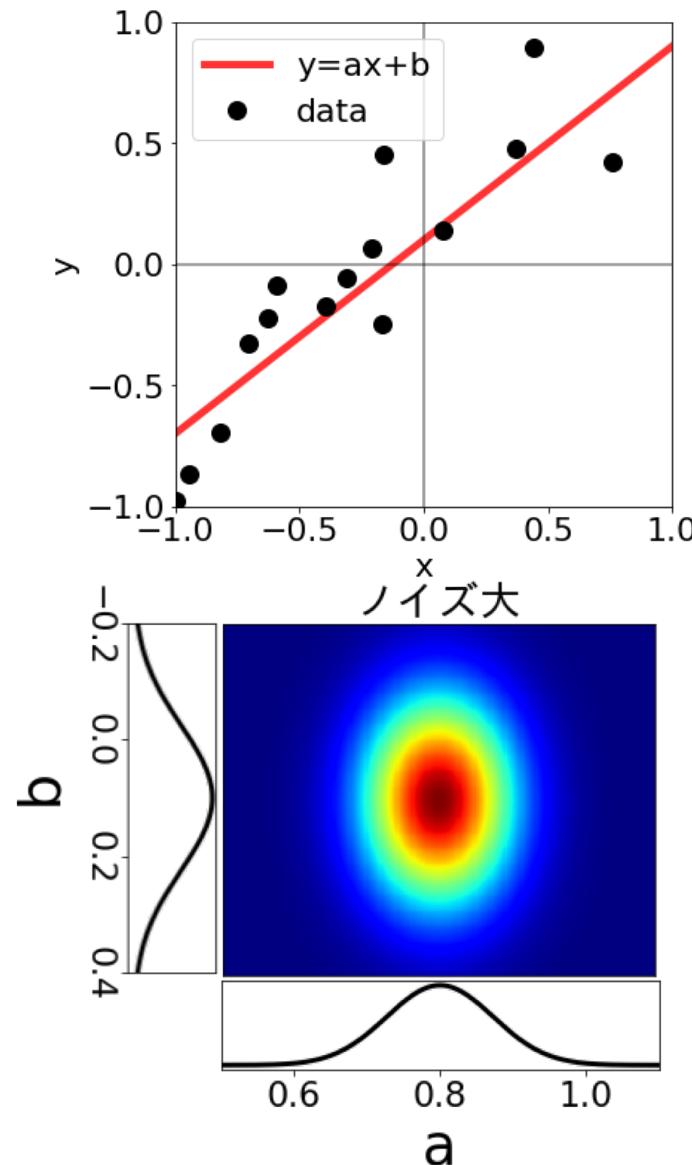
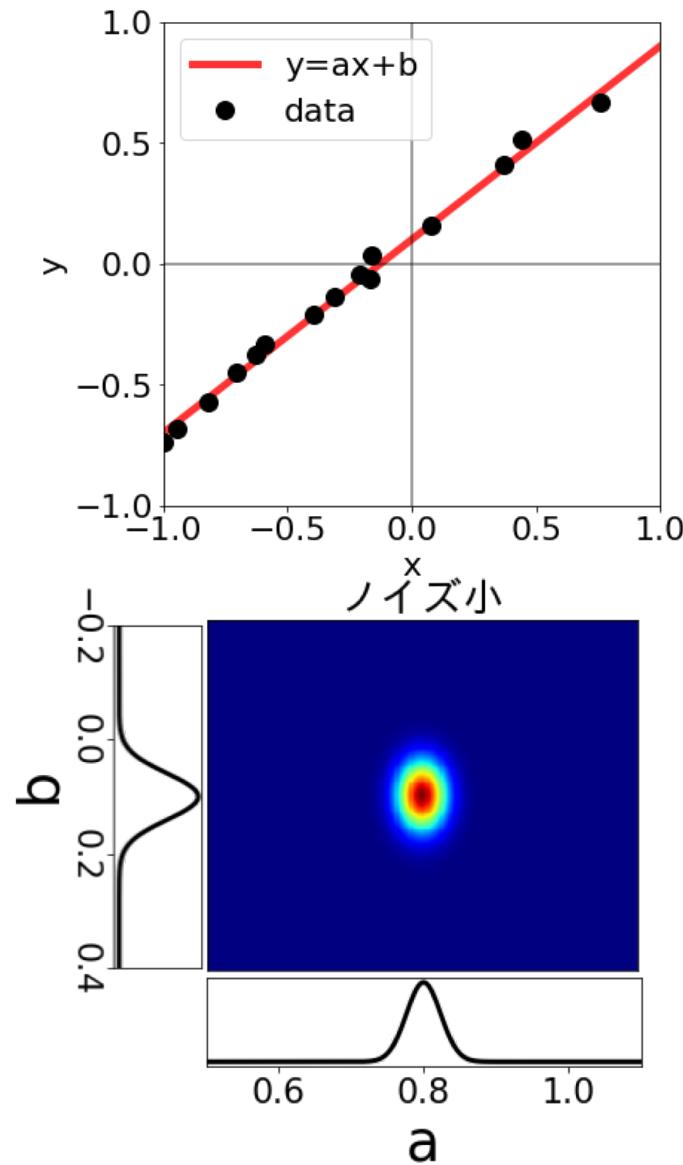
$y=ax+b$ の取り扱いを通じて

現状でも用いられている最も簡単な例

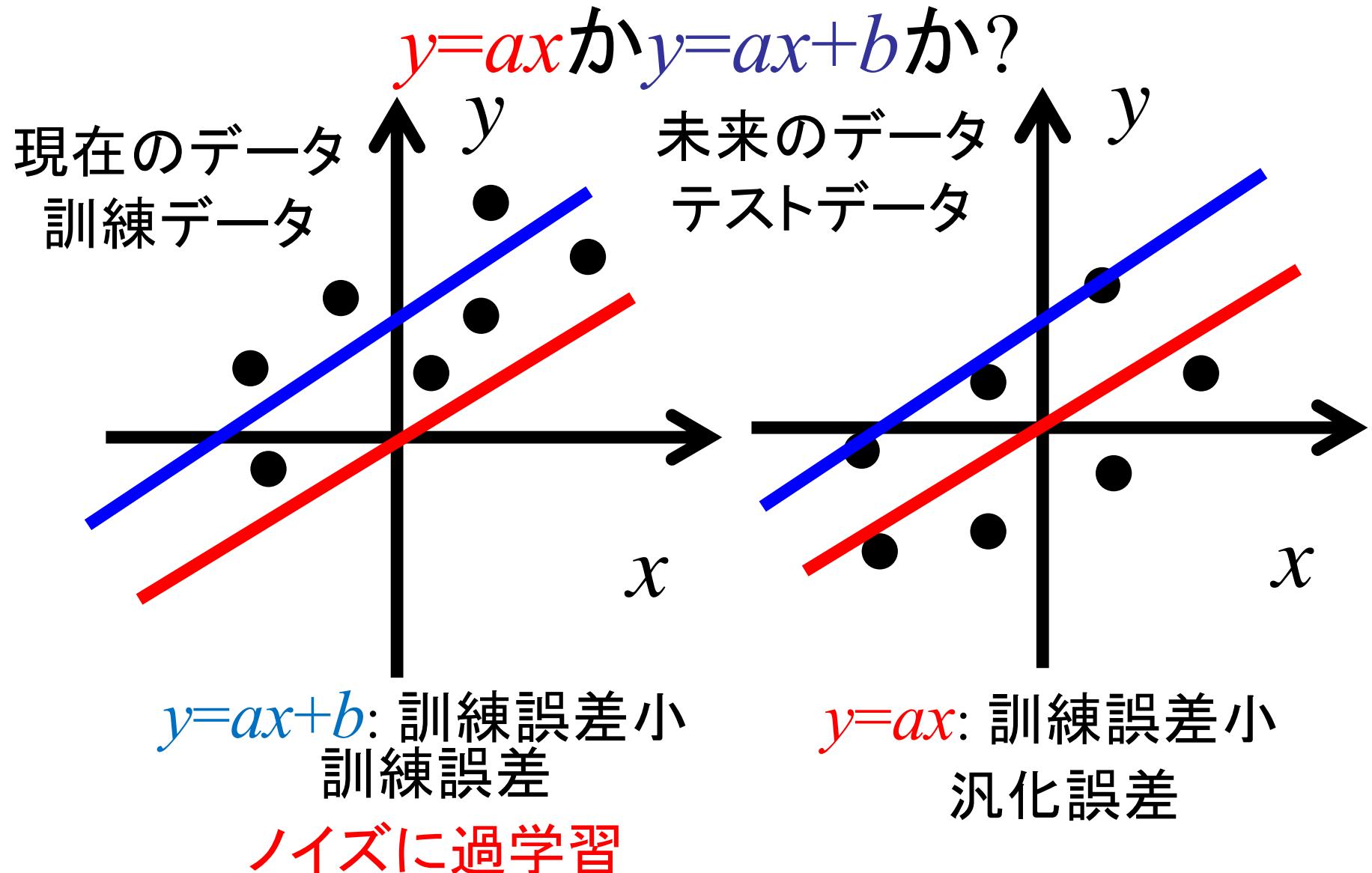


傾き a : 系の線形応答、バネ定数、電気伝導度、誘電率

神器1: パラメータの事後確率推定 (4/4)

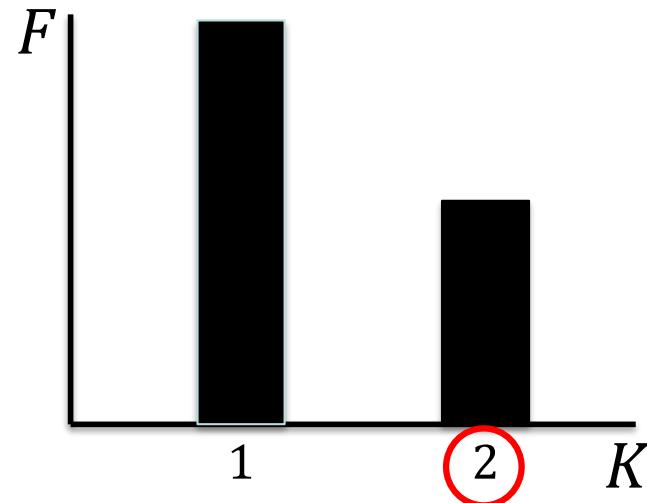
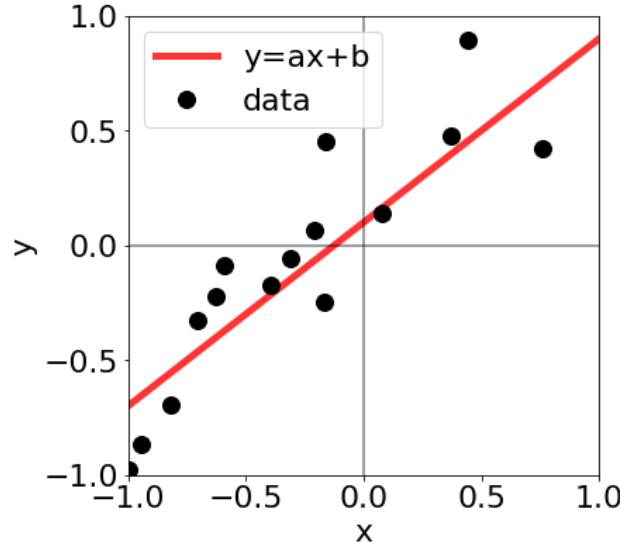


神器2: ベイズ的モデル選択



モデル選択できる理由: 汎化誤差は観測ノイズに依存する

モデル選択: 自由エネルギー差



- $K = 1 : y = ax$
- $K = 2 : y = ax + b$

$$F(K=1) = N \left\{ \frac{1}{\sigma^2} E(a_0) + \frac{\log N}{2N} \right\}$$

$$F(K=2) = N \left\{ \frac{1}{\sigma^2} E(a_0, b_0) + \frac{\log N}{N} \right\}$$

データのみからモデルを選択できる

内容

- ・自己紹介
- ・異分野共創の目的と実践
 - ・データ駆動科学の三つのレベル
 - ・ベイズ的スペクトル分解を例にして
- ・可解線形回帰モデル
- ・NMRの緩和モード分解
- ・SPring-8全ビームラインベイズ化計画
- ・まとめと今後の展開

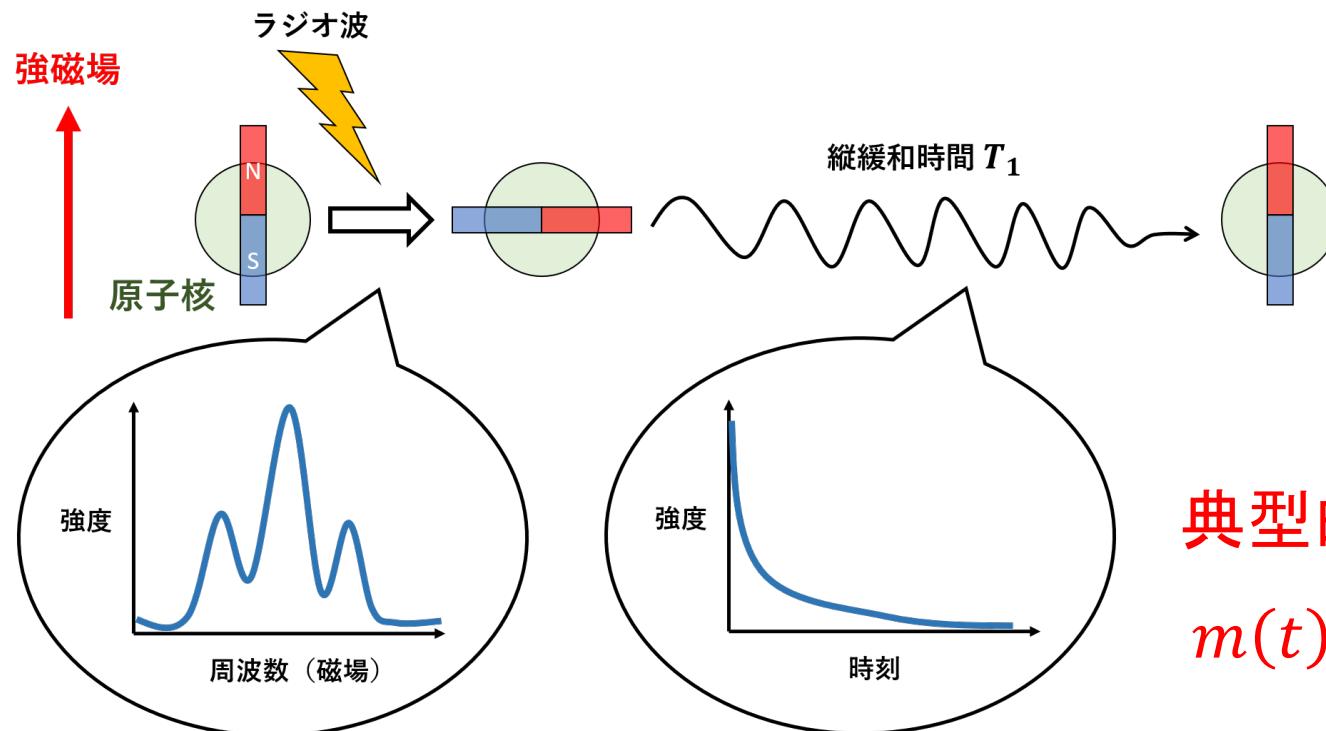
核磁気共鳴法へのデータ駆 動科学的手法の開発

上田朔^A, 片上舜^A, 吉田章吾^B, 中井祐介^B,
水戸毅^B, 水牧仁一朗^C, 岡田真人^A

A 東大新領域, B 兵庫県立大理学研究科, C JASRI

Ueda, Katakami, Yoshida, Koyama, Nakai, Mito,
Mizumaki and Masato Okada, "Bayesian
approach to T_1 analysis in NMR spectroscopy
with applications to solid state physics", *Journal
of the Physical Society of Japan*.92, 054002 (2023)

NMR測定



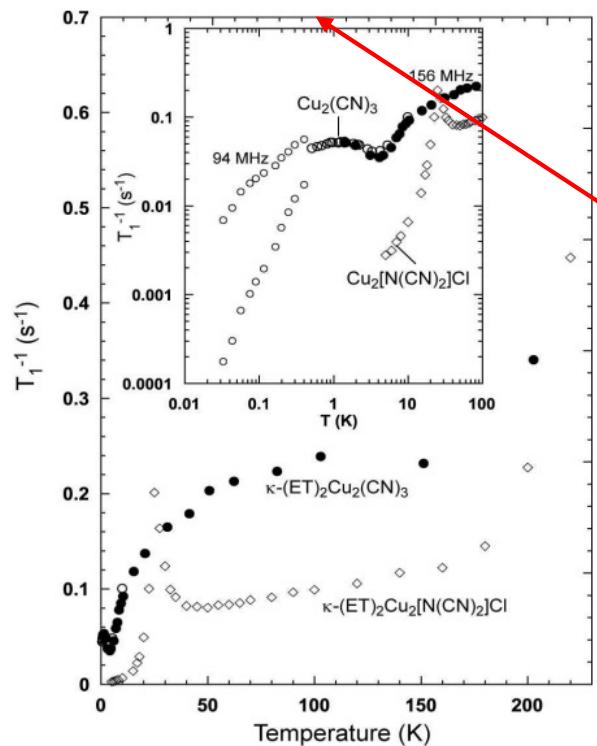
典型的には
 $m(t) \propto \exp\left(-\frac{t}{T_1}\right)$

NMRスペクトル

核スピン格子緩和曲線

核スピン格子緩和 – 多成分の場合

例：量子スピン液体のNMR測定では、高温側では单一の指数関数でフィッティングできていた核スピン格子緩和曲線が、低温では2つの指数関数の和に分裂することが報告されている。



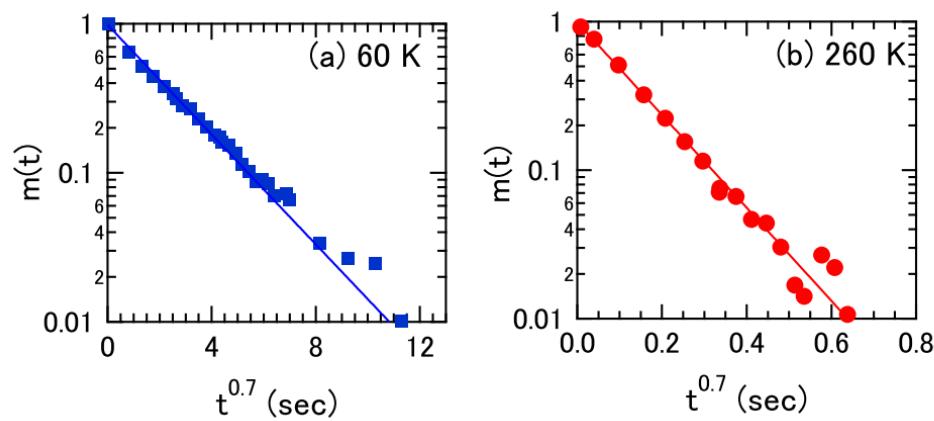
有機モット絶縁体スピン液体を報告したY. Shimizu et al. では、試料が2相に分かれていると解釈している。

0.4 Kより低温側で緩和時間 T_1 が分裂

Y Shimizu et al., Phys. Rev. Lett. 91, 107001 (2003)

核スピン格子緩和 – stretched exponentialの場合

粉末試料を使う場合や試料のdisorderが原因で緩和率 T_1^{-1} が一つの値の周りに連続的に分布している場合、現象論的に stretched exponential 関数 $\exp\left(-\left(\frac{t}{T_1}\right)^\beta\right)$ でフィッティングする場合もある。



常圧下における半導体SmS(粉末)の核スピン格子緩和曲線。T Koyama et al. は $\beta = 0.7$ のstretched exponential関数を使ってフィッティングし、得られた T_1 の温度依存性などを議論した。

核スピン格子緩和のフォワードモデル

単一の指数関数でうまくフィッティングできない緩和曲線が実験で得られた場合、①多成分の緩和を仮定するか②stretched exponentialを仮定するかで解釈が変わってしまう。

そこで、ベイズ推論の枠組みで最も妥当なモデルを選択することを試みる。

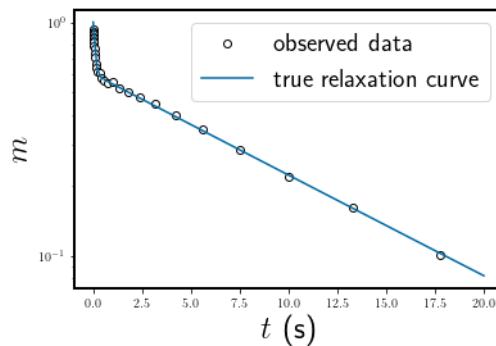
$$m(t) = \sum_{k=1}^K A_k \exp\left(-\left(W_k t\right)^{\beta_k}\right) + C + \epsilon$$

↓ ↓ ↓ ↓ ↓
緩和振幅 緩和率 Stretching オフセットノイズ
($= \frac{1}{T}$) 指数

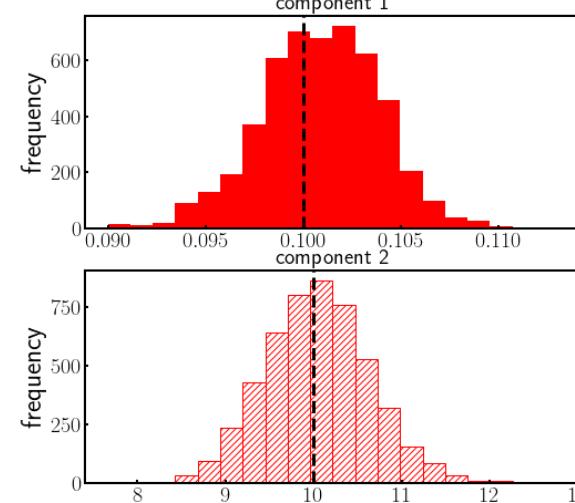
モデル数 K の事前分布は一様とする。観測データ $\mathcal{D} = \{(t_i, m_i)\}_{i=1,2,\dots,N}$ が与えられたとき、事後確率 $p(K | \mathcal{D})$ が最大となるモデル数を選択する。

数値実験(多成分: $K = 3$)

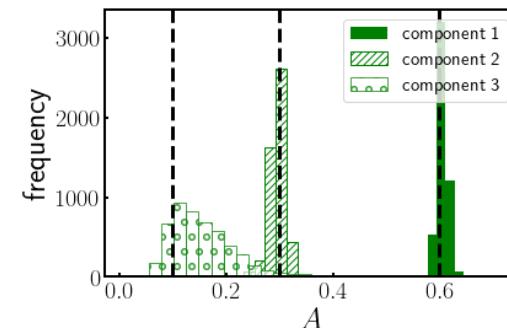
生成データ



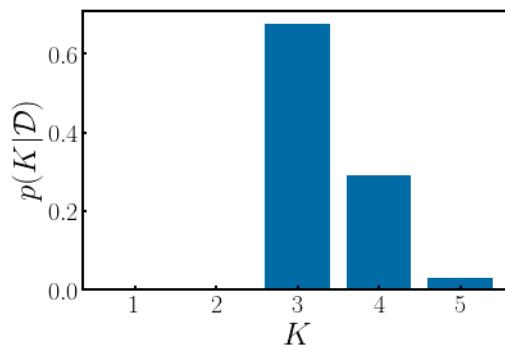
緩和率



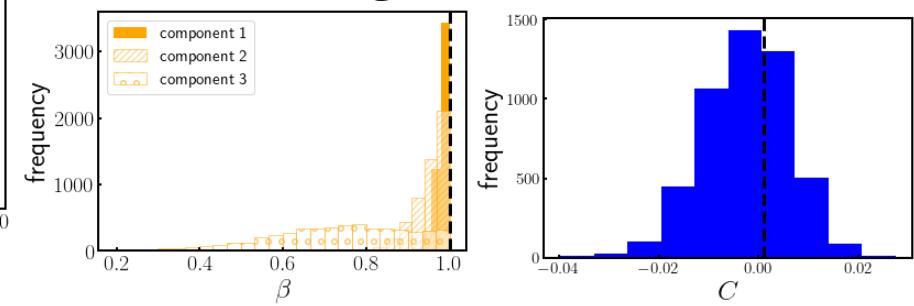
緩和振幅



モデル選択

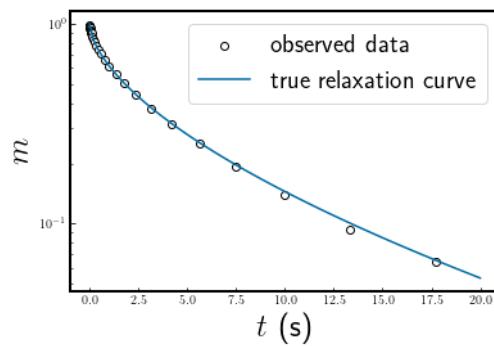


Stretching指数 オフセット

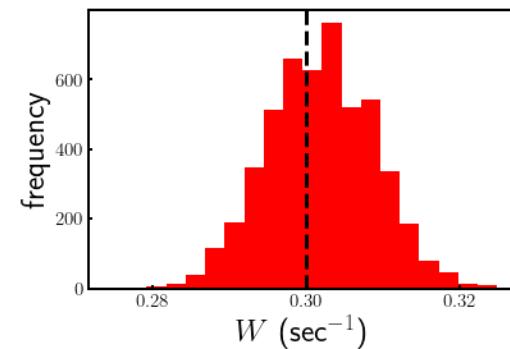


数値実験(単一のstretching exponential)

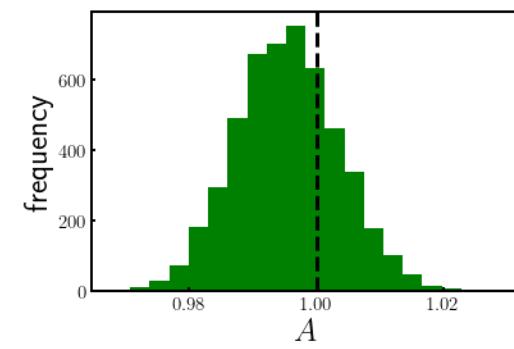
生成データ



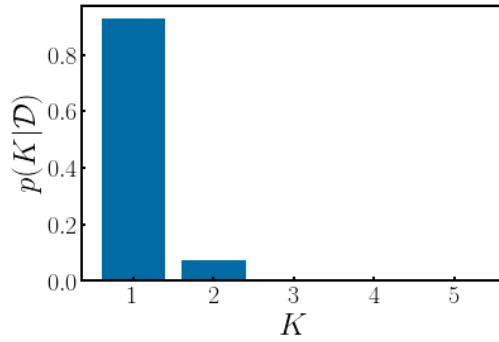
緩和率



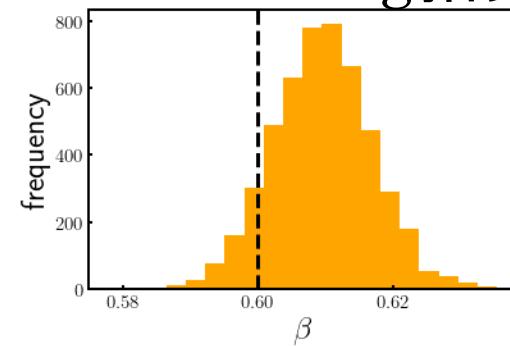
緩和振幅



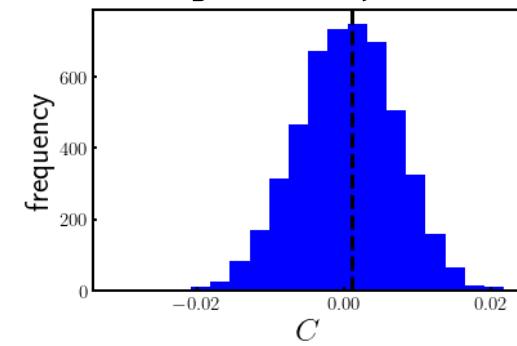
モデル選択



Stretching指数

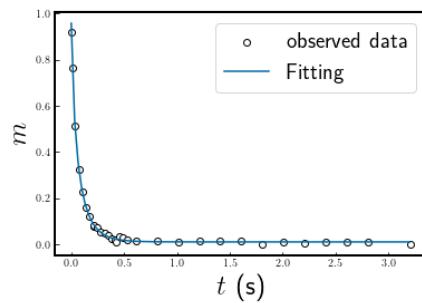


オフセット

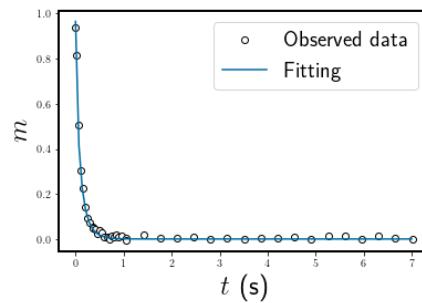


SmS の実験データ(常圧)

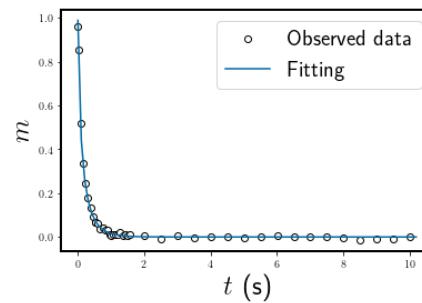
260 K



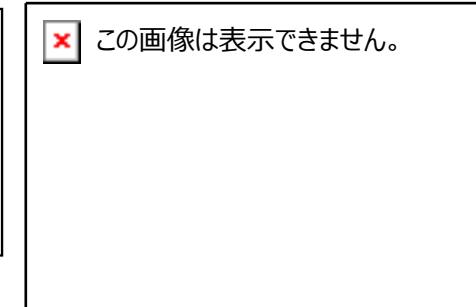
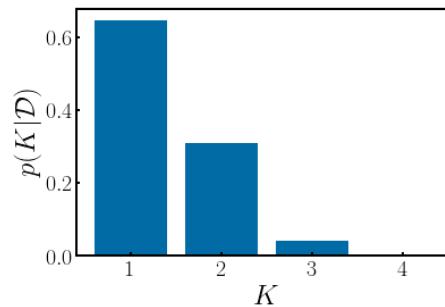
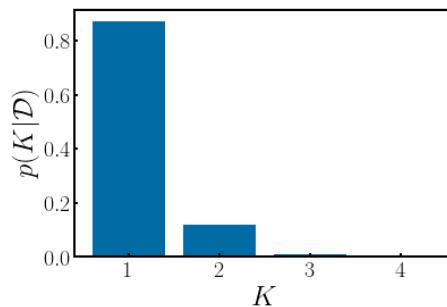
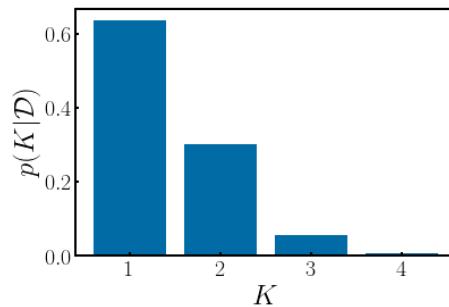
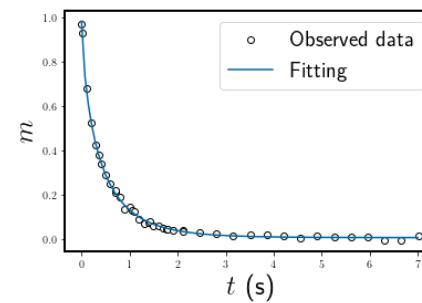
200 K



140 K



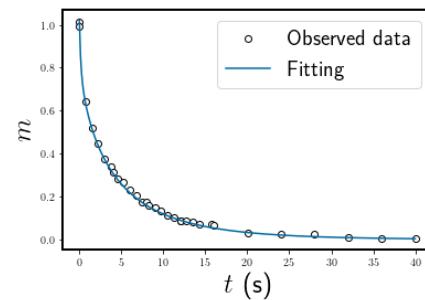
100 K



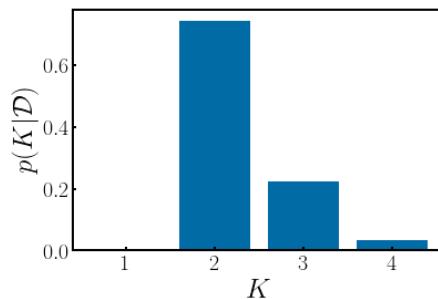
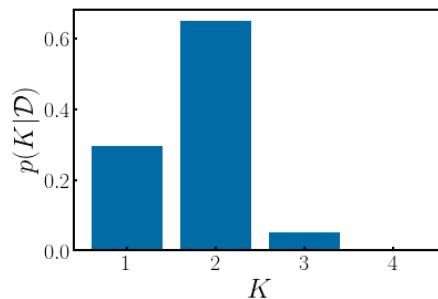
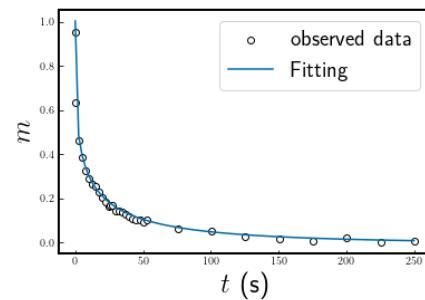
SmS の実験データ(常圧)

9

60 K



30 K



100 Kより高温側では単一の stretched exponentialが選択される。(データを報告した T Koyama et al. が行った解析を支持)

一方、低温側では2つの緩和成分があるという結果が得られた。(物理的解釈は次回以降)

- ・試料の乱れや複数の緩和成分の共存を考慮した核スピン格子緩和のモデルを構築した。

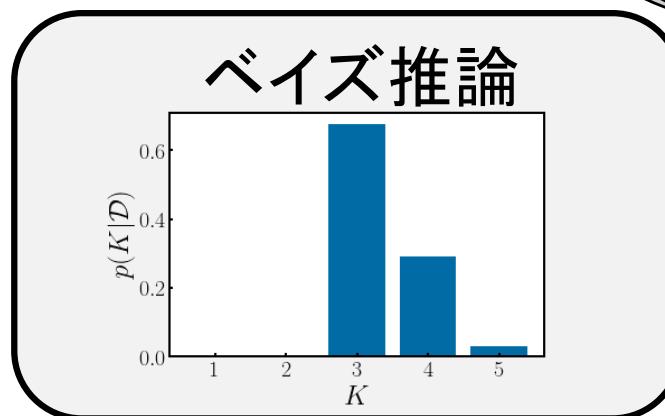
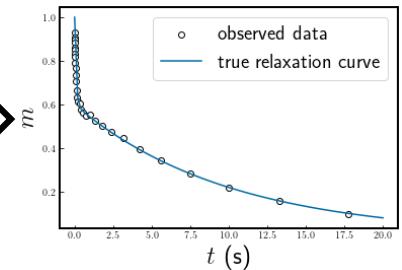
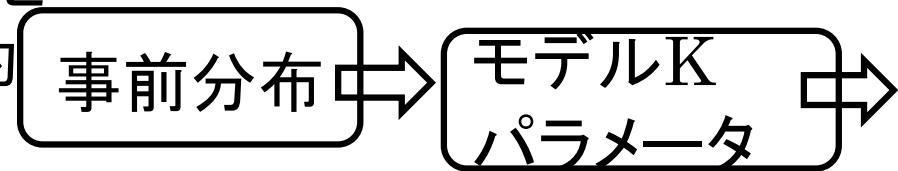
- ・NMR測定における核スピン格子緩和曲線の物理的解釈をサポートできる。

- ・ β の適切な事前分布に関する物理的な事前知識がないため $[0,1]$ 上の一様分布を使ったが、事前分布の改善が望まれる。

- ・今後は半導体SmSのNMR測定実験に本手法を適用し、物理的解釈を議論する。

NMRまとめ

核スピン格子緩和曲線



内容

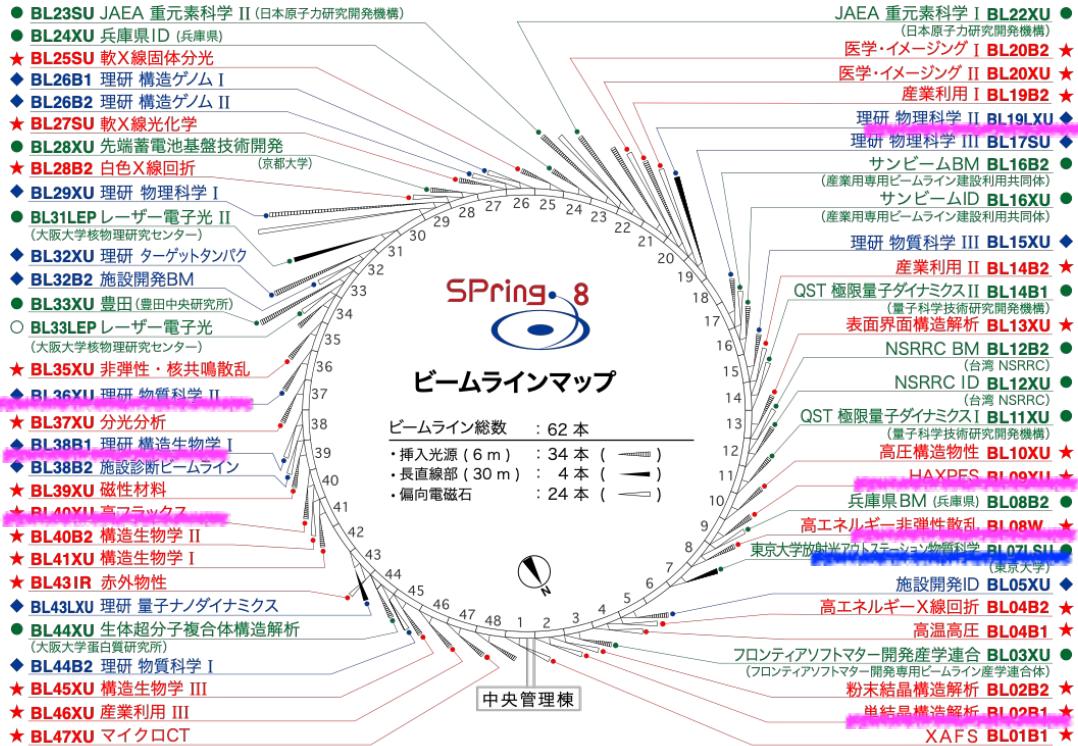
- ・自己紹介
- ・異分野共創の目的と実践
 - ・ベイズ的スペクトル分解を例にして
- ・可解線形回帰
- ・NMRの緩和モード分解
- ・SPring-8全ビームラインベイズ化計画
- ・まとめと今後の展開

内容

- ・自己紹介
- ・異分野共創の目的と実践
 - ・データ駆動科学の三つのレベル
 - ・ベイズ的スペクトル分解を例にして
- ・可解線形回帰モデル
- ・NMRの緩和モード分解
- ・SPring-8全ビームラインベイズ化計画
- ・まとめと今後の展開

SPring-8全ビームラインベイズ化計画

敬称略



赤色BLが共用BL(JASRI担当): 計26本

2024年内に14BL/26のベイズ化が完了
 理事長賞受賞の波及効果により、
 SPring-8全体のミッションとなり、
 ベイズ化実績によりBLが評価される体制へ

情報と放射光研究者のマッチング

メスバウアー

岡田研学生+筒井

小角散乱

BL08B2
BL19B2

岡田研学生+桑本

XAS測定

BL37XU
BL39XU

岡田研学生+水牧

放射光ユーザーへの展開

時分割XRD

横山優一+河口彰吾、沙織
ユーザー: 公立大、東工大

年度	2021	2022	2023
導入	2	8	14
全BL	26	26	26

SPring-8

- ・アメリカのAdvanced Photon Source (APS),ヨーロッパのEuropean Synchrotron Radiation Facility(ESRF)と合わせて, **世界3大放射光施設**.
- ・理研はSPring-8を「データ創出基盤」であると言っている. **年間延べ1万人**が利用.
- ・APSやESRFにおいてベイズ計測は導入されていない.
- ・放射光におけるベイズ計測に関しては**日本が最先端**である.

SPring-8全ビームラインベイズ化計画

- ・通常では系統的手法がない、**モデル選択とデータ統合**をベイズ計測で系統的に取り扱う
- ・フラッグシップ戦略：ベイズ計測をSPring-8に導入し、身近(近くにくるな症候群)な計測と他の大型計測施設への**起爆剤**とする。
- ・2023年度JASRI理事長賞JASRIデータ駆動科学グループ横山優一氏受賞を契機に、全BLにベイズ計測利用の加速へ
- ・2024年中に14BL/26のベイズ化完了

SPring-8全ビームライン ベイズ化計画の波及効果

- ・ フラッグシップ戦略もあり、追従施設が続出
- ・ SPring-8/JASRI: 2023年3月7日
- ・ あいちSR: 2023年10月30日
- ・ 日本放射光学会 若手研究会: 2024年9月2日
- ・ 台湾(NSRRC): 2024年9月4日
- ・ 佐賀LS: 2024年10月16日
- ・ 広大HiSOR: 2024年11月18日
- ・ PF: 2025年2月6日

内容

- ・自己紹介
- ・異分野共創の目的と実践
 - データ駆動科学の三つのレベル
 - ベイズ的スペクトル分解を例にして
- ・可解線形回帰モデル
- ・NMRの緩和モード分解
- ・SPring-8全ビームラインベイズ化計画
- ・まとめと今後の展開

本セミナーの開催趣旨

物質科学を始めとした計測・解析分野と数学・統計学の研究者が
互いのシーズとニーズを共有し、**相互発展の可能性を探索**



工学研究院／超顕微解析研究センター



志賀雅亘



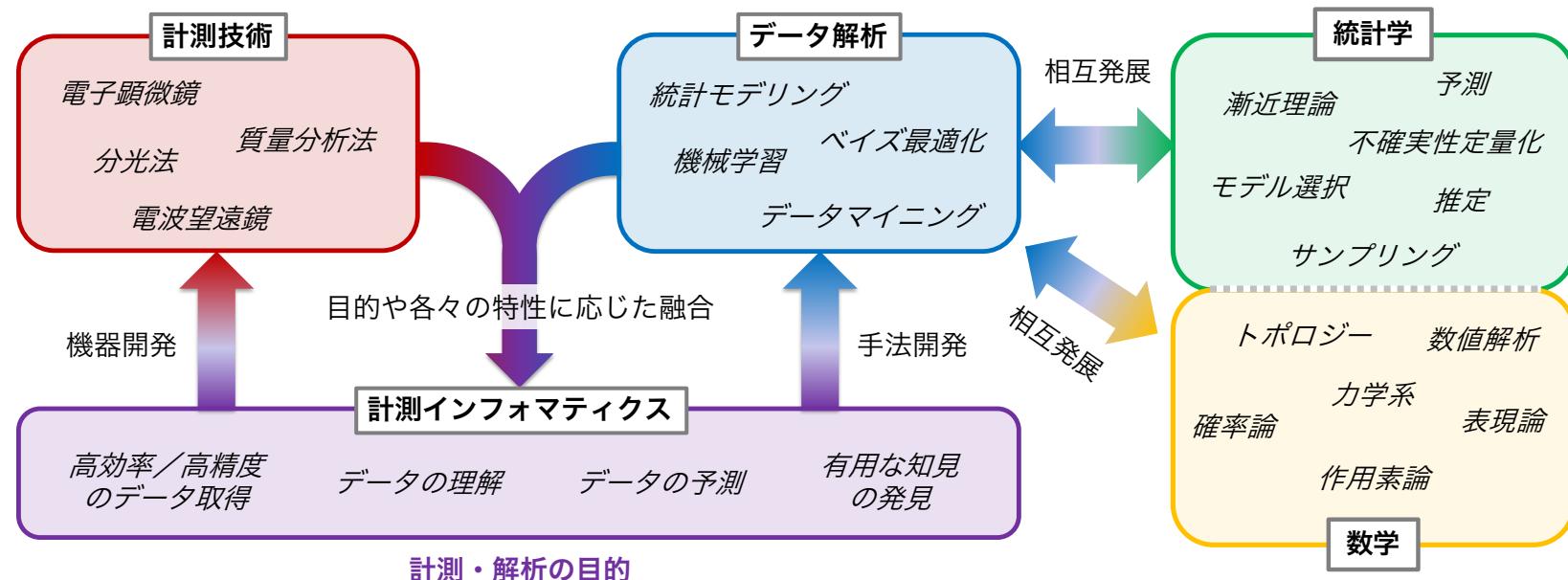
山本知一

マス・フォア・インダストリ研究所



徳田悟

and YOU!



徳田悟先生(九大マスフォアインダストリ)のスライドを引用

まとめ

- ・自己紹介
- ・異分野共創の目的と実践
 - ・データ駆動科学の三つのレベル
 - ・ベイズ的スペクトル分解を例にして
- ・可解線形回帰モデル
- ・NMRの緩和モード分解
- ・SPring-8全ビームラインベイズ化計画
- ・まとめと今後の展開