

統計学の科学応用とその周辺

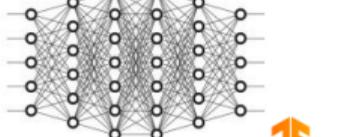
奥野彰文^{1,2,3}

¹統計数理研究所 ²総合研究大学院大学 ³理化学研究所 (AIP/CBS)

自己紹介



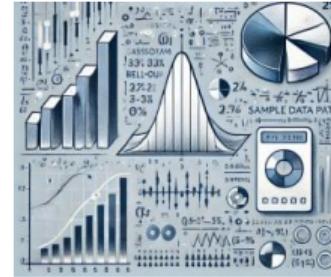
おくの あきふみ
奥野 彰文
博士 (情報学)



PyTorch TensorFlow

K Keras

ニューラルネット (深層学習)

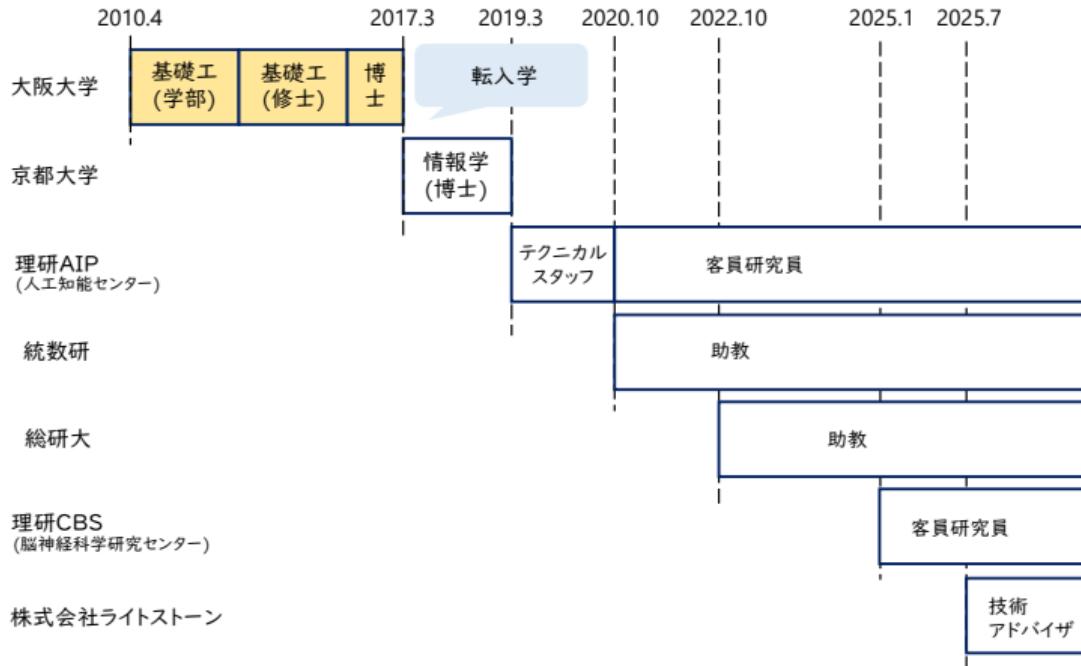


(DALL-E3によるイメージ図)

伝統的な統計学

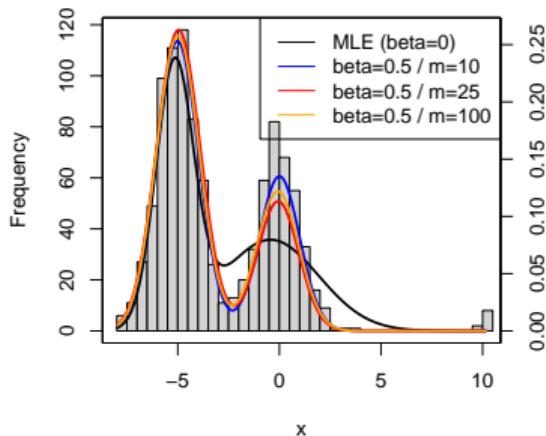
小サンプル領域での機械学習手法研究

学生時代、阪大基礎工にいました

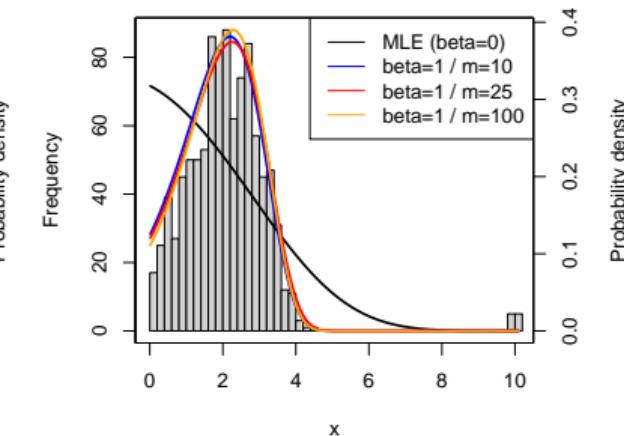


下平研究室. 2014年に基礎工学部賞, 2016年に基礎工学研究科賞.

- ▶ フラフラと色々やっている統計学/機械学習手法の研究者です。



(a) 混合正規分布



(b) ゴンペルツ分布

Figure: Okuno (AISM2024) Fig.1 より転載。

「一般の確率モデルに対するロバストダイバージェンスの最小化」

- ▶ マニアックな研究が好き。
- ▶ 特に非線形モデル（ニューラルネット）での統計解析を研究している。

昨日の集会では別のニッチな話をしていました

- ▶ 計算代数によるニューラルネットの局所解全列挙¹
- ▶ 「パーセプトロンの2乗損失+リッジ正則化」は区分多項式なので、推定方程式が(区分的に)代数方程式になり、局所最小解を全列挙できる².

The one-dimensional connected set of local minima lies entirely within the interior region. This solution set is characterized as the collection of points $\psi = (b_{11}, b_{21}, c_1, c_2)$ simultaneously satisfying all the constraints shown below: that is, inequalities

$$\begin{aligned} c_1 - \frac{17b_{11}}{100} &> 0, \quad c_2 - \frac{17b_{21}}{100} > 0, \quad \frac{11b_{11}}{25} + c_1 > 0, \quad \frac{11b_{21}}{25} + c_2 > 0, \\ c_1 - B_{11} &< 0, \quad c_2 - B_{21} < 0, \quad c_1 - \frac{2b_{11}}{5} < 0, \quad c_2 - \frac{2b_{21}}{5} < 0, \quad c_1 - \frac{71b_{11}}{100} < 0, \quad c_2 - \frac{71b_{21}}{100} < 0, \end{aligned}$$

and equations

$$\begin{aligned} 0 &= b_{11} + R_1 c_1^7 + R_2 c_1^5 c_2^2 + R_3 c_1^5 + R_4 c_1^3 c_2^4 + R_5 c_1^3 c_2^2 + R_6 c_1^3 + R_7 c_1 c_2^6 + R_8 c_1 c_2^4 + R_9 c_1 c_2^2 - R_{10} c_1, \\ 0 &= b_{21} + R_{11} c_1^6 c_2 + R_{12} c_1^4 c_2^3 + R_{13} c_1^4 c_2 + R_{14} c_1^2 c_2^5 + R_{15} c_1^2 c_2^3 + R_{16} c_1^2 c_2 + R_{17} c_2^7 + R_{18} c_2^5 + R_{19} c_2^3 - R_{20} c_2, \\ 0 &= c_1^8 + 4c_1^6 c_2^2 + R_{21} c_1^6 + 6c_1^4 c_2^4 + R_{22} c_1^4 c_2^2 + R_{23} c_1^4 + 4c_1^2 c_2^6 + R_{24} c_1^2 c_2^4 + R_{25} c_1^2 c_2^2 - R_{26} c_1^2 + c_2^8 \\ &\quad + R_{27} c_2^6 + R_{28} c_2^4 - R_{29} c_2^2 - R_{30}, \end{aligned}$$

¹*代数統計のつもりでいます

²Fukasaku, Kabata, and Okuno[†], arXiv:2508.17783 ; [†]責任著者

手法研究者 meets 異分野連携研究 (科学研究)



天文学



核融合プラズマ



計算代数など

(推進中)

企業共同研究

- ▶ 素人が少しずつ何かをやろうとしているところ.
- ▶ まだ始めたばかりですが、少しずつ進めています.

“理想”の手法研究



手法研究



すごい手法の提案

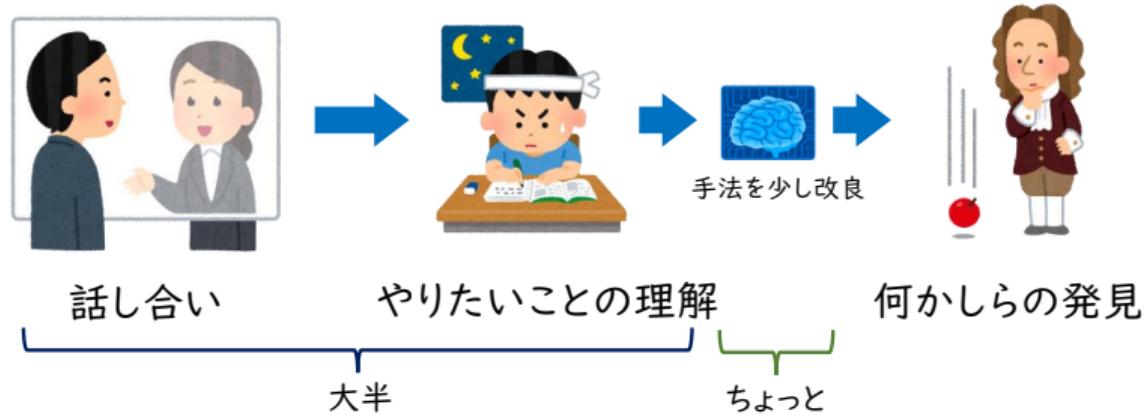


すごい発見

- ▶ …という感じでトントン拍子にはいかない。
- ▶ 「手法としての差分」が役立つ問題は、そんなに簡単に見つからない。

現實

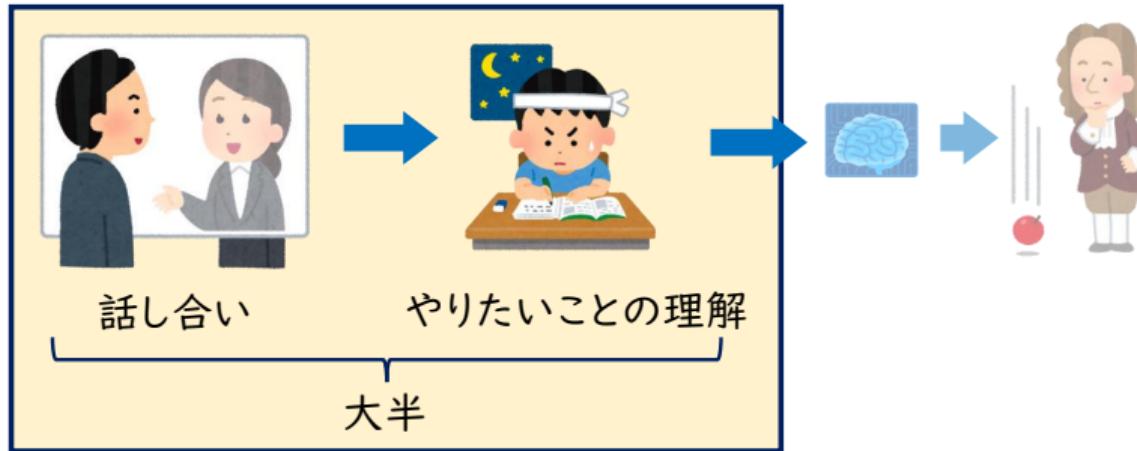
- ▶ 「相手の言っていることを理解する」まで大半の時間が消費される。



- ▶ AI for Science?
 - ▶ MA³ for Science...

3 Multivariate Analysis

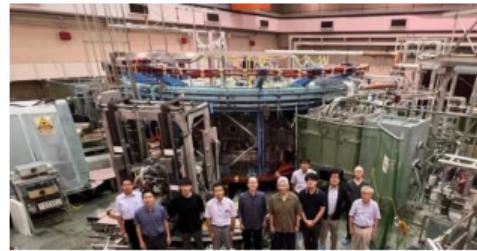
一番重要なこと



- ▶ 「いかにして前半部分の時間を用意するか」が一番重要.
- ▶ ひとたび（手法の言葉で）目的が明確化されれば、誰がやってもだいたい同じ.
- ▶ 翻訳して問題に落とし込むまでが腕の見せ所.

プラズマ物理学者との関わり

共同研究プロジェクト「統計数理核融合」(2022年6月~)

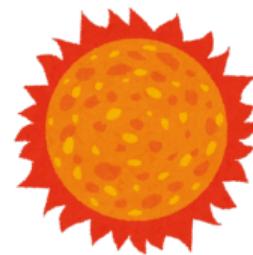


<https://statplasma.github.io>

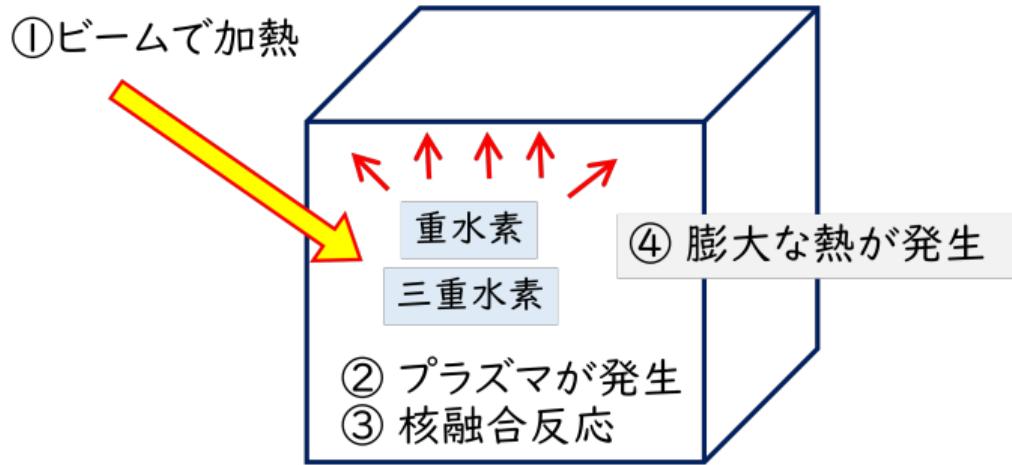
- ▶ ROIS戦略的研究プロジェクト (2022年6月~2025年3月, 代表: 三分一史和)
- ▶ NINS-Open Mix Lab (2025年4月~, 代表: 横山雅之)
- ▶ 定期的な集会の重要性を感じています。

背景1：プラズマとは何か？

- ▶ 固体・液体・気体に続く物質の第4の状態と呼ばれ、
気体に更にエネルギーを加えた不安定な活性状態を指します。
- ▶ 太陽や雷・大気圏上層のオーロラなど、自然界のいたるところで見られます。
- ▶ 「装置内に人工的にプラズマを発生させて維持したい」という研究段階です。

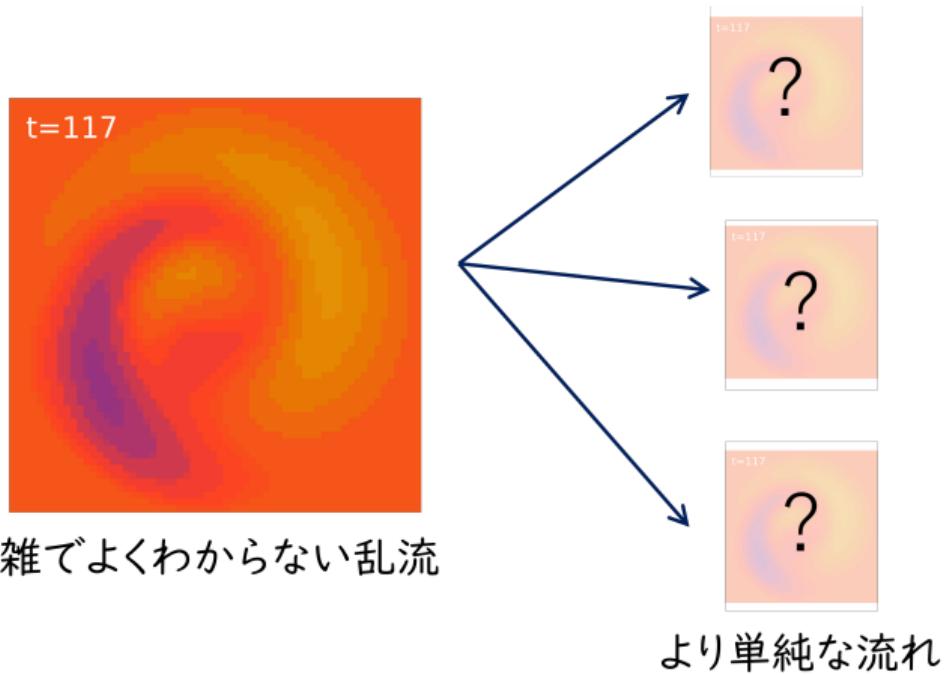


背景2：核融合発電とは何か？



- ▶ 出てきた熱を使って発電する
- ▶ 一方、熱がすごすぎて装置が溶ける ⇒ 磁場を利用した閉じ込め
- ▶ プラズマ状態を維持（制御）できないと核融合反応が起こせない

乱流を理解するために、「分解」をする



分解にかかる問題点2つ

- ▶ 分解の自由度の高さ : 2次元の流れ $\phi(x, y, t)$ を分解するとき, 時刻 t 方向もあるので, (何の指針も無しだと) 分解の自由度が高すぎる.
- ▶ 総エネルギーの分解 : 時刻 t での総エネルギーは

$$E[\phi](t) = \iint \left\{ \left(\frac{\partial \phi(x, y, t)}{\partial x} \right)^2 + \left(\frac{\partial \phi(x, y, t)}{\partial y} \right)^2 \right\} dx dy. \quad (1)$$

- ▶ $E[\phi_1 + \phi_2](t) = E[\phi_1](t) + E[\phi_2](t)$ となるような分解 $\phi = \phi_1 + \phi_2$ にしたい.
- ▶ 難に分解すると総エネルギーが分解されない….

特異値分解

- ▶ 亂流(静電ポテンシャル) $\phi(x, y, t)$ を独立な要素 ϕ_1, ϕ_2, \dots に分解:

$$\phi(x, y, t) = \sum_{j \geq 0} \phi_j(x, y, t).$$

特異値分解 (a.k.a., 主成分分析)

$$\phi_j(x, y, t) = s_j \Psi_j(x, y) h_j(t). \quad (2)$$

s_j は特異値(要素 j の重要度), $\Psi_j(x, y)$ は空間構造, $h_j(t)$ は時間発展を表す.

- ▶ プラズマ業界では(2)型の分解を特に Proper Orthogonal Decomp. (POD) と呼ぶ.
- ▶ $E[\phi_j + \phi_k](t) = E[\phi_j](t) + E[\phi_k](t)$ を制約に入れる POD がある.
- ▶ 実際にはシミュレーションの観測値から離散的に計算する.

より詳細な構造 $\Psi_j(x, y)$ のリスト

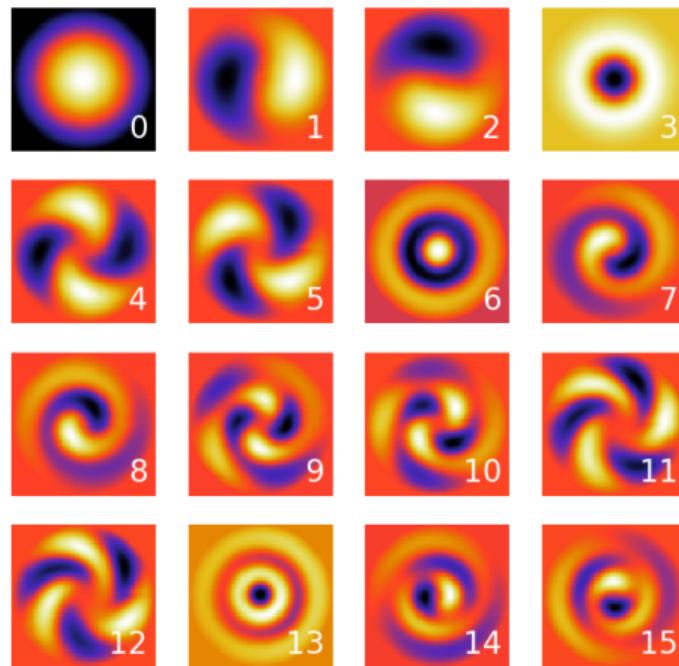
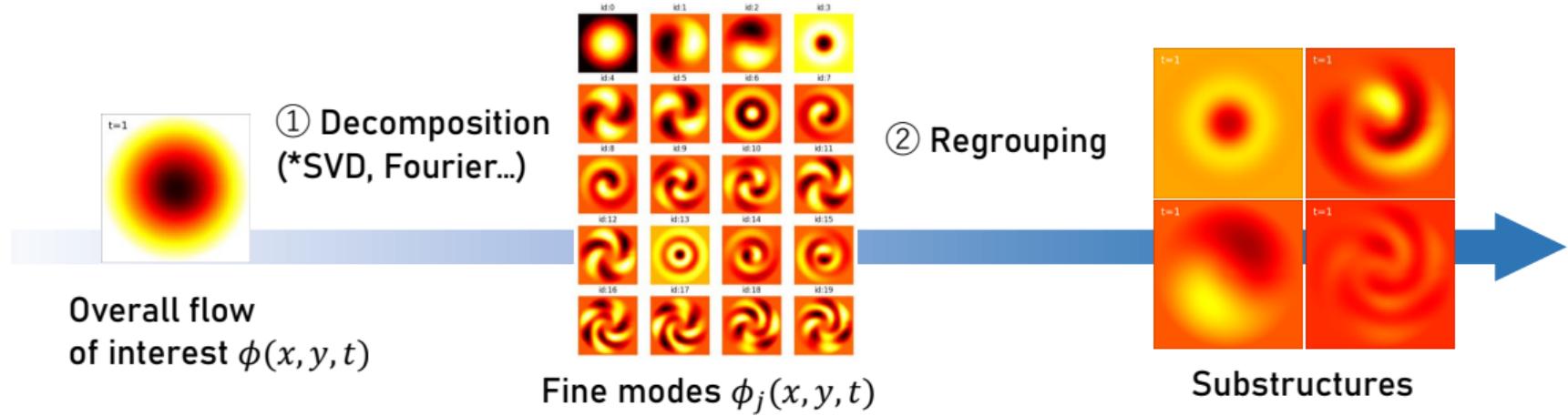


Figure: 背景構造 ($j = 0$), Zonal ($j = 3, 13, \dots$), 亂流 (それ以外)

手続きの概要



- ▶ 「要素の分解」 (SVD/フーリエ/...)
- ▶ 「まとめあげ」 (クラスタリング/特殊法...)

まずは、とりあえずやってみた研究

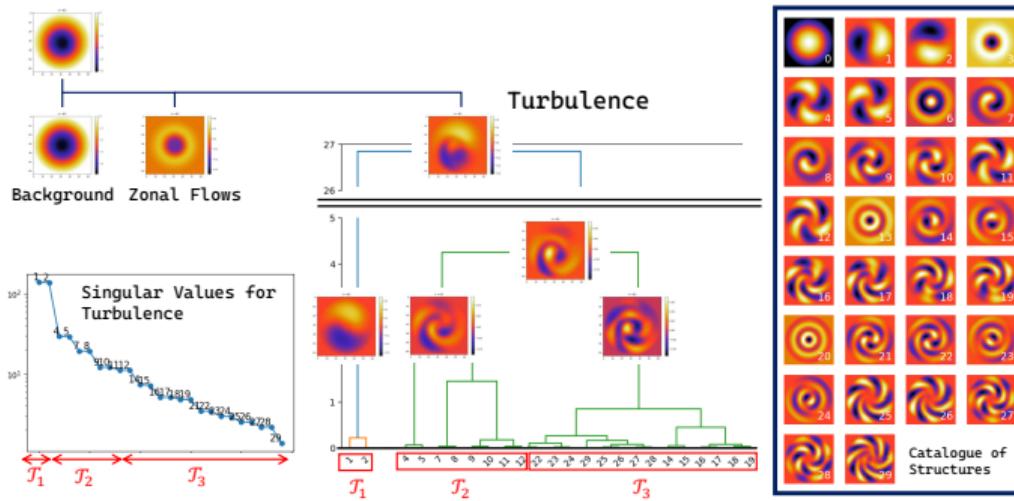
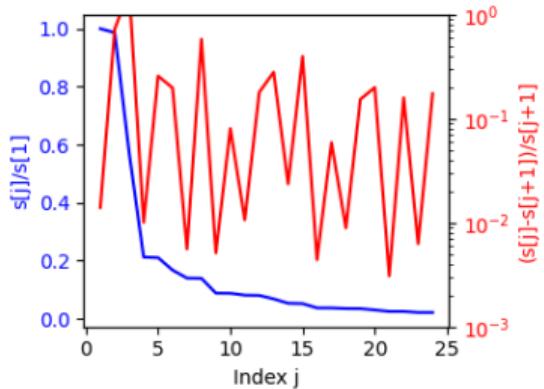


Figure: 亂流の分解と階層的な分類. 実際には階層型クラスタリングをかけただけ

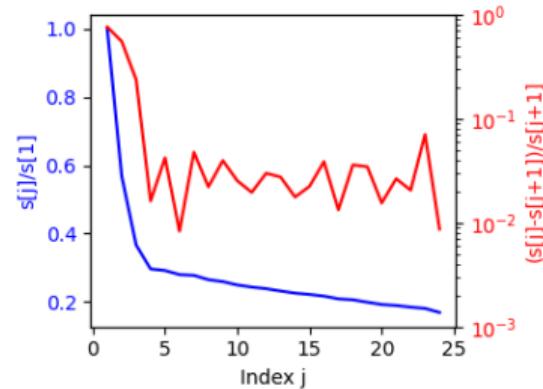
- ▶ Okuno, Kodahara, and Sasaki (Plasma and Fusion Research:RC 2024)
- ▶ コミュニケーションと、数理的な定式化が一番難しい。

実際の乱流はもっと難しい

- ▶ SVDなどで分解したモードが重複した特異値を持つ=自由度は低め.



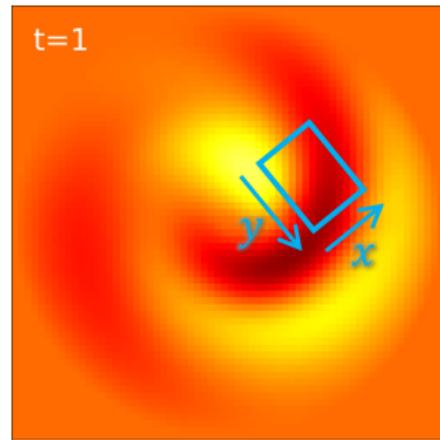
(a) 前節の乱流の特異値（自由度低め）



(b) 今回扱う乱流の特異値（自由度高め）

- ▶ より複雑な乱流の分解になる.
- ▶ 何かしら物理的な情報を入れて考えたい.

利用する座標系



微小四角形領域を切り取る. (x : radial, y : poloidal).

- ▶ zonal成分を引っ張り出す :

$$\phi_{ZF}(x, y, t) = \frac{1}{y_{\max} - y_{\min}} \int \phi(x, y, t) dy.$$

残りの成分を乱流とする

乱流成分を

$$\phi_{\text{turb}}(x, y, t) = \phi(x, y, t) - \phi_{\text{ZF}}(x, y, t)$$

として、これを分解したい。どうやるか？

- ▶ poloidal velocity of the zonal flowを

$$V[\phi_{\text{ZF}}](x, t) = \int \frac{\partial \phi_{\text{ZF}}(x, y, t)}{\partial x} dy$$

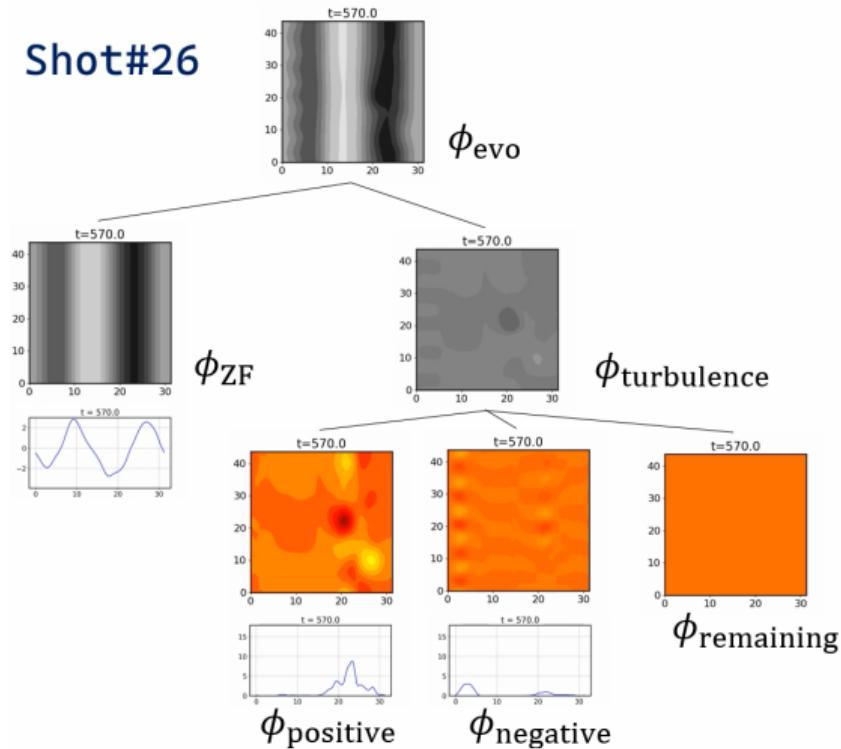
として、運動エネルギー密度

$$I[\phi](x, t) = \int \left\{ \left(\frac{\partial \phi(x, y, t)}{\partial x} \right)^2 + \left(\frac{\partial \phi(x, y, t)}{\partial y} \right)^2 \right\} dy. \quad (3)$$

がこのvelocityと類似するもの ϕ_{pos} 、逆になるもの ϕ_{neg} 、その他 ϕ_{rem} に分解する。

最終的に得たい分解はこんな感じ

Shot#26



エネルギー“密度”の分解

▶ 再掲：エネルギー密度

$$I[\phi](x, t) = \int \left\{ \left(\frac{\partial \phi(x, y, t)}{\partial x} \right)^2 + \left(\frac{\partial \phi(x, y, t)}{\partial y} \right)^2 \right\} dy. \quad (4)$$

▶ 渦を分解する以上は、この密度も分解したい：

$$I[\phi_j + \phi_k](x, t) = I[\phi_j](x, t) + I[\phi_k](x, t) + \underbrace{\nu(x, t)}_{\text{Interaction}} \quad (5)$$

が、単純なPOD (2) だと相互作用項 $\nu(x, t)$ が残る。困る。

(査読でツッコミを受けたので) 識者向けの補足

- ▶ 既存研究にあるとおり, 勾配を使って内積

$$\langle \phi_j, \phi_k \rangle = \int \nabla \phi_j \cdot \nabla \phi_k$$

を定義したPOD (2) はエネルギー $E[\phi](t) = \int I[\phi](x, t) dx$ を分解できるが, (x方向の) エネルギー密度 $I[\phi](x, t)$ は分解できない.

エネルギー密度の分解ではPODでも相互作用項 $\nu(x, t)$ が残るのだが, 総エネルギーを考えるときは x についても積分をるので, 分解可能性が成り立っていて矛盾しない:

$$E[\phi_j + \phi_k](t) - \{E[\phi_j](t) + E[\phi_k](t)\} = \int \nu(x, t) dx = \langle \phi_j, \phi_k \rangle = 0$$

ではどうするか?

- ▶ 直交テンソル分解 $\phi(x, y, t) = \sum_j s_j \alpha_j(x) \beta_j(y) \gamma_j(t)$. 将来やりたい.
- ▶ (今回) 部分的なフーリエ基底展開 :

$$\phi_{\text{turb}}(x, y, t) = \sum_{j \geq 0} \phi_j(x, y, t), \quad \phi_j(x, y, t) = \lambda_j(x, t) e_j(y), \quad (6)$$

(ただし $\lambda_j(x, t) = \int \phi_{\text{turb}}(x, y, t) e_j(y) dy$). 離散的にはテンソル積で高速計算できる.

このとき, エネルギー密度はモードごとに分解される :

$$I[\phi_j + \phi_k](x, t) = I[\phi_j](x, t) + I[\phi_k](x, t).$$

状況の整理

(1) エネルギー密度が分解可能なモード $\{\phi_j(x, y, t)\}_j$ が得られた：

$$\phi_{\text{turb}}(x, y, t) = \sum_j \phi_j(x, y, t), \quad I[\phi_j + \phi_k](x, t) = I[\phi_j](x, t) + I[\phi_k](x, t).$$

(2) 互いに素な添え字集合 $\mathcal{J}_{\text{pos}}, \mathcal{J}_{\text{neg}}, \mathcal{J}_{\text{rem}} \subset \{1, 2, \dots\}$ をいい感じに作り、

$$\phi_{\text{pos}}(x, y, t) = \sum_{j \in \mathcal{J}_{\text{pos}}} \phi_j(x, y, t), \quad \phi_{\text{neg}}(x, y, t) = \sum_{j \in \mathcal{J}_{\text{neg}}} \phi_j(x, y, t)$$

のエネルギー密度 $I[\phi_{\text{pos}}](x, t), I[\phi_{\text{neg}}](x, t)$ が poloidal velocity of ZF $V[\phi_{\text{ZF}}](x, t)$ と正/負の相関を持つようにしたい。

類似度の設計

- ▶ 各時点 t で、エネルギー密度 $I[\phi](x, t)$ と $V[\phi_{ZF}](x, t)$ の類似度を以下で定義：

$$\begin{aligned}s[\phi](t) &= s^{(0)}[\phi](t) + s^{(2)}[\phi](t), \\s^{(0)}[\phi](t) &= \int V[\phi_{ZF}](x, t) I[\phi](x, t) dx, \\s^{(2)}[\phi](t) &= \int \frac{\partial^2 V[\phi_{ZF}](x, t)}{\partial x^2} \frac{\partial^2 I[\phi](x, t)}{\partial x^2} dx.\end{aligned}$$

- ▶ こうすると、類似度 $s, s^{(0)}, s^{(2)}$ もモードごとに分解できる：

$$s[\phi_j + \phi_k](t) = s[\phi_j](t) + s[\phi_k](t), \quad (7)$$

すごくテクニカルな話

行列

$$S = \begin{pmatrix} s[\phi_1](t_1) & s[\phi_2](t_1) & \cdots & s[\phi_J](t_1) \\ s[\phi_1](t_2) & s[\phi_2](t_2) & \cdots & s[\phi_J](t_2) \\ \vdots & \vdots & & \vdots \\ s[\phi_1](t_T) & s[\phi_2](t_T) & \cdots & s[\phi_J](t_T) \end{pmatrix} = (\mathbf{s}[\phi_1], \mathbf{s}[\phi_2], \dots, \mathbf{s}[\phi_J])$$

を最初に一回計算すると、構成された

$$\tilde{\phi} = \sum_{j \in \mathcal{J}} \phi_j$$

とzonal flowとの類似度（の時系列変化）は

$\sum_{j \in \mathcal{J}} \mathbf{s}[\phi_j] \in \mathbb{R}^T$ で計算できる。毎回内積計算(積分)しなくてよい。

当初の思惑

- ▶ 単純に、最終的な類似度を

$$\tilde{s}[\tilde{\phi}] = \sum_{t \in T} s[\tilde{\phi}](t) = \sum_{t \in T} \sum_{j \in \mathcal{J}} s[\phi_j](t)$$

とすると、

$$\mathcal{J}_{\text{pos}} = \left\{ j : \sum_{t=1}^T s[\phi_j](t) > 0 \right\}$$

$$\mathcal{J}_{\text{neg}} = \left\{ j : \sum_{t=1}^T s[\phi_j](t) < 0 \right\}$$

となってかっこよく終了・・・！のはずでしたが、得られた分解は微妙でした。

学習スコアの設計

色々スコアを作って試行錯誤した結果、

$$\begin{aligned} M(\mathcal{J}_{\text{pos}}, \mathcal{J}_{\text{neg}}, \mathcal{J}_{\text{rem}}) = & \int \tanh \left(s^{(0)}[\phi_{\text{pos}}](t) \right) dt \\ & + \int \tanh \left(-s^{(0)}[\phi_{\text{neg}}](t) \right) dt \\ & - \lambda_0 \rho \left(s^{(0)}[\phi_{\text{rem}}](t) \right) \\ & + \int \tanh \left(s^{(2)}[\phi_{\text{pos}}](t) \right) dt \\ & + \int \tanh \left(-s^{(2)}[\phi_{\text{neg}}](t) \right) dt \\ & - \lambda_2 \rho \left(s^{(2)}[\phi_{\text{rem}}](t) \right), \end{aligned}$$

を最大化するのが良かった。 $\rho(f(t))$ は $f(t)$ の時間方向の標準偏差。

遺伝的アルゴリズム

!pip install GeneticAlgorithm

- ▶ 生物が進化して環境に適合する様子を模したアルゴリズム⁴



- ▶ プラズマの“遺伝子” $\mathcal{J}_{\text{pos}}, \mathcal{J}_{\text{neg}}, \mathcal{J}_{\text{rem}}$ を交配し、スコア $M(\mathcal{J}_{\text{pos}}, \mathcal{J}_{\text{neg}}, \mathcal{J}_{\text{rem}})$ を最大化するよう進化させる。

⁴外部向け説明

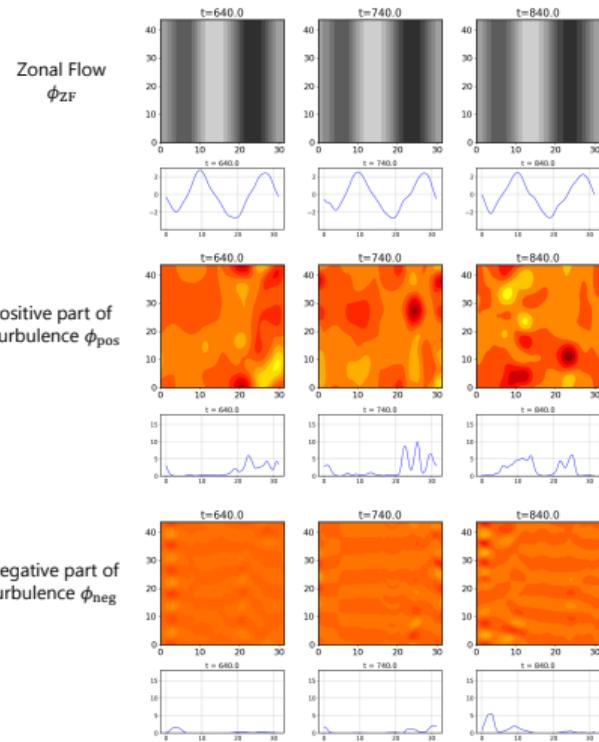


Figure: shot26: Temporal evolution (at $t = 640, 740, 840$) of the zonal flow and positive/negative substructures.

総運動エネルギーの変化

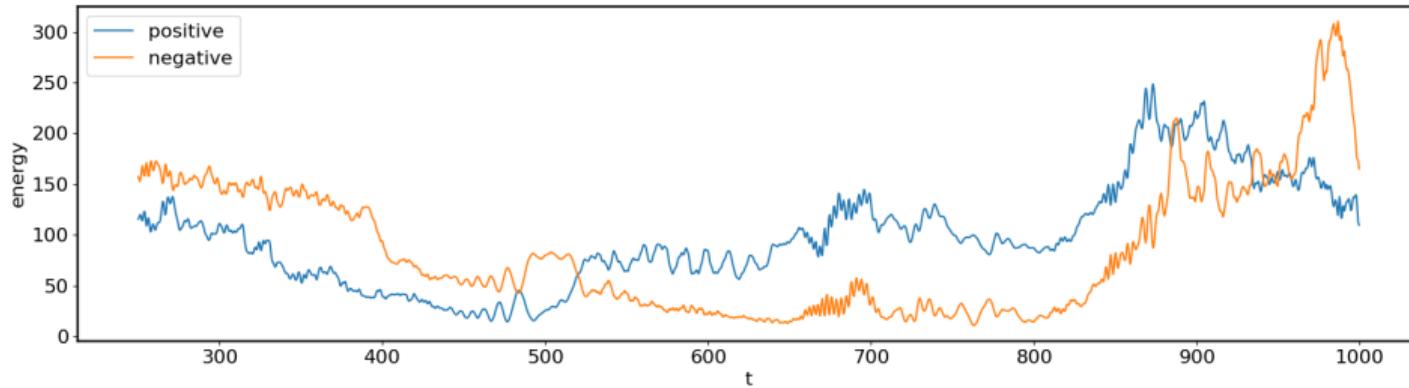


Figure: **shot26**: Temporal evolution of the kinetic energy for the positive/negative substructures. It suggests that the phases of the positive and negative substructures appear to be inverted over relatively long time spans.

- ▶ $\phi_{\text{pos}}(x, t), \phi_{\text{neg}}(x, t)$ の総エネルギー量が完全に逆の位相になっている。

総運動エネルギーの変化（似ている別shot）

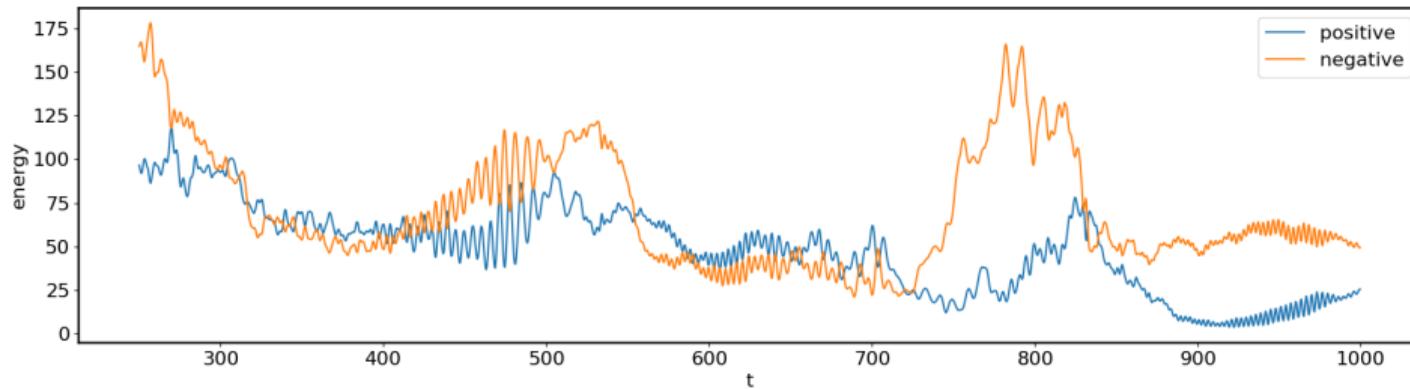


Figure: shot 27

色々計算した結果

ビデオを流す。

- ▶ Shot18: Negativeが強い
- ▶ Shot26: Positive/Negativeのバランス型
- ▶ Shot32: Positiveが強い

- ▶ 採択された : Okuno and Sasaki (Physics of Plasmas, 2025)
- ▶ 共同研究者が言っていた通りの結果が自動で再現できるようになった。
- ▶ 別の類似度で計算するともっと面白いらしい(?)

相互交流は今後も続きます...

これまでに開催・参加したイベント (*対面での開催のみ抜粋)

日付	イベント名	開催場所
2025年12月1日-4日	プラズマ・核融合学会年会 [写真1, 写真2, 写真3, 写真4]	京都工芸繊維大学松ヶ崎キャンパス
2025年9月10日	統計学会・核融合企画セッション [写真1, 写真2]	関西大学千里山キャンパス
2025年6月17日-18日	NINS OML キックオフ会合 [写真1, 写真2, 写真3]	統計数理研究所
2025年3月8日	日本統計学会春季集会 参加 [写真1, 写真2]	筑波大学東京キャンパス文京校舎
2025年2月13日-14日	戦略PJT最終シンポジウム [写真1, 写真2, 写真3, 写真4]	核融合科学研究所
2025年1月31日	個別研究打ち合わせ [詳細]	京都大学宇治キャンパス
2024年12月9日-13日	13th ITER International School 参加 [写真1, 写真2]	名古屋プライムセントラルタワー
2024年11月17日-20日	プラズマ・核融合学会年会参加 [写真1, 写真2]	タワーホール船堀
2024年9月2日	統計学会・核融合企画セッション [写真1, 写真2]	東京理科大学神楽坂キャンパス
2024年6月16日	合同研究打ち合わせ [写真1]	福島県郡山市
2024年6月13日-14日	核融合エネルギー連合講演会参加 [写真1]	八戸市公民館

<https://statplasma.github.io>

その他のプロジェクト

天文学

- ▶ Kohei Hattori, Akifumi Okuno, and Ian U. Roederer. “Finding r-II sibling stars in the Milky Way with the Greedy Optimistic Clustering algorithm”. *Astrophysical Journal* (2023).
- ▶ Akifumi Okuno and Kohei Hattori. “A greedy and optimistic clustering for leveraging individual covariate uncertainty”. *Annals of the Institute of Statistical Mathematics* (2025).



天文学者と統計学者の意見交換会@国立天文台

鉱物化学

- ▶ 東大地震研共同研究 (2024-B-01, 代表: 板野敬太)
「データサイエンスを活用した地球物理・化学データ解析とモデリングの精緻化」



- ▶ ROIS戦略プロジェクト (採択済): 秋田大-極地研-統数研.
 - ▶ JpGU (日本地球惑星科学連合) 招待講演.
- このほかまだ非公開トピックがいくつか..

基礎研究者か？応用研究者か？

- ▶ 統計学者は基礎/応用両方の立場から共同研究がありうる。



- ▶ 両方を体験して思うところはたくさんあります。
 - ▶ どれくらい個別案件のドメイン知識をつけるべきかなど、いろいろ模索中…

宣传 1

1月 19 シンポジウム「AI4S 2026：機械学習・統計数理が切り拓くAI for Scienceの新展開」

(月) 10:00 | 千代田区一ツ橋2丁目1-2 学術総合センター2F | By 情報・システム研究機構 統計数理研究所



テクノロジー/サイエンス テクノロジー サイエンス データベース AI 生成AI

AI4S 2026 ~機械学習・統計数理が切り拓くAI for Scienceの新展開~

【日時】

2026年1月19日 (月)

第一部 10:00 -17:30 (シンポジウム) *受付開始: 9:30

第二部 18:00 -20:00 (意見交換会・ポスター展示)

チケット

第一部 (無料)

第二部 (2,000円: 当日現金払) *申込時にはチケットの金額は無料と表示されますが、当日現金2,000円をお支払いください。

日時

2026/1/19 (月)

10:00 - 20:00 GMT+09:00

[カレンダーに追加](#)

【場所】

一橋講堂

〒101-8439 東京都千代田区一ツ橋 2丁目1-2 学術総合センター2階

アクセス: <https://www.hit-u.ac.jp/hall/>

販売期限: 2026/1/14 (水) 23:59

[チケットを申し込む](#)

<https://peatix.com/event/4583168/>

第2回：計算技術による学際的統計解析ワークショップ

開催概要

高度な計算技術を活用した統計手法研究を目指し、学際的な研究者交流を目的とします。口頭発表はオンライン配信予定（質疑応答は対面を最優先します）。第1回の開催情報は[こちら](#)。参加登録は[こちら](#)。

- 日時：2026年2月16日・17日
- 場所：[統計数理研究所 D305室 \(予定\)](#)
- オーガナイザ：[奥野彰文](#)（統数研/総研大/理研）

スケジュール（予定）

(*は全て招待講演)

2月16日

13:00-13:30

[奥野彰文](#)（統計数理研究所）

統計手法研究と計算技術

13:30-14:30

[深谷猛](#)（北海道大学）*

TRA

<https://okuno.net/events/ISACT2026>

宣伝 3

統計サマーセミナーとは

統計サマーセミナーは、統計学または関連分野に在籍する大学院生および若手研究者 (young statisticians' group; ysg) が中心に運営するセミナーで、1971年以来（コロナ禍により中止になった2020年をのぞき）ほぼ毎年開催されています。その長い歴史とともに、多くの研究者が本セミナーに関わり、現在も第一線で活躍しています。

日程・場所（※予定）

2026年8月5日（水）～8月7日（金）

海と空と風の宿 ホテル明山荘（愛知県 蒲郡市）

参加登録は2026年4月以降開始予定

招待講演者（※予定）

清水 昌平 先生（大阪大学 産業科学研究所）

青柳 美輝 先生（日本大学 理工学部）

幹事

奥野 彰文（統計数理研究所 統計基盤数理研究系）

高畠 哲也（大阪大学大学院 基礎工学研究科）

<https://okuno.net/events/ysg2026>

統計学者は何をするべきか?

(複数の意味で) 使えるものは何でも使えるよいと思う.



何かやりたい人は是非ご連絡を : okuno@ism.ac.jp